

Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources

Gene-Wei Li,^{1,2,3,*} David Burkhardt,^{2,4} Carol Gross,^{2,4,5} and Jonathan S. Weissman^{1,2,3,*}

¹Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute

²California Institute of Quantitative Biosciences

³Center for RNA Systems Biology

⁴Department of Microbiology and Immunology

⁵Department of Cell and Tissue Biology

University of California, San Francisco, San Francisco, CA 94158, USA

*Correspondence: gene-wei.li@ucsf.edu (G.-W.L.), weissman@cmp.ucsf.edu (J.S.W.)

<http://dx.doi.org/10.1016/j.cell.2014.02.033>

SUMMARY

Quantitative views of cellular functions require precise measures of rates of biomolecule production, especially proteins—the direct effectors of biological processes. Here, we present a genome-wide approach, based on ribosome profiling, for measuring absolute protein synthesis rates. The resultant *E. coli* data set transforms our understanding of the extent to which protein synthesis is precisely controlled to optimize function and efficiency. Members of multiprotein complexes are made in precise proportion to their stoichiometry, whereas components of functional modules are produced differentially according to their hierarchical role. Estimates of absolute protein abundance also reveal principles for optimizing design. These include how the level of different types of transcription factors is optimized for rapid response and how a metabolic pathway (methionine biosynthesis) balances production cost with activity requirements. Our studies reveal how general principles, important both for understanding natural systems and for synthesizing new ones, emerge from quantitative analyses of protein synthesis.

INTRODUCTION

Protein biosynthesis is by far the largest consumer of energy during cellular proliferation; translation by ribosomes is estimated to account for ~50% of the energy consumption of a rapidly growing bacterial cell and ~30% of that for a differentiating mammalian cell (Buttgereit and Brand, 1995; Russell and Cook, 1995). The tremendous cost associated with protein synthesis makes it a key step for regulating diverse cellular functions. Therefore, determining how a cell allocates its synthesis capacity for each protein provides foundational information for systems biology.

A fundamental question is whether it is necessary for the cell to exert tight control over the synthesis of individual protein components. For example, the levels of stoichiometric components of protein complexes could be established by differential degradation of excess subunits (Blikstad et al., 1983; Lehnert and Lodish, 1988), rather than by precise synthesis. Moreover, precise control of steady-state protein abundance may not be critical for the performance of cellular circuits. The architectures of several signaling and metabolic pathways have been shown to be robust against variation in protein levels through posttranslational feedback (Alon et al., 1999; Barkai and Shilo, 2007; Batchelor and Goulian, 2003; Hart et al., 2011; Shinar et al., 2007; von Dassow et al., 2000). It remains to be explored whether these posttranslational mechanisms are the dominant strategy for maintaining proper functions or are simply fail-safe mechanisms added on to fine-tuned protein synthesis. More generally, defining such design principles is key to both understanding and manipulating quantitative behavior of a cell.

Efforts to monitor protein synthesis rates at the global level have mainly relied on pulsed metabolic labeling followed by 2D gel electrophoresis or, more recently, by mass spectrometry (Dennis, 1974; Lemaux et al., 1978; Schwanhäusser et al., 2009). Although relative changes in synthesis rates for the same protein are attainable (Selbach et al., 2008), absolute rates are more difficult to evaluate. Additionally, the precision of pulsed metabolic labeling is limited by requirement for nutrient shifts, which affect instantaneous rates of protein synthesis. Alternative methods for expression profiling by determining global mRNA levels (e.g., by high-density microarrays or RNA sequencing [RNA-seq]) do not report the extensive regulation present at the level of translation. These constraints point to a need for a label-free method with unbiased and deep coverage of cellular proteins.

Ribosome profiling—deep sequencing of ribosome-protected mRNA fragments—directly captures protein synthesis in natural settings (Ingolia et al., 2009). It is a general tool for monitoring expression as well as enabling identification of novel translational events (Brandman et al., 2012; Brar et al., 2012; Ingolia et al., 2011; Li et al., 2012; Oh et al., 2011; Stern-Ginossar

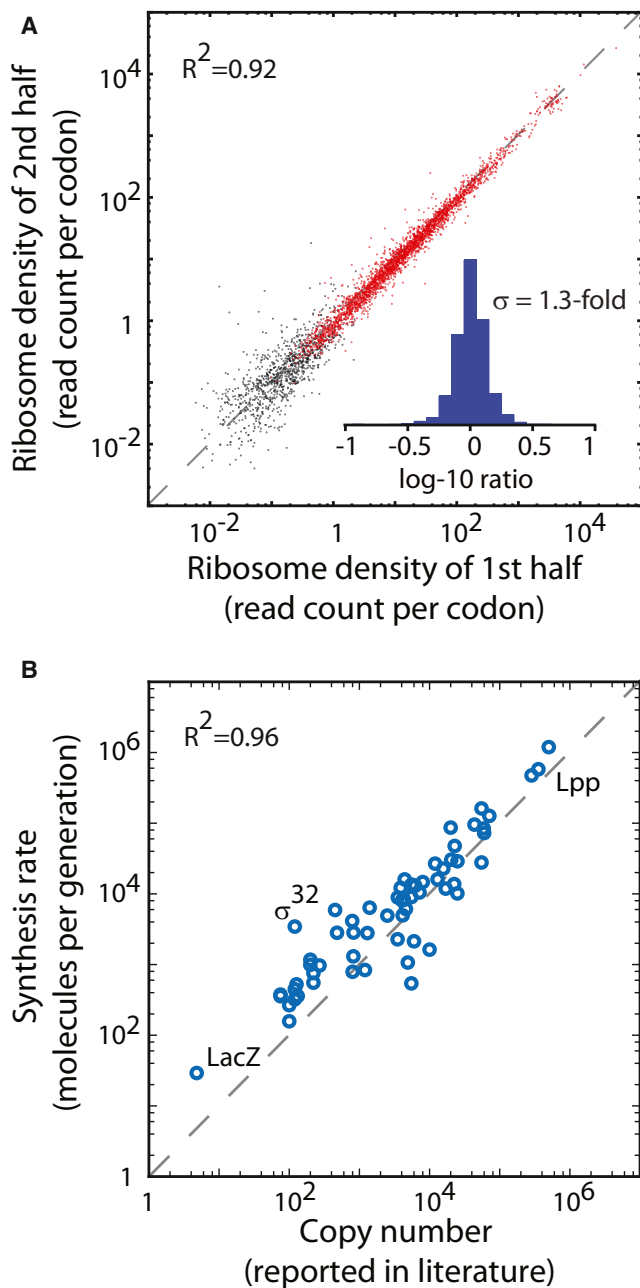


Figure 1. Absolute Quantification of Protein Synthesis Rates

(A) Effect of translational pausing on average ribosome density. Average ribosome density is plotted for the first and second half of each gene. The Pearson correlation for genes with at least 64 reads aligned to both halves (red) is $R^2 = 0.92$. The inset shows the distribution of the fold difference between the second and the first halves ($n = 2,870$; SD, 1.3-fold).

(B) Agreement between published protein copy numbers and absolute synthesis rates. The copy numbers of 62 proteins that have been individually quantified in the literature are plotted against the absolute protein synthesis rates (Pearson correlation, $R^2 = 0.96$).

See also [Figures S1](#) and [S2](#) and [Tables S1](#) and [S2](#).

[et al., 2012](#)). Here, we exploited the ability of ribosome profiling to provide quantitative measurements of absolute protein synthesis rates, covering >96% of cellular proteins synthesized in a single experiment. For stable proteins in bacteria, we then estimated absolute protein copy numbers.

This analysis revealed precise tuning of protein synthesis rates at the level of translation, including a broadly used “proportional synthesis” strategy in which components of multiprotein complexes are synthesized with ratios that quantitatively reflect their subunit stoichiometry. Optimized translation rates are also prevalent among members of functional modules—differential expression pertinent to their functional hierarchy, i.e., when the activity of one member is controlled by the other, was widely observed in our data set. The protein copy numbers inferred from synthesis rates also revealed rules that govern the abundance of transcription factors (TFs) and allowed quantitative characterization for the methionine (Met) biosynthesis pathway, for which we identified a bottleneck enzyme whose expression level is optimized for maximal growth rate. More broadly, our approach and data sets provide a foundation for quantitative understanding of both cellular physiology and precise biological engineering.

RESULTS

Genome-wide Measurement of Absolute Protein Synthesis Rates and Protein Copy Numbers

The ribosome-profiling approach involves freezing of cellular translation followed by digestion of all mRNA regions that are not protected by the ribosome ([Ingolia et al., 2009, 2012](#)). Each ribosome-protected mRNA fragment is then identified by massively parallel next-generation sequencing ([Ingolia et al., 2009, 2012](#)). Because each ribosome is producing one protein molecule, the rate of protein synthesis is proportional to the ribosome density of a given gene as measured by the footprint density (number of footprint per unit length of the gene), provided that all ribosomes complete a full-length protein and have similar average rates of elongation across genes. Both criteria are broadly met in our data set. During exponential growth in *E. coli*, there is little drop-off in ribosome density for the vast majority of genes ([Li et al., 2012](#); [Oh et al., 2011](#)) ([Figure 1A](#)). The few genes that display large drop-off could represent novel events of translational regulation ([Figure S1A](#) available online). We have previously demonstrated that rare codons are generally translated at similar speed as abundant codons, indicating that differences in codon usage between transcripts do not cause differences in the average rates of elongation ([Ingolia et al., 2011](#); [Li et al., 2012](#)). Moreover, sequence-dependent pausing of ribosomes ([Li et al., 2012](#)) does not appear to broadly distort the average density of ribosomes along a message because similar ribosome densities are observed in the first and second halves of each gene. Most genes differ by <30% (SD of the mean; [Figure 1A](#)). Additionally, correcting for sequence- and position-specific variation in elongation rates has only a modest effect on average ribosome density ([Figure S1](#)). Together, these results indicate that local variations in translation speed do not strongly impact synthesis rates measurements based on average ribosome density.

To broadly evaluate the rates of protein synthesis, we performed ribosome profiling in *E. coli* grown in different growth conditions with high sequencing depth (90 million fragments per sample) using a modified protocol that enables more complete capture of footprints (Experimental Procedures). Within each data set, synthesis rates were calculated as the average ribosome density in the gene body, with correction factors for elevated ribosome density at internal Shine-Dalgarno sequences and toward the beginning of open reading frames (Extended Experimental Procedures). The corrections were small (Figure S1D) but were nonetheless important for the quantitative analysis described below. We determined the absolute rates of synthesis (in units of molecules produced per generation) by normalizing the average ribosome density for each protein in the proteome by the total amount of proteins synthesized during the cell doubling time (Experimental Procedures). For growth in a rich defined medium (Neidhardt et al., 1974), we evaluated 3,041 genes, which account for >96% of total proteins synthesized. A similar number of genes were evaluated for glucose-supplemented minimal media. All of these genes have >128 ribosome footprint fragments sequenced, with an error of less than 1.3-fold across biological replicates. The lowest expression rate among these genes corresponds to approximately ten molecules per generation. The complete list of protein synthesis rates can be obtained at <http://ecoliwiki.net/tools/proteome/> (Table S1).

We validated our results by comparing our data against published measures of specific protein copy numbers for *E. coli*. Because the overwhelming majority of proteins are long lived compared to the cell cycle during exponential growth (Larrabee et al., 1980), the absolute copy number of a protein can be estimated as the synthesis rate multiplied by generation time (21.5 min in rich defined media; see Experimental Procedures). We compiled a list of 62 proteins that have been quantified individually in 21 independent laboratories (Table S2). Although each measurement is associated with its own uncertainty, we argue that collectively they represent the current standard for quantification. Our results agreed well with these published copy numbers with a Pearson correlation coefficient of $R^2 = 0.96$ (Figure 1B). Deviations from the identity line in Figure 1B likely reflect a biological phenomenon. For example, the strongest outlier is σ^{32} , the heat shock TF that is known to be actively degraded (Grossman et al., 1987). Our measures based on synthesis rates thus provide an upper bound for the protein levels for the small subset of proteins that are rapidly degraded. Differences in growth conditions and strain backgrounds contribute to other small differences between literature values and our results (see Extended Experimental Procedures). Existing efforts to globally quantify protein abundance in *E. coli* using mass spectrometry or fluorescent reporter show less concordance and dynamic range (Figure S2). In conclusion, our genome-wide synthesis rate measurements and the resulting estimate of protein abundance are supported by classic biochemical measurements across five orders of magnitude of protein abundance.

Proportional Synthesis of Multiprotein Complexes

We next used our measurements to evaluate the extent to which fine-tuned synthesis rates are a general feature of cellular physiology, focusing initially on members of stable multiprotein

complexes with known stoichiometry. The subunits of these complexes require balanced steady-state levels because excess components are often prone to misfolding or aggregation (Tyedmers et al., 2010). Although quality control mechanisms for removing uncomplexed proteins exist (Shemorry et al., 2013), it was unclear whether the stoichiometry balance is generally established first at the synthesis level.

We first examined the F_0F_1 ATP synthase complex, which consists of eight subunits, each with different stoichiometry, expressed from a single polycistronic transcript (the “ATP operon”). Despite sharing the same message, the ribosome density of each open reading frame is clearly distinct (Figure 2A) and qualitatively agrees with the differential synthesis rates previously reported by Brusilow et al. (1982) and Quax et al. (2013). Remarkably, the synthesis rates quantitatively reflect the stoichiometry of the complex; the ATP operon has evolved to synthesize the appropriate ratio of subunit proteins, ranging from 1- to 10-fold.

Rather than the ATP operon being a specialized case, we found that tuning of synthesis rates to the subunit stoichiometry, or “proportional synthesis,” is a broadly used strategy for protein complexes. We systematically assembled a list of stable multiprotein complexes with well-characterized stoichiometry in *E. coli* (Table S3). Of the 64 complexes (comprising 212 different proteins) that are expressed in our growth conditions, 59 (92%) adhere to proportional synthesis. The majority (55%) is synthesized at levels that are indistinguishable from the stoichiometry (smaller than the experimental uncertainty of 1.3-fold difference). The ratio of synthesis rates exceeds the ratio of stoichiometry by a factor of two in only five complexes (Figure S3D), and this small number of exceptions could suggest dominant control at the level of degradation or the existence of dynamic subcomplexes, as in the case of the outer-membrane protein assembly complex (β -barrel assembly machine [BAM]) (Rigel et al., 2013).

Proportional synthesis applies to both cytosolic and membrane proteins. For complexes with more than two components, the agreement between synthesis rates and subunit stoichiometry is plotted in Figures 2B and S3. We also observed very similar synthesis rates for complexes with two equimolar subunits (Figures 2C and S3A–S3C). Notably, proportional synthesis is robust against temperature; similar ratios in synthesis rates were observed both at 37°C and at 10°C (Figure S4A). Furthermore, both abundant and scarce proteins have evolved strict tuning of synthesis rates because the expression levels of these complexes range over four orders of magnitude.

Proportional synthesis in *E. coli* is predominantly achieved through translational, rather than transcriptional, control. The majority of multiprotein complexes encode their subunits on a single polycistronic mRNA, with each subunit translated from its own initiation site (47 out of 64 complexes; Figures 2B, 2C, and S3A). RNA-seq analysis confirms that the mRNA levels of the genes in these operons are similar, whereas the different translation efficiency (synthesis rate per mRNA) reflects the stoichiometry (Figures S4B and S4C; Table S4). Moreover, gene order does not explain differential synthesis rates (Figures 2A, 2C, and S4D), consistent with our previous observation that translation rates among genes in the same operon are only weakly correlated (inset in Figure 2C) (Oh et al., 2011). Protein

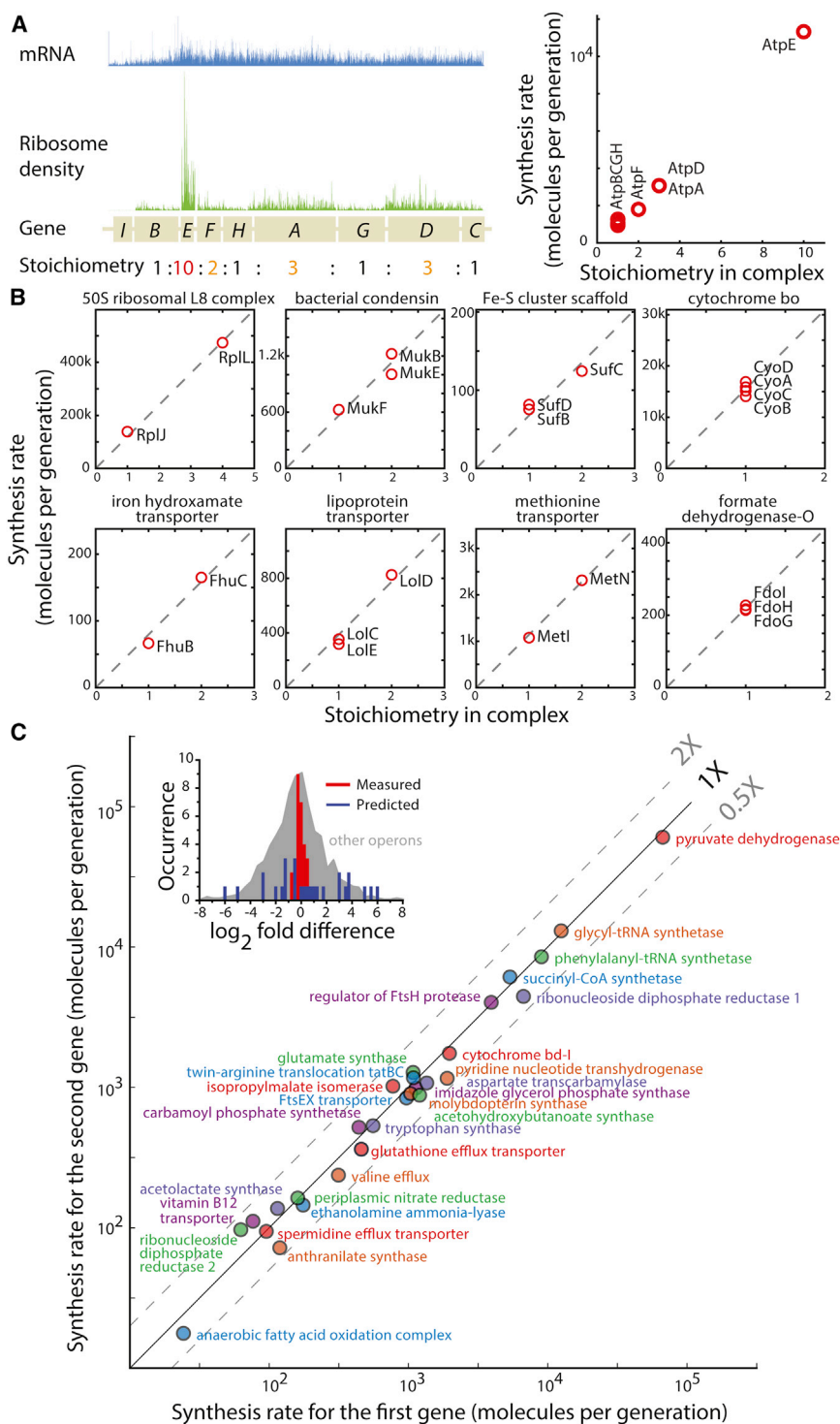


Figure 2. Proportional Synthesis of Multi-protein Complexes

(A) Translation rates reflecting subunit stoichiometry for the ATP operon. Eight subunits of the F_0F_1 ATP synthase are expressed from a polycistronic mRNA, whose level as measured by RNA-seq is shown in blue. Each subunit is associated with different levels of ribosome density (green), and the average density is proportional to the subunit stoichiometry (right).

(B) Proportional synthesis for a diverse range of complexes. Synthesis rates are plotted as a function of the subunit stoichiometry for multi-protein complexes whose subunits are encoded in the same operon. Complexes with different subunit stoichiometry or more than two subunits are included here (also see C). The dashed line indicates the best fit that crosses the origin.

(C) Proportional synthesis for complexes with two equimolar subunits. Each complex is plotted for the synthesis rates of the two subunits, with the earlier (later) gene in the operon on the horizontal (vertical) axis. A total of 28 equimolar and cotranscribed complexes, covering 4 orders of magnitude in expression level, are plotted here. Inset shows the histogram of fold difference between the synthesis rates of the two subunits. Our experimental results are shown in red, and the predicted values based on a thermodynamic model considering the sequence surrounding translation initiation sites are shown in blue (Salis et al., 2009). The distribution of the differences in translation rates for all other operons is shown in gray. (B) and (C) show complexes whose subunits are encoded on a single polycistronic operon. See Figures S3B and S3C for examples of proportional synthesis involving distinct transcripts. See also Figures S3, S4, and S6 and Tables S3 and S4.

ure 2C) (Salis et al., 2009). Translational autoregulation (Nomura et al., 1984), coupling (Baughman and Nomura, 1983), or specific RNA secondary structures (McCarthy and Gualerzi, 1990) are factors that could contribute to precise tuning of synthesis rates. Our discovery of proportional synthesis in polycistronic messages should help guide efforts to dissect the molecular mechanism of translation initiation quantitatively, as well as aid the precise engineering of synthetic biological networks.

The use of translational control and polycistronic operons to achieve proportional

synthesis rates are generally determined by the frequency of translation initiation (Andersson and Kurland, 1990). However, our current understanding of what determines translation initiation rates is highly incomplete because existing models for either the strength of ribosome-binding site or the Shine-Dalgarno sequence alone do not predict proportional synthesis (Fig-

proportional synthesis has important potential advantages. In particular, setting the ratios of subunit expression levels exclusively at the translational level greatly simplifies transcriptional regulation; the cell needs only to control the overall expression of the complex, and not the relative amounts within the complex. Additionally, sharing the same polycistronic mRNA reduces

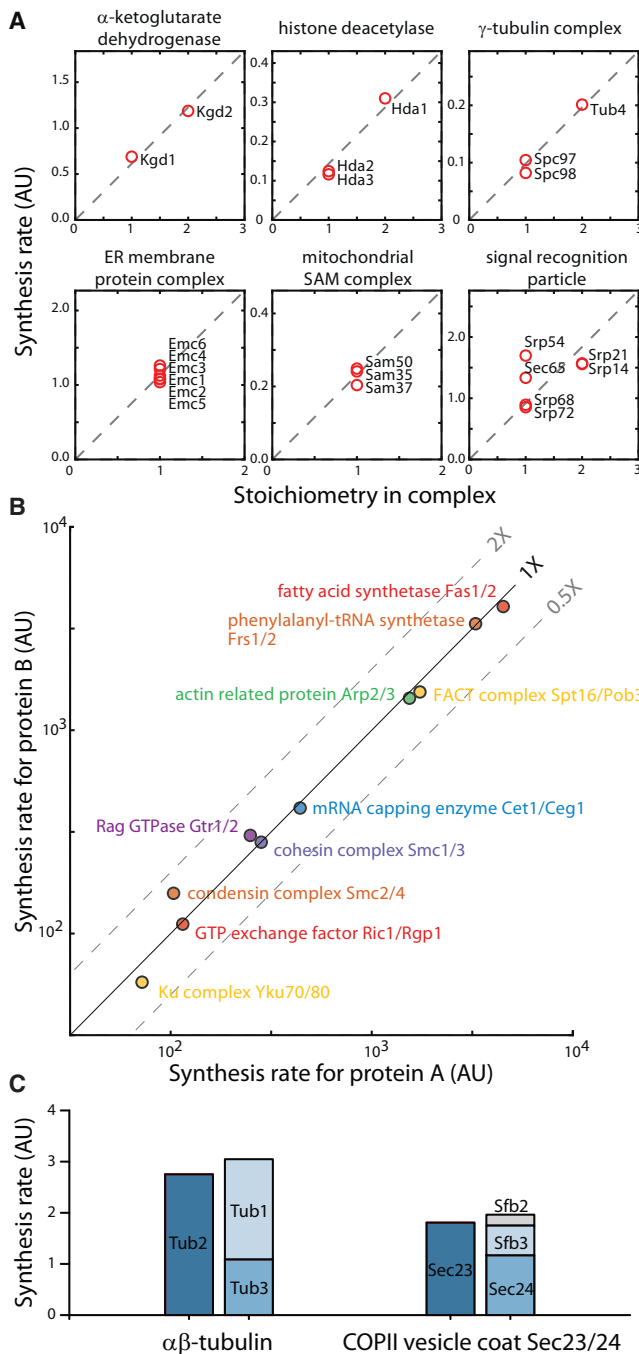


Figure 3. Proportional Synthesis for Complexes in Yeast

(A) Proportional synthesis for multiprotein complexes in *S. cerevisiae*. Synthesis rates are plotted as a function of the subunit stoichiometry for complexes with more than two subunits. For the signal recognition particle, four subunits (Srp14/Srp21/Srp68/Srp72) are synthesized according to their stoichiometry, and the other two are exceptions.

(B) Proportional synthesis for heterodimeric complexes in *S. cerevisiae*. Each complex is plotted for the synthesis rate of the two subunits.

(C) Proportional synthesis for complexes with paralogous subunits. For each complex, the subunits that can substitute each other are plotted in the same column.

stochastic imbalance among components of the complex. Because transcription originates from a single gene locus and is thus inherently noisy (Li and Xie, 2011), the ratio of proteins encoded on different mRNAs would be subject to much higher noise levels (Elowitz et al., 2002; Swain, 2004). The use of polycistronic mRNAs circumvents this issue, but translational tuning becomes necessary to achieve different expression levels.

Evidence for Proportional Synthesis in Budding Yeast

We found evidence that the budding yeast *S. cerevisiae* also exhibits tightly controlled synthesis of stably associated protein complexes, as indicated by our analysis of a subset of highly characterized complexes (Figures 3A and 3B). Genomic duplication events in *S. cerevisiae* have led to numerous paralogous genes, which in some cases can substitute for each other in multiprotein complexes. Interestingly, we found that proportional synthesis is maintained by tuning the synthesis rates for duplicated genes that encode the same subunit. For example, the two α -tubulin genes together are translated at a similar rate as the single β -tubulin gene (Figure 3C). Similarly, for the COPII Sec23/24 heterodimer, the production rate of Sec23 matches that of Sec24 and its two homologs (Sfb2 and Sfb3) combined (Figure 3C). A notable exception for proportional synthesis is the signal recognition particle, for which four subunits are translated at a 1:1:2:2 ratio, and the other two subunits are in excess (Figure 3A). It has also been shown that vertebrates produce uneven amounts of α - versus β -spectrin and immunoglobulin light chains versus heavy chains (Blikstad et al., 1983; Lehnert and Lodish, 1988; Shapiro et al., 1966). Understanding the rationale behind the unequal synthesis in these exceptions could provide insights into their physiological functions.

Yeasts must employ distinct mechanisms to achieve proportional synthesis because subunits are encoded on different mRNAs in eukaryotes. For example, the dynamics of nuclear localization of TFs and their affinity to promoter sites could provide independent control for complex levels and subunit ratios (Cai et al., 2008). Given the fundamentally different molecular mechanisms for prokaryotic and eukaryotic expression, these observations argue that proportional synthesis is a result of convergent evolution that maximizes protein synthesis efficiency while minimizing the adverse effects of having uncomplexed subunits.

The broad use of proportional synthesis has important implications for the effect of aneuploidy. Most genes do not possess feedback mechanisms for controlling their expression levels (Springer et al., 2010). Thus, a sudden change in gene dosage would lead to a large imbalance of subunits (Papp et al., 2003). Because cells normally do not face large imbalances in the synthesis rate of multiprotein complexes, aneuploidy would lead to a strong challenge to the protein folding and chaperone networks, consistent with the findings of Amon and coworkers that general proteotoxic stress is a hallmark of aneuploidy (Oromendia et al., 2012; Torres et al., 2008).

Taken together, our findings argue that the relative expression of members of multiprotein complexes is primarily determined at the synthesis level and that targeted degradation of excess subunits is a secondary layer of control. Indeed, components of multiprotein assemblies whose uncomplexed subunits have been

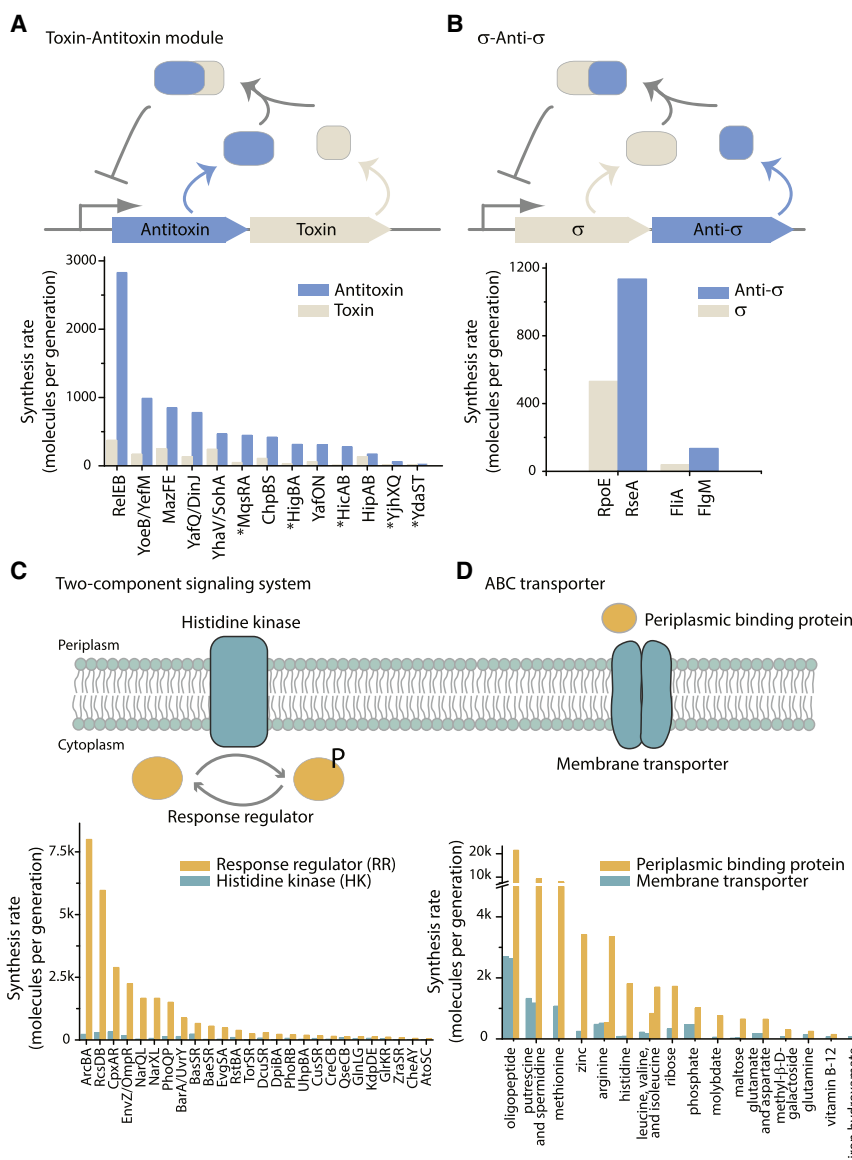


Figure 4. Hierarchical Expression for Functional Modules

(A) Synthesis rates for TA modules. *E. coli* contains 12 type II TA systems that are each expressed from a polycistronic mRNA. (The order of genes differs among systems.) The antitoxin protein binds to and inhibits the toxin protein, while repressing its own transcription. The synthesis rates for each system are plotted (bottom). Modules with the toxin gene preceding the antitoxin gene in the operon are marked with an asterisk.

(B) Synthesis rates for σ -anti- σ modules. The anti- σ binds to and inhibits the σ , preventing transcription from the promoter driven by the corresponding σ . The synthesis rates for each system are plotted (bottom).

(C) Synthesis rates for two-component signaling systems. Bacterial two-component signaling system consists of a membrane-bound HK and the cognate RR. The synthesis rates for 26 two-component systems in *E. coli* are plotted (bottom).

(D) Synthesis rates for ABC transporters. An ABC transporter consists of a core membrane transporter, an ATP-binding domain, and the corresponding periplasmic-binding proteins. The synthesis rates for each transporter are plotted (bottom).

shown to be degraded, including the ribosomal L8 complex and the SecYEG translocon in *E. coli* and Fas1/2 in *S. cerevisiae*, also show proportional synthesis (Akiyama et al., 1996; Petersen, 1990; Schüller et al., 1992).

Hierarchical Expression of Functional Modules

Stable protein complexes are only one of a wide range of functional modules that are organized into operons in bacteria, leading us to ask whether translational control also sets expression of other types of functional modules. Because our genome-wide ribosome-profiling data set covers many different modules in the same functional class, we can use our data to identify common expression pattern strategies that are selected through evolution. Our studies of several different modules identified a second pattern: hierarchical expression, in which components are differentially expressed according to their hierarchical role.

Bacterial toxin-antitoxin (TA) modules are widely utilized two-gene systems that control cellular survival (Yamaguchi et al., 2011). The role of antitoxin is to bind to and inhibit its cognate toxin. *E. coli* contains at least 12 type II TA systems, each consisting of a toxin protein and an antitoxin protein in a bicistronic operon (Yamaguchi et al., 2011). For every well-characterized type II TA system, we found that the antitoxin is synthesized at a much higher rate than the toxin (Figure 4A), which would allow *E. coli* to produce a sufficient amount of antitoxin to avoid triggering cell death or growth arrest during unstressed growth. The hierarchical expression between antitoxin and toxin is irrespective of their relative order in the operon (Figure 4A). Because most toxins target global translation, the translational control observed for hierarchical expression of TA modules may provide insight into how the system switches to a toxin-dominated state via translational feedback—a central question in antibiotic persistence (Gerdes and Maisonneuve, 2012).

σ /anti- σ modules are conceptually similar to TA modules. Both are usually encoded in the same operon, and anti- σ inhibits the transcriptional activity of the σ by direct binding. Interestingly, anti- σ s, like antitoxins, are produced in excess compared to σ s (Figure 4B). In both cases, the uncomplexed antagonists (antitoxins and anti- σ s) are also subject to regulated degradation (Ades et al., 1999; Yamaguchi et al., 2011). Thus, the hierarchical expression would not be evident by measuring

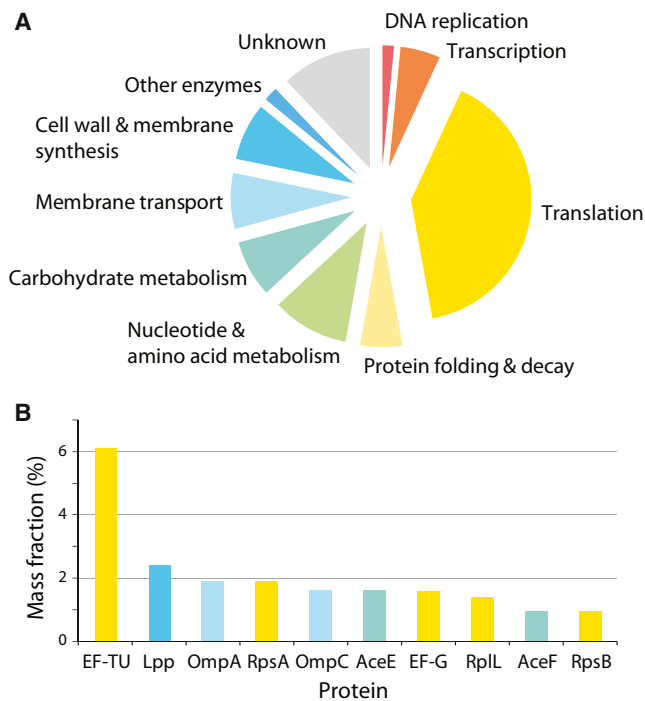


Figure 5. Composition of the *E. coli* Proteome

(A) Breakdown of the proteome by functions. The mass fraction of the proteome that is devoted to specific biological functions is plotted as a pie chart. The copy numbers were estimated for *E. coli* grown in rich defined medium (Experimental Procedures).

(B) Ten proteins with the largest mass fraction in the proteome. The color used for each protein corresponds to the biological function indicated in (A).

protein levels, even though cells ensure an excess of inhibitor during synthesis.

Translationally controlled hierarchical expression appears to be common for a diverse range of functional modules. ATP-binding cassette (ABC) transporters are comprised of core transmembrane proteins and corresponding substrate-binding periplasmic proteins. Whereas the core membrane complex components follow the proportional synthesis principle elucidated above (Figures 2B and 2C), we found that the periplasmic-binding proteins are always in large excess (Figure 4D), suggesting that substrate binding is slower than transport across the membrane. Two-component signaling systems, consisting of a histidine kinase (HK) and its substrate, a response regulator (RR), also exhibit hierarchical translation. For each of the 26 two-component systems in *E. coli*, the substrate is synthesized at a much higher level than the kinase (Figure 4C). Using mathematical modeling and experimental validation, it has been demonstrated that a large excess of a RR relative to an HK promotes robustness against variations in RR and HK levels (Batchelor and Goulian, 2003; Shinar et al., 2007). Here, we show that this strategy is universally employed for all two-component systems.

Taken together, these results show that hierarchical expression within operons is a key design principle for many diverse functional modules. As illustrated in the four examples above, the same hierarchy of expression levels is repetitively used for

the same type of module, pointing to a common quantitative property that is critical for the execution of each task. The examples here are certainly an incomplete list; more quantitative design principles could be uncovered by identifying commonalities among similar systems in such genome-wide data sets.

Bacterial Proteome Composition

Because the large majority of proteins are stable in *E. coli* (Larabee et al., 1980), our protein synthesis rate data provide a comprehensive view of proteome composition, allowing us to probe how cells allocate resources (Figure 5). By far, the largest fraction of the protein synthesis capacity is dedicated to making the machinery needed for further translation (41% for growth in rich media and 21% in minimal media), whereas transcription-related proteins account for only 5%. This disparity illustrates the importance of understanding the translational control systems that allow cells to allocate their translational capacity. The ability to monitor the partitioning of protein synthesis capacity under different conditions will provide a critical tool for quantitative characterization of cellular physiology.

The expression level of every protein in the cell is subject to two opposing constraints: the requirement of its function, and the cost associated with having an excess that consumes limited resources, such as protein synthesis capacity, quality control machineries, and space (Dekel and Alon, 2005). Our data set opens up the possibility of broadly investigating how these competing constraints govern protein expression levels. We select two specific cellular functions (TFs and Met biosynthesis) for further study.

Copy Numbers of TFs Reveal Their Mode of Action

The bacterial chromosome is densely covered with TFs that bind DNA both specifically and nonspecifically (Li et al., 2009). The crowded space on DNA imposes constraints on the abundance of TFs because overcrowding by nonspecifically associated DNA-binding proteins could drastically reduce the overall binding kinetics (Hammar et al., 2012; Li et al., 2009). Thus, although higher concentrations of any given TF would allow it to find its cognate DNA sites more rapidly (von Hippel, 2007), too many TFs in total would mask binding sites. Based on our protein-abundance estimates, we found that the average distance between DNA-binding proteins is only ~36 bp on the *E. coli* chromosome (assuming that most DNA-binding proteins are associated with DNA nonspecifically and are randomly distributed throughout the genome; see Extended Experimental Procedures), which is close to the theoretically optimal density for rapid binding (Li et al., 2009). How cells allocate the limited space on DNA to maximize rapid regulation by each TF remained obscure.

Our data indicate that the ~200 well-characterized TFs in *E. coli* show a wide variation in level—more than 60% of the TFs are found to have an upper bound of fewer than 100 monomers per genome equivalent (Figures 6A and 6B). A low copy number for a TF implies a slow association rate to DNA, which could lead to slow transcriptional responses (Winter et al., 1981). For example, single-molecule imaging in vivo previously revealed that it takes 6 min for one Lac repressor to find a single binding site in a cell (Elf et al., 2007). Compared to the cell doubling time, which can be as short as 20 min, the binding

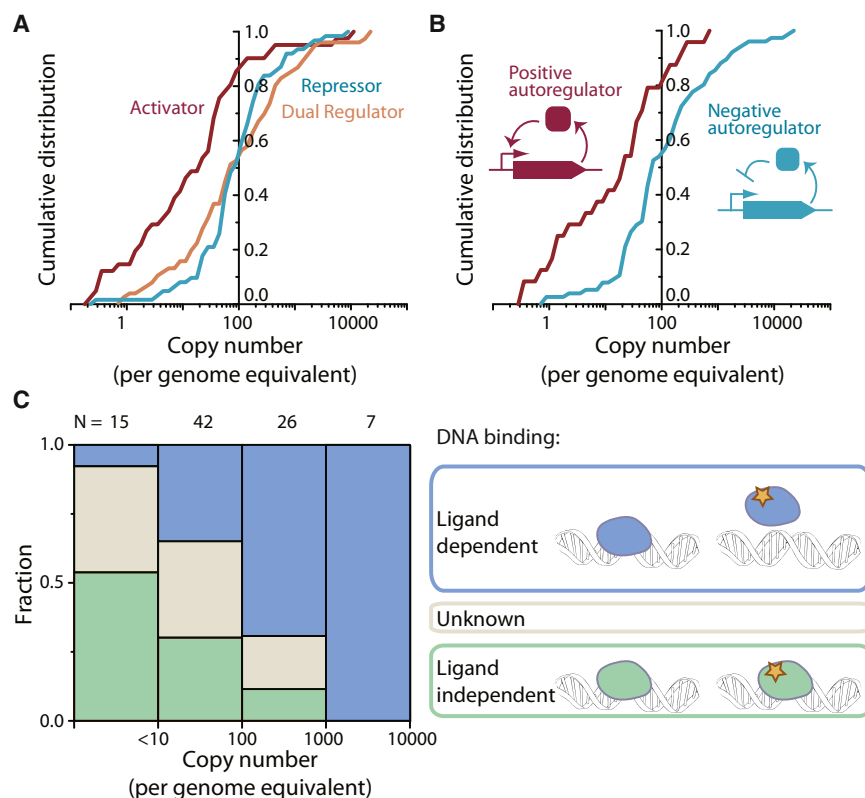


Figure 6. Abundance of TFs

(A) Cumulative distribution of abundance for transcriptional activators, repressors, and dual regulators. The cumulative distribution for each class of TF is plotted as a function of the copy number per genome equivalent.

(B) Cumulative distribution of abundance for autoregulators. The cumulative distributions for positive and negative autoregulators are plotted as a function of the copy number per genome equivalent.

(C) Ligand dependence of target binding. Among TFs whose abundance falls into a given range, the fraction that binds to the target site in a ligand-dependent way is shown in blue, and the fraction that binds to the target site independent of ligands is shown in green. The number of TFs analyzed is indicated above each bin.

See also [Table S5](#).

kinetics for a low copy number TF would make it difficult to achieve timely regulation. This can be circumvented with the use of TFs that are always bound to their target but whose ability to recruit RNA polymerase depends on the presence of ligands because the kinetics of regulation would be determined by diffusion of the small ligand rather than by diffusion of the bulky and far less abundant protein. We therefore hypothesize that the low copy number TFs have evolved to bind to DNA independent of their activity.

To test this hypothesis, we mined the literature for the biochemical properties of 102 TFs in *E. coli* ([Table S5](#)). We found that abundant TFs bind to DNA only in response to ligands ([Figure 6C](#)). By contrast, the large majority of low abundance TFs bind to the target sites independent of the corresponding ligands ([Figure 6C](#)). Therefore, cells optimize the limited space on DNA and the need for rapid regulation by requiring that TFs with low abundance always bind to their target sites. This mode of DNA binding for low copy number TFs also supports the model that TFs have evolved to occupy their target sites in native environments ([Savageau, 1977](#); [Shinar et al., 2006](#)). This class of TFs can be exploited to build transcriptional circuits with fast response time without incurring extra synthesis cost and nonspecific interactions. A potential downside, however, is increased gene expression noise due to stochastic TF dissociation.

Precise Control of Enzyme Production Required for Met Biosynthesis

The expression of metabolic enzymes similarly faces two constraints: the requirement for function, and the cost of synthesis.

Metabolic control analysis suggests that enzymes are generally made in excess amounts, such that small changes in the level for each enzyme have moderate effects on the output ([Fell, 1997](#)). On the other hand, the pools of bacterial enzymes in related metabolic pathways are strictly dependent on growth rates ([You et al., 2013](#)), arguing for precise control of expression based on cellular need. Thus, the principal determinant of expression remained obscure. Here, we show that our quantification of the proteome composition makes it possible to globally analyze the relationship between the levels of metabolic enzymes and their actual reaction fluxes.

We focused on the well-characterized L-Met biosynthetic pathway for *E. coli* grown in media devoid of Met. We first calculated the cellular demand for this pathway ($31,000 \text{ s}^{-1}$ Met per cell), i.e., the rate of Met consumption by protein synthesis, by summing up the absolute rates of protein synthesis we determined for each protein multiplied by the number of Met residues in that protein. The other major pathway that consumes Met, which is the synthesis of S-adenosyl-L-Met, was estimated to contribute to a small fraction of the overall flux ([Feist et al., 2007](#)) (see also [Extended Experimental Procedures](#)). We then compared the rate of Met consumption with the maximum velocity (V_{\max}) for its biosynthetic pathway. For each reaction in the pathway, we calculated V_{\max} by multiplying the enzyme abundance we determined by its published turnover number (k_{cat}) ([Schomburg et al., 2002](#)). The V_{\max} varies by more than one order of magnitude among the reactions in Met biosynthesis, suggesting that most reactions do not operate at saturating substrate concentration. The last step that is catalyzed by MetE has among the smallest V_{\max} ([Figure 7A](#)), suggesting that it may be a bottleneck in this pathway. Remarkably, we found that the maximal Met production rate allowed by MetE (V_{\max} , $34,000 \text{ s}^{-1}$ per cell) matches the Met consumption rate. Therefore, under these growth conditions, MetE-catalyzed conversion

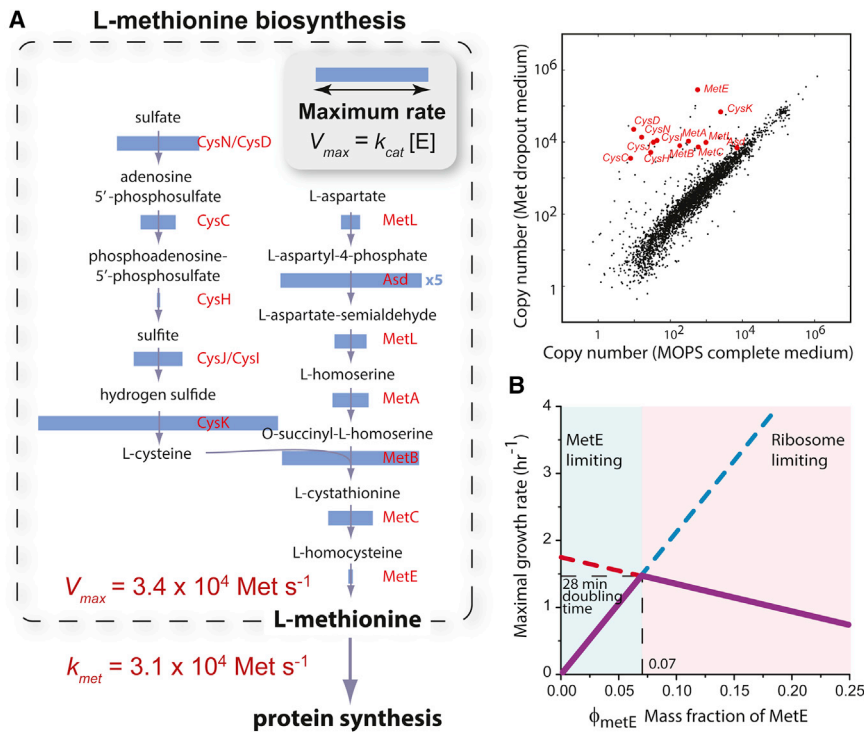


Figure 7. Quantitative Analysis of the Met Biosynthesis Pathway

(A) Maximal reaction rates for the intermediate steps. For each step of the pathway, the V_{max} , inferred from the enzyme abundance in vivo and the k_{cat} measured in vitro, is shown as the width of the blue bar. The last step that is catalyzed by the enzyme MetE has a V_{max} of 34,000 Met/s/cell, whereas the flux of Met into protein synthesis is 31,000 Met/s/cell. The scatterplot on the right shows upregulation of these enzymes in media without Met. MOPS, 3-(N-morpholino)propane-sulfonic acid.

(B) Model predicting the optimal MetE level. In a model that considers the cost and benefit of MetE expression, the maximal growth rate is plotted as a function of the mass fraction of MetE in the proteome. The cost due to competition with new ribosome synthesis is shown in red, and the benefit from increased Met flux is shown in blue. The maximal growth rate is highest (28 min) when the mass fraction of MetE is $\sim 7\%$. This prediction agrees with experimental results. See also Figure S5.

of L-homocysteine to L-Met is a bottleneck step that operates at maximal velocity with saturating substrate concentration.

Given that Met biosynthesis by MetE is limiting the overall rate of protein synthesis, why do cells not simply make more MetE protein? MetE is a large and slow enzyme, whose production consumes $\sim 8\%$ of the total protein synthesis capacity in media devoid of Met. We investigated whether the cost of increasing MetE production further would outweigh its benefit. To do so, we constructed a simple analytical model for the effect of MetE expression on growth rate (Figure 7B; Experimental Procedures). The model considers the cost and benefit of MetE synthesis independently and allows us to evaluate the level of synthesis where the trade-off between cost and benefit is optimized. The benefit of producing MetE arises from our observation that it is a bottleneck for the Met supply for protein synthesis. Hence, devoting more protein synthesis capacity to MetE increases growth rate linearly (Experimental Procedures). The cost of producing excess proteins, independent of their function, comes from competition for ribosomes—an effect that has been widely studied for *E. coli* (Dekel and Alon, 2005; Dong et al., 1995; Scott et al., 2010). To evaluate this cost, we used the well-validated numerical relationship described by Scott et al. (2010).

These two constraints predict that the fastest growth rate, a 28 min doubling time, is achieved at an optimal MetE level of 7% of protein synthesis capacity (Figure 7B). Remarkably, these predictions were in close agreement with the actual values observed for cells lacking Met: 27 min doubling time and 8% of protein synthesis capacity devoted to MetE. We verified experimentally that both decrease and increase in MetE production lead to slower growth (Figure S5). Therefore, the expression of the key enzyme MetE is accurately tuned to allow the highest possible growth

rate. Furthermore, the cost of expressing MetE is the main determinant for the slower growth rate when Met is limiting.

Our quantitative analysis of the Met pathway revealed a bottleneck step and its relationship to fitness. The same approach should be applicable for a broad range of cellular and engineered metabolic pathways, for which the control points are still largely unknown. In addition, the global analysis of maximum reaction velocity (V_{max}) can be used in concert with flux balance analysis (Price et al., 2004; Schuetz et al., 2012) to identify possible routes of metabolic flux at a given condition. More broadly, the global quantification of absolute enzyme concentration provides a transformative tool for studying cellular metabolism.

DISCUSSION

We illustrate here the capacity to measure absolute synthesis rates for cellular proteins and its utility for deciphering the logic behind the design principles of biological networks. We identify the rules underlying the observed synthesis rates for many distinct classes of proteins. These include proportional synthesis for multiprotein complexes and hierarchical expression for common functional modules, both of which are made possible by finely tuned rates of translation initiation. We anticipate that there are many more principles embedded in this and similar data sets that will both elucidate the regime in which biochemical reactions operate, and provide a foundation for rational design of synthetic biological systems.

Our genome-wide data set on protein synthesis rates also allows in-depth analysis of how cells optimize the use of limited resources. Specifically, these data revealed strategies for allocating limited space on DNA and limited protein synthesis capacity—TFs can be kept at low abundances without kinetic penalties by prebinding to target sites, and the synthesis rate

of a key enzyme that limits metabolic flux in the Met biosynthetic pathway is optimized to achieve a maximal growth rate. Limited resources of various kinds pose constant challenges to all cells. Our approach reveals how the translational capacity of a cell is allocated in the face of these challenges, greatly expanding our ability to perform systems level analyses that were previously limited to selected proteins and pathways.

Although our studies illustrate the role of precisely tuned protein synthesis rates in bacteria, our knowledge of how this translational control is achieved remains highly limited. Understanding the control of translation initiation is both of fundamental importance and a prerequisite for quantitative design in synthetic biology. Yet, our current approaches for predicting translation rates, based on the strength of Shine-Dalgarno site and computed RNA structure (Salis et al., 2009), fail to accurately account for the observed differences in translation initiation rates (Figure S6). Empirical measures of mRNA structures as they exist in the cell, in combination with our measures of translation efficiency (Table S4), could be a key tool in addressing this deficiency.

Although we focus on bacterial cells in this work, our approach to globally measure absolute protein synthesis rates has broader applicability. Any species that is amenable to ribosome profiling and has an annotated genome can be subject to this line of investigation; the growing list currently includes both Gram-negative and Gram-positive bacteria, budding yeast, nematodes, fruit fly, zebrafish, and mammals. For eukaryotes and multicellular organisms, our approach will likely reveal a distinct set of principles and constraints for optimizing the allocation of biosynthetic capacities. Furthermore, the breakdown of these principles under stress conditions, such as aneuploidy and temperature and chemical shock, will provide critical insight into the modes of failure and their rescue mechanisms.

EXPERIMENTAL PROCEDURES

Ribosome Profiling

Bacterial cells grown in specified liquid media were harvested by rapid filtration followed by flash freezing in liquid nitrogen. Ribosome-protected mRNA footprints were extracted from pulverized lysates as previously described (Li et al., 2012; Oh et al., 2011). Different from previous procedures, a wider range of mRNA footprint sizes (~15–45 nt long) was selected on a denaturing polyacrylamide gel. The mRNA fragments were converted to a cDNA library as previously described (see Extended Experimental Procedures) (Ingolia et al., 2009; Li et al., 2012; Oh et al., 2011). Deep sequencing was performed by Illumina HiSeq 2000.

Analysis for Absolute Synthesis Rates

Counts of ribosome footprints for each gene were first corrected for the elevated density toward the start codon. A metagene analysis for the relative density as a function of the distance to start codons was used as a calibration. The resulting counts were corrected for the elevated ribosome density downstream from internal Shine-Dalgarno sequences. For each position on the gene, the affinity of the upstream hexameric sequence to the anti-Shine-Dalgarno sequence was used to calibrate the distance-corrected counts (Li et al., 2012). The calibration curve was obtained empirically by fitting the observed average ribosome occupancy of hexameric sequences as a function of the hybridization energy to the anti-Shine-Dalgarno sequence. The resulting ribosome density was averaged within the gene body, excluding the first five and the last five codons.

The relative ribosome density was converted to absolute protein synthesis rates using the total weight of cellular protein. The relative synthesis rate of a protein, as measured by its corrected ribosome density compared to that of

all proteins, was multiplied by the weight of total proteins per cell—a proxy for the amount of proteins synthesized in a cell cycle. The weight of total proteins per cell was estimated by dividing the amount of proteins per unit volume of cell culture, which was measured using the Lowry method with BSA as standard after trichloroacetic acid precipitation, by the number of cells per unit volume, which was measured by counting colony-forming units after serial dilution. The absolute synthesis rates listed in Table S1 are also available through PortEco (Hu et al., 2014).

Model for Cost and Benefit of MetE

In order to understand the amount of MetE expressed in the medium without Met, we constructed a quantitative model to predict the optimal level of MetE and growth rate. The model considers the cost and benefit of MetE synthesis on growth rate. The cost function is based on previous observations that synthesis of excess proteins competes with that of new ribosomal proteins, which in turn leads to slower growth rate (Scott et al., 2010). Based on the work by Scott et al. (2010), this relationship is $\lambda = \lambda_0(1 - ((\phi_{m/c} + \phi_E)/\phi_C))$, where λ is the growth rate, λ_0 is the growth rate when Met is not limiting, ϕ_E is the mass fraction of MetE, $\phi_{m/c}$ is the mass fraction of all other enzymes in the Met and cysteine biosynthetic pathways, and ϕ_C is the phenomenological fitting parameter that was established in their work. The benefit function is based on our observation that the level of MetE determines that rate of Met synthesis and its consumption by protein synthesis: $N_E k_{cat} = f_{met} N_R k_e$. N_E , N_R are the numbers of MetE and translation ribosome, respectively. k_{cat} , k_e are the k_{cat} of MetE and translation elongation rate, respectively. f_{met} is the fraction of translated codons that encodes Met. Rewriting this equation using ϕ_E and λ gives $\lambda = k_{cat} \phi_E / f_{met} l_E$, where l_E is the number of amino acid residues in MetE. These two functions relating the growth rate and the mass fraction of MetE are plotted in Figure 7C.

ACCESSION NUMBERS

Data are available at Gene Expression Omnibus with accession number GSE53767.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.02.033>.

ACKNOWLEDGMENTS

We thank L. Qi for providing material for CRISPRi knockdown; R. Milo, K.C. Huang, J. Elf, J. Dunn, G. Brar, O. Brandman, C. Jan, J. Rabinowitz, and members of the J.S.W. and C.G. labs for discussions; and P. Choi and H. Chen for critical reading of the manuscript. We also thank J. Lund and E. Chow for help on sequencing and C. Reiger and M. DeVera for administrative support. This research was supported by the Helen Hay Whitney Foundation (to G.-W.L.), NIH Pathway to Independence Award (GM105913 to G.-W.L.), NIH Center for RNA Systems Biology (to J.S.W.), and Howard Hughes Medical Institute (to J.S.W.).

Received: September 26, 2013

Revised: December 31, 2013

Accepted: February 11, 2014

Published: April 24, 2014

REFERENCES

- Ades, S.E., Connolly, L.E., Alba, B.M., and Gross, C.A. (1999). The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes Dev.* 13, 2449–2461.
- Akiyama, Y., Kihara, A., Tokuda, H., and Ito, K. (1996). FtsH (HflB) is an ATP-dependent protease selectively acting on SecY and some other membrane proteins. *J. Biol. Chem.* 271, 31196–31201.

- Alon, U., Surette, M.G., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. *Nature* 397, 168–171.
- Andersson, S.G., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54, 198–210.
- Barkai, N., and Shilo, B.Z. (2007). Variability and robustness in biomolecular systems. *Mol. Cell* 28, 755–760.
- Batchelor, E., and Goulian, M. (2003). Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proc. Natl. Acad. Sci. USA* 100, 691–696.
- Baughman, G., and Nomura, M. (1983). Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. *Cell* 34, 979–988.
- Blikstad, I., Nelson, W.J., Moon, R.T., and Lazarides, E. (1983). Synthesis and assembly of spectrin during avian erythropoiesis: stoichiometric assembly but unequal synthesis of alpha and beta spectrin. *Cell* 32, 1081–1091.
- Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C.C., Li, G.W., Zhou, S., King, D., Shen, P.S., Weibezahn, J., et al. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* 151, 1042–1054.
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557.
- Brusilow, W.S., Klionsky, D.J., and Simoni, R.D. (1982). Differential polypeptide synthesis of the proton-translocating ATPase of *Escherichia coli*. *J. Bacteriol.* 151, 1363–1371.
- Buttgereit, F., and Brand, M.D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *Biochem. J.* 312, 163–167.
- Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455, 485–490.
- Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436, 588–592.
- Dennis, P.P. (1974). In vivo stability, maturation and relative differential synthesis rates of individual ribosomal proteins in *Escherichia coli* B/r. *J. Mol. Biol.* 88, 25–41.
- Dong, H., Nilsson, L., and Kurland, C.G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* 177, 1497–1504.
- Elf, J., Li, G.W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316, 1191–1194.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186.
- Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., and Palsson, B.O. (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.
- Fell, D. (1997). *Understanding the Control of Metabolism* (London: Portland Press).
- Gerdes, K., and Maisonneuve, E. (2012). Bacterial persistence and toxin-antitoxin loci. *Annu. Rev. Microbiol.* 66, 103–123.
- Grossman, A.D., Straus, D.B., Walter, W.A., and Gross, C.A. (1987). Sigma 32 synthesis can regulate the synthesis of heat shock proteins in *Escherichia coli*. *Genes Dev.* 1, 179–184.
- Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G., and Elf, J. (2012). The lac repressor displays facilitated diffusion in living cells. *Science* 336, 1595–1598.
- Hart, Y., Madar, D., Yuan, J., Bren, A., Mayo, A.E., Rabinowitz, J.D., and Alon, U. (2011). Robust control of nitrogen assimilation by a bifunctional enzyme in *E. coli*. *Mol. Cell* 41, 117–127.
- Hu, J.C., Sherlock, G., Siegele, D.A., Aleksander, S.A., Ball, C.A., Demeter, J., Gouni, S., Holland, T.A., Karp, P.D., Lewis, J.E., et al. (2014). PortEco: a resource for exploring bacterial biology through high-throughput data and analysis tools. *Nucleic Acids Res.* 42 (Database issue), D677–D684.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550.
- Larrabee, K.L., Phillips, J.O., Williams, G.J., and Larrabee, A.R. (1980). The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J. Biol. Chem.* 255, 4125–4130.
- Lehnert, M.E., and Lodish, H.F. (1988). Unequal synthesis and differential degradation of alpha and beta spectrin during murine erythroid differentiation. *J. Cell Biol.* 107, 413–426.
- Lemaux, P.G., Herendeen, S.L., Bloch, P.L., and Neidhardt, F.C. (1978). Transient rates of synthesis of individual polypeptides in *E. coli* following temperature shifts. *Cell* 13, 427–434.
- Li, G.W., and Xie, X.S. (2011). Central dogma at the single-molecule level in living cells. *Nature* 475, 308–315.
- Li, G.W., Berg, O.G., and Elf, J. (2009). Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* 5, 294–297.
- Li, G.W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538–541.
- McCarthy, J.E., and Gualerzi, C. (1990). Translational control of prokaryotic gene expression. *Trends Genet.* 6, 78–85.
- Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. *J. Bacteriol.* 119, 736–747.
- Nomura, M., Gourse, R., and Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annu. Rev. Biochem.* 53, 75–117.
- Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., et al. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* 147, 1295–1308.
- Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* 26, 2696–2708.
- Papp, B., Pál, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197.
- Petersen, C. (1990). *Escherichia coli* ribosomal protein L10 is rapidly degraded when synthesized in excess of ribosomal protein L7/L12. *J. Bacteriol.* 172, 431–436.
- Price, N.D., Reed, J.L., and Palsson, B.O. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–897.
- Quax, T.E., Wolf, Y.I., Koehorst, J.J., Wurtzel, O., van der Oost, R., Ran, W., Blombach, F., Makarova, K.S., Brouns, S.J., Forster, A.C., et al. (2013). Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep.* 4, 938–944.
- Rigel, N.W., Ricci, D.P., and Silhavy, T.J. (2013). Conformation-specific labeling of BamA and suppressor analysis suggest a cyclic mechanism for β -barrel assembly in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 110, 5151–5156.
- Russell, J.B., and Cook, G.M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.* 59, 48–62.
- Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950.
- Savageau, M.A. (1977). Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl. Acad. Sci. USA* 74, 5647–5651.
- Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49.

- Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012). Multidimensional optimality of microbial metabolism. *Science* 336, 601–604.
- Schüller, H.J., Förtsch, B., Rautenstrauss, B., Wolf, D.H., and Schweizer, E. (1992). Differential proteolytic sensitivity of yeast fatty acid synthetase subunits alpha and beta contributing to a balanced ratio of both fatty acid synthetase components. *Eur. J. Biochem.* 203, 607–614.
- Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9, 205–209.
- Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. *Science* 330, 1099–1102.
- Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63.
- Shapiro, A.L., Scharff, M.D., Maizel, J.V., and Uhr, J.W. (1966). Synthesis of excess light chains of gamma globulin by rabbit lymph node cells. *Nature* 211, 243–245.
- Shemorry, A., Hwang, C.S., and Varshavsky, A. (2013). Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* 50, 540–551.
- Shinar, G., Dekel, E., Tlusty, T., and Alon, U. (2006). Rules for biological regulation based on error minimization. *Proc. Natl. Acad. Sci. USA* 103, 3999–4004.
- Shinar, G., Milo, R., Martínez, M.R., and Alon, U. (2007). Input output robustness in simple bacterial signaling systems. *Proc. Natl. Acad. Sci. USA* 104, 19931–19935.
- Springer, M., Weissman, J.S., and Kirschner, M.W. (2010). A general lack of compensation for gene dosage in yeast. *Mol. Syst. Biol.* 6, 368.
- Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T., Hein, M.Y., Huang, S.X., Ma, M., Shen, B., Qian, S.B., Hengel, H., et al. (2012). Decoding human cytomegalovirus. *Science* 338, 1088–1093.
- Swain, P.S. (2004). Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.* 344, 965–976.
- Torres, E.M., Williams, B.R., and Amon, A. (2008). Aneuploidy: cells losing their balance. *Genetics* 179, 737–746.
- Tyedmers, J., Mogk, A., and Bukau, B. (2010). Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* 11, 777–788.
- von Dassow, G., Meir, E., Munro, E.M., and Odell, G.M. (2000). The segment polarity network is a robust developmental module. *Nature* 406, 188–192.
- von Hippel, P.H. (2007). From “simple” DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* 36, 79–105.
- Winter, R.B., Berg, O.G., and von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor—operator interaction: kinetic measurements and conclusions. *Biochemistry* 20, 6961–6977.
- Yamaguchi, Y., Park, J.H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet.* 45, 61–79.
- You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C.W., Wang, Y.P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* 500, 301–306.

EXTENDED EXPERIMENTAL PROCEDURES

Strain and Growth Conditions

E. coli K-12 strain MG1655 was used for this study. All cultures were based on MOPS media with 0.2% glucose (Teknova), with either full supplement (Neidhardt et al., 1974), full supplement without L-methionine, or no supplement. An overnight liquid culture was diluted 400-fold into 200 ml fresh media. The culture was kept in a 2.8-l flask at 37°C with aeration (180 rpm) until OD₆₀₀ reached 0.3. For the experiment at 10°C, the culture in M9 complete media with 0.2% glucose and amino acids except L-methionine was transferred from 37°C to 10°C with continuous shaking when OD₆₀₀ reached 0.12. Cells were harvested 50 hr later when OD₆₀₀ reached 0.43. The doubling time at 37°C is 21.5 ± 0.4 min in fully supplemented MOPS media, 26.5 ± 1.1 min in the methionine dropout medium, and 56.3 ± 0.5 min in minimal medium. The results presented in this work are based on MOPS complete media unless otherwise mentioned.

Ribosome Profiling

Bacterial ribosome profiling was performed as described in detail previously (Li et al., 2012; Oh et al., 2011) with the following modifications. 200 ml of cell culture was rapidly filtered at 37°C by passing through a nitrocellulose filter with 220 nm pore size (GE MicronSep). Cell pellets were rapidly collected using a prewarmed metal table crumber, flash frozen in liquid nitrogen, and combined with 650 µl of frozen droplets of lysis buffer (10 mM MgCl₂, 100 mM NH₄Cl, 20 mM Tris pH 8.0, 0.1% NP-40, 0.4% Triton X-100, 100 U/ml DNase I, 1 mM chloramphenicol). Cells and lysis buffer were pulverized in 10 ml canisters (Retsch) prechilled in liquid nitrogen using QIAGEN TissueLyser II (5 cycles of 3 min at 15 Hz). Pulverized lysate was thawed on ice and clarified by centrifugation at 20,000 rcf for 10 min at 4°C. 5 mM CaCl₂ was added to the clarified lysate containing 0.5 mg of RNA which was then digested with 750 U of micrococcal nuclease (Roche) at 25°C for 1 hr. The reaction was quenched by adding EGTA to 6 mM and moved on ice.

The monosome fraction, following nuclease digestion to create footprints, was collected using sucrose gradient and hot-phenol extraction. Ribosome-protected mRNA fragments were isolated by size excision on a denaturing polyacrylamide gel (15%, TBE-Urea, Invitrogen). Fragments with size ranging from 15 to 45 nucleotides were excised from the gel. The 3' end of footprints was dephosphorylated using 20 units of T4 polynucleotide kinase (New England Biolabs) at 37°C for one hour. Five picomoles of footprints were ligated to 1 µg of 5' adenylated and 3'-end blocked DNA oligo (5rApp/CTGTAGGCACCATCAAT/3 ddc, Integrated DNA Technologies) using truncated T4 RNA ligase 2 K277Q at 37°C for 2.5 hr. The ligated product was purified by size excision on a 10% TBE-Urea polyacrylamide gel (Invitrogen). cDNA was generated by reverse transcription using Superscript III (Invitrogen) at 50°C for 30 min with primer o225-Link1 (5phos/GATCGTCCGACTGTAGAACTCTGAACCTGTCGGTGGTCCCGTATCATT/iSp18/CACTCA/iSp18/CAAGCAGAAGACGGCATAACGAATTGATGGTGCCTACAG), and isolated by size excision on a 10% TBE-Urea polyacrylamide gel (Invitrogen).

Single-stranded cDNA was circularized using CirLigase (Epicenter) at 60°C for 2 hr. Ribosomal RNA fragments were removed using biotin-linked DNA oligos (5Biosg/TCATCTCCGGGGGTAGAGCACTGTTTCG, 5Biosg/GGCTAAACCATGCACCGAAGCTGCG GCAG, 5Biosg/AAGGCTGAGGCGTGATGACGAGGCACT, 5Biosg/CGGTGCTGAAGCAACAAATGCCCTGCTT) and MyOne Streptavidin C1 Dynabeads (Invitrogen). After being purified using isopropanol precipitation, the remaining cDNA was amplified using Phusion DNA polymerase (Finnzymes) with o231 primer (CAAGCAGAAGACGGCATAACGA) and indexing primers (AATGATACGGCGACC ACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNGCAGAGTTTCAGAGTTC). After 8-10 rounds of PCR amplification, the product was selected by size excision on a 8% TB polyacrylamide gel (Invitrogen).

Sequencing was performed on an Illumina HiSeq 2000. Bowtie v. 0.12.0 was used for sequence alignment to the reference genomes NC_000913.fna obtained from NCBI Reference Sequence Bank. The footprint reads with size between 20 to 42 nucleotides in length were mapped to the genome using the center-weighted approach; for each footprint read, the center residues that are at least 10 nucleotides away from either ends were given the same score, which is weighted by the length of the fragment. The data set is deposited in the Gene Expression Omnibus (GEO) under accession number GSE53767.

Ribosome profiling data for *S. cerevisiae* S288C were obtained from Brandman et al. (2012). For paralogous genes in Figure 3C, we used regions with at least one nucleotide difference within a 28-nucleotide window to calculate the synthesis rates.

mRNA Sequencing

Pulverized cell lysate described in the previous section was collected and RNA was extracted using hot phenol extraction. Ribosomal RNA was removed by subtractive hybridization using MICROExpress (Ambion). Small RNAs were removed using MEGAclear purification kit (Ambion). The remaining mRNA was fragmented using alkaline hydrolysis (50 µl of 44 mM NaHCO₃, 6 mM Na₂CO₃, 1 mM EDTA) at 95°C for 23 min. RNA was immediately precipitated. Fragment size between 20-40 nucleotides were selected on a denaturing polyacrylamide gel (15%, TBE-UREA, Invitrogen). These fragments were then ligated and converted to DNA as described in the previous section.

To estimate mRNA levels, we calculated the number of sequencing reads mapped to a gene, divided by the length of the gene to yield the number of reads corresponding to the message per thousand bases of message per million sequencing reads (RPKM). Translation efficiency was calculated by dividing the protein synthesis rate by the estimate for mRNA level. Table S5 lists mRNA level and translation efficiency in arbitrary units.

Average Ribosome Density and Correction for Translation Elongation Rate

Average ribosome density was calculated for reads mapped to the gene excluding the first and last five codons to remove effects of translation initiation and termination. A number of genes with unconventional translational events were treated differently, as described below. (1) For genes with translational frameshift (*prfB* and *dnaX*), only the density after the frameshift event was used. The *dnaX* gene uses ~50% frameshift to code for two stable proteins—the larger tau subunit and the smaller gamma subunit. The density for the tau subunit was subtracted from the density before the frameshift to give the density for the gamma subunit. (2) For selenoproteins (FdhF, FdoG, FdnG), we observe higher ribosome density before the codon for selenocysteine, suggesting that most ribosomes terminate at the selenocysteine codon, which is also a stop codon recognized by release factor 2. Only the density after the selenocysteine codon was used in our calculation. (3) For proteins translated without a stop codon, such as the alternative ribosome rescue factor (ArfA), only the density before the end of the transcript was considered. (4) For proteins translated with a known ribosome stalling site (SecM and TnaC), the ribosome density around the stalling site was not included in the average. (5) For proteins with nearly identical coding sequences, such as TufA and TufB, GadA and GadB, YnaE and YdfK, LdrA and LdrC, YbfD and YhhI, TfaR and TfaQ, RzoD and RzoR, PinR and PinQ, we considered the pair as the same protein and calculated the average ribosome density together.

To test whether the overall measure of ribosome density for the entire gene averages out local variation in elongation rates and sequencing biases, we compared the density for the first half and second half of the gene. Genes with at least 20 codons and 64 reads in either halves are included in the analysis. The result is shown in Figure 1A. The lack of bias toward the first half confirms the finding that there is little drop-off in ribosome density for most genes (Li et al., 2012; Oh et al., 2011). Genes with special translational events, as discussed above, were excluded from this analysis.

These small effects can be further corrected using the knowledge of the sequence features that drive variations in elongation rates. We first considered the elevated density observed for the first 50–100 codons (Oh et al., 2011). For each gene, we calculated the local ribosome density as a function of length, which was measured as the average ribosome density in a window of 50 codons relative to that in the first 50 codons. The median values of this function among all genes, except those with unconventional translation events and those with less than 128 reads mapped, were fitted with an exponential decay function with an offset. The fitting result was independent of the window size. The resulting function was used to adjust the ribosome density for all genes, similar to the method used by Ingolia et al. (2009), to remove the elevated density at the beginning of open reading frames.

We next correct for the elevated density due to ribosome pausing. We have previously reported that interactions between mRNA and the 16S rRNA cause ribosome pausing at internal Shine-Dalgarno (SD) sequences (Li et al., 2012). We have also found that the affinity between a hexanucleotide sequence and the ribosomal anti-Shine-Dalgarno (aSD) site strongly predicts the duration of ribosome pausing (Li et al., 2012). Here we used this information to correct for the effect of ribosome pausing. We fitted a linear function for the average ribosome occupancy downstream from a hexanucleotide sequence with respect to its affinity to the aSD site. This function is then used to adjust the ribosome occupancy at each position in each gene; at each position, the measured occupancy was divided by the expected pause duration based on the strongest hexanucleotide sequence at 6–11 bases upstream. The adjusted ribosome occupancy is no longer correlated with the SD-aSD interaction.

Finally, we remove the residual variations that are not accounted for using 90% Winsorization (Tukey, 1962). Namely, the top and bottom 5% of the ribosome occupancy for each gene were removed from the calculation for average ribosome density. The results before and after all three corrections listed here are shown in Figure S1. These corrections together only have moderate effects, as the difference between uncorrected and corrected density is typically below 11% (standard deviation of the mean).

Total Protein Measurement and the Conversion to Absolute Synthesis Rates

Ribosome profiling and the analyses above allow direct comparison of the relative synthesis rates among proteins. To obtain the absolute synthesis rate, we normalized the results by the total weight of proteins synthesized per cell cycle. Because the majority of the proteins in *E. coli* are long-lived compared to the generation time during rapid growth, the total weight of proteins synthesized per cell cycle can be approximated as the total weight of proteins per cell.

To measure the total weight of proteins per cell, we grew cells in the same way as those used for ribosome profiling experiments. When OD₆₀₀ reached the same level, we counted the number of cells per unit volume by serial dilution and plating on LB-agar plates. At the same time, 1 ml of culture harvested using centrifugation for 30 s. After removing 950 μ l supernatant, cells were resuspended and added to 950 μ l ice-cold PBS with 0.017% deoxycholate and on ice for 5 min. We then added 113 μ l of 100% trichloroacetic acid and incubate on ice overnight. Protein precipitation was collected by centrifugation at 20,000 rcf for 15 min at 4°C. The amount of proteins was quantified using the Lowry method with Peterson's modification (Sigma-Aldrich). To establish a standard curve, serial dilution of bovine serum albumin was made in the same culture media, and precipitated in the same way. The total weight of protein per cell (*P*) is calculated as the amount of protein per OD₆₀₀ per ml of culture divided by the number of cells per OD₆₀₀ per ml of culture. We measured 680 fg of protein per cell for culture in MOPS complete media, 450 fg per cell in MOPS complete media without methionine, and 238 fg per cell in MOPS glucose minimal media. These results on protein content and the corresponding estimate for the number of ribosomes (Table S1) at various growth rates are consistent with the estimate by Bremer and Dennis (1996).

To obtain the absolute synthesis rates, we first used the corrected ribosome density of each protein relative to that of all proteins to estimate its mass fraction in the proteome:

$$\phi_i = \frac{MW_i RD_i}{\sum_j MW_j RD_j},$$

where ϕ_i , MW_i , RD_i , are the mass fraction, molecular weight, and ribosome density of protein i , respectively. The synthesis rate for protein i (k_i) is given by

$$k_i = \frac{\phi_i P}{MW_i} = \frac{RD_i P}{\sum_j MW_j RD_j},$$

where k_i has the unit of molecules per generation. One generation is 21.5 ± 0.4 min in fully supplemented MOPS media. For stable proteins, k_i is also the copy number. The results are listed in [Table S1](#).

Literature Mining for Published Protein Copy Numbers

We validated our results by comparing with a list of protein copy numbers that have been individually characterized using classic assays. To compile such a list, we combined a community-based approach and computer-aided search. We consulted bacteriologists for relevant publications to their knowledge. We also utilized search engines such as PubMed and Google with relevant keywords. To obtain an unbiased list, every publication we could identify that measured specific protein abundance in *E. coli* was included. Results from other high-throughput studies were excluded in order to avoid bias toward any one method.

The published quantification used various strain backgrounds, growth media, temperature, and growth phases. If the same protein has been reported for multiple conditions, we chose the values that were measured in the condition closest to ours (MG1655, MOPS complete media with 0.2% glucose, 37°C, and exponential phase growth). Because most of the published quantification was based on cells with slower growth rates and lower cell mass, the protein copy numbers are in general slightly lower than our estimates in rich media. For copy numbers that were reported as molecules per genome equivalent, we multiplied the number by four to approximate our growth condition. The detailed protein information, strain background, media, temperature, growth phase, and PubMed ID for the original publication were listed in [Table S2](#).

Systematic Analysis of Complex Stoichiometry

To our knowledge, there is no curated database for multi-protein complexes and their stoichiometry. We systematically created a list based on the references available in the Ecocyc Database ([Keseler et al., 2013](#)). We first obtained a list of proteins that have been annotated as either “subunit” or “component.” We then inspected the literature to confirm whether each protein is a stoichiometric component of a larger complex. Because we set out to analyze the synthesis rates for obligate members of stable complexes, several criteria were used for the selection. (1) The subunit has no additional roles outside the complex that have been reported. Several ribosomal proteins do not meet this criterion (even though their synthesis rates are closely matched) because the free proteins function separately in translational repression or ribosome assembly. (2) The complex is formed in the default state, rather than as a response to signals. For example, the DNA repair enzyme UvrABC is excluded because the assembly depends on damaged DNA. (3) The stoichiometry of the complex has been well documented. We identified 64 multi-protein complexes that meet these criteria and are expressed in our data set (more than 128 sequenced reads per protein). The gene names and the corresponding stoichiometry is listed in [Table S3](#). For the budding yeast *S. cerevisiae*, we performed a small-scale analysis on multi-protein complexes using the same criteria.

Among the 64 complexes in *E. coli*, 47 have all the components encoded in the same operon, and their synthesis rates are shown in [Figure 2](#). The operon structure is based on the experimentally validated annotation in Ecocyc, and confirmed by our RNA-seq data. The rest of the complexes have members expressed from at least two different mRNAs, and their synthesis rates are shown in [Figure S3](#). The exceptions to proportional synthesis are also shown in [Figure S3C](#).

Predicted Translation Rates Using RBS Calculator

To calculate the predicted translation rates based on the model established by [Salis et al. \(2009\)](#), we used the RBS Calculator downloaded from <http://www.github.com/hsalis/Ribosome-Binding-Site-Calculator-v1.0>. For each gene, we used the nucleotide sequence from 35 bases upstream to 35 bases downstream from the translation start site as the input, and obtained the predicted strength of ribosome binding site, as well as predicted Shine-Dalgarno sequence and mRNA accessibility. For complexes with two equimolar subunits that are expressed from the same operon, the predicted fold-difference between the translation rates is $e^{-\beta(\Delta G_2 - \Delta G_1)}$, where $\beta = 0.45$ and ΔG_1 and ΔG_2 are the predicted strength of ribosome binding sites for gene 1 and gene 2, respectively.

DNA-Binding Properties of Transcription Factors

To test our hypothesis that low abundance transcription factors (TFs) always bind to their cognate sites independent of ligands, we mined the literature to determine the mode of DNA binding of TFs and compared with our results on TF copy numbers. *E. coli*

has > 400 annotated TFs, most of which are putative with little biochemical characterization. To focus on well-characterized TFs, we based our analysis on TFs that have been shown to regulate their own transcription. For each of these 102 TFs, we first identified the ligand that binds directly to regulate the activity of the TF. We then searched for evidence for whether the binding of ligands alter the ability to bind to the cognate DNA binding sites. TFs that do not have known ligands are not included in Figure 6. Table S4 lists the ligands, mode of DNA binding, and the PubMed ID for the reference that showed whether DNA-binding is affected by ligands.

To estimate the average spacing between DNA-binding proteins on the chromosome, we divided the length of the *E. coli* chromosome by the total amount of proteins (per genome equivalent) that are annotated with DNA binding activity. Proteins whose main function is not associated with DNA binding, such as alanyl-tRNA synthetase, are not considered in the calculation. The total number of DNA-binding proteins is divided by two because most of them bind to DNA as dimers. A large fraction of these proteins consists of nucleoid-associated proteins and the RNA polymerase (~70%). In this estimation, we assumed that the vast majority of these proteins are nonspecifically associated with DNA in vivo. Indeed, the concentrations for nucleoid-associated proteins and the RNA polymerase are higher than the reported nonspecific dissociation constants (Li et al., 2009), indicating that they are bound to the chromosome. This estimation gives ~56 base pairs center-to-center distance between neighboring DNA-binding proteins. The average spacing between adjacent proteins is 36 base pairs if the average footprint size on DNA is assumed to be 20 base pairs (Li et al., 2009).

Quantitative Analysis of the Methionine Biosynthesis Pathway

The description of the pathway and the corresponding enzymes were obtained from the Ecocyc Database (Keseler et al., 2013). The methionine synthesis pathway acquires the backbone of the amino acid from aspartate, and sulfur from cysteine. Although both aspartate and cysteine are supplied in our media, we noticed that the enzyme involved in cysteine biosynthesis were also induced. We also noticed that cysteine codons, in addition to the methionine codon, have slightly elevated ribosome occupancy, suggesting an insufficient pool of cysteine in the cell. Therefore, we also included the sulfur assimilation pathway from sulfate in the analysis.

Several reactions in the pathway can be catalyzed by more than one enzyme. For example, for the last step of pathway, *E. coli* has two homocysteine transmethylase: cobalamin-dependent MetH and cobalamin-independent MetE. In our methionine drop-out MOPS media, MetH expression was not induced compared to the level in the complete media, and was 0.6% of the level of MetE. Further, cells with MetE knockdown are unable to grow in media without methionine (see below), suggesting that MetH is not contributing significantly to methionine biosynthesis in this condition. Therefore, we only included MetE and not MetH in the analysis. Similarly for other reactions, we only included enzymes that were upregulated.

In addition to providing building blocks for protein synthesis, L-methionine is also converted to S-adenosyl-L-methionine (SAM). In eukaryotes, the demand for SAM is high due to its involvement in the synthesis of phosphatidylcholine (PC), which is a major component of the cell membrane (Giovannelli et al., 1985; Hirata and Axelrod, 1978). However, this demand is nonexistent in *E. coli* because it does not carry enzymes that synthesize PC (Sohlenkamp et al., 2003), and its lipid composition lacks PC (Oursel et al., 2007). The other pathways that utilize SAM were estimated to account for a small fraction of the methionine synthesis rate (0.4%) (Feist et al., 2007). We therefore consider protein synthesis as the major consumption for L-methionine.

The turnover number (k_{cat}) for each enzyme was obtained from the BRENDA database and the references therein (Schomburg et al., 2002). The only enzyme whose turnover number has not been reported for *E. coli* is sulfate adenylyltransferase (CysND). We instead used the measurement for a different proteobacteria, *Thiobacillus denitrificans*. The maximal reaction velocity (V_{max}) for each step was calculated as the product of k_{cat} and our estimate on the enzyme copy number. Because we provide an upper bound of the copy number, the maximal reaction velocity is also an upper bound of the actual flux.

The smallest V_{max} was found for two reactions in the pathway, catalyzed by MetE and CysH, respectively. Whereas the V_{max} for MetE matches the methionine consumption rate, the V_{max} by CysH is even smaller, raising the possibilities that either (1) there is an alternative for sulfur assimilation that has not been characterized, or (2) the reported k_{cat} for CysH in vitro is lower than that in vivo. To distinguish between these possibilities, we first tested whether a *cysH* null strain (from the KEIO collection) can grow on various sulfur sources. The *cysH* null strain was unable to grow in media with sulfate as the only sulfur source, but was able to grow when supplemented with 40 mM MOPS (3-morpholinopropane-1-sulfonic acid). Therefore, *E. coli* has an alternative pathway for sulfur assimilation from MOPS, which could explain the extra flux needed in our MOPS-based media.

We then tested which genes are responsible for this uncharacterized pathway, by constructing double-deletion strains with *cysH* and either genes that are upregulated in the Met drop-out media, or enzymes that are known to utilize other types of sulfonic acids. We found that the ability to grow on MOPS was lost when both *cysH* and *tauD*, which encodes a taurine dioxygenase, were deleted. Therefore, TauD is an essential enzyme for this alternative pathway. However, for wild-type cells in the Met drop-out media, TauD is not upregulated and is only expressed at a very low level (<50 copies/cell), suggesting that this novel pathway is not active in this condition. Therefore, it is likely that CysH is still the main pathway for cysteine biosynthesis, and that the published k_{cat} for CysH in vitro does not reflect the actual turnover number in vivo.

Analytical Model for Optimal Level of MetE Expression

To understand why cells produce the measured level of MetE when it is limiting methionine and protein synthesis, we modeled the growth rate as a function of the amount of MetE synthesized. The model takes into account the cost and benefit of MetE synthesis separately. The cost function is based on previous observations that synthesis of excess proteins competes with that of other

proteins (Scott et al., 2010). As a result of decreased amount of ribosomes that are necessary for auto-catalysis, growth rate decreases. The work by Hwa and coworkers established the relationship between the growth rate and the mass fraction of excess proteins (Scott et al., 2010). Here, we used a modified version for the methionine biosynthetic pathway, and express the predicted growth rate (λ) as a function of the mass fraction of MetE in the proteome (ϕ_E):

$$\lambda = \lambda_0 \left(1 - \frac{\phi_{m/c} + \phi_E}{\phi_C} \right),$$

where $\lambda_0 = 1.93 \text{ hr}^{-1}$ is the growth rate in methionine-supplemented media, $\phi_C = 0.48$ is the phenomenological parameter obtained by Scott et al., $\phi_{m/c} = 0.045$ is the mass fraction of all enzymes except MetE in the methionine and cysteine biosynthesis pathways. We chose to fix $\phi_{m/c}$ while varying ϕ_E because these other enzymes appear to be made in excess capacity. This cost function is plotted in red in Figure 7.

The benefit function is based on our observation that the rate of Met synthesis at maximal MetE activity is equal to the rate of Met consumption by protein synthesis:

$$N_E k_{cat} = f_{met} N_R k_e,$$

where N_E , N_R are the numbers of MetE and translating ribosomes, respectively. k_{cat} , k_e are the turnover number of MetE and translation elongation rate, respectively. f_{met} is the fraction of translated codons that encodes methionine. We can re-write this equation in terms of the growth rate λ and the mass fraction of MetE ϕ_E . To do so, we first notice that, when the majority of proteins are long-lived compared to the cell doubling time, the mass fraction of MetE is equivalent to the fraction of translating ribosomes making MetE, which is given by

$$\phi_E = \frac{k_{in}^E \frac{l_E}{k_e}}{N_R},$$

where k_{in}^E is the translation initiation rate of MetE and l_E/k_e is the time it takes to synthesize MetE (l_E is the length of MetE polypeptide). The translation initiation rate of MetE is proportional to the amount of MetE in a cell:

$$N_E = \frac{k_{in}^E}{\lambda},$$

Using these relations, the production/consumption equation becomes

$$\lambda = \frac{k_{cat}}{f_{met} l_E} \phi_E.$$

Notice that this relationship is only dependent on well-established parameters: the published turnover number ($k_{cat} = 0.12 \text{ s}^{-1}$), the fraction of Met codons (2.7%), and the length of MetE (753 amino acids). This benefit function is plotted in blue in Figure 7.

Combining the cost and benefit functions, we predict that the maximal growth rate can be achieved when the mass fraction of MetE is

$$\phi_E = \frac{\phi_C - \phi_{m/c}}{1 + \frac{k_{cat} \phi_C}{f_{met} l_E \lambda_0}} = 0.069.$$

And the maximal growth rate is

$$\lambda = \frac{k_{cat}}{f_{met} l_E} \frac{\phi_C - \phi_{m/c}}{1 + \frac{k_{cat} \phi_C}{f_{met} l_E \lambda_0}} = 1.47 \text{ hr}^{-1},$$

which corresponds to a doubling time of 28 min. These predictions match what we observed for cells grown in the Met drop-out media.

MetE Repression and Overexpression

To test the prediction that the MetE level is optimized in wild-type cells to maximize growth rate in media without methionine, we constructed strains with either MetE repression or overexpression and measured the effect on growth rate. For repression, we used the recently described CRISPRi system, which uses a DNA binding protein, dCas9 to block transcription elongation

(Qi et al., 2013). With a short guide RNA (sgRNA), dCas9 represses transcription with high sequence specificity in *E. coli*. We designed a sgRNA with a short guide region (18 nucleotides) targeting +182–198 bases of the *metE* gene on the nontemplate strand. The sgRNA is expressed from a tetracycline-inducible promoter on plasmid pJW1423, which is derived from pgRNA-bacteria (Addgene #44251) (Qi et al., 2013). To compare the effect on growth rate at various induction levels, we used a control sgRNA targeting RFP, which is absent in the cell, expressed from pgRNA-bacteria (Qi et al., 2013). Tetracycline repressor (TetR) and dCas9 are expressed from pdCas9-bacteria (Addgene #44249) (Qi et al., 2013). The dCas9 and sgRNA plasmids were simultaneously transformed into MG1655 and selected under chloramphenicol and carbenicillin.

Full induction of the sgRNA targeting *metE* inhibits cell growth in media lacking methionine, likely due to impaired methionine biosynthesis. This result also indicates that the other methionine synthase, MetH, is not functional in this growth condition. To test whether even modest reduction in MetE level affects growth as our model predicts, we measured growth rates with either uninduced basal expression level, or induced with 10 nM anhydrotetracycline. Overnight culture in methionine dropout media without anhydrotetracycline was diluted 1:1000 into 20 ml fresh media. The culture was kept at 37°C in waterbath shaker, and OD₆₀₀ was measured until it reaches 0.4. The effect of knockdown is confirmed by qPCR (see below). The results are plotted in Figure S6.

For overexpression, ectopic MetE was expressed from a tetracycline-inducible promoter (plasmid pJW1424) in MG1655. Because the endogenous *metE* transcript is regulated by sRNA, which could limit the ability to over-produce MetE proteins (Boysen et al., 2010), we replaced the native 5' UTR with a synthetic sequence. Due to this difference in 5' UTR, the induction of MetE protein levels may not be proportional to the changes in its mRNA levels. We measured growth rates at 37°C with either 432 nM or 4.32 μM anhydrotetracycline. The effect of the overexpression is confirmed by qPCR.

While these results provide strong qualitative agreement with our model (Figure S6), we caution that quantitative measurements are much more difficult due to the lack of nonperturbative tools for fine-tuning protein expression. Both the knockdown and overexpression systems we used requires accessory proteins, whose expression itself affects growth. Therefore, we are careful to compare the effect of *metE* mRNA level on growth rate should be relative to controls in the presence of these accessory proteins, and not to the wild-type for which our quantitative model is based upon.

Quantitative PCR for *metE* mRNA Levels

RNA was extracted following the RNAsnap protocol (Stead et al., 2012). 1 ml of culture at OD₆₀₀ = 0.3 was pelleted by centrifugation for 30 s. The pellet was added to 100 μl of RNA extraction solution (95% formamide, 1% beta-mercaptoethanol, 18 mM EDTA, 0.026% SDS), and incubated at 95°C for 7 min. After centrifugation at 21 kg for 5 min, the supernatant was transferred to new tubes with 150 μl Tris buffer, 25 μl sodium acetate, and 825 μl ethanol. RNA was precipitated by incubating at –80°C for > 1 hr, centrifugation at 4°C for 30 min, and the pellet was washed by 250 μl ice-cold 100% ethanol and resuspended in 500 μl Tris buffer pH 7. Remaining debris was removed by centrifugation and RNA was precipitated again and resuspended in 20 μl.

10 μg RNA in 20 μl was treated with 2 μl DNase I (10 U/μl, Roche) in supplied buffer at 37°C for 30 min, followed by 75°C for 10 min. RNA was purified using Zymo Clean and Concentrate Columns. cDNA was generated using M-MuLV reverse transcriptase (NEB) with random hexamers. Quantitative PCR was performed in triplicates using DyNAmo HS CYBR Green qPCR kit (Thermo) on Roche LightCycler. Standard curves were generated using concentrated cDNA samples. Primers targeting *cyoA* cDNA was used for normalization. Relative *metE* levels were compared to controls (empty vector for overexpression and sgRNA targeting RFP for repression).

SUPPLEMENTAL REFERENCES

Ariake, L., Valgepea, K., Peil, L., Nahku, R., Adamberg, K., and Vilu, R. (2012). Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*. *J. Proteomics* 75, 5437–5448.

Boysen, A., Møller-Jensen, J., Kallipolitis, B., Valentin-Hansen, P., and Overgaard, M. (2010). Translational regulation of gene expression by an anaerobically induced small non-coding RNA in *Escherichia coli*. *J. Biol. Chem.* 285, 10690–10702.

Bremer, H., and Dennis, P.P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, F.C. Neidhardt, ed. (Washington, DC: American Society Microbiology), pp. 1553–1569.

Giovannelli, J., Mudd, S.H., and Datko, A.H. (1985). Quantitative analysis of pathways of methionine metabolism and their regulation in *lemna*. *Plant Physiol.* 78, 555–560.

Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. (2008). Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J. Proteome Res.* 7, 1118–1130.

Hirata, F., and Axelrod, J. (1978). Enzymatic synthesis and rapid translocation of phosphatidylcholine by two methyltransferases in erythrocyte membranes. *Proc. Natl. Acad. Sci. USA* 75, 2348–2352.

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M.J., and Frishman, D. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 9, 102.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41 (Database issue), D605–D612.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 25, 117–124.

- Oursel, D., Loutelier-Bourhis, C., Orange, N., Chevalier, S., Norris, V., and Lange, C.M. (2007). Lipid composition of membranes of *Escherichia coli* by liquid chromatography/tandem mass spectrometry using negative electrospray ionization. *Rapid Commun. Mass Spectrom.* *21*, 1721–1728.
- Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., and Aebersold, R. (2009). Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* *138*, 795–806.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* *152*, 1173–1183.
- Sohlenkamp, C., López-Lara, I.M., and Geiger, O. (2003). Biosynthesis of phosphatidylcholine in bacteria. *Prog. Lipid Res.* *42*, 115–162.
- Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B., and Kushner, S.R. (2012). RNAsnap™: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.* *40*, e156.
- Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* *329*, 533–538.
- Tukey, J.W. (1962). The future of data analysis. *Ann. Math. Statist.* *33*, 1–324.

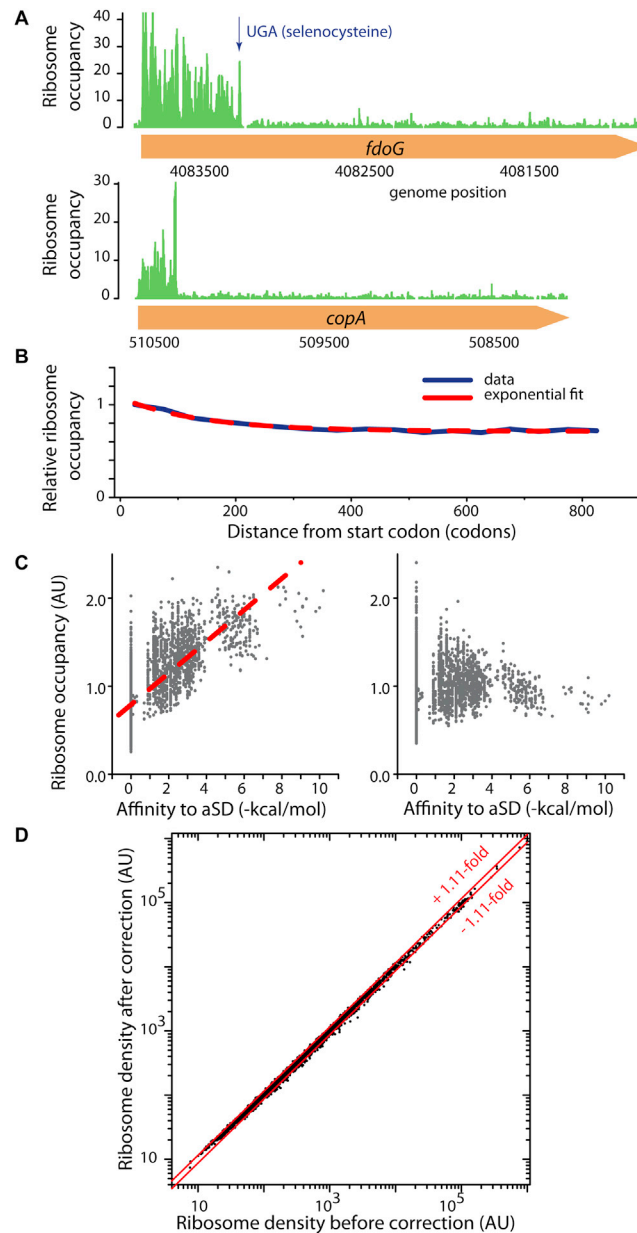


Figure S1. Adjustment to Ribosome Density Based on Sequence- and Position-Specific Variation in Translation Elongation Rates, Related to Figure 1A

(A) Atypical genes with large drop-off in ribosome occupancy. For selenoproteins, e.g., FdoG, we observed reduced ribosome occupancy after the selenocysteine codon. Only the region after the selenocysteine codon was used to calculate the average ribosome density. Abrupt decrease in ribosome occupancy on a few other genes, such as *copA*, could indicate novel translational events.

(B) Correction for elevated ribosome occupancy toward the beginning of open reading frames. The slight increase in occupancy (blue) was modeled as an exponential function (red). The fitting parameters were used to adjust the position-dependent ribosome occupancy.

(C) Correction for translational pausing induced by internal Shine-Dalgarno-sequences. The average ribosome occupancy downstream from a hexanucleotide sequence is plotted against its affinity to the anti-Shine-Dalgarno sequence. The observed relationship (left) was fitted with a linear function (red). The fitting parameters were used to adjust the sequence-dependent ribosome occupancy, so that the result is independent of the strength of Shine-Dalgarno sequences (right).

(D) Effects of the corrections for local variation in translation elongation rates. For each gene, the average ribosome density before and after corrections is plotted. The standard deviation for the differences is 1.11-fold.

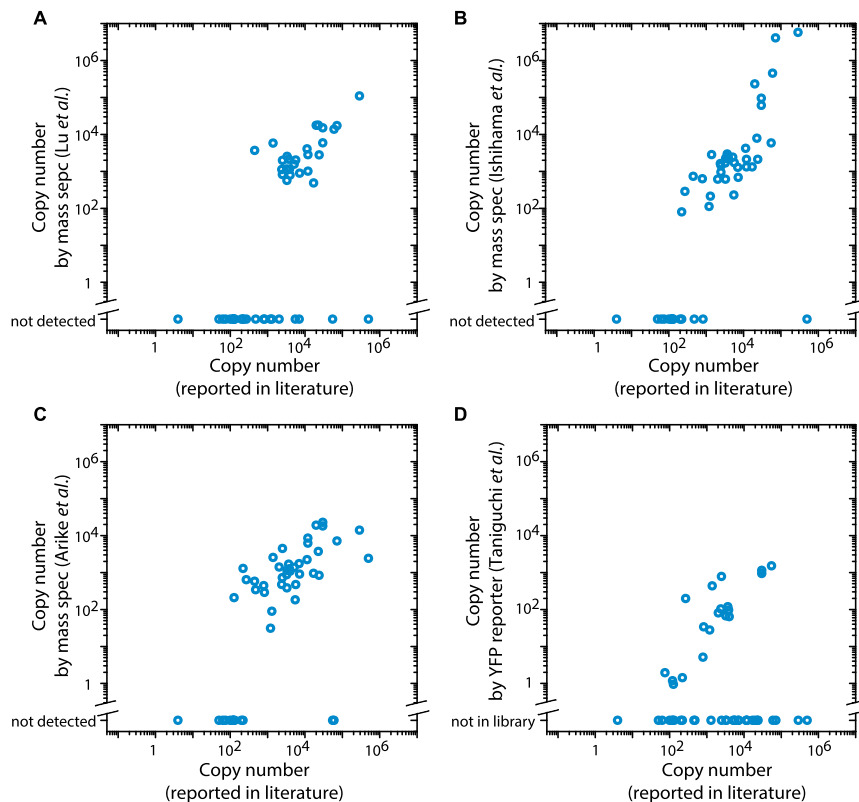


Figure S2. Comparison of Published Quantitative Proteomics Measurements and Individually Measured Protein Copy Number, Related to Figure 1B

(A) Proteomics data using absolute protein expression (APEX) profiling based on mass spectrometry (Lu et al., 2007).

(B) Proteomics data using exponentially modified protein abundance index (emPAI) based on mass spectrometry (Ishihama et al., 2008).

(C) Proteomics data using intensity-based absolute quantification (iBAQ) based on mass spectrometry (Arike et al., 2012). We note that the data in (A)–(C) were obtained using label-free quantification. Current development in other absolute quantification methods using isotopic labeling and synthetic peptides as standards could provide improvements in accuracy and coverage (Hanke et al., 2008; Picotti et al., 2009).

(D) Proteomics data using a YFP-fusion library (Taniguchi et al., 2010). The library was constructed for ~25% of the genome. The measurements were performed at a lower growth rate (150 min per doubling) compared to other reports, which gave rise to lower protein abundance in general.

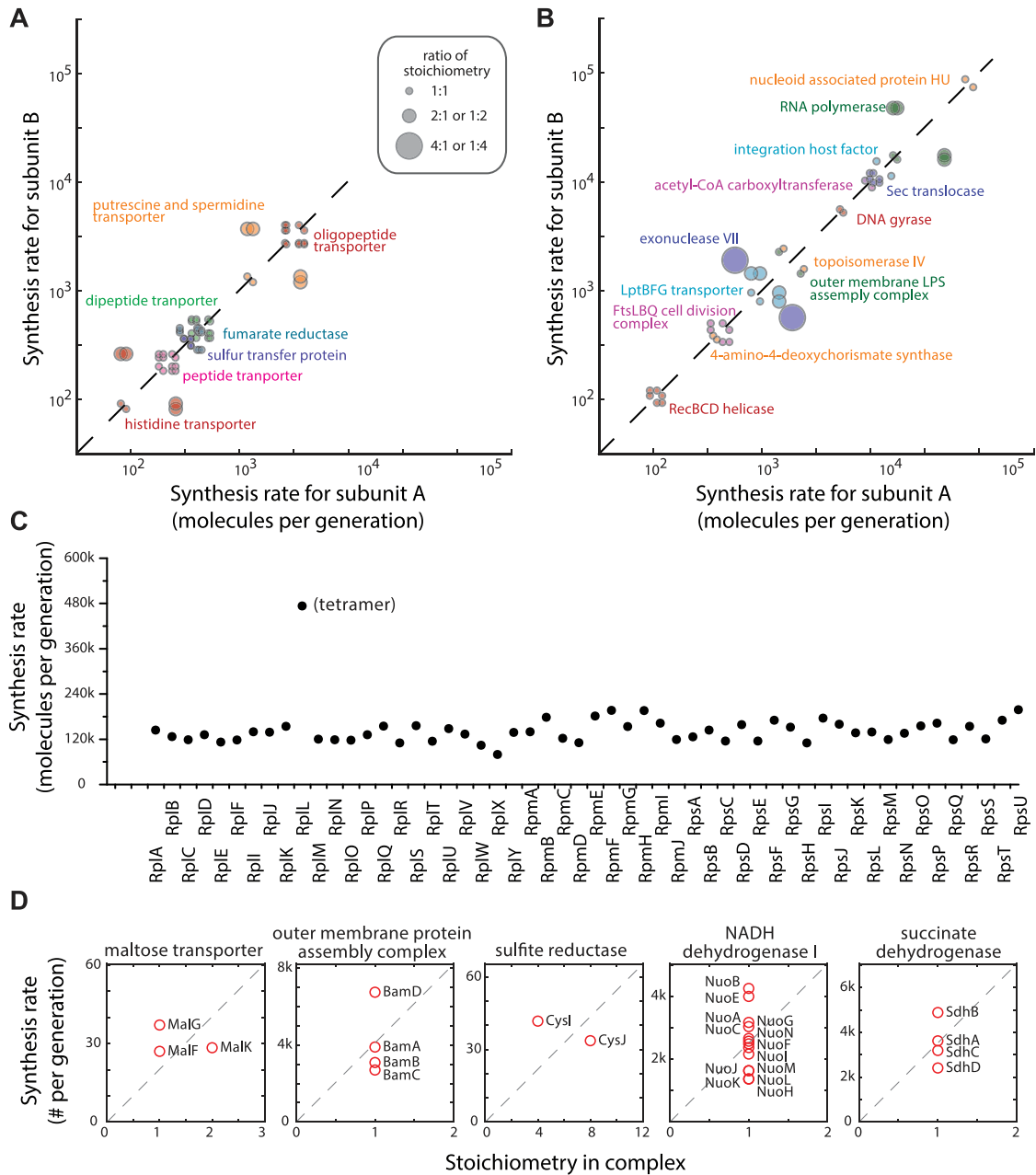


Figure S3. Proportional Synthesis for Other Multiprotein Complexes, Related to Figure 2

(A) Proportional synthesis for complexes whose members are encoded in the same operon. Complexes not included in Figure 2B are shown here. The synthesis rate for each pair of subunits in the complex is plotted, with the identity of the complex indicated by the color code. The size of the symbol reflects the ratio of stoichiometry between the pair. Each pair is plotted twice with different order.

(B) Proportional synthesis for complexes whose members are encoded in more than one operon. The size of symbols is the same as in (A). Inset shows synthesis rates for ribosomal proteins. For some of the ribosomal protein with equal stoichiometry, proportional synthesis may be achieved by a combination of translational coupling and auto-regulation.

(C) Proportional synthesis for ribosomal proteins. All proteins, except RplL (L7/L12), have the stoichiometry of one per ribosome.

(D) Exceptions to proportional synthesis. Five complexes do not follow proportional synthesis out of 64 complexes. The synthesis rates relative to the stoichiometry are plotted here. Subunits of the maltose transporter and the BAM complex are translated from different mRNA, whereas the other three complexes are translated from the same polycistronic mRNA.

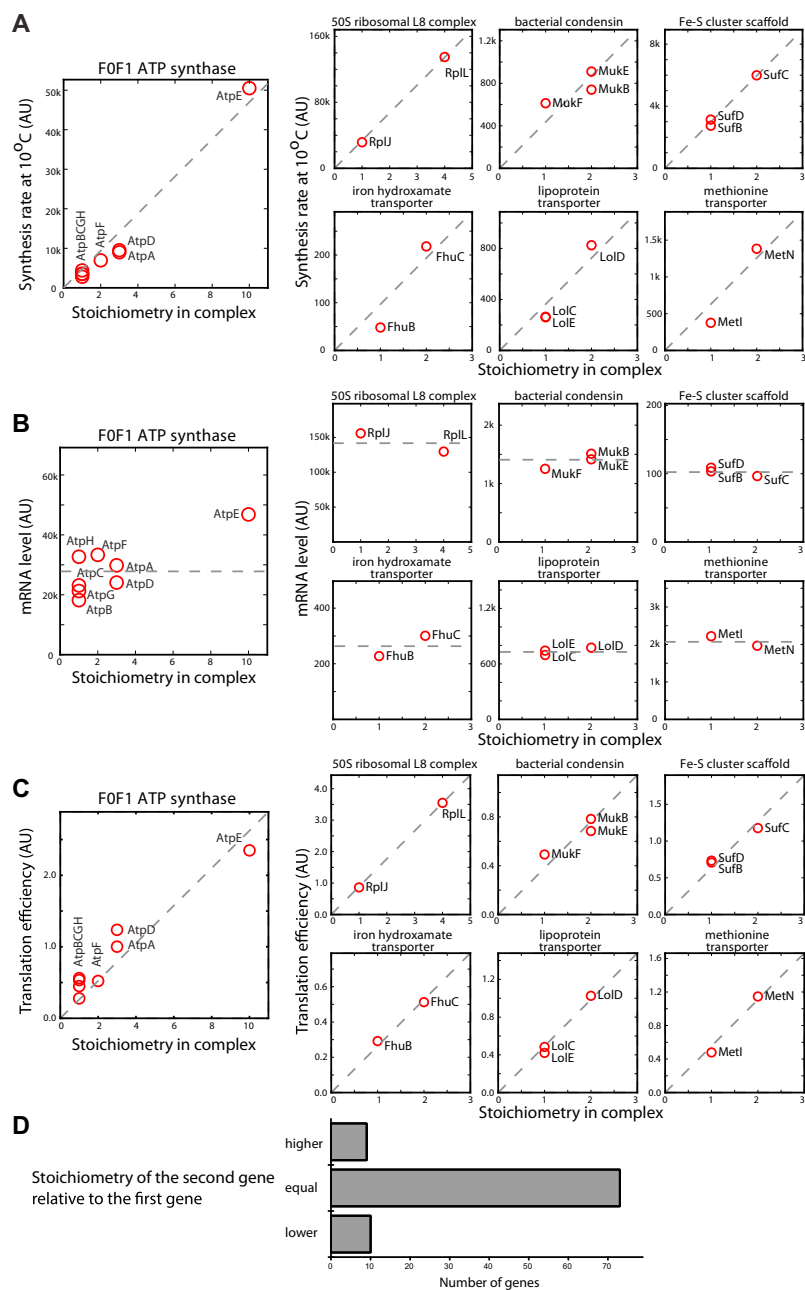


Figure S4. Proportional Synthesis at 10°C, mRNA Levels, and Gene Order, Related to Figure 2

(A) Proportional synthesis at 10°C. Synthesis rates relative to stoichiometry are plotted for complexes expressed from the same operon. Experiment was performed at 50 hr after shifting the culture to 10°C. The dashed line indicates the best-fit that crosses the origin.

(B) mRNA levels for multi-protein complexes. The average transcript level for each gene is plotted against the stoichiometry in the complex. The dashed lines indicate the average transcript levels. Complexes with subunits expressed from the same polycistronic operon with uneven stoichiometry are shown here. Small variation in mRNA level could be due to alternative transcription start site or differential degradation. Overall, the mRNA levels are similar across different subunit and are not proportional to the stoichiometry.

(C) Translation efficiency for multi-protein complexes. The rate of protein synthesis per mRNA, as measured by protein synthesis rates (from ribosome profiling) divided by mRNA levels (from mRNA-seq), is plotted against the stoichiometry in the complex. The dashed line indicates the best-fit that crosses the origin. In combination with (B), the difference in protein production is mainly determined at the translational level.

(D) Gene order and ratio of translation rates. For the complexes analyzed in this work, the relative stoichiometry between the gene products and their order in the operon is shown in the histogram. The preceding gene product has similar likelihood to have higher and lower stoichiometry relative to the following gene product.

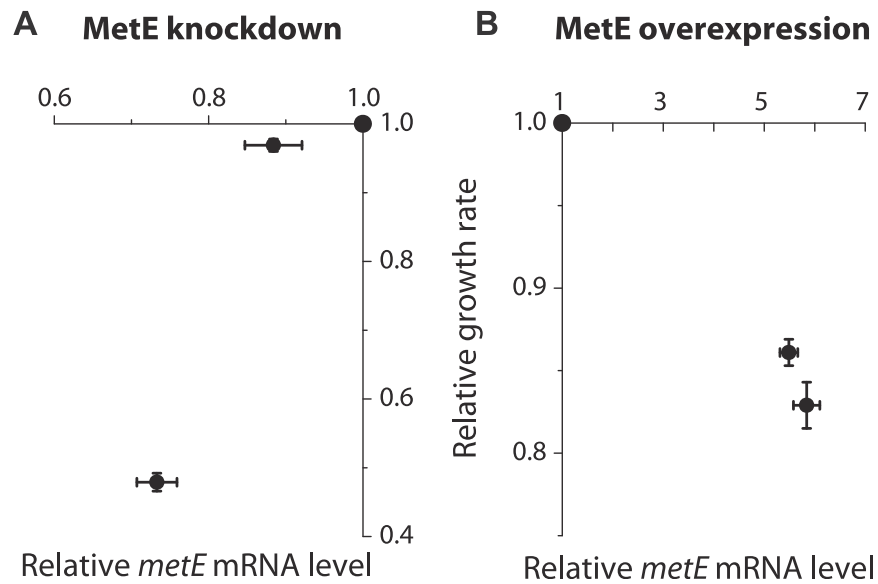


Figure S5. Effect of MetE Level on Growth Rate, Related to Figure 7

(A) Effect of MetE knockdown. The growth rate relative to control is plotted as a function of *metE* mRNA level. The transcription of *metE* is reduced using CRISPRi knockdown. Both growth rates and levels of *metE* mRNA are relative to control experiment with sgRNA targeting RFP instead of *metE*. (x-error bars: standard deviation of the mean, N = 3, y-error bars: standard error)

(B) Effect of MetE overexpression. The growth rate relative to control is plotted as a function of *metE* mRNA level. The transcription of ectopic *metE* is induced using a tetracycline-inducible promoter. Both growth rates and levels of *metE* mRNA are relative to control experiments with empty vector. (x-error bars: standard deviation of the mean, N = 3, y-error bars: standard error).

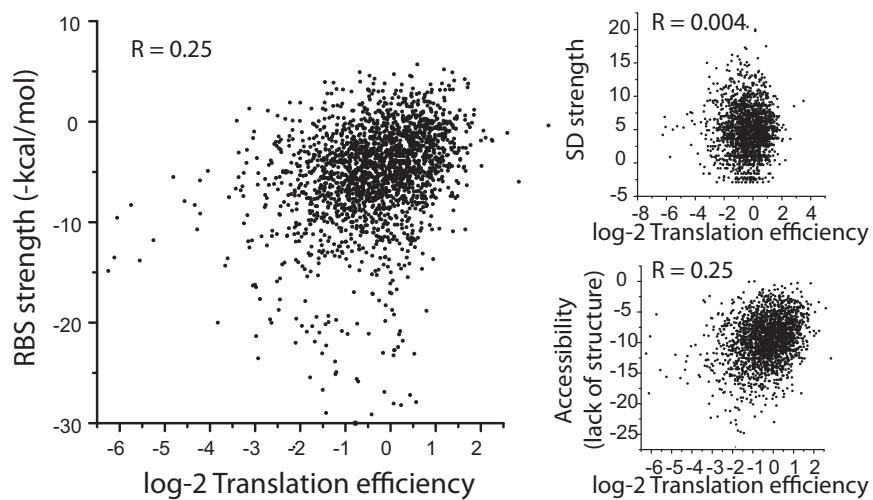


Figure S6. Predicted Strength of Ribosome-Binding Sites and Observed Translation Efficiency, Related to Figure 2

Prediction using the sequence near the translational start site was based on the model established by Salis et al. (2009). Translation efficiency was estimated from the ribosome footprint density relative to the mRNA level. The small degree of correlation is mostly explained by the predicted secondary structure of mRNA, and not by the strength of Shine-Dalgarno sequences.