

# Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria

Sheng Hui<sup>1</sup>, Josh M Silverman<sup>2,3</sup>, Stephen S Chen<sup>2,3</sup>, David W Erickson<sup>1</sup>, Markus Basan<sup>1</sup>, Jilong Wang<sup>1</sup>, Terence Hwa<sup>1,4,\*</sup> & James R Williamson<sup>2,3,\*\*</sup>

## Abstract

A central aim of cell biology was to understand the strategy of gene expression in response to the environment. Here, we study gene expression response to metabolic challenges in exponentially growing *Escherichia coli* using mass spectrometry. Despite enormous complexity in the details of the underlying regulatory network, we find that the proteome partitions into several coarse-grained sectors, with each sector's total mass abundance exhibiting positive or negative linear relations with the growth rate. The growth rate-dependent components of the proteome fractions comprise about half of the proteome by mass, and their mutual dependencies can be characterized by a simple flux model involving only two effective parameters. The success and apparent generality of this model arises from tight coordination between proteome partition and metabolism, suggesting a principle for resource allocation in proteome economy of the cell. This strategy of global gene regulation should serve as a basis for future studies on gene expression and constructing synthetic biological circuits. Coarse graining may be an effective approach to derive predictive phenomenological models for other 'omics' studies.

**Keywords** growth physiology; metabolic network; microbiology; quantitative proteomics; systems biology

**Subject Categories** Quantitative Biology & Dynamical Systems; Genome-Scale & Integrative Biology; Metabolism

**DOI** 10.15252/msb.20145697 | Received 18 August 2014 | Revised 11 December 2014 | Accepted 23 December 2014

**Mol Syst Biol.** (2015) **11**: 784

## Introduction

One of the most extensively studied questions in biology is how cells alter gene expression to deal with changes in their environment. A widely held view, supported by a mountain of observations, is the idea that cells handle challenges to growth-limiting

perturbations, for example, nutrient limitation, by increasing the amount of enzymes devoted to overcoming the limited process, in analogy with 'supply and demand' (Hofmeyr & Cornish-Bowden, 2000). This qualitative picture has been widely articulated in conceptual models from early on (Hinshelwood, 1944) to the present and is supported by analyses of 2D gel experiments (O'Farrell, 1975), microarrays (Brown & Botstein, 1999), deep sequencing (Ingolia *et al*, 2009), mass spectrometry (Aebersold & Mann, 2003), and other high-throughput measurements of gene expression (Ghaemmaghami *et al*, 2003; Taniguchi *et al*, 2010). For example, cells grown in minimal media increase the level of amino acid synthesis enzymes compared to rich media, and cells grown in the presence of translation inhibitors increase the synthesis of ribosomes (Dennis, 1976; Tao *et al*, 1999; Boer *et al*, 2003). Despite these results, there is little in the way of a quantitative understanding of resource allocation even in the simplest cells (Chubukov & Sauer, 2014).

Recently, it was shown that simple genetic circuits respond to changes in the physiological state of a cell in different ways, based upon the details of their defined regulation (Klump *et al*, 2009). At the molecular level, a cell's response to an applied limitation is the outcome of a complex interaction of metabolites, transcription factors, promoters, and other factors, conspiring to produce the observed pattern of gene expression. It is therefore unclear how the behavior of single genes under defined and specific regulation can be generalized to shifts in global gene expression arising from environmental changes. Many elementary questions remain unaddressed. In response to a growth-limiting perturbation, by how much does the cell adjust its composition to deal with the limiting process(es)? Does the cell handle limitation in the supply of a given nutrient by adjusting operons related to the specific shortage, or is gene expression organized according to some higher schema? Can the effect of different types of growth limitations be meaningfully compared? From the perspective of analysis, can cellular response, with changes in thousands of quantities as revealed by 'omics' experiments, be summarized by simple quantitative measures beyond statistical analysis? In characterizing the state of a gas, useful quantitative

<sup>1</sup> Department of Physics, University of California at San Diego, La Jolla, CA, USA

<sup>2</sup> Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA

<sup>3</sup> Department of Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA

<sup>4</sup> Section of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, CA, USA

\*Corresponding author. Tel: +1 858 534 7263; E-mail: hwa@ucsd.edu

\*\*Corresponding author. Tel: +1 858 784 8740; E-mail: jrwill@scripps.edu

measures are macroscopic quantities such as pressure and temperature, not the statistical clustering of the trajectories of molecules in the gas. In systems biology, might similar measures exist to provide meaningful quantitative characterization of cellular responses?

Early studies of bacterial physiology identified a number of relations between the cell growth rate and quantities such as chromosome copy number, cell mass, and ribosome content (Schaechter *et al*, 1958; Bremer & Dennis, 2009). Despite the incredible complexity of ribosome biogenesis and its regulation, the proportion of translational machinery among all proteins can be captured by a simple linear relation with the cell growth rate (Bennett & Maaloe, 1974). These observations hint to a quantitative framework underlying the intuitive ‘supply and demand’ picture. The hint is the balance between the flux of amino acids synthesized into proteins by ribosomes and the flux of molecular building blocks from catabolic and biosynthetic reactions culminating in amino acids that are consumed by the ribosomes. This highlights an attractive possibility. If enzymes are regulated as subsets according to their shared purpose, as the hundred or so genes involved in translation are, it may be possible to capture their collective behavior quantitatively as is possible for the translational machinery. Rather than focusing upon the molecular details of hundreds of enzymes as they facilitate myriad reactions, the enzymes of a functional group might instead be profitably viewed as a single effective coarse-grained enzyme that catalyzes interconversion between major metabolic pools, such as carbon precursors to amino acids. In this view, proteome-wide response to nutrient limitations may be characterized quantitatively as adjustments to the concentrations of coarse-grained enzymes. This coarse-grained view of the proteome yields a simple picture that is amenable to mathematical analysis. Recently, the coarse-graining approach has been used to address the effects of protein overexpression (Scott *et al*, 2010), cAMP-mediated catabolite repression (You *et al*, 2013), growth bistability in response to antibiotics (Deris *et al*, 2013), and methionine biosynthesis (Li *et al*, 2014). But these studies focused on the expression of only a few genes, declared to be proxies for hundreds of proteins (Scott *et al*, 2010; You *et al*, 2013), or isolated in a backdrop of changing proteome (Deris *et al*, 2013; Li *et al*, 2014). There has been no study of its global applicability and, indeed, no work to predict quantitative proteome composition from physiological state.

Toward this end, it is our aim to quantitatively characterize global gene expression under various modes of growth limitation and to interrogate the intuitive ideas regarding resource allocation quantitatively. Samples were collected for *E. coli* cells growing exponentially in a variety of growth conditions: under titration of carbon import and nitrogen assimilation, and in the presence of varying amounts of translation inhibitor. Using quantitative mass spectrometry, the relative concentrations of ~1,000 enzymes were measured across the set of growth-limiting conditions. Analysis of the enzyme concentrations reveals six groups of enzymes with distinct modes of gene expression in response to the applied limitations. An enrichment analysis of gene ontology terms appearing in these groups shows that each group consists of enzymes with uniform purpose, such as translation and catabolism. The cell up-regulates relevant groups to counteract the imposed limitation,

confirming the qualitative expectations based on supply and demand. A key to this analysis is the concept of an ‘effective concentration’ for each coarse-grained enzyme, obtained as the fractional abundance of the sum of all its enzyme components among all expressed proteins in each condition. The concentration of the coarse-grained enzymes was estimated using coarse-grained spectral counts as a proxy for protein abundance (Malmström *et al*, 2009). Strikingly, the concentrations of these coarse-grained enzymes correlated linearly with the growth rate. These data, together with the intrinsic constraints of finite resource allocation, led to the construction of a self-consistent, flux-matching model of the proteome that not only quantitatively accounts for all the observed data but also predicts proteome composition in novel environments involving combinatorial modes of growth limitation.

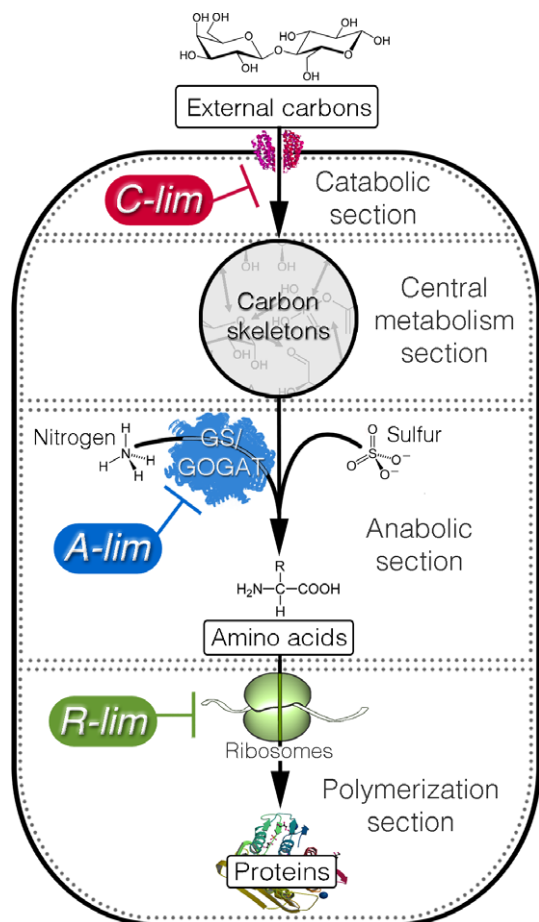
## Results

### Growth limitations

To probe gene expression, cell growth was perturbed by imposing three different modes of growth limitation at crucial bottlenecks in the metabolic network. A coarse-grained metabolic flow diagram for protein production by *E. coli* growing in minimal medium is shown in Fig 1. Four *metabolic sections* act in concert to convert external carbon sources to proteins, incorporating nitrogen and sulfur elements during the process. Following the work of You *et al* (You *et al*, 2013), growth limitation was imposed on three of the four metabolic sections. The limitation imposed on the catabolic section (C-limitation or *C-lim*) was implemented by titrating the expression of lactose permease for cells growing on lactose (Supplementary Fig S1). The limitation on the anabolic section (A-limitation or *A-lim*) was realized by titrating a key enzyme (GOGAT) in the ammonia assimilation pathway (Supplementary Fig S2). Such ‘titratable uptake systems’ have been characterized in detail and found comparable to other modes of growth limitations such as those derived from continuous culture or microfluidic devices (You *et al*, 2013). To impose growth limitation on the polymerization sections, sublethal amounts of a translation inhibitor antibiotic, chloramphenicol, were supplied to the growth medium to inhibit translation by ribosomes (R-limitation or *R-lim*). The collective response of the *E. coli* proteome to these applied growth limitations was monitored using quantitative mass spectrometry.

### Quantitative proteomic mass spectrometry

Proteomic mass spectrometry is a powerful tool for quantifying changes in global protein expression patterns (Aebersold & Mann, 2003; Ong & Mann, 2005; Bantscheff *et al*, 2007; Han *et al*, 2008). As shown below, mass spectrometry also has the advantage of reliably detecting small changes in protein levels, with precision comparable to that of enzymatic assays. Metabolic labeling with <sup>15</sup>N (Oda *et al*, 1999) provides relative quantitation of unlabeled proteins with respect to labeled proteins across growth conditions of interest. Each experimental sample in a series is mixed in equal amount with a known labeled standard sample as reference, and the relative change of protein expression in the experimental sample is obtained for each protein.



**Figure 1. Coarse-grained metabolic flow of protein production and the three modes of growth limitation.**

Through the (carbon) catabolic section, the cells take up external carbon sources and break them down into the set of standard carbon skeletons (pyruvate, oxaloacetate, etc.). The carbon skeletons are interconvertible through the central metabolism section. The anabolic section synthesizes amino acids from the carbon skeletons and other necessary elements such as ammonia and sulfur. The amino acids are then assembled into proteins by the polymerization section. The three modes of growth limitation were imposed on the metabolic sections as shown. The C-limitation (C-lim) and A-limitation (A-lim) were carried out with strains constructed for titrating the catabolic and anabolic flux, respectively; see Supplementary Figs S1 and S2, and Supplementary Table S1. The R-limitation (R-lim) was realized for the WT strain by supplying the growth medium with various levels of an antibiotic, chloramphenicol.

### Accuracy and precision

The accuracy and precision of quantifying relative protein expression levels was determined from a standard curve using samples of unlabeled and  $^{15}\text{N}$ -labeled purified ribosomes. The observed relative levels, measured by ratios of the labeled to the unlabeled ribosomal proteins (or  $^{15}\text{N}/^{14}\text{N}$ ), agree extremely well with the expected values over a range of about two orders of magnitude (Fig 2A). To assess the accuracy and precision for a whole-cell lysate with a much more complex proteome, labeled and unlabeled cells were mixed in fixed ratios and measured with quantitative mass spectrometry. The relative changes in protein levels can be precisely determined over the range of ratios from 0.1 to 10, as shown in Fig 2B. The effective precision of relative protein

quantification is  $\pm 18\%$ , based on analysis of the 1:1 sample (Supplementary Fig S3). Thus, subtle changes in proteome composition that are much  $< 2$ -fold can be precisely determined. Furthermore, the relative quantitation using quantitative mass spectrometry agrees extremely well with a traditional biochemical measurement of ribosome content (Supplementary Fig S4A) and also with quantitation of LacZ using a  $\beta$ -galactosidase assay (Supplementary Fig S4B).

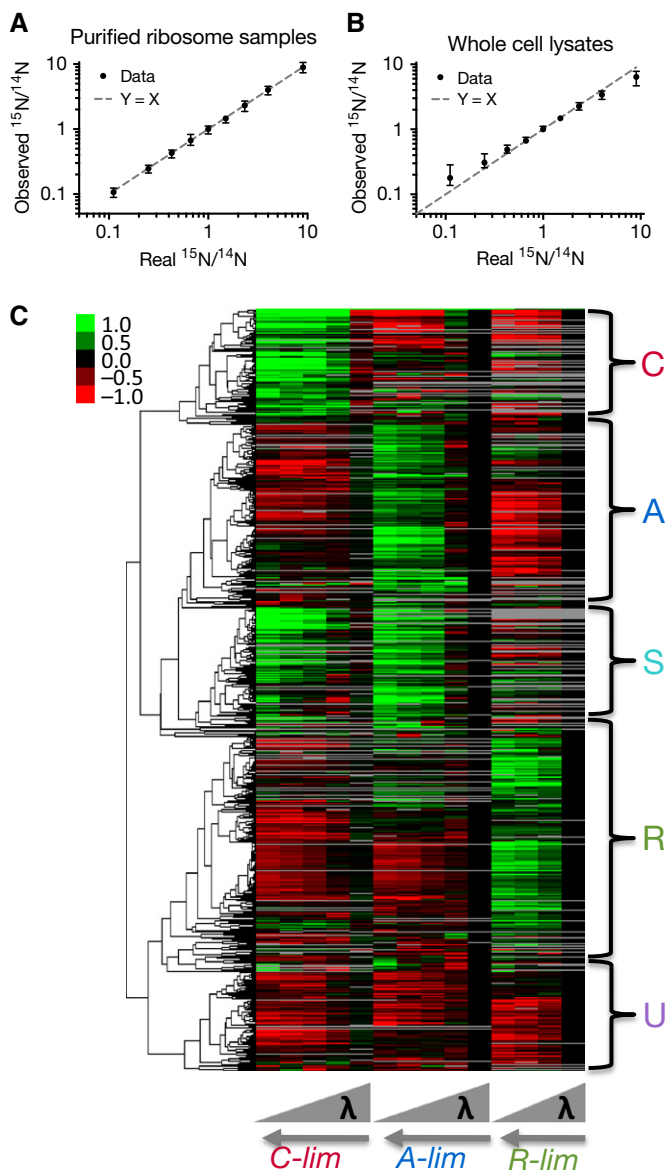
### Datasets and protein coverage

For the C-, A-, and R-limitations, a series of cultures were prepared with varying growth rates. For the C-limitation series, controlled inducible expression of the lacY gene gave doubling times from 40 to 92 min (five conditions), for the A-limitation series, controlled expression of GOGAT gave doubling times from 43 to 91 min (five conditions), and for the R-limitation series, inhibition of protein synthesis with chloramphenicol gave doubling times from 42 to 147 min (four conditions), as detailed in Supplementary Table S1. Samples from each of the fourteen cultures were collected, and the relative protein levels were determined using mass spectrometry, as described in the Materials and Methods. For C-, A-, and R-limitations, the numbers of proteins with reliable expression data are 856, 898, and 756, respectively. Most proteins present in one dataset are present in others, with 616 proteins shared in all three datasets and a total of 1,053 unique proteins in any dataset. Due to a highly non-uniform distribution of protein abundance, our experiments are estimated to cover  $\sim 80\%$  of the total proteome by mass and are validated using absolute abundance estimated by a recent experiment using ribosome profiling (Li *et al*, 2014); see Supplementary Fig S5. For data analysis, the combined datasets were represented as a matrix of 1,053 proteins across the 14 growth conditions (Supplementary Table S2), graphically shown in Fig 2C.

### Clustering analysis of protein expression trends

A qualitative global analysis of the data was performed with hierarchical clustering using the Pearson correlation as a distance metric (Materials and Methods), and the resulting dendrogram is shown on the expression matrix in Fig 2C. Five major clusters are apparent, characterized by different trends in the three limitation series. The cluster where protein levels increase as growth rate is reduced under C-limitation, but decrease under A- and R-limitations, represents proteins that specifically respond to C-limitation and is designated as the C-cluster. The A-cluster is defined by increased protein levels under A-limitation, but decreased levels under C- and R-limitations, responding specifically to A-limitation. Similarly, the cluster where proteins levels increase in response to R-limitation, but decrease under C- and A-limitations, specifically respond to R-limitation and is designated as the R-cluster. The S-cluster is defined by protein levels that increase under both A- and C-limitations. Finally, the cluster for proteins that generally do not respond specifically to any of the three modes of growth limitation is designated as the U-cluster.

The clustering analysis is useful for providing an overview of the trends in the proteomic data, and revealing the qualitative responses of proteins to the different modes of growth limitation: Most proteins respond specifically to a single mode of growth limitation with the exception of the S-cluster. These clusters suggest that proteome levels are strongly coordinated based on the environmental



**Figure 2. The quantitative protein mass spectrometry.**

- A** Observed values versus real values for ratios of  $^{15}\text{N}$  ribosomal proteins to  $^{14}\text{N}$  ribosomal proteins. Black dots are the mean values, with error bars representing the range of the values for all ribosomal proteins. The dashed line represents perfect agreement between the observed values and real values.
- B** Observed values versus real values for ratios of  $^{15}\text{N}$  proteins to  $^{14}\text{N}$  proteins from whole-cell lysates. Black dots are the median values for more than 600 proteins. The error bar for each median value indicates the quartiles. The dashed line represents perfect agreement between the observed values and real values. Additional characterizations are shown in Supplementary Figs S3 and S4.
- C** The expression matrix and clustering results. The matrix is composed of 1,053 proteins (rows) and 14 conditions (columns); see Supplementary Table S2. The first five columns are for C-limitation, the next five columns for A-limitation, and the last four columns for R-limitation. For each mode of growth limitation, the growth rate increases from left to right. The matrix is log<sub>2</sub>-transformed, with expression values at the standard condition as zero (see Materials and Methods), represented as black color. Red color indicates negative values, green color positive values, and gray color missing entries. A dendrogram generated by clustering analysis is shown on the left of the expression matrix (see Materials and Methods), with the five major clusters shown on the right side of the matrix. The data are estimated to cover ~80% of the proteome; see Supplementary Fig S5.

(Supplementary Fig S6A), and is further supported by comparison with a quadratic fit of the data (Supplementary Fig S6B). Possible causes for the occurrence of low  $R^2$  values include limited method precision (Supplementary Fig S7A) and weak growth rate dependence for some genes (Supplementary Fig S7B and C). The approximate linear nature of the protein abundance data suggests that the results may be simplified using a coarse-grained analysis, by summing over the absolute abundance of individual proteins in a cluster (since the sum of linear functions is still linear).

For a protein exhibiting linear growth rate dependence, a negative slope corresponds to a higher expression level at slower growth rate, referred to as the ‘upward’ response ( $\uparrow$ ), while a positive slope corresponds to a lower expression level at slower growth rate, referred to as the ‘downward’ response ( $\downarrow$ ). Given that a protein has either upward or downward response under each of the three modes of growth limitation (C-, A-, and R-limitation), it has to belong to one of the  $2^3 = 8$  groups:  $C^\uparrow A^\downarrow R^\downarrow$ ,  $C^\uparrow A^\uparrow R^\downarrow$ ,  $C^\downarrow A^\downarrow R^\downarrow$ ,  $C^\downarrow A^\uparrow R^\downarrow$ ,  $C^\uparrow A^\downarrow R^\uparrow$ ,  $C^\uparrow A^\uparrow R^\uparrow$ ,  $C^\downarrow A^\downarrow R^\uparrow$ , and  $C^\downarrow A^\uparrow R^\uparrow$ , where the group names are indicated by the upward or downward response under each of the three modes of growth limitation. For example, the  $C^\uparrow A^\downarrow R^\downarrow$  group consists of proteins that have upward response under C-limitation and downward responses under both the A- and R-limitation. The membership of proteins in the resulting eight groups is given in Supplementary Table S2 and graphically shown in Supplementary Fig S8. Due to the precision limitations of the method, proteins exhibiting small change under a specific growth limitation are subject to misclassification. To examine the effect of this misclassification on our results, we carried out a probabilistic classification, by assigning a protein to one of the eight groups according to a probability (see Supplementary Text S2 for details). The analysis shows a very limited effect misclassification has on the binary classification.

The collective behavior of a protein group can be approximated by coarse graining, effectively summing the absolute

stress and that the response of the proteome to the environment might be amenable to a quantitative coarse-graining analysis.

### Coarse-grained proteome sectors

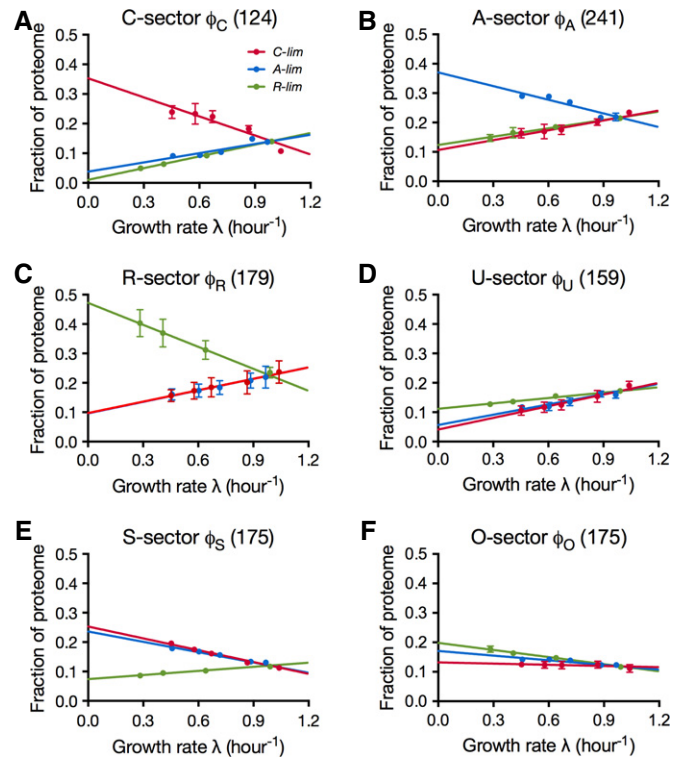
Extensive analysis of a number of exemplary reporters of catabolic and biosynthetic gene expression revealed strikingly linear growth rate dependence in the expression of these genes (You *et al.*, 2013). The prevalence of linear growth rate dependence has been described in omics studies of both proteins (Pedersen *et al.*, 1978) and mRNAs (Brauer *et al.*, 2008). Visual inspection of the expression data of individual proteins in Fig 2C (see Supplementary Dataset S1 for individual plots) suggested that many exhibited a linear trend, and the coefficient of determination ( $R^2$ ) for the expression of each protein was calculated for each mode of growth limitation. The cumulative distribution of  $R^2$  for each mode of growth limitation shows that linear dependence on growth rate is widespread in our data



protein abundance of proteins in the same group. Among the methods for quantifying absolute protein abundance from proteomic mass spectrometry data (Beynon *et al*, 2005; Ishihama *et al*, 2005, 2008; Lu *et al*, 2006; Silva *et al*, 2006; Vogel & Marcotte, 2008; Schmidt *et al*, 2011; Muntel *et al*, 2014), the method of spectral counting takes the number of peptides recorded for each protein as proxy for the absolute abundance of the protein (Malmström *et al*, 2009). While spectral counting provides a crude estimate of the absolute protein abundance for individual proteins (Bantscheff *et al*, 2007), it gives a much more reliable approximation for groups of proteins. For a protein group comprising more than ~5% of the total proteome, spectral counting produces estimates with < 20% error (Supplementary Fig S9A). The comparison of spectral counting data for ribosomal proteins with estimates based on biochemical measurements and the ribosome profiling results (Li *et al*, 2014) is in good agreement (Supplementary Fig S9B). By applying the spectral counting method, the proteome fractions for the nine protein groups defined in Supplementary Table S2 were determined for each of the three series of growth limitations (Supplementary Fig S10). It is clear from Supplementary Fig S10 that some groups occupy significant fractions of the proteome while others are minor constituents. Ranked by the extent the fraction varies (indicated by the difference between the maximal and minimal intercepts on the  $y$ -axis), the top three groups are  $C^{\uparrow}A^{\downarrow}R^{\downarrow}$ ,  $C^{\downarrow}A^{\uparrow}R^{\downarrow}$ , and  $C^{\downarrow}A^{\downarrow}R^{\uparrow}$ . These consist of proteins that only respond upward to the C-, A-, and R-limitation and are referred to as the C-, A-, and R-sector, respectively (Fig 3A–C). The  $C^{\downarrow}A^{\downarrow}R^{\downarrow}$  group includes proteins that are uninduced by any of the three applied limitations, and is referred to as the U-sector (Fig 3D). Another significant protein sector is the  $C^{\uparrow}A^{\uparrow}R^{\downarrow}$  group, which is composed of proteins that have upward response to both the A- and C-limitations, and referred to as the S-sector for general starvation; see Fig 3E. The three remaining groups (i.e.,  $C^{\uparrow}A^{\uparrow}R^{\uparrow}$ ,  $C^{\downarrow}A^{\uparrow}R^{\uparrow}$ , and  $C^{\downarrow}A^{\downarrow}R^{\uparrow}$  groups) are small, with most of the data at or below 5% of the proteome, below the accuracy of the spectral counting method (Supplementary Fig S9A). The three small groups were placed together into the O-sector (Fig 3F). In summary, the proteome is coarse-grained into 6 ‘sectors’: C-, A-, R-, U-, S-, and O-sectors with distinct growth rate dependences as shown in Fig 3, with complete data for all fractions shown in Supplementary Table S3. In contrast, the results obtained for randomly shuffled expression matrices do not show significant growth rate dependence (Supplementary Fig S11).

### Qualitative proteome responses to growth limitations

To elucidate the biological functions for each proteome sector, a Gene Ontology (GO) analysis was carried out using an *abundance-based* GO term enrichment to identify a small number of GO terms that best represent the abundant proteins in a sector. To reach such a list of GO terms, instead of calculating a single score of one measure (e.g., enrichment) for each GO term as in many GO analyses, we have taken a multi-step procedure to search for the best representing GO terms by examining a number of different measures such as coverage and overlap. The procedure leads to only a few GO terms accounting for more than 60% of the proteome in the sector; see Supplementary Text S3.

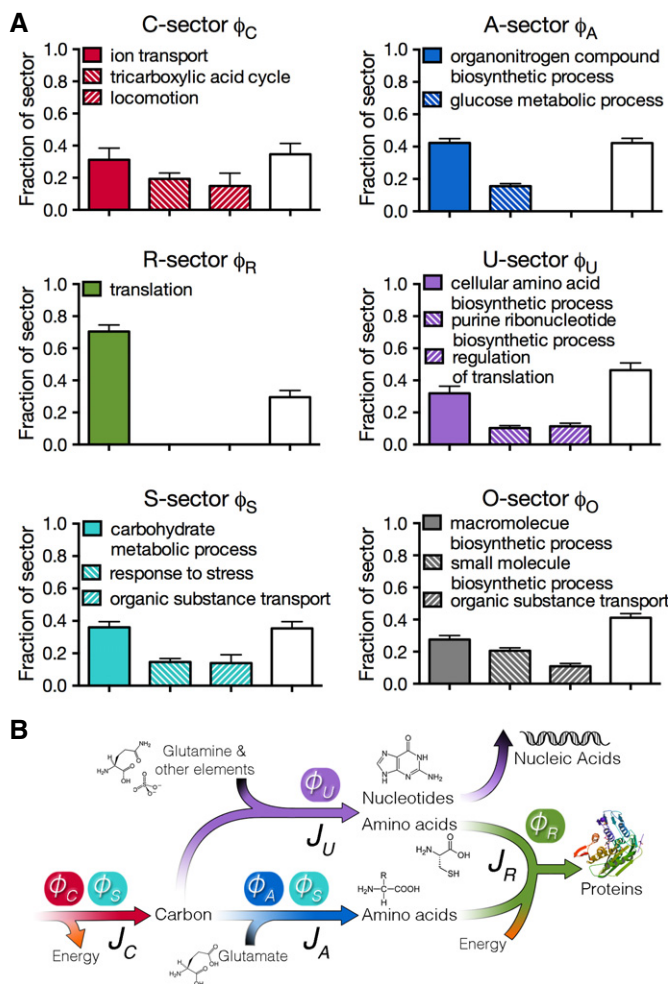


**Figure 3. The coarse-grained proteome sectors.**

A–F Coarse-grained responses of the C-, A-, R-, U-, S-, and O-sectors to the three modes of growth limitation. As indicated in (A), the red symbols in each panel are for C-limitation, the blue for A-limitation, and the green for R-limitation. The error bars indicate the standard deviation of triplicate mass spectrometry runs. Error bars smaller than the corresponding symbols are not shown (see Supplementary Fig S10 on the different degrees of variability associated with different sectors.) On each plot, the number in the title indicates the number of proteins in that sector, and colored lines are best linear fits of the data represented by symbols of the same colors; see Supplementary Table S3 for the data on proteome fraction and Supplementary Table S4 for parameters of the fitted lines.

The results of the analysis are summarized in Fig 4A, with each bar graph describing the major proteome composition for each sector. Sixty percent of the mass fraction of each sector could be accounted for by at most three terms, providing a simple interpretation of the functional significance of the sectors. For example, a single GO term, ‘translation’, describes more than 70% of the proteins in the R-sector. Since the R-limitation inhibits translation rate, the term suggests a strategy by which the cell specifically counteracts the applied growth limitation by increasing the abundance of ‘translational’ proteins (Scott *et al*, 2010).

The GO terms best describing the C-sector are ‘ion transport’, ‘tricarboxylic acid cycle’, and ‘locomotion’, pointing to a mode of carbon scavenging (by moving and increasing carbon uptake) and carbon saving (by increasing the efficiency of energy generation using the tricarboxylic acid cycle) to counteract the imposed carbon limitation. For the A-sector, the most abundant term is ‘organonitrogen compound biosynthetic process’. A closer look reveals that most of the terms are related to biosynthesis of amino acids (Supplementary Table S5). Again, similar to the



**Figure 4. Abundance-based GO analysis.**

**A** Composition of proteome sectors. Each bar graph shows the results of the abundance-based Gene Ontology analysis for each of the six sectors. Each bar indicates the mass fraction the corresponding GO term accounts for within a sector. The empty bar in each graph indicates the remaining fraction of the sector not accounted for by the GO terms listed. The results were calculated based on triplicate runs of all samples. Each bar height indicates the mean result and the standard deviation is shown as the error bar. See Supplementary Table S9 for the list of proteins represented by each bar in each sector, and Supplementary Text S3 for details of the method.

**B** Association between metabolic fluxes and proteome sectors. There are four fluxes  $J_C$ ,  $J_A$ ,  $J_U$ , and  $J_R$ , represented by the arrows, replenishing the pools of carbon precursors, amino acids, other building blocks, and macromolecules, respectively. The  $\phi_s$  on top of the fluxes represents the corresponding proteome fractions that carry the fluxes. Note that the S-sector proteins contribute to both  $J_C$  and  $J_A$ .

responses to R- and C-limitations, the finding here suggests that the cell tries to counteract the imposed A-limitation, which specifically limits the biosynthesis of amino acids (Supplementary Fig S2). Interestingly, ‘glucose metabolic process’ proteins also constitute a significant fraction (~15%) of the A-sector, possibly reflecting the important role of glycolysis in generating precursors for amino acid biosynthesis.

The U-sector consists of proteins that are not up-regulated by any of the growth limitations. More than one-third of the U-sector is

accounted for by the term ‘cellular amino acid biosynthetic process’. The categories of proteins associated with this term are primarily related to cysteine, methionine, and tryptophan biosynthesis (Supplementary Table S5). This is not surprising for cysteine and methionine synthesis since sulfur is not limited in any of the growth limitations imposed here. The same logic applies to the case of tryptophan synthesis enzymes because tryptophan biosynthesis is not directly affected by the particular mode of A-limitation that was applied (trans-amination, see Supplementary Fig S2), nor by the other two limitations. The second abundant term of the U-sector is ‘purine ribonucleotide biosynthetic process’, which is again not targeted by our mode of A-limitation. The third term is ‘regulation of translation’, where a single ribosomal protein RpsA accounts for the majority of the mass fraction. It is surprising that unlike most of other ribosomal proteins, RpsA was not grouped into the R-sector. This is likely a misclassification due to statistical fluctuation. In summary, the U-sector is composed of enzymes that make up a diverse group of building blocks including some amino acids and purine ribonucleotides, with the common trait that they were not specifically limited by the three growth limitations tested.

The major term of the S-sector is ‘carbohydrate metabolic process’, revealing the sector’s role in central metabolism- and energy-related activities. The other two terms are ‘response to stress’ and ‘organic substance transport’. These terms suggest the possible ‘multiple-purpose’ nature of the S-sector proteins that are mobilized in response to starvation conditions via either C- or A-limitation. This notion is best illustrated by the term ‘organic substance transport’, consisting mostly of transporters for peptides and amino acids which can clearly be used to counteract both C- and A-limitations (Supplementary Table S5). The interpretation of the O-sector is less obvious, with the top three terms as ‘macromolecule biosynthetic process’, ‘small molecule biosynthetic process’, and ‘organic substance transport’, reflecting diverse activities of proteins in this sector. Due to the way the O-sector is defined (it results from lumping together three small groups), it is likely that it also includes proteins with weak growth rate dependencies.

In summary, the GO analysis reveals that the R-sector consists mostly of the translational machinery, the C-sector engages in carbon scavenging and saving, the A-sector makes nitrogen-containing building blocks consisting mostly of amino acids, and the U-sector produces other building blocks including sulfur-containing amino acids and purine nucleotides. In exponentially growing cells, these coarse-grained enzymes carry steady fluxes of biomass. As illustrated in Fig 4B, these four metabolic fluxes are denoted as  $J_R$ ,  $J_C$ ,  $J_A$ , and  $J_U$ , respectively, representing a coarse-grained metabolism. The S-sector shares functions with both the C- and A-sectors, thus carrying both  $J_C$  and  $J_A$  fluxes.

### Flux matching

The growth rate dependences of the proteome sectors shown in Fig 3 are well described by linear relations (Supplementary Table S4). Closer scrutiny of the data and the fits in Fig 3 suggests additional simplicity in the structure of the responses. In particular, the downward responses in Fig 3 (positive slopes) are similar for each sector, and such responses are referred to as ‘general’ responses as they are not distinguishable between at least two different modes of limitations. On the other hand, the upward response of each of the

C-, A-, and R-sectors is specific to only the C-, A-, and R-limitation, respectively, and such a response is referred to as a ‘specific’ response. The only exception is the S-sector, which has similar upward responses to both C- and A-limitations, and the O-sector, which is essentially growth rate-independent. These features suggest that there is a fundamental principle underlying the proteome response to environmental challenges.

As summarized in Fig 4B, the GO analysis provides a strong motivation to construct a quantitative flux model for the growth rate dependence of the fluxes associated with each of the proteome sectors. Based on the analysis of the data in Fig 3, the proteome is partitioned into six sectors, or ‘coarse-grained enzymes’,  $\phi_\sigma$ , each of which is comprised of a basal level,  $\phi_{\sigma,0}$ , and a growth rate-dependent component,  $\Delta\phi_\sigma(\lambda)$ , that is,

$$\phi_\sigma(\lambda) = \phi_{\sigma,0} + \Delta\phi_\sigma(\lambda). \quad (1)$$

In our flux model, we make the central assumption that the flux processed by a proteome sector  $\sigma$ ,  $J_\sigma$ , is proportional to the growth rate-dependent component of the corresponding proteome fraction,  $\Delta\phi_\sigma$ , that is,

$$J_\sigma = k_\sigma \cdot \Delta\phi_\sigma \quad (2)$$

where  $k_\sigma$  is a coarse-grained kinetic coefficient describing the efficiency of the metabolic sector  $\sigma$ . The model is an extension of a similar model proposed in a previous work based on the growth rate dependences of a few reporter genes (You *et al*, 2013). The flux of a sector can be defined as the sum of the metabolic products that flow out from the terminal enzymes per unit time, multiplied by a stoichiometric factor that reflects the composition of the material. For the collection of enzymes that we term the R-sector, the flux is clearly the proteins translated by ribosomes, while for the A-sector, it is largely amino acids. Some proteins, such as those involved in chemotaxis, do not directly handle flux in batch culture but are nonetheless coregulated as part of the C-sector, presumably reflecting their role in facilitating carbon flux in *E. coli*'s natural environment. As shown below, the model comprising of equations (1, 2) can quantitatively account for all of the observations summarized in Fig 3.

The ‘downward’ general responses in Fig 3 can be exemplified by the R-sector, where the total protein synthesis flux through the ribosomes is given by  $J_R$ . The R-sector fraction of the proteome ( $\phi_R$ ) is given by

$$J_R = k_R \cdot \Delta\phi_R, \quad (3)$$

where  $k_R$  is the corresponding enzyme kinetic parameter (given by the peptide elongation rate (Scott *et al*, 2010)). In combination with the stoichiometric requirement of the flux for cell growth,  $c_R \cdot J_R = \lambda$ , where  $c_R$  is the stoichiometric coefficient (Varma & Palsson, 1994), the growth rate-dependent proteome fraction for the R-sector is given by

$$\Delta\phi_R(\lambda) = \lambda/v_R, \quad (4)$$

where  $v_R = k_R c_R$  is an effective rate constant for the R-sector. Upon applying the C- or A-limitation, the peptide elongation is not

affected,  $v_R$  is constant and equation (4) describes a linear relation between  $\phi_R$  and  $\lambda$ , which is the ‘general response’. Note that this model explicitly predicts identical general responses for the R-sector under C- and A- limitations (equation S1 of Supplementary Table S6), in good agreement with the data of Fig 3.

Similarly, for the U-sector:

$$\Delta\phi_U(\lambda) = \lambda/v_U. \quad (5)$$

The downward lines of the U-sector in Fig 3 are produced by equation (5) as long as none of the growth limitations affects the value of  $v_U$ , and none of the metabolic processes catalyzed by the U-sector is affected. Thus, the model predicts identical general responses for the U-sector under C-, A-, and R- limitations (equation S2 of Supplementary Table S6), which is consistent with the data in Fig 3.

The responses of the C-, A-, and S- sectors are more complex since the S-sector is composed of proteins that provide both  $J_C$  and  $J_A$  fluxes (Fig 4). This effect is modeled by considering two lists of proteins, called  $\tilde{C}$  and  $\tilde{A}$ , each responding specifically to C- and A-limitation, respectively. We apply the same linear relation between proteome fractions  $\phi_{\tilde{C}}$ ,  $\phi_{\tilde{A}}$  and with the fluxes, that is,

$$\begin{aligned} \phi_{\tilde{C}}(\lambda) &= \phi_{\tilde{C},0} + \lambda/v_C \\ \phi_{\tilde{A}}(\lambda) &= \phi_{\tilde{A},0} + \lambda/v_A \end{aligned}$$

Then S-sector proteins are composed of those proteins that are common to both  $\tilde{C}$  and  $\tilde{A}$ , while C- and A-sector proteins are those unique in  $\tilde{C}$  and  $\tilde{A}$ , respectively (see Supplementary Fig S12). This is modeled as the follows,

$$\begin{aligned} \Delta\phi_C(\lambda) &= (1-f) \cdot \lambda/v_C, \\ \Delta\phi_A(\lambda) &= (1-f) \cdot \lambda/v_A, \\ \Delta\phi_S(\lambda) &= f \cdot [\lambda/v_C + \lambda/v_A], \end{aligned} \quad (6)$$

with  $f$  being the fraction of  $\tilde{C}$ - and  $\tilde{A}$ -sector proteins that are in common (see Supplementary Fig S12). These relations describe the general responses of the C-, A-, and S-sectors (equations S3–S5 of Supplementary Table S6), for growth limitations that do not affect  $v_C$  or  $v_A$ . Note that in a previous proteome partition model (You *et al*, 2013) based on measurements of a few reporter genes, the hypothesized C- and A-sectors correspond respectively to the  $\tilde{C}$ - and  $\tilde{A}$ -sectors here, whereas the possibility of the S-sector was not anticipated. Finally, we assume the existence of a growth rate-independent sector and identify it with the O-sector, that is,  $\phi_O(\lambda) = \phi_{O,0}$  (equation S6 of Supplementary Table S6) with

$$\Delta\phi_O(\lambda) = 0. \quad (7)$$

### Constraint of finite proteome resources

A striking result of this flux model is that the ‘specific’ upward responses of the C-, A-, R-, and S-sectors in Fig 3 can also be produced by equations (4–7), without introducing any additional parameters. For example, under R-limitation, the value of  $v_R$



changes in response to the limitation, and consequently, the growth rate dependence of  $\varphi_R$  can no longer be obtained from equation (4). However,  $\Delta\varphi_R(\lambda)$  can be obtained from the important constraint  $\sum_{\sigma} \phi_{\sigma}(\lambda) = 1$ , or equivalently

$$\sum_{\sigma} \Delta\phi_{\sigma}(\lambda) = \phi_{\max}, \quad (8)$$

where  $\phi_{\max} \equiv 1 - \sum_{\sigma} \phi_{\sigma,0}$ . Since under R-limitation only  $v_R$  is reduced, all other  $s_{\sigma}$  sectors still follow the general responses. Using equations (4–8), the expression for the specific response of the R-sector becomes:

$$\Delta\phi_R(\lambda) = \phi_{\max} - \lambda/\kappa_R, \quad (9)$$

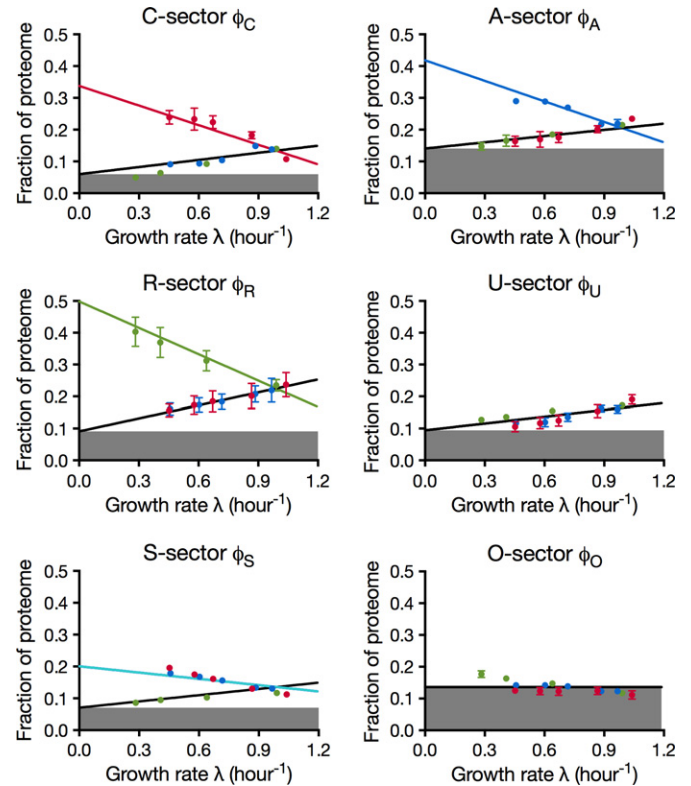
with  $\kappa_R^{-1} \equiv v_C^{-1} + v_A^{-1} + v_U^{-1}$  (equation S7 of Supplementary Table S6). Note that both parameters appearing in equation (9) are determined completely in terms of the parameters already introduced, and an important feature of the flux model is that there is no additional parameter for the specific responses once the general responses are established. In a similar manner, the specific responses of the C-, A-, and S-sectors are obtained in terms of  $\phi_{\max}$  and the  $v_s$ , with no additional parameters; see equations (S8–S10) of Supplementary Table S6 with derivation given in Supplementary Fig S12.

In summary, the linear equations in Supplementary Table S6 describe the prediction of the simple flux model (equations 1, 2) on the partitioning of the proteome as a function of growth rate under the three different modes of growth limitation. Although the model contains only 10 adjustable parameters, the quality of the fit of the model to the data (lines in Fig 5; Supplementary Table S7) is comparable to the 24-parameter fit for each individual response (Fig 3; Supplementary Table S4).

## Two global parameters

The straightforward meanings of the remaining 10 parameters are illustrated by the cartoon in Fig 6. The top pie chart in Fig 6 represents the proteome fractions for the sectors under the glucose standard condition, with the growth rate-independent fraction of the proteome,  $\varphi_Q$  (gray area in the top pie chart) being  $\phi_Q = \sum_{\sigma} \phi_{\sigma,0} = 1 - \phi_{\max}$ . The growth rate-dependent component includes the remainder of every sector, shown as colored wedges, whose proteome fractions make up the rest of the pie,  $\varphi_{\max}$ .

The bottom three pie charts in Fig 6 describe the responses of the proteome to each of the three modes of growth limitation in the extreme case  $\lambda \rightarrow 0$  according to the model. Under extreme R-limitation, the R-sector fraction  $\Delta\varphi_R$  approaches  $\varphi_{\max}$ , while under C-limitation,  $\varphi_{\max}$  is partitioned into  $\Delta\varphi_{C,\max} = (1 - f) \cdot \varphi_{\max}$  and  $\Delta\varphi_{S,\max} = f \cdot \varphi_{\max}$ , and under A-limitation,  $\varphi_{\max}$  is partitioned into  $\Delta\varphi_{A,\max} = (1 - f) \cdot \varphi_{\max}$  and  $\Delta\varphi_{S,\max} = f \cdot \varphi_{\max}$ . Note that the growth rate-dependent responses  $\Delta\varphi_{\sigma}(\lambda)$  are described effectively by only two parameters,  $\varphi_{\max}$  and  $f$ .  $\varphi_{\max}$  provides a cap on the magnitude of the growth rate-dependent component of each sector. The best-fit value,  $\varphi_{\max} \approx 40\%$ , is in quantitative agreement with previous estimates based on the ribosomal content (Scott *et al.*, 2010) and a few reporters (You *et al.*, 2013).



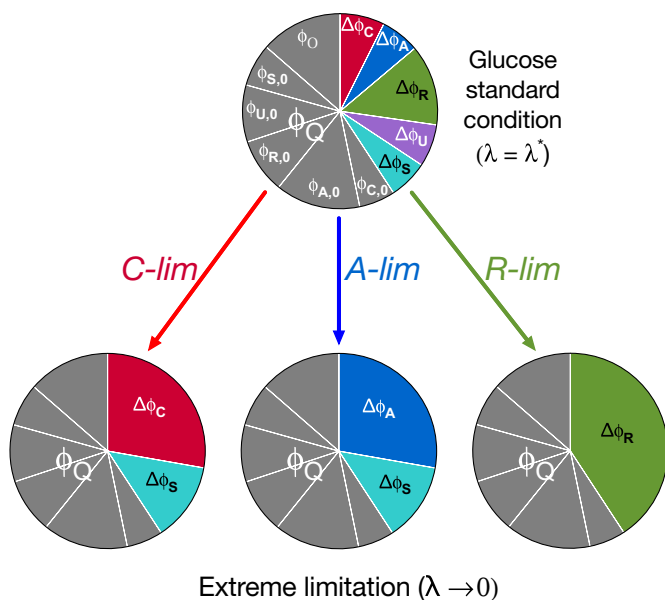
**Figure 5. Performance of the proteome-based flux model.**

The data points are identical to those in Fig 3. The lines here are the result of a global fit to the predictions of the flux-based proteome model (Supplementary Table S6). The growth rate-independent component of each sector ( $\varphi_{\sigma,0}$ ) is represented as the height of the filled area in the corresponding plot. See Supplementary Table S7 for parameters of the fitted lines.

## Model prediction and testing

Among the 10 parameters of the model, the four values of  $v_{\sigma}$ s are dependent on the growth medium, while the  $\varphi_{\sigma,0}$ s as well as the constant  $f \approx 0.32$  are expected to be medium independent for a given strain. All of the data described so far (summarized in Fig 3) were obtained using glucose minimal medium as the standard condition (with the  $v_{\sigma}$ s taking on the values  $v_{\sigma}^*$ ), with each mode of growth limitation corresponding to varying one of the  $v_{\sigma}$ s away from  $v_{\sigma}^*$ . The proteome flux model also makes explicit predictions on the response of the proteome under combinatorial modes of growth limitation, corresponding to varying multiple  $v_{\sigma}$ s. The effect of varying multiple  $v_{\sigma}$ s can be treated as simply repeating the single mode of growth limitations for different standard conditions. This prediction was tested by repeating the proteomic experiments under C- and A-limitations using a different standard condition, for growth in the glycerol minimal medium (Supplementary Table S8). Compared to the standard condition with glucose minimal medium, the glycerol minimal medium should differ by only the value of  $v_C$ , which is fixed by the growth rate for the glycerol standard condition (Supplementary Table S7). Using this new value of  $v_C$ , together with the values of the other nine parameters obtained from the glucose data, the model describes the new data remarkably well (Fig 7; Supplementary Table S9). Thus, the model can describe



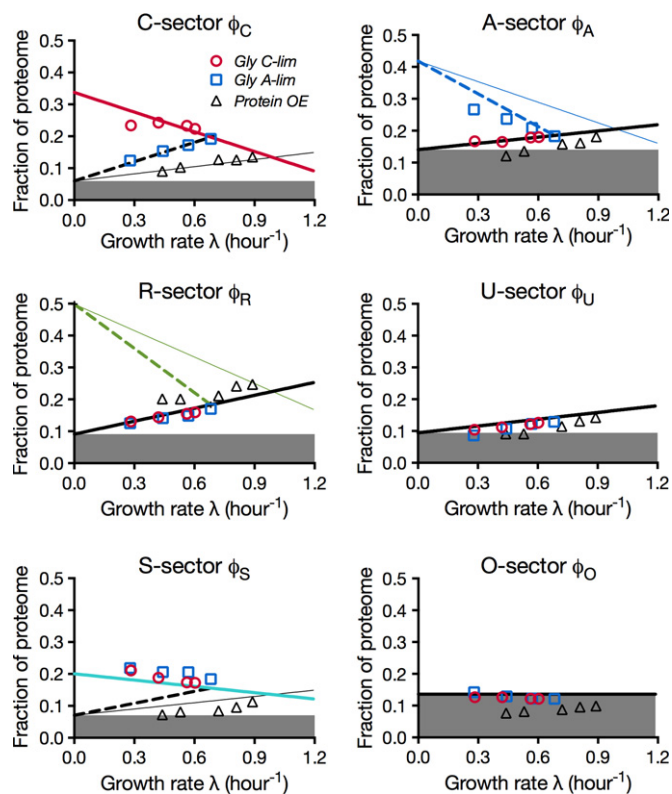


**Figure 6. Representation of the proteome responses under extreme growth limitations and interpretation of the model parameters.** The growth rate-independent component of the protein is represented as  $\phi_Q$  (the entire gray area), which is composed of the growth rate-independent components of the C-, A-, R-, U-, and S-sectors, and the O-sector. See Supplementary Table S7 for their values. The growth rate-dependent part of a sector  $\sigma$  is labeled as  $\Delta\phi_\sigma$ , distinguished by the different colors. The colored wedges in the top pie chart show the sizes of these sectors,  $\Delta\phi_\sigma(\lambda^*)$ , under the glucose standard condition (with growth rate  $\lambda^*$ ). Their values are as follows:  ${}^2\varphi_C = 0.07$ ,  ${}^2\varphi_A = 0.06$ ,  ${}^2\varphi_R = 0.13$ ,  ${}^2\varphi_U = 0.07$ , and  ${}^2\varphi_S = 0.06$ . The pie charts at the bottom show the sizes of these sectors under the three modes of growth limitations in the extreme limit  $\lambda \rightarrow 0$ . These sizes are governed by two parameters,  $\varphi_{\max} = 1 - \varphi_Q$  and  $f \approx 0.32$ .

experiments in different standard conditions, an important benchmark for its ability to capture proteome responses to combinatorial limitations.

As a further test of the model, and specifically, the notion of general response, we studied another way of growth limitation by expressing useless proteins (Scott *et al.*, 2010), which reduces the proteome fraction available for the six sectors, that is, it reduces  $\varphi_{\max}$ . This was achieved by using a LacZ-overexpressing strain (NQ1389) grown on glucose minimal medium (Supplementary Table S8). Since the applied growth limitation does not single out any metabolic sector, our model predicts that every proteome sector except the O-sector should exhibit general responses (equations 4–7), with the same slope  $v_{\sigma S}$  as those obtained from C-, A-, and R-limitations. In addition, the y-offset of every sector,  $\varphi_{\sigma,0}$  in equation (1), should remain unchanged. (Mathematically, we expect only  $\varphi_{\max}$  to be reduced while all other nine model parameters to remain constant.)

The proteome fraction data (Supplementary Table S9) in response to LacZ overexpression are shown as the black triangles in Fig 7. We see that there is generally a congruence of the black triangles and the black solid lines (general responses with glucose minimal medium as the standard condition) as predicted, except for the O-sector which also showed substantial reduction as growth rate is decreased. This suggests that perhaps a good share of the proteins that got classified into the O-sector actually exhibit growth



**Figure 7. Proteome fractions under growth limitations with respect to the glycerol standard condition, and under growth limitation by expressing useless proteins.**

Proteome fractions  $\varphi_\sigma$  for C- and A-limitation under glycerol standard condition are shown as the red and blue circles, respectively, for each of the six sectors; see Supplementary Table S9 for values. All thick lines are model predictions for responses under the glycerol standard condition. Thick solid lines describe responses which are predicted to be unchanged between the glucose and glycerol standard conditions, because these lines do not involve the parameter  $v_C$ , which has a new value for the new standard condition according to the model. Thick dashed lines describe responses which are predicted to be unique for the glycerol standard condition, due to their dependence on the value of parameter  $v_C$ . See Supplementary Fig S12 describing the dashed lines for the C-, A-, R-, and S-sectors. For comparison, the four respective proteome responses under glucose standard condition are also shown as thin solid lines. All solid lines are from Fig 5. Note that the new value of  $v_C$  is determined from the growth rate of cells in glycerol standard condition (Supplementary Table S7). Thus, all predictions for the glycerol standard condition were generated with no adjustable parameters. Proteome fractions under growth limitation by LacZ overexpression are shown as the black triangles for the six sectors; see Supplementary Table S9 for values.

rate dependence which were obscured by noise under C, A-, and R-limitations.

## Discussion

Understanding the principles of the global regulation of gene expression is a major goal of systems biology. However, the intricacy of genetic regulatory networks makes this goal difficult to realize through bottom-up analysis. Quantitative measurements of the concentrations of thousands of proteins, mRNA, and metabolites in cells have recently become possible. These techniques invite revealing measurements of the differential composition of the cell as a

function of growth condition, as well as a framework capable of describing the resource allocation of the cell. Coarse-graining procedures offer a means to capture the allocation of proteome resources by the cell using such data, without detailed knowledge of thousands of enzyme rates, binding constants, and regulatory relationships.

Here, we measured the quantitative response of ~1,000 proteins in *E. coli* as cells are progressively limited in three broad manners: limiting carbon uptake through the lactose importer, limiting nitrogen assimilation through the GS-GOGAT pathway, and limiting protein translation using sublethal amounts of chloramphenicol. Analysis of the individual protein concentrations in the three limitation series suggests six distinct sectors of the proteome, and abundance-based GO term enrichment reveals a functional coherence across the enzymes of each sector that is largely orthogonal to the functions of the other sectors. Note that the sectors are revealed by the nature of growth limitations applied and we expect other sectors to emerge under other growth limitations.

### Coarse-grained proteome sectors

During balanced exponential growth, a constant flux of matter from environmental nutrients cascades through the metabolic network of the cell to form biomass. In contrast to bottom-up descriptions of the metabolic network as an object of great complexity (e.g., the KEGG map), our results reveal an enzymatic network that is simply coarse-grained according to the functional grouping of the proteome sectors. Strikingly, the mass fractions of the various proteome sectors increase or decrease in approximately linear fashion with the change in cell growth rate, which serves as a quantitative measure of the degree of the applied limitation. For instance, the C-sector, consisting mostly of carbon catabolic proteins, increases linearly in response to the limitation of carbon influx.

The control of proteome partition is likely orchestrated by sophisticated regulatory networks that integrate information from multiple signaling molecules. Some of these signals are well known, for example, ppGpp directs ribosome synthesis in accordance with the level of amino acid depletion (Ross *et al*, 2013), cAMP-Crp coordinates catabolic protein expression in accordance with the availability of alpha-keto acids (You *et al*, 2013), and the PII/NtrBC system determines the degree of nitrogen assimilation in accordance with the availability of glutamine (Reitzer, 2003). However, many mysteries remain. The coherent response of the proteins in the anabolic sector is well beyond what is known to be controlled by nitrogen regulatory system, and the enzymes for amino acid synthesis and nucleic acid synthesis are clearly distinguished in their responses. Further, a substantial number of proteins are in the S-sector which responds to both C- and A-limitations; yet little similarity can be seen based on their promoter regions. It is possible that major pleiotropic regulators are yet to be discovered or that the roles of some existing pleiotropic regulators are to be reappraised (as has been done recently for cAMP-Crp (You *et al*, 2013), a well-characterized regulator whose function was long thought to be understood). The simple behaviors of the proteome sectors revealed in this work are molecular phenotypes that can be relied upon in future studies to identify the coordinators of such coherent responses. Importantly, a number of high abundance proteins of as yet unknown function reside within the coarse-grained functional sectors we identified.

In recent years, a number of studies have characterized the growth rate dependence of relative mRNA abundance in Baker's yeast under various nutrient-limiting conditions in chemostat (Regenberg *et al*, 2006; Castrillo *et al*, 2007; Levy *et al*, 2007; Brauer *et al*, 2008; Airoidi *et al*, 2009). Brauer *et al*, Castrillo *et al*, and Regenberg *et al* report a large group of mRNA that increase with carbon limitation, and are characterized by the enriched GO terms cellular carbohydrate metabolism, cellular macromolecule catabolism, transport, and response to stress, recalling our C- and S-sectors (Regenberg *et al* normalize their data such that 42 ORFs with negative correlations with the growth rate become growth rate independent. With this in mind, growth rate-independent groups should obtain negative  $\lambda$  dependence.). Notably, Levy *et al* report a decrease in ribosomal protein mRNA synthesis as the cell exits exponential growth, while Airoidi *et al* successfully predict growth rate in *S. cerevisiae* from the behavior of a few reporter mRNA, which comports with our finding that the majority of proteins change with growth rate in a characteristic fashion (Supplementary Text S4).

### Proteome fraction as a quantitative measure

The similarities mentioned above suggest that the principles we uncover here may be fundamental to metabolism in a variety of cell types. However, we question the effectiveness of using mRNA measurements to infer protein concentrations which protein activities depend on (see Supplementary Text S5). A primary lesson drawn from the linear relation between the ribosome and growth rate is that cells operate at saturated translational capacity (Maaloe, 1979; Scott *et al*, 2010). As a result, mRNA and protein levels are not expected to couple tightly due to mRNA competition for the limited number of ribosomes, a phenomenon also known as ribosome queueing (Mather *et al*, 2013). Indeed, studies that compare mRNA and protein levels from *H. sapiens* to *E. coli* generally report poor correlations that fail to provide predictive power (Pearson correlation  $r_p \approx 0.5$ ) (Maier *et al*, 2009; Taniguchi *et al*, 2010; Vogel *et al*, 2010). Further, mRNA levels are typically reported as a fraction of total fluorescence intensity, that is, as a fraction of total mRNA, without keeping track of the change in total mRNA concentration across different growth conditions [e.g., cell volume can change many fold under nutrient limitation (Schaechter *et al*, 1958)], so that mRNA measurements do not necessarily correlate with mRNA concentration (Klump *et al*, 2009). Also, coarse graining requires knowledge of absolute concentrations, as can be provided by methods such as quantitative mass spectrometry and ribosome profiling (Li *et al*, 2014; Muntel *et al*, 2014), without which the true cost of gene expression is difficult to quantify. For example, a 50% increase in proteome mass fraction from 20% to 30% should not be compared to a 50% increase from 0.1% to 0.15%.

So far, the cost associated with making proteins has been quantified mostly for useless proteins (Andrews & Hegeman, 1976; Dong *et al*, 1995; Kurland & Dong, 1996; Scott *et al*, 2010; Shachrai *et al*, 2010) and, in some cases, proteins with specific functions (Dekel & Alon, 2005). Several recent studies have found that the protein cost plays a key role in understanding regulatory strategies in metabolism (Molenaar *et al*, 2009; Wessely *et al*, 2011; Flamholz *et al*, 2013). Genomescale computational models that integrate protein

resource allocation to existing constraint-based models (e.g., stoichiometry constrained models) have been proposed as a step forward in unraveling intricate relations between growth, metabolism, and gene expression (Goelzer & Fromion, 2011; O'Brien *et al*, 2013). The quantitative data provided in this study will hopefully stimulate further quantitative studies along these directions.

### Open questions on regulatory mechanisms

Though mass spectrometry and ribosome profiling are capable of providing absolute abundances and are thus crucial to coarse-graining analysis of the proteome, there are deeper, regulatory layers to resource allocation (requiring other techniques) that remain unaddressed. As mentioned above, one urgent question raised by our results is the molecular basis for the coordination of genes in the A- and S-sectors, which could be addressed by transcription factor profiling and quantitative metabolomics. Significant action of a riboswitch or small RNA in translational regulation would be invisible to mass spectrometry and ribosome profiling; here, RNA-seq would be an obvious method to use. Finally, growth conditions with significant protein degradation (e.g., during growth transition; Kuroda, 2001) would lead to chronic overestimates of protein level by ribosome profiling and underestimates of translational cost by mass spectrometry. In these conditions, both methods could be rescued by an accurate determination of individual protein degradation rates (e.g., by pulse labeling).

### Principles of resource allocation

Despite the lack of detailed regulatory information, we showed that a phenomenological model that stipulates flux matching in the flow of material between the sectors of the coarse-grained reaction network, along with the constraint of a finite proteome, is sufficient to quantitatively capture the observed sector behavior over a range of growth rates, with only a few parameters. These governing principles imply that the flux through each sector  $\sigma$  is carried by a mass fraction  $\phi_\sigma$  whose size is determined by the cost of supplying flux through the given sector under the given mode of growth limitation, which is given by  $v_\sigma^{-1} \equiv \partial\phi_\sigma/\partial\lambda$ . In this way, the cell's proteome management is analogous to the economic concept of 'division of labor' (Hayek, 1945), with finite capital allocated according to an effective pricing system given by the  $v_{\sigma,s}$  (Lovell, 2004; Mankiw, 2011). When a sector is specifically challenged, such as the C-sector under carbon uptake limitation, the price to carry flux through the C-sector,  $\partial\phi_C/\partial\lambda$ , is increased while the price to carry flux through the other sectors remains the same. This requires an increased investment of capital, for example, proteome fraction, to carry the requisite flux  $J_C$ . These results provide a quantitative framework to buttress the common depictions of cell metabolism, putting the conceptual device of supply and demand on rigorous footing.

### Possible origins of the growth rate-independent sectors

While the growth rate-dependent components (the  ${}^2\phi_{\sigma,s}$ ) closely follow economic principles, much of the growth rate-independent component ( $\phi_Q$ ) comprises of offsets of the identified proteome sectors, that is, the  $\phi_{\sigma,0s}$  as shown in Fig 6. A variety of possible mechanisms have been proposed for the R-sector offset: A favorite

early model was the existence of a fraction of non-translating ribosomes; see Scott *et al* (2010) and references there. Zaslaver *et al* (2009) obtained an offset from *ad hoc* optimization scheme, while Klumpp *et al* (2013) proposed another mechanism based on the growth rate dependence of tRNA. For the offsets of the metabolic sectors, the simplest mechanism could be the biophysical difficulty to tightly repress gene expression, since a zero offset requires protein synthesis to be completely turned off at zero growth rate. It was shown that abundance in the growth rate-independent sector directly diminishes the maximal growth rate (Scott *et al*, 2010). However, this seemingly wasteful allocation of proteome resources may serve a purpose that transcends the simple economics of steady-state growth. For example, keeping substantial offsets on hand may help bacteria adapt more quickly to varying nutrient conditions (Kjeldgaard *et al*, 1958; Koch & Deppe, 1971; Dennis & Bremer, 1974). Competing considerations may well arise at very slow growth, or in starvation conditions, adding to the principles of proteome management revealed by this work.

This work specifically examined cells kept at moderate growth rates, and it is unclear when or whether the observed linear relations cease to hold at slower growth. We emphasize that our goal here is to describe the data by a minimal model with predictive power. We do not rule out nonlinear generalizations of the model presented here.

## Materials and Methods

Detailed bacterial growth protocol, procedures used for strain construction, total RNA and total protein measurements, and  $\beta$ -galactosidase assay are described in the Supplementary Text S1.

### Growth conditions

All growth media used in this study were based on the MOPS-buffered minimal medium used by Cayley *et al* (1989) with slight modifications. See Supplementary Text S1 for the composition of the base medium. The lactose minimal medium and the glucose minimal medium had 0.2% (w/v) lactose and 0.2% (w/v) glucose in addition to the base medium, respectively. For the C-limitation growth, 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and various concentrations (0–500  $\mu$ M) of the inducer 3-methylbenzyl alcohol (3MBA) were added to the lactose minimal medium. For the A-limitation growth, various concentrations of IPTG (30–100  $\mu$ M) were added to the glucose minimal medium. Various concentrations of chloramphenicol (0–8  $\mu$ M) were used for the glucose minimal medium for the R-limitation growth. For the C-limitation growth with NQ399, 0.2% (w/v) glycerol was added to the MOPS base medium, in addition to 1 mM IPTG and various concentrations (0–500  $\mu$ M) of 3MBA. The same glycerol minimal medium with no 3MBA and various amounts of IPTG was used for the A-limitation on glycerol.

### $^{15}\text{N}$ -labeled proteomic mass spectrometry

#### Sample preparation

1.8 ml of cell culture at  $\text{OD}_{600} = 0.4\text{--}0.5$  during the exponential phase of the experimental culture (defined above) was collected by

centrifugation. The cell pellet was re-suspended in 0.2 ml water and fast-frozen on dry ice.

Aliquot of the  $^{15}\text{N}$  reference cell sample (or labeled cell sample) was mixed with each of the  $^{14}\text{N}$  cell samples (or non-labeled cell samples), which contained the same amount of proteins. Each aliquot of the  $^{15}\text{N}$  samples contained about 100  $\mu\text{g}$  of proteins. Each of the  $^{14}\text{N}$  cell samples also contained about 100  $\mu\text{g}$  proteins. For each mode of growth limitation, a  $^{15}\text{N}$  reference cell sample was made in such a way that it contained cell samples from both the fastest and slowest growth conditions under that growth limitation. The mixed reference is used to avoid the composition of proteins in the reference cell sample be biased by a particular growth medium.

Proteins were precipitated by adding 100% (w/v) trichloroacetic acid (TCA) to 25% final concentration. Samples were let stand on ice for a minimum of 1 h. The protein precipitates were spun down by centrifugation at 16,000  $g$  for 10 min at 4°C. The supernatant was removed and the pellets were washed with cold acetone. The pellets were dried in a Speed-Vac concentrator.

The pellets were dissolved in 80  $\mu\text{l}$  100 mM  $\text{NH}_4\text{HCO}_3$  with 5% acetonitrile (ACN). Then, 8  $\mu\text{l}$  of 50 mM dithiothreitol (DTT) was added to reduce the disulfide bonds before the samples were incubated at 65°C for 10 min. Cysteine residues were modified by the addition of 8  $\mu\text{l}$  of 100 mM iodoacetamide (IAA) followed by incubation at 30°C for 30 min in the dark. The proteolytic digestion was carried out by the addition of 8  $\mu\text{l}$  of 0.1  $\mu\text{g}/\mu\text{l}$  trypsin (Sigma-Aldrich, St. Louis, MO) with incubation overnight at 37°C.

The peptide solutions were cleaned by using the PepClean C-18 spin columns (Pierce, Rockford, IL). After drying in a Speed-Vac concentrator, the peptides were dissolved into 10  $\mu\text{l}$  sample buffer (5% ACN and 0.1% formic acid).

#### Mass spectrometry

The peptide samples were analyzed on an AB SCIEX TripleTOF 5600 system (AB SCIEX, Framingham, MA) coupled to an Eksigent NanoLC Ultra system (Eksigent, Dublin, CA). The samples (2  $\mu\text{l}$ ) were injected using an autosampler. The samples were first loaded onto a Nano cHiPLC Trap column 200  $\mu\text{m} \times 0.5$  mm ChromXP C18-CL 3  $\mu\text{m}$  120 Å (Eksigent) at a flow rate of 2  $\mu\text{l}/\text{min}$  for 10 min. The peptides were then separated on a Nano cHiPLC column 75  $\mu\text{m} \times 15$  cm ChromXP C18-CL 3  $\mu\text{m}$  120 Å (Eksigent) using a 120-min linear gradient of 5–35% ACN in 0.1% formic acid at a flow rate of 300 nl/min. MS1 settings: mass range of  $m/z$  400–1,250 and accumulation time 0.5 s. MS2 settings: mass range of  $m/z$  100–1,800, accumulation time 0.05 s, high sensitivity mode, charge state 2–5, selecting anything over 100 cps, maximal number of candidate/cycle 50, and excluding former targets for 12 s after each occurrence.

#### Protein identification

The raw mass spectrometry data files generated by the AB SCIEX TripleTOF 5600 system were converted to Mascot generic format (mgf) files, which were submitted to the Mascot database searching engine (Matrix Sciences, London, UK) against the *E. coli* SwissProt database to identify proteins. The following parameters were used in the Mascot searches: maximum of two missed trypsin cleavage, fixed carbamidomethyl modification, variable oxidation modification, peptide tolerance  $\pm 0.1$  Da, MS/MS tolerance  $\pm 0.1$  Da, and

1+, 2+, and 3+ peptide charge. All peptides with scores less than the identity threshold ( $P = 0.05$ ) were discarded.

#### Relative protein quantitation

The raw mass spectrometry data files were converted to the .mzML and .mgf formats using conversion tools provided by AB Sciex. The .mgf files were used to identify sequencing events against the Mascot database. Finally, spectra for peptides from the Mascot search were quantified using least-squares Fourier transform convolution implemented in house (Sperling *et al*, 2008). Briefly, data were extracted for each peak using a retention time and  $m/z$  window enclosing the envelope for both the light and heavy peaks. The data are summed over the retention time, and the light and heavy peaks amplitudes are obtained from a fit to the entire isotope distribution, yielding the relative intensity of the light and heavy species. The ratio of the non-labeled to labeled peaks was obtained for each peptide in each sample.

The relative protein quantitation data for each protein in each sample mixture was then obtained as a ratio by taking the median of the ratios of its peptides. No ratio (i.e., no data) was obtained if there was only one peptide for the protein. The uncertainty for each ratio was defined as the two quartiles associated with the median. To filter out data with poor quality, the ratio was removed for the protein in that sample if at least one of its quartiles lied outside of 50% range of its median; Furthermore, ratios were removed for a protein in all the sample mixtures in a growth limitation if at least one of the ratios has one of its quartiles lying outside of the 100% range of the median.

Since the ratios are all defined relative to the same reference sample, they represent the relative change of the expression of the protein across all the non-labeled cell samples and are referred as 'relative expression data'.

#### Absolute protein quantitation

The spectral counting data used for absolute protein quantitation were extracted from Mascot search results. For our  $^{15}\text{N}$  and  $^{14}\text{N}$  mixture samples, only the  $^{14}\text{N}$  spectra were counted. The absolute abundance of a protein was calculated by dividing the total number of spectra of all peptides for that protein by the total number of  $^{14}\text{N}$  spectra in the sample.

#### Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno *et al*, 2014) via the PRIDE partner repository with the dataset identifier PXD001467. Plots of individual proteins under the three growth limitations are available as in Supplementary Dataset S1.

#### Data analysis

##### Expression matrices

For each of the growth limitation, the relative expression data can be represented in the form of an expression matrix. For example, under C-limitation, the expression matrix is  $N \times 5$ , where  $N$  is the number of proteins and 5 is the number of growth rates. To focus on proteins with high-quality data, a protein entry (i.e., a row in the matrix) is removed if the number of nonempty data elements for the protein is  $< 3$ . As described in the main text, the sizes of the three



final expression matrices are  $856 \times 5$ ,  $898 \times 5$ , and  $756 \times 4$ , respectively, for the C-, A-, and R-limitation.

#### Scaling of the expression matrices

Because different  $^{15}\text{N}$  reference samples were used for different modes of growth limitation, it is convenient to rescale the relative expression data, so that for each protein the value is set to 1 under a 'glucose standard condition', which was the condition of WT NCM3722 cells growing in glucose minimal medium. Note that for both the A-limitation and R-limitation, the unlimited condition (or the fastest growth condition) was exactly the standard condition. For the C-limitation, however, the standard condition was not one of the growth conditions. The growth rate of the standard condition was between the fastest growth condition (with a doubling time of 40 min) and the second fastest growth condition (with a doubling time of 48 min). Assuming protein expression follows a linear relation under C-limitation, the expression level for the standard condition was determined by extrapolating the expression levels for the two neighboring growth rates.

#### Clustering analysis of the expression data

After scaling, the three expression matrices were merged into a  $1,053 \times 14$  expression matrix, with 14 ( $= 5 + 5 + 4$ ) for the total number of growth conditions and 1,053 for the total number of unique proteins. The pairwise distance ( $d$ ) used for clustering was defined as  $d = 1 - \rho$ , where  $\rho$  is the Pearson correlation:

$$\rho = \frac{\sum_i X_i \cdot Y_i}{\sqrt{\sum_i X_i^2} \cdot \sqrt{\sum_i Y_i^2}},$$

where  $X_i$  and  $Y_i$  are the log<sub>2</sub>-transformed relative expression data for two proteins at same growth condition  $i$ . The Matlab (The Mathworks, Natick, MA) function 'linkage.m' was used to carry out a hierarchical clustering with the option of 'unweighted average distance'. The results were written in the format for the cluster viewing software Java TreeView (Saldanha, 2004).

#### Measure of the quality of a fit

We used the coefficient of determination,  $R^2$  as a measure of fit quality. Assuming a dataset has values  $y_i$ , and the predicted values  $f_i$  based on the fit,  $R^2$  is defined as  $R^2 \equiv 1 - \sum_i (y_i - f_i)^2 / \sum_i (y_i - \bar{y})^2$ , where  $\bar{y}$  is the mean of the values  $y_i$ . The value of  $R^2$  ranges from 0 to 1, with larger number meaning high quality of fit. For a linear fit,  $R^2$  indicates the degree of linearity of the data.

**Supplementary information** for this article is available online: <http://msb.embopress.org>

#### Acknowledgments

We benefited from helpful discussions with numerous colleagues through the course of this work, including Hirota Mori, Kenji Nakahigashi, Steen Pedersen, Alexander Schmidt, Matthew Scott, Yang Shen, John Yates, and members of the Hwa and Williamson labs. We particularly thank Sherry Niessen, Elizabeth Valentine, and Catherine Wong who helped with mass spectrometry, and Jessica Lynn Figueroa for artistic contributions to the figures. This work is supported by the NIH to JRW (R37-GM053357) and to TH (R01-GM095903), by the NSF through the Center for Theoretical

Biological Physics (PHY0822283), and by the Simons Foundation (330378). JW acknowledges support of a scholarship from the China Scholarship Council. JMS acknowledges the support of a fellowship from the ARCS Foundation. The data deposition to the ProteomeXchange Consortium was supported by PRIDE Team, EBI.

#### Author contributions

SH, JMS, TH, and JRW designed the study. SH, JMS, SSC, and DWE developed mass spectrometry method. MB and JW constructed the key strains. SH acquired experimental data and developed the abundance-based gene ontology analysis algorithm. SH, JMS, SSC, TH, and JRW analyzed the data. SH, JMS, TH, and JRW formulated the model and wrote the manuscript.

#### Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198–207
- Airoidi EM, Huttenhower C, Gresham D, Lu C, Caudy AA, Dunham MJ, Broach JR, Botstein D, Troyanskaya OG (2009) Predicting cellular growth from gene expression signatures. *PLoS Comput Biol* 5: e1000257
- Andrews KJ, Hegeman GD (1976) Selective disadvantage of non-functional protein synthesis in *Escherichia coli*. *J Mol Evol* 8: 317–328
- Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389: 1017–1031
- Bennett P, Maaloe O (1974) The effects of fusidic acid on growth, ribosome synthesis and RNA metabolism in *Escherichia coli*. *J Mol Biol* 90: 541
- Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* 2: 587–589
- Boer VM, de Winder JH, Pronk JT, Piper MD (2003) The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J Biol Chem* 278: 3265–3274
- Brauer M, Huttenhower C, Airoidi E, Rosenstein R, Matese J, Gresham D, Boer V, Troyanskaya O, Botstein D (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* 19: 352
- Bremer H, Dennis PP (2009) Modulation of chemical composition and other parameters of the cell at different exponential growth rates. In *EcoSal- Escherichia coli and Salmonella Cellular and Molecular Biology*, Böck A, Curtiss III R, Kaper JB, Karp PD, Neidhardt FC, Nyström T, Slauch JM, Squires CL, Ussery D, Schaechter E (eds), pp 1–95. Washington, DC: ASM Press
- Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21: 33–37
- Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst D, O'Donoghue K, Hester SS, Dunkley TP, Hart SR, Swainston N, Li P et al (2007) Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol* 6: 4
- Cayley S, Record MT, Lewis BA (1989) Accumulation of 3-(N-morpholino) propanesulfonate by osmotically stressed *Escherichia coli* K-12. *J Bacteriol* 171: 3597–3602

- Chubukov V, Sauer U (2014) Environmental dependence of stationary phase metabolism in *Bacillus subtilis* and *Escherichia coli*. *Appl Environ Microbiol* 12: 327–340
- Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436: 588–592
- Dennis PP, Bremer H (1974) Differential rate of ribosomal protein synthesis in *Escherichia coli* B/r. *J Mol Biol* 84: 407–422
- Dennis PP (1976) Effects of chloramphenicol on the transcriptional activities of ribosomal RNA and ribosomal protein genes in *Escherichia coli*. *J Mol Biol* 108: 535–546
- Deris JB, Kim M, Zhang Z, Okano H, Hermsen R, Groisman A, Hwa T (2013) The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science* 342: 1237435
- Dong H, Nilsson L, Kurland CG (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J Bacteriol* 177: 1497–1504
- Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci USA* 110: 10039–10044
- Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* 425: 737–741
- Goelzer A, Fromion V (2011) Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim Biophys Acta* 1810: 978–988
- Han X, Aslanian A, Yates JR III (2008) Mass spectrometry for proteomics. *Curr Opin Chem Biol* 12: 483–490
- Hayek FA (1945) The use of knowledge in society. *Am Econ Rev* 35: 519–530
- Hinshelwood CN (1944) Bacterial growth. *Biol Rev* 19: 150–163
- Hofmeyr JS, Cornish-Bowden A (2000) Regulating the cellular economy of supply and demand. *FEBS Lett* 476: 47–51
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324: 218–223
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 4: 1265–1272
- Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genom* 9: 102
- Kjeldgaard NO, Maaloe O, Schaechter M (1958) The transition between different physiological states during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol* 19: 607–616
- Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139: 1366–1375
- Klumpp S, Scott M, Pedersen S, Hwa T (2013) Molecular crowding limits translation and cell growth. *Proc Natl Acad Sci USA* 110: 16754–16759
- Koch A, Deppe C (1971) In vivo assay of protein synthesizing capacity of *Escherichia coli* from slowly growing chemostat cultures. *J Mol Biol* 55: 549
- Kurland CG, Dong H (1996) Bacterial growth inhibition by overproduction of protein. *Mol Microbiol* 21: 1–4
- Kuroda A (2001) Role of inorganic polyphosphate in promoting ribosomal protein degradation by the Lon protease in *E. coli*. *Science* 293: 705–708
- Levy S, Ihmels J, Carmi M, Weinberger A, Friedlander G, Barkai N (2007) Strategy of transcription regulation in the budding yeast. *PLoS ONE* 2: e250
- Li G-W, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157: 624–635
- Lovell MC (2004) *Economics with Calculus*. Hackensack, NJ: World Scientific Publishing Co
- Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2006) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25: 117–124
- Maaloe O (1979) Regulation of the protein synthesizing machinery – ribosomes, tRNA, factors, and so on. In *Biological Regulation and Development Volume 1 – Gene expression*, pp. 1–56.
- Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583: 3966–3973
- Malmström J, Beck M, Schmidt A, Lange V, Deutsch EW, Aebersold R (2009) Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* 460: 762–765
- Mankiw NG (2011) *Principles of Economics*, 6th edn. Stamford, CT: Gengage Learning
- Mather WH, Hasty J, Tsimring LS, Williams RJ (2013) Translational cross talk in gene networks. *Biophys J* 104: 2564–2572
- Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5: 323
- Muntel J, Fromion V, Goelzer A, Maass S, Mader U, Buttner K, Hecker M, Becher D (2014) Comprehensive absolute quantification of the cytosolic proteome of *Bacillus subtilis* by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MSE). *Mol Cell Proteomics* 13: 1008–1019
- O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 9: 693
- Oda Y, Huang K, Cross FR, Cowburn D, Chait BT (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA* 96: 6591–6596
- O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250: 4007–4021
- Ong S, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1: 252–262
- Pedersen S, Bloch P, Reeh S, Neidhardt F (1978) Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* 14: 179–190
- Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, Bro C, Hansen LK, Brunak S, Nielsen J (2006) Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol* 7: R107
- Reitzer L (2003) Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu Rev Microbiol* 57: 155–176
- Ross W, Vrentas CE, Sanchez-Vazquez P, Gaal T, Gourse RL (2013) The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. *Mol Cell* 50: 420–429
- Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20: 3246–3248
- Schaechter M, Maaloe O, Kjeldgaard N (1958) Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *Microbiology* 19: 592
- Schmidt A, Beck M, Malmström J, Lam H, Claassen M, Campbell D, Aebersold R (2011) Absolute quantification of microbial proteomes at different states by directed mass spectrometry. *Mol Syst Biol* 7: 510
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330: 1099–1102

- Shachrai I, Zaslaver A, Alon U, Dekel E (2010) Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Mol Cell* 38: 758–767
- Silva JC, Gorenstein MV, Li G-Z, Vissers JPC, Geromanos SJ (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 5: 144–156
- Sperling E, Bunner AE, Sykes MT, Williamson JR (2008) Quantitative analysis of isotope distributions in proteomic mass spectrometry using least-squares fourier transform convolution. *Anal Chem* 80: 4906–4917
- Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, Xie XS (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329: 533–538
- Tao H, Bausch C, Richmond C, Blattner F, Conway T (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* 181: 6425
- Varma A, Palsson BO (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnology* 12: 994–998
- Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz P-A, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus H-J, Albar JP, Martínez-Bartolomé S et al (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 32: 223–226
- Vogel C, Marcotte EM (2008) Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. *Nat Protoc* 3: 1444–1451
- Vogel C, Abreu Rde S, Ko D, Le S-Y, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM, Penalva LO (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* 6: 400
- Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C (2011) Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. *Mol Syst Biol* 7: 515
- You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, Wang Y-P, Lenz P, Yan D, Hwa T (2013) Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* 500: 301–306
- Zaslaver A, Kaplan S, Bren A, Jinich A, Mayo A, Dekel E, Alon U, Itzkovitz S (2009) Invariant distribution of promoter activities in *Escherichia coli*. *PLoS Comput Biol* 5: e1000545



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.