Published in final edited form as:
Cell. 2014 April 24; 157(3): 624–635. doi:10.1016/j.cell.2014.02.033.

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,*]

[1]Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute, University of California, San Francisco, CA 94158, USA.

[2]California Institute of Quantitative Biosciences, University of California, San Francisco, CA 94158, USA.

[3]Center for RNA Systems Biology, University of California, San Francisco, CA 94158, USA.

[4]Department of Microbiology and Immunology, University of California, San Francisco, CA 94158, USA.

[5]Department of Cell and Tissue Biology, University of California, San Francisco, CA 94158, USA.

Abstract

Quantitative views of cellular functions requires 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* dataset transforms our understanding of the extent to which protein synthesis is precisely controlled to optimize function and efficiency. Members of multi-protein 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 used to optimize 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

* To whom correspondence should be addressed. gene-wei.li@ucsf.edu, weissman@cmp.ucsf.edu.

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 post-translational 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 post-translational 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 two-dimensional gel electrophoresis, or more recently by mass spectrometry (Dennis, 1974; Lemaux et al., 1978; Schwanhausser et al., 2009). While 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-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 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 and verified 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 multi-protein 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 dataset. The protein copy numbers inferred from synthesis rates also revealed rules that govern the abundance of transcription factors, and allowed quantitative characterization for the

methionine biosynthesis pathway, for which we identified a bottleneck enzyme whose expression level is optimized for maximal growth rate. More broadly, our approach and datasets 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., 2012; Ingolia et al., 2009). Each ribosome-protected mRNA fragment is then identified by massively parallel next-generation sequencing (Ingolia et al., 2012; Ingolia et al., 2009). 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 dataset. 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) (Fig 1A). The few genes that display large drop-off could represent novel events of translational regulation (Fig. S1A). 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, as similar ribosome densities are observed in the first and second halves of each gene. Most genes differ by <30% (standard deviation of the mean, Fig 1A). Additionally, correcting for sequence- and position-specific variation in elongation rates has only modest effect on average ribosome density (Fig. 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 (Methods). Within each dataset, 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 towards the beginning of open reading frames (Methods). The corrections were small (Fig. 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 (Methods). 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 correspond to ~10 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 times generation time (21.5 min in rich defined media, see Methods). 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 $R^2 = 0.96$ (Fig. 1B). Deviations from the identity line in Fig. 1B likely reflect biological phenomenon. For example, the strongest outlier is σ^{32}, the heat shock transcription factor 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 Methods). Existing efforts to globally quantify protein abundance in *E. coli* using mass spectrometry or fluorescent reporter show less concordance and dynamic range (Fig. S2). In conclusion, our genome-wide synthesis rate measurements and the resulting estimate of protein abundance are supported by classic biochemical measurements across 5 orders of magnitude of protein abundance.

Proportional synthesis of multi-protein 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, as 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 8 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 (Fig. 2A), and qualitatively agrees with the differential synthesis rates previously reported (Brusilow et al., 1982; 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 multi-protein 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%) are 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 (Fig. S3D), and these small number of exceptions could suggest dominant control at the level of degradation or the existence of dynamic sub-complexes, as in the case of the outer membrane protein assembly complex (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 Fig. 2B and Fig. S3. We also observed very similar synthesis rates for complexes with two equimolar subunits (Fig. 2C and Fig. S3A-C). Notably, proportional synthesis is robust against temperature; similar ratios in synthesis rates were observed both at 37°C and at 10°C (Fig. S4A). Furthermore, both abundant and scarce proteins have evolved strict tuning of synthesis rates, as the expression levels of these complexes ranges over four orders of magnitude.

Proportional synthesis in *E. coli* is predominantly achieved through translational, rather than transcriptional control. The majority of multi-protein complexes encode their subunits on a single polycistronic mRNA, with each subunit translated from its own initiation site (47/64 complexes, Fig. 2B-C and Fig. 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 (Fig. S4BC and Table S4). Moreover, gene order does not explain differential synthesis rates (Fig. 2A and 2C and Fig. S4D), consistent with our previous observation that translation rates among genes in the same operon are only weakly correlated (inset, Fig. 2C) (Oh et al., 2011). Protein 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 as existing models for either the strength of ribosome binding site or the Shine-Dalgarno sequence alone do not predict proportional synthesis (Fig. 2C) (Salis et al., 2009). Translational auto-regulation (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 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 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 (Fig. 3A-B). Genomic duplication events in *S. cerevisiae* have led to numerous paralogous genes, which in some cases can substitute for each other in multi-protein 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 (Fig. 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 (Fig. 3C). A notable exception for proportional synthesis is the signal recognition particle, for which four subunits are translated at 1:1:2:2 ratio and the other two subunits are in excess (Fig. 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, as subunits are encoded on different mRNAs in eukaryotes. For example, the dynamics of nuclear localization of transcription factors 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 changes 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 co-workers 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 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; Schuller 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 dataset covers many different modules in the same functional class, we can use our data to identify common expression patterns strategies that are selected through evolution. Our studies of several different modules identified a 2nd pattern: hierarchical expression, in which components are differentially expressed according to their hierarchical role.

Bacterial toxin-antitoxin modules (TA) 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 (Fig. 4A), which would allow *E. coli* to produce 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 (Fig. 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 (Fig. 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 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 (Fig. 2B-C), we found that the periplasmic binding proteins are always in large excess (Fig. 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 (Fig. 4C). Using mathematical modeling and experimental validation, it has been demonstrated that large excess of RR relative to 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 datasets.

Bacterial proteome composition

Because the large majority of proteins are stable in *E. coli* (Larrabee et al., 1980), our protein synthesis rate data provides a comprehensive view of proteome composition, allowing us to probe how cells allocate resources (Fig. 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 dataset opens up the possibility of broadly investigating how these competing constraints govern protein expression levels. We select two specific cellular functions (transcription factors and methionine biosynthesis) for further study.

Copy numbers of transcription factors reveal their mode of action

The bacterial chromosome is densely covered with transcription factors (TFs) that bind DNA both specifically and non-specifically (Li et al., 2009). The crowded space on DNA imposes constraints on the abundance of TFs, as overcrowding by non-specifically 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 basepairs on the *E. coli* chromosome (assuming most DNA-binding proteins are associated with DNA nonspecifically and 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 indicates 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 (Fig. 6A-B). 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 six

minutes 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 minutes, the binding 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, as 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 (Fig. 6C). By contrast, the large majority of low abundance TFs bind to the target sites independent of the corresponding ligands (Fig. 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 methionine 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 of 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-methionine biosynthetic pathway for *E. coli* grown in media devoid of methionine (Met). We first calculated the cellular demand for this pathway (31,000 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 times the number of methionine residues in that protein. The other major pathway that consumes Met, which is the synthesis of *S*-adenosyl-L-methionine, was estimated to contribute to a small fraction of the overall flux (Feist et al., 2007) (see also Methods). 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 maximum velocity 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} (Fig.

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 s^{-1} per cell) matches the Met consumption rate. Therefore, under these growth conditions, MetE catalyzed conversion of L-homocysteine to L-methionine is a bottleneck step that operates at maximal velocity with saturating substrate concentration.

Given that methionine 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 ~8% of the total protein synthesis capacity in media devoid of methionine. 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 (Fig. 7B, Methods). The model considers the cost and benefit of MetE synthesis independently, and allows us to evaluate the level of synthesis where the tradeoff between cost and benefit is optimized. The benefit of producing MetE arises from our observation that it is a bottleneck for the methionine supply for protein synthesis. Hence, devoting more protein synthesis capacity to MetE increases growth rate linearly (Methods). 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 and Hwa (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 (Fig. 7B). Remarkably, these predictions were in close agreement with the actual values observed for cells lacking methionine: 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 (Fig. 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 multi-protein 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 datasets which will both elucidate the regime in which biochemical reactions operate, and provide a foundation for rational design of synthetic biological systems.

Our genome-wide dataset 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—transcription factors can be kept at low abundances without kinetic penalties by pre-binding to target sites, and the synthesis rate of a key enzyme that limits metabolic flux in the methionine 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.

While 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 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 (Fig. 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, zebra fish, and mammals. For eukaryotes and multi-cellular 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 footprints size (~15 to 45 nucleotides long) were selected on a denaturing polyacrylamide gel. The mRNA fragments were converted to 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 HiSeq2000. Data are available at Gene Expression Omnibus with accession number GSE53767.

Analysis for absolute synthesis rates

Counts of ribosome footprints for each gene were first corrected for the elevated density towards 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 standards 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 methionine, 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, this relationship is $\lambda=\lambda_0\left(1-\frac{\phi_{m/c}+\phi_E}{\phi_C}\right)$, where λ is the growth rate, λ_0 is the growth rate when methionine is not limiting, ϕ_E is the mass fraction of MetE, $\phi_{m/c}$ is the mass fraction of all other enzymes in the methionine and cysteine biosynthetic pathways, and ϕ_C is the phenomenological fitting parameter that were established in their work. The benefit function is based on our observation that the level of MetE determines that rate of methionine 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 turnover number of MetE and translation elongation rate, respectively. f_{met} is the fraction of translated codons that encodes methionine. Re-writing this equation using ϕ_E and λ gives $\lambda=\frac{k_{cat}\phi_E}{f_{met}l_E}$, 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 Fig. 7C.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank L. Qi for providing material for CRISPRi knockdown, R Milo, KC Huang, J Elf, J Dunn, G Brar, O Brandman, C Jan, Josh Rabinowitz and members of the Weissman lab and the Gross lab for discussions, P Choi and H Chen for critical reading of the manuscript. We also thank Jessica Lund and Eric Chow for help on sequencing, and Christopher Reiger and Manny DeVera for administrative support. This research was supported by the Helen Hay Whitney Foundation (to GWL), NIH Pathway to Independence Award (GM105913, to GWL), NIH Center for RNA Systems Biology (to JSW), and Howard Hughes Medical Institute (to JSW).

REFERENCES

Ades SE, Connolly LE, Alba BM, Gross CA. The Escherichia coli sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. Genes Dev. 1999; 13:2449–2461. [PubMed: 10500101]

Akiyama Y, Kihara A, Tokuda H, Ito K. FtsH (HflB) is an ATP-dependent protease selectively acting on SecY and some other membrane proteins. J Biol Chem. 1996; 271:31196–31201. [PubMed: 8940120]

Alon U, Surette MG, Barkai N, Leibler S. Robustness in bacterial chemotaxis. Nature. 1999; 397:168–171. [PubMed: 9923680]

Andersson SG, Kurland CG. Codon preferences in free-living microorganisms. Microbiol Rev. 1990; 54:198–210. [PubMed: 2194095]

Barkai N, Shilo BZ. Variability and robustness in biomolecular systems. Mol Cell. 2007; 28:755–760. [PubMed: 18082601]

Batchelor E, Goulian M. Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. Proc Natl Acad Sci U S A. 2003; 100:691–696. [PubMed: 12522261]

Baughman G, Nomura M. Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in Escherichia coli. Cell. 1983; 34:979–988. [PubMed: 6354472]

Blikstad I, Nelson WJ, Moon RT, Lazarides E. Synthesis and assembly of spectrin during avian erythropoiesis: stoichiometric assembly but unequal synthesis of alpha and beta spectrin. Cell. 1983; 32:1081–1091. [PubMed: 6220807]

Brandman O, Stewart-Ornstein J, Wong D, Larson A, Williams CC, Li GW, Zhou S, King D, Shen PS, Weibezahn J, et al. A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell. 2012; 151:1042–1054. [PubMed: 23178123]

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science. 2012; 335:552–557. [PubMed: 22194413]

Brusilow WS, Klionsky DJ, Simoni RD. Differential polypeptide synthesis of the proton- translocating ATPase of Escherichia coli. J Bacteriol. 1982; 151:1363–1371. [PubMed: 6213603]

Buttgereit F, Brand MD. A hierarchy of ATP-consuming processes in mammalian cells. Biochem J. 1995; 312(Pt 1):163–167. [PubMed: 7492307]

Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature. 2008; 455:485–490. [PubMed: 18818649]

Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005; 436:588–592. [PubMed: 16049495]

Dennis PP. In vivo stability, maturation and relative differential synthesis rates of individual ribosomal proteins in Escherichia coli B/r. J Mol Biol. 1974; 88:25–41. [PubMed: 4613843]

Dong H, Nilsson L, Kurland CG. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. J Bacteriol. 1995; 177:1497–1504. [PubMed: 7883706]

Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. Science. 2007; 316:1191–1194. [PubMed: 17525339]

Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297:1183–1186. [PubMed: 12183631]

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. 2007; 3:121. [PubMed: 17593909]

Fell, D. Understanding the control of metabolism. Portland Press; London: 1997.

Gerdes K, Maisonneuve E. Bacterial persistence and toxin-antitoxin loci. Annu Rev Microbiol. 2012; 66:103–123. [PubMed: 22994490]

Grossman AD, Straus DB, Walter WA, Gross CA. Sigma 32 synthesis can regulate the synthesis of heat shock proteins in Escherichia coli. Genes Dev. 1987; 1:179–184. [PubMed: 3315848]

Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. The lac repressor displays facilitated diffusion in living cells. Science. 2012; 336:1595–1598. [PubMed: 22723426]

Hart Y, Madar D, Yuan J, Bren A, Mayo AE, Rabinowitz JD, Alon U. Robust control of nitrogen assimilation by a bifunctional enzyme in E. coli. Mol Cell. 2011; 41:117–127. [PubMed: 21211727]

Hu JC, Sherlock G, Siegele DA, Aleksander SA, Ball CA, Demeter J, Gouni S, Holland TA, Karp PD, Lewis JE, et al. PortEco: a resource for exploring bacterial biology through high-throughput data and analysis tools. Nucleic Acids Res. 2014; 42:D677–684. [PubMed: 24285306]

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc. 2012; 7:1534–1550. [PubMed: 22836135]

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009; 324:218–223. [PubMed: 19213877]

Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011; 147:789–802. [PubMed: 22056041]

Larrabee KL, Phillips JO, Williams GJ, Larrabee AR. The relative rates of protein synthesis and degradation in a growing culture of Escherichia coli. J Biol Chem. 1980; 255:4125–4130. [PubMed: 6989832]

Lehnert ME, Lodish HF. Unequal synthesis and differential degradation of alpha and beta spectrin during murine erythroid differentiation. J Cell Biol. 1988; 107:413–426. [PubMed: 3166462]

Lemaux PG, Herendeen SL, Bloch PL, Neidhardt FC. Transient rates of synthesis of individual polypeptides in E. coli following temperature shifts. Cell. 1978; 13:427–434. [PubMed: 350413]

Li GW, Berg OG, Elf J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. Nat Phys. 2009; 5:294–297.

Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012; 484:538–541. [PubMed: 22456704]

Li GW, Xie XS. Central dogma at the single-molecule level in living cells. Nature. 2011; 475:308–315. [PubMed: 21776076]

McCarthy JE, Gualerzi C. Translational control of prokaryotic gene expression. Trends Genet. 1990; 6:78–85. [PubMed: 2183416]

Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. J Bacteriol. 1974; 119:736–747. [PubMed: 4604283]

Nomura M, Gourse R, Baughman G. Regulation of the synthesis of ribosomes and ribosomal components. Annu Rev Biochem. 1984; 53:75–117. [PubMed: 6206783]

Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell. 2011; 147:1295–1308. [PubMed: 22153074]

Oromendia AB, Dodgson SE, Amon A. Aneuploidy causes proteotoxic stress in yeast. Genes Dev. 2012; 26:2696–2708. [PubMed: 23222101]

Papp B, Pal C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003; 424:194–197. [PubMed: 12853957]

Petersen C. Escherichia coli ribosomal protein L10 is rapidly degraded when synthesized in excess of ribosomal protein L7/L12. J Bacteriol. 1990; 172:431–436. [PubMed: 2403546]

Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol. 2004; 2:886–897. [PubMed: 15494745]

Quax TE, Wolf YI, Koehorst JJ, Wurtzel O, van der Oost R, Ran W, Blombach F, Makarova KS, Brouns SJ, Forster AC, et al. Differential translation tunes uneven production of operon-encoded proteins. Cell Rep. 2013; 4:938–944. [PubMed: 24012761]

Rigel NW, Ricci DP, Silhavy TJ. Conformation-specific labeling of BamA and suppressor analysis suggest a cyclic mechanism for beta-barrel assembly in Escherichia coli. Proc Natl Acad Sci U S A. 2013; 110:5151–5156. [PubMed: 23479609]

Russell JB, Cook GM. Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol Rev. 1995; 59:48–62. [PubMed: 7708012]

Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009; 27:946–950. [PubMed: 19801975]

Savageau MA. Design of molecular control mechanisms and the demand for gene expression. Proc Natl Acad Sci U S A. 1977; 74:5647–5651. [PubMed: 271992]

Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. Nucleic Acids Res. 2002; 30:47–49. [PubMed: 11752250]

Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. Multidimensional optimality of microbial metabolism. Science. 2012; 336:601–604. [PubMed: 22556256]

Schuller HJ, Fortsch B, Rautenstrauss B, Wolf DH, Schweizer E. 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. 1992; 203:607–614. [PubMed: 1735446]

Schwanhausser B, Gossen M, Dittmar G, Selbach M. Global analysis of cellular protein translation by pulsed SILAC. Proteomics. 2009; 9:205–209. [PubMed: 19053139]

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. Science. 2010; 330:1099–1102. [PubMed: 21097934]

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008; 455:58–63. [PubMed: 18668040]

Shapiro AL, Scharff MD, Maizel JV, Uhr JW. Synthesis of excess light chains of gamma globulin by rabbit lymph node cells. Nature. 1966; 211:243–245. [PubMed: 5965538]

Shemorry A, Hwang CS, Varshavsky A. Control of protein quality and stoichiometries by N- terminal acetylation and the N-end rule pathway. Mol Cell. 2013; 50:540–551. [PubMed: 23603116]

Shinar G, Dekel E, Tlusty T, Alon U. Rules for biological regulation based on error minimization. Proc Natl Acad Sci U S A. 2006; 103:3999–4004. [PubMed: 16537475]

Shinar G, Milo R, Martinez MR, Alon U. Input output robustness in simple bacterial signaling systems. Proc Natl Acad Sci U S A. 2007; 104:19931–19935. [PubMed: 18077424]

Springer M, Weissman JS, Kirschner MW. A general lack of compensation for gene dosage in yeast. Mol Syst Biol. 2010; 6:368. [PubMed: 20461075]

Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, et al. Decoding human cytomegalovirus. Science. 2012; 338:1088–1093. [PubMed: 23180859]

Swain PS. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. J Mol Biol. 2004; 344:965–976. [PubMed: 15544806]

Torres EM, Williams BR, Amon A. Aneuploidy: cells losing their balance. Genetics. 2008; 179:737–746. [PubMed: 18558649]

Tyedmers J, Mogk A, Bukau B. Cellular strategies for controlling protein aggregation. Nat Rev Mol Cell Biol. 2010; 11:777–788. [PubMed: 20944667]

von Dassow G, Meir E, Munro EM, Odell GM. The segment polarity network is a robust developmental module. Nature. 2000; 406:188–192. [PubMed: 10910359]

von Hippel PH. From “simple” DNA-protein interactions to the macromolecular machines of gene expression. Annu Rev Biophys Biomol Struct. 2007; 36:79–105. [PubMed: 17477836]

Winter RB, Berg OG, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. Biochemistry-Us. 1981; 20:6961–6977.

Yamaguchi Y, Park JH, Inouye M. Toxin-antitoxin systems in bacteria and archaea. Annu Rev Genet. 2011; 45:61–79. [PubMed: 22060041]

You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, Wang YP, Lenz P, Yan D, Hwa T. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature. 2013

Highlights

- Global measurement for absolute rates of protein synthesis using ribosome profiling
- Majority of protein complexes are precisely made in proportion to stoichiometry
- Rates of synthesis for individual proteins are optimized for growth and function
- Copy number estimates for stable proteins provide basis for quantitative biology

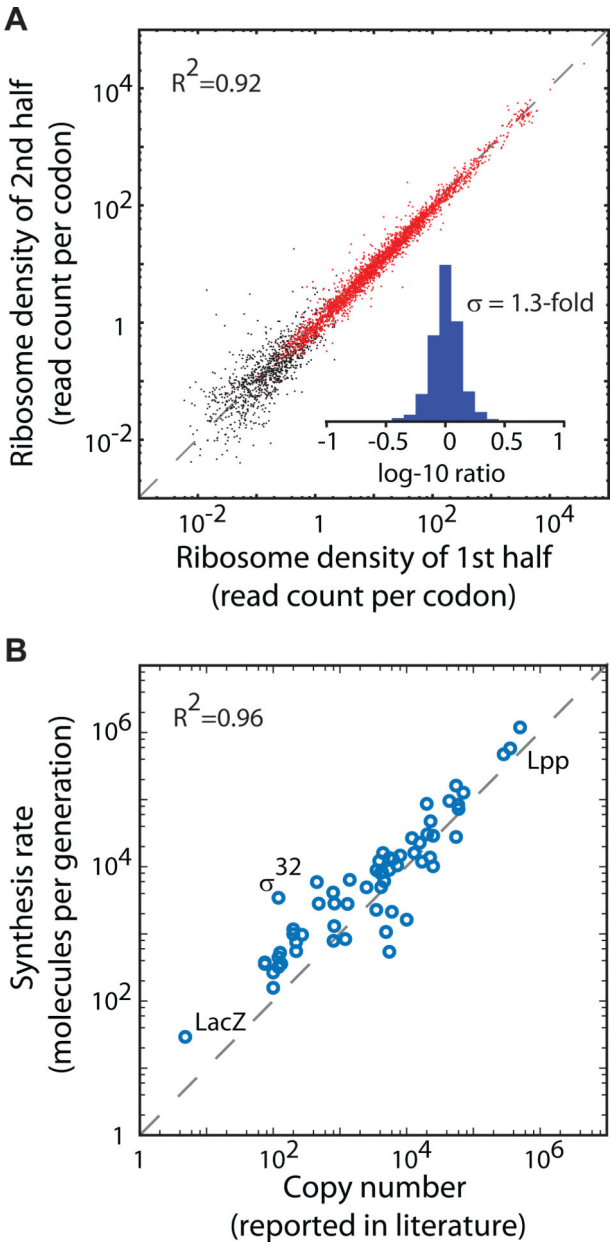

Figure 1. Absolute Quantification of Protein Synthesis Rates

(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 which have been individually quantified in the literature are plotted against the absolute protein synthesis rates (Pearson correlation $R^2 = 0.96$).

See also Figure S1, Figure S2, Table S1, and Table S2

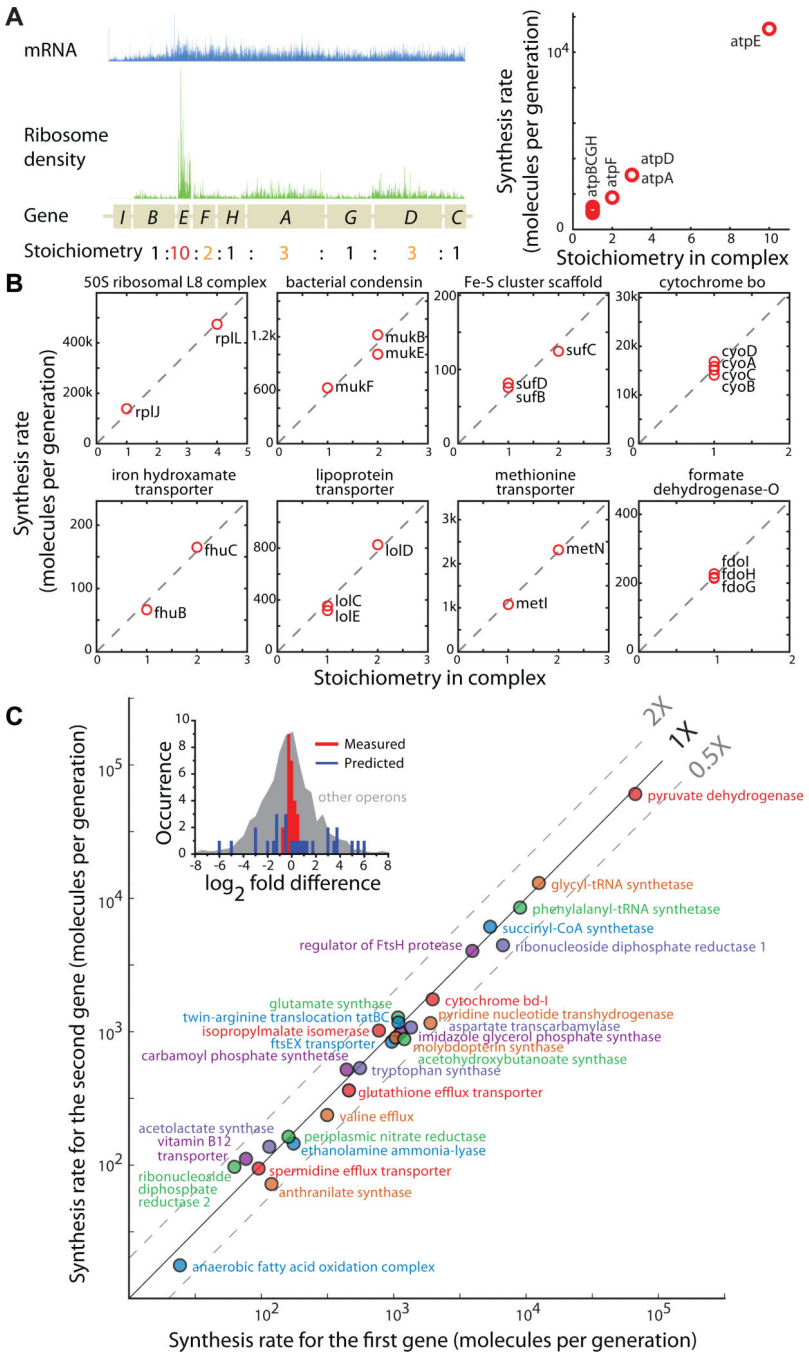

Figure 2. Proportional Synthesis of Multi-Protein Complexes

(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 panel (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. 28 equimolar and co-transcribed 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. Panels B and C show complexes whose subunits are encoded on a single polycistronic operon. See Fig. S3BC for examples of proportional synthesis involving distinct transcripts.

See also Figure S3, Figure S4, Figure S6, Table S3, and Table S4.

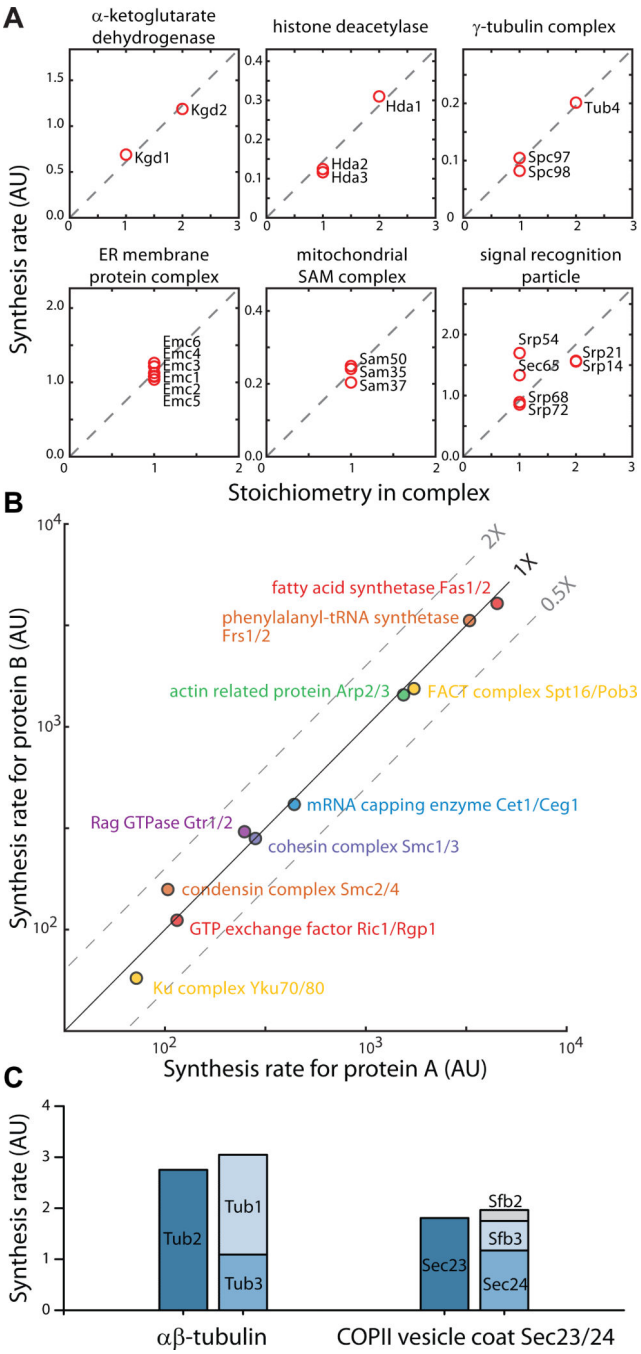

Figure 3. Proportional Synthesis for Complexes in Yeast

(A) Proportional synthesis for multi-protein 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 (Srp14/21/68/72) 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.

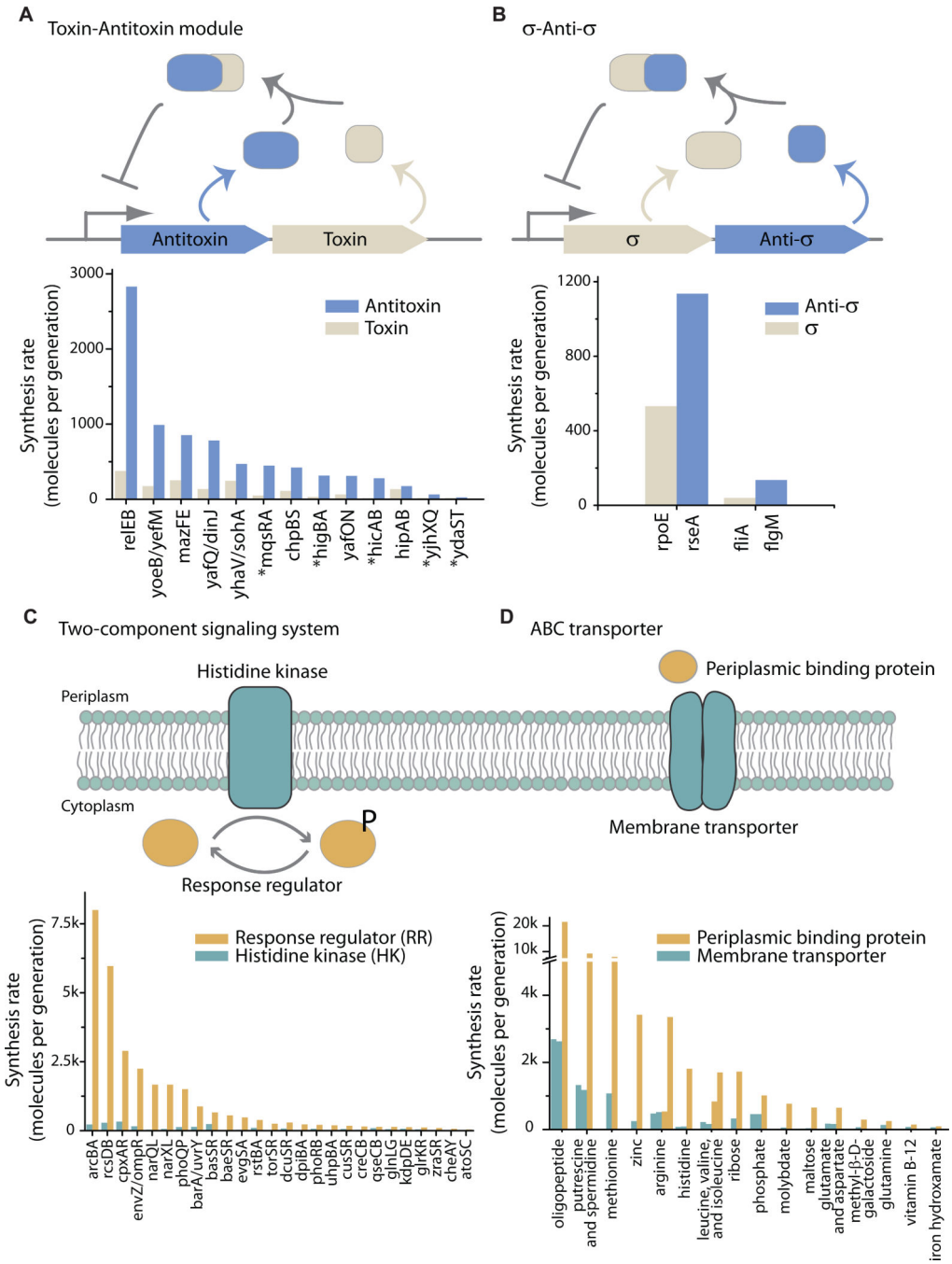

Figure 4. Hierarchical Expression for Functional Modules

(A) Synthesis rates for toxin-antitoxin (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 anti-toxin 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 is marked by asterisk.

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

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

(D) Synthesis rates for ATP-binding cassette (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).

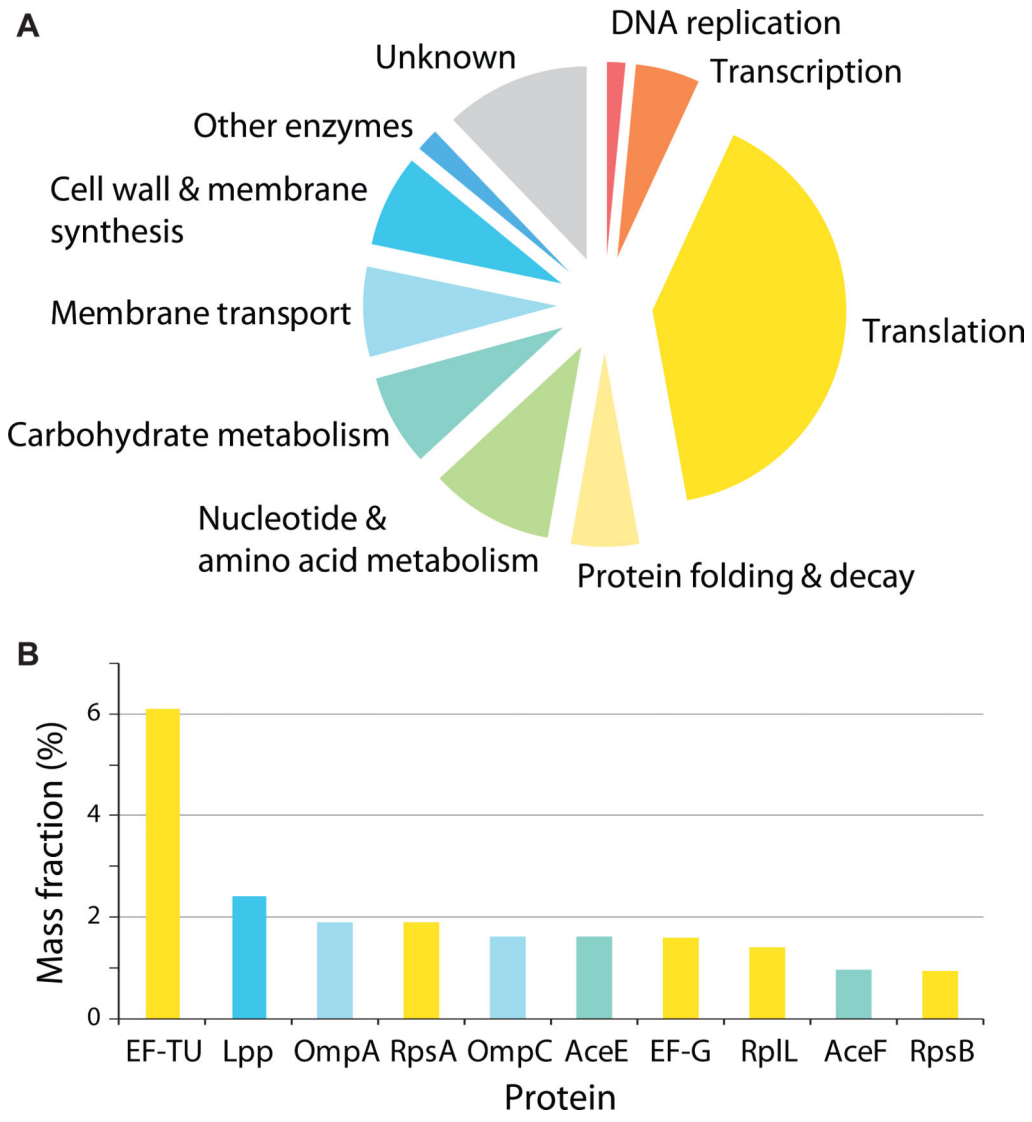

Figure 5. Composition of the *E. coli* Proteome

(A) Break down 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 (Methods).

(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.

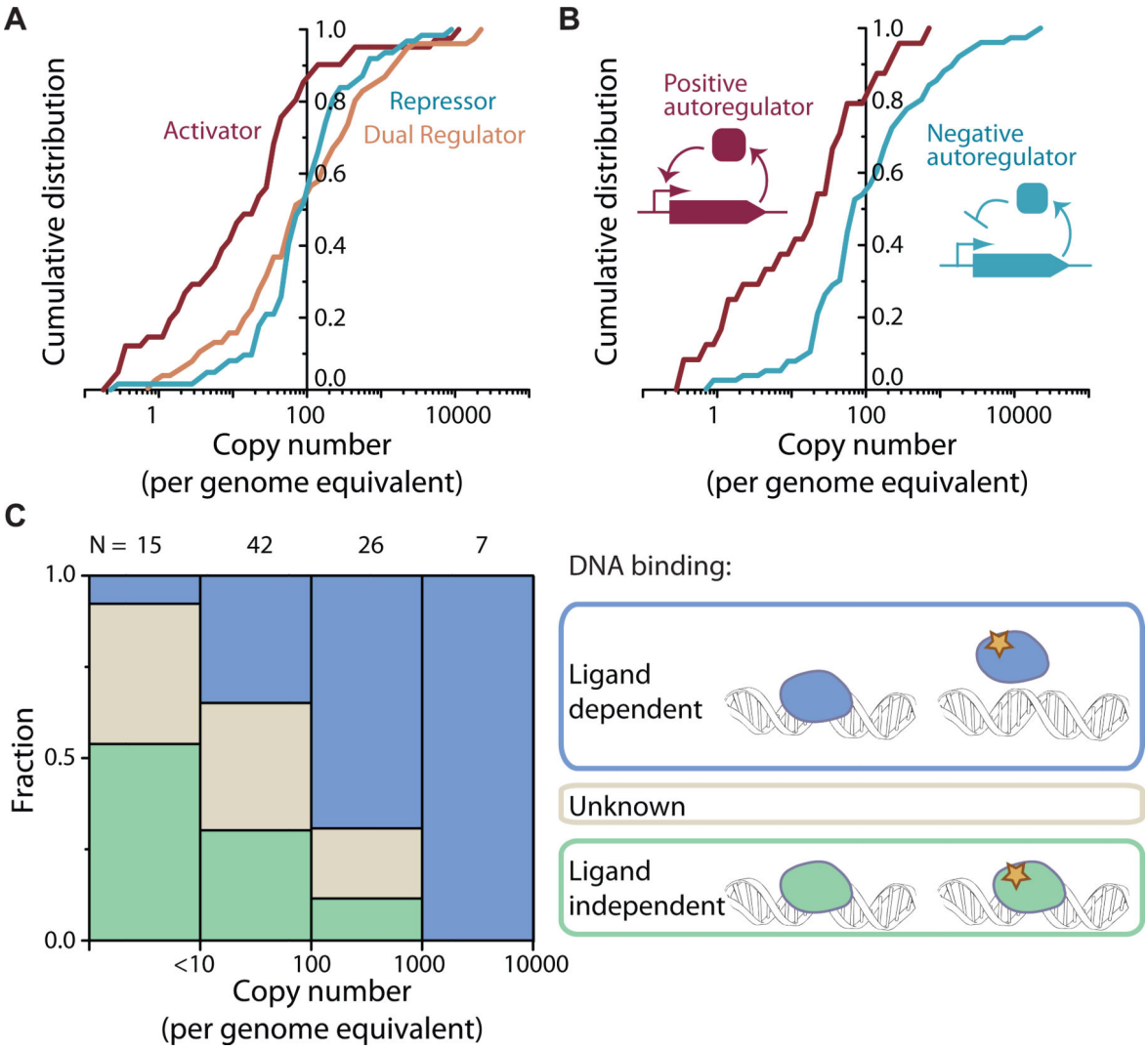

Figure 6. Abundance of Transcription Factors (TFs)

(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 fall 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 transcription factors analyzed is indicated above each bin.

See also Table S5.

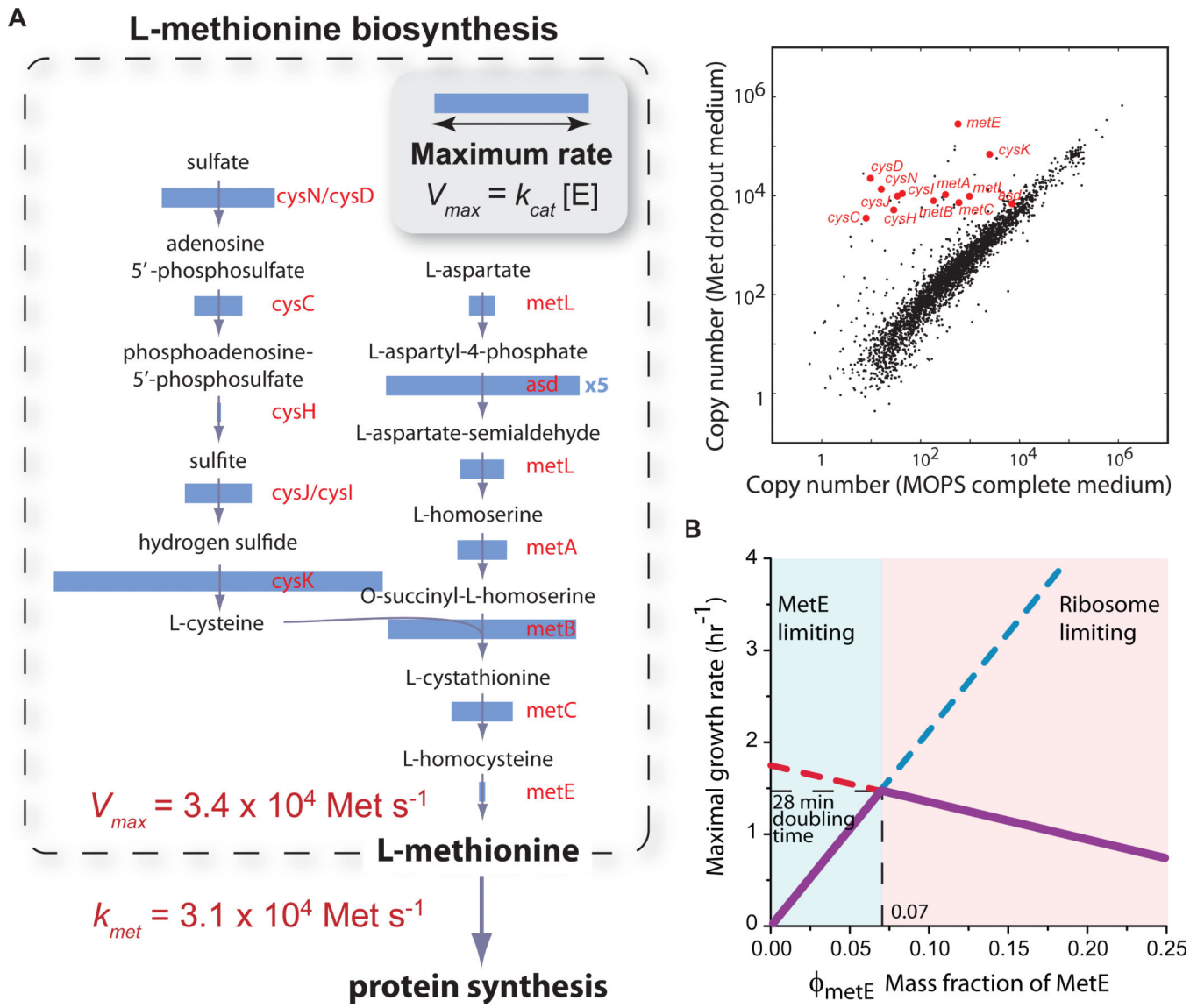

Figure 7. Quantitative Analysis of the Methionine Biosynthesis Pathway

(A) Maximal reaction rates for the intermediate steps. For each step of the pathway, the maximal reaction rate (V_{max}), inferred from the enzyme abundance *in vivo* and the turnover number measured *in vitro*, is shown as the width of the blue bar, unless no *in vitro* data were available. The last step that is catalyzed by the enzyme MetE has V_{max} = 34,000 Met/s/cell, whereas the flux of methionine into protein synthesis is 31,000 Met/s/cell. The scatter plot on the right shows up-regulation of these enzymes in media without methionine.

(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 methionine flux is shown in blue. The maximal growth rate is highest (28 min) when the mass fraction of MetE is ~7%. This prediction agrees with experimental results.

See also Figure S5.