UNIVERSITY OF CALIFORNIA, SAN DIEGO

Allocation of Proteome Resources in *E. coli*

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Physics (Biophysics)

by

Sheng Hui

Committee in charge:

Professor Terence Hwa, Chair
Professor Alexander Groisman
Professor Suckjoon Jun
Professor Milton H. Saier
Professor Immo E. Scheffler

2014

The dissertation of Sheng Hui is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2014

DEDICATION

To my mother and Dora

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FILES

Table2_2.xlsx

Table2_6.xlsx

ACKNOWLEDGEMENTS

First, I want to thank my family for their support, patience, or love, even though it has been a challenge for them to really comprehend why it is taking me so long. I am especially indebted to my mother, whose love I will never be able to give back. I pray she can see her son successfully completed his study.

Thank you Dora for coming with me to San Diego. You have been amazing over these years and I couldn't have asked for more support and love.

I would like to acknowledge the help and support of everyone in the Hwa group, especially Dr. Terry Hwa for showing me the rigorous way of doing science, starting from exploratory stage to writing a paper. Thanks for your detailed comments in emails, presentations, and manuscripts. I have learned a lot from you. I thank Dr. Conghui You for patiently teaching me molecular biology skills when I knew nothing about biological experiments.

I am indebted to Dr. Jamie Willamson for the precious opportunity to work with your lab. I have benefited a lot from the collaboration. I thank all the people in the Willamson lab, especially Josh Silverman, Stephen Chen, and Dr. Elizabeth Valentine for a lot of their help.

In addition, I owe gratidude to Dr. Alex Groisman, Dr. Suckjoon Jun, Dr. Herbert Levine, Dr. Milton Saier, and Dr. Immo Scheffler for serving on my thesis committee and providing valuable comments on my work.

With permission from the coauthors, chapters 2-3 of this dissertation contain work

from a manuscript that has been submitted for publication with the working title, "Metabolic partitioning of the bacterial proteome as a principle of global gene regulation" 2014, Hui, S; Silverman, J; Chen, S; Erickson, D; Wang, J; Hwa, T; Williamson, J.

With permission from the coauthors, Chapters 4-5 of this dissertation contain work from a manuscript that has been submitted for publication with the working title, "Coordination of carbon supply and proteomic resources for energy metabolism results in acetate overflow" 2014, Basan, M; Hui, S; Shen, Y; Zhang, Z; Williamson, J; Hwa, T. The dissertation author and Markus Basan are the co-first authors of the work.

VITA

2002	B.Sc., Hong Kong Baptist University
2005	M.Phil., Hong Kong Baptist University
2005-2006	Research Assistant, Hong Kong Baptist University
2006-2007	Ph.D. student, Rutgers University
2014	Ph.D., University of California, San Diego

Publications

M. Basan*, **S. Hui*,** Y. Shen, Z. Zhang, J. R. Williamson and T. Hwa. Coordination of energy metabolism with carbon supply and proteomic resources results in acetate overflow. In preparation.

S. Hui, J. M. Silverman, D. W. Erickson, J. Wang, T. Hwa*, and J. R. Williamson*. Metabolic partitioning of bacterial proteome as a principle of global regulation of gene expression. In submission.

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

G. Alexe, R. Vijaya Satya, M. Seiler, D. Platt, T. Bhanot, **S. Hui**, M. Tanaka, A. J. Levine and G. Bhanot. PCA and Clustering Reveal Alternate mtDNA Phylogeny of N and M Clades. *J. Mol. Evol.* 67, 465-87 (2008)

S. Hui and L.-H. Tang, Ground state and glass transition of the RNA secondary structure. *Eur. Phys. J. B* 53, 77-84 (2006)

ABSTRACT OF THE DISSERTATION

Allocation of Proteome Resources in *E. coli*

by

Sheng Hui

Doctor of Philosophy in Physics (Biophysics)

University of California, San Diego, 2014

Professor Terence Hwa, Chair

Protein mass is a major constituent of bacterial cell dry weight. An exponentially growing cell diverts most of its nutrient uptake to protein synthesis. The optimal allocation of a cell's total proteins, or proteome, to various cellular functions is a major concern for the cell growing in different growth conditions. Obligatory relations between ribosomal proteome fraction and growth rate impose further constraints on proteome resources allocation. In this dissertation, I am interested in understanding the global strategy of proteome resources allocation and exploring its physiological consequences in *E. coli*.

To understand the global strategy of proteome allocation, we apply a series of metabolic challenges to probe the responses of the proteome of exponentially growing *E. coli* using quantitative mass spectrometry. Despite the enormous complexity in the details

of the underlying regulatory network, the proteome partitions into several coarse-grained sectors whose total abundances exhibit linear relations with the growth rate. These growth-rate dependent proteome fractions comprise about half of the proteome by mass, and their mutual dependencies can be characterized quantitatively by a simple proteome-based flux model involving only two effective parameters. The success and apparent generality of this simple model is due to the tight coordination between proteome partition and metabolism, suggesting a principle for resource-allocation in proteome economy. This general strategy of global gene regulation should serve as a basis for future studies on gene expression and constructing synthetic biological circuits.

We also show that proteome resources allocation has important consequences in cell physiology. An important yet mysterious phenomenon in bacterial physiology is the overflow metabolism, where the cell excretes acetate in the case of *E. coli*. We observed a set linear relations between acetate excretion rate and growth rate under different modes of growth limitation. By including both the carbon and proteome resources in a simple resource allocation model, we can quantitatively describe the observed linear relations. Key model parameters of protein costs were determined directly using flux analysis and protein mass spectrometry measurements. These results suggest optimal allocation of carbon and proteome resources as a possible driving force for the occurrence of overflow metabolism.

Chapter 1
Introduction

1.1 Proteome and proteome fraction

The proteome of a bacterium can refer to the total collection of different proteins of the bacteria, with each protein having a unique sequence. It can also refer to the total protein mass of a bacterial cell. In this dissertation, I use the latter definition for proteome. For an exponentially growing bacterial cell such as a *E. coli* cell, its proteome is the major component of the cell dry mass (Neidhardt et al., 1990). The process of protein synthesis also consumes most of the energy produced by the cell (Neidhardt et al. 1990). The cell has to optimally allocate its proteome resources to various cellular functions in different growth environments, which is exemplified by the bacterial growth laws as the obligatory relations between proteome resources allocated to ribosomes and growth rate (Scott et al., 2010). Here we study the global strategy of the proteome resources allocation, and understand an important physiological phenomenon from the perspective of proteome resources allocation.

One emerging view of cell physiology is based on the partitioning of the bacterial proteome into a set of functional sectors, which are groups of genes connected by their common purpose, such as carbon catabolism, amino acid synthesis, or translation (Scott et al., 2010; You et al., 2013). In this view, the size and composition of these sectors are mutually constrained by their codependence on the finite resources of the cell, in

particular proteome resources, as well as by their respective metabolic roles in utilizing those resources. The proteome fraction of a sector, the total mass of sector proteins relative to the total cell protein mass, naturally emerges as a quantity of central importance, which is further supported by the early studies of bacterial physiology (Bennett and Maaloe, 1974; Bremer and Dennis, 2009; Schaechter et al., 1958). The interdependent set of constraints on proteome sector size suggests that the cell is subject to a principle of proteome economy. Adjustments to the environment such as a shift in the abundance or quality of carbon and nitrogen sources, or the presence of drug, will change the demand for the functions of each proteome sector. The cell must adapt to such a shift in metabolic demands by changing the sector mass fractions, effectively remodeling the distribution of resources across the functional sectors of the proteome.

1.2 The strategy of proteome resources allocation

In Chapter 2, we describe the strategy of proteome resources allocation by *E. coli* in the exponential growth phase, where the cell produces a steady flux of matter from nutrient to biomass through the processes supported by each sector at steady state. Using two strains engineered with either titratable carbon uptake or titratable amino acid synthesis, and using the protein synthesis inhibitor chloramphenicol on wild type cells, three key metabolic sections are modulated directly by specifically limiting their flux capacity with respect to a common reference condition. The global coordination of gene expression was monitored across three series of growth limitation modes (carbon influx, nitrogen assimilation, translation rate) using quantitative mass spectrometry (Aebersold and Mann, 2003; Bantscheff et al., 2007; Chen et al., 2012; Han et al., 2008; Ong and

Mann, 2005) in conjunction with hierarchical clustering. Most of the ~1,000 observed proteins exhibit a linear change with growth rate in each of the three series. Proteins were grouped using a binary classifier into one of eight groups according to the observed increases or decreases in mass fraction under each growth limitation. Six distinct sectors of the proteome are readily identified, composed of several hundred proteins each. The proteome fraction of each functional sector exhibits a striking linear relationship parameterized by the growth rate of the cell. Here, the careful selection of limitation conditions and abundance of data allows for quantitative analysis to reveal such functional groups and capture their responses accurately by simple mathematics. As detailed in Chapter 3, using an abundance-based term enrichment to analyze the gene ontology descriptions of member proteins, the function of each proteome sector is found to be well described by just two or three descriptive terms, suggesting that the proteins of a particular sector are characterized by a uniformity of purpose. A simple proteome-based flux model is then formulated based on the mutual constraints and flux matching (Orth et al., 2010; Sauer, 2006; Varma and Palsson, 1994; Yuan et al., 2008), capturing the observed behavior of the sector mass fractions. The model is further validated by the successful prediction of sector behavior in additional test conditions involving multiple modes of growth limitation.

The coordinated expression of ~1,000 observed proteins, which might be expected to exhibit an arbitrarily complex behavior, can be described by an incredibly simple linear strategy that coordinates the sectors to balance flux. For *E. coli* in balanced growth, the proteome fraction for each sector obeys the laws of supply and demand according to an orderly proteome economy.

1.3 A physiological consequence of proteome resources allocation

In Chapter 4, we study the consequence of the proteome resources allocation in an important physiological phenomenon in *E. coli*. Under anaerobic conditions, organisms ranging from bacteria to mammalian cells excrete large quantities of fermentation products such as acetate or lactate. Strikingly, the excretion of these fermentation products occurs widely even in the presence of oxygen in fast-growing bacteria and fungi (De Deken, 1966; Wolfe, 2005), as well as mammalian cells including stem cells, immune cells, and cancerous cells (Vander Heiden et al., 2009; Weinberg, 2013). This seemingly wasteful phenomenon, in which fermentation is used instead of the more efficient respiration process for energy generation, is generally referred to as "overflow metabolism" (or the Warburg effect in the case of cancer (Vander Heiden et al., 2009; Weinberg, 2013)). Various rationalizations of overflow metabolism have been proposed over the years (Molenaar et al., 2009; Valgepea et al., 2010; Vander Heiden et al., 2009; Wolfe, 2005). However, experimental tests of the proposed hypotheses as well as systematic studies of the origin of overflow metabolism are lacking.

In Chapter 4, we provide a quantitative, physiological study of overflow metabolism for the bacterium *E. coli*. We report an intriguing set of distinct linear relations between the rate of acetate excretion and the rate of steady state growth for *E. coli* in different nutrient environments and different degrees of induced stresses. These results suggest that overflow metabolism in these diverse settings arises from a common origin, namely from the coordination of carbon resources with proteome resources. A simple resources allocation model, based on a recently established proteome partition

model (Scott et al., 2010) can quantitatively account for all of the observed behavior, assuming only an increased protein cost of respiration relative to fermentation. As described in details in Chapter 5, this hypothesis is confirmed by direct measurements and analysis of metabolic fluxes and protein costs using protein mass spectrometry. Our findings suggest that incorporation of protein resources allocation may constitute a key ingredient for understanding metabolic phenomena and their regulation.

Chapter 2

The responses of proteome to various growth limitations

2.1 Introduction

In this chapter, I describe a systems and quantitative approach to probe the responses of proteome to different modes of growth limitation, in order to grasp the global strategy the cell takes to manage its proteome.

2.2 Systematic growth perturbation

To probe responses of the proteome, 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 Figure 2.1. 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 (Figure 2.2). The limitation on the anabolic section (A-limitation or *A-lim*) was realized by titrating a key enzyme (GOGAT) in the ammonia

assimilation pathway (Figure 2.3). To impose growth limitation on the polymerization sections, sub-lethal 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 spectroscopy.

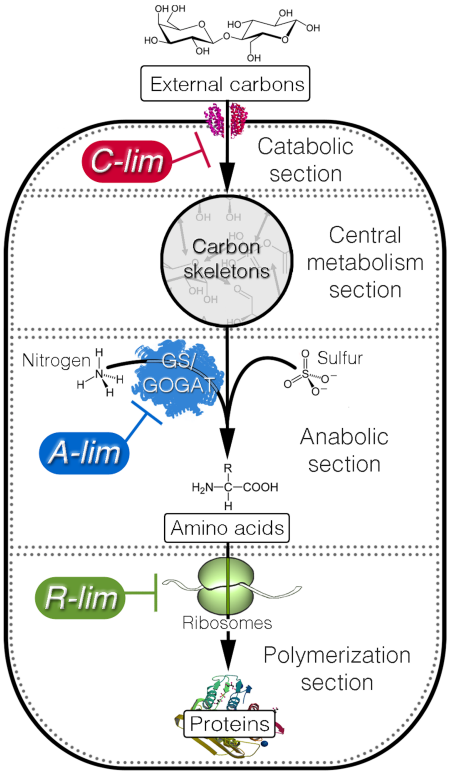


Figure 2.1. Coarse-grained metabolic flow of protein production and the three modes of growth limitation.

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. The R-limitation (R-lim) was realized for the WT strain by supplying the growth medium with various levels of an antibiotic, chloramphenicol.

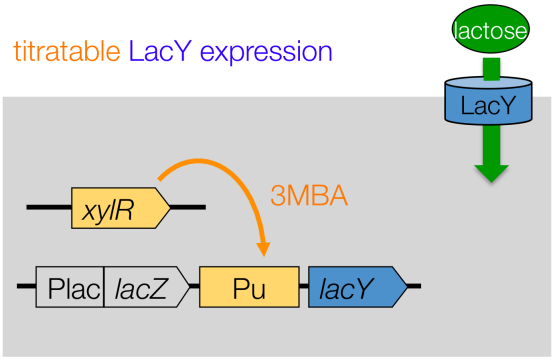


Figure 2.2. C-limitation by titrating lactose uptake.

LacY (or lactose permease) is the only transporter that allows E. coli to grow on lactose as the sole carbon source. We therefore sought to control lactose uptake by titrating the expression of LacY using the strain NQ381. The strain was constructed by inserting a titratable Pu promoter from Pseudomonas putida between the lacZ stop codon and lacY start codon. The expression of the Pu promoter is activated by the regulator XylR upon induction by 3-methylbenzyl alcohol (3MBA). Strain NQ381 was grown in lactose minimal medium, supplemented with 1 mM IPTG and various levels of 3MBA (0-500 µM) to stimulate XylR and titrate the expression of LacY.

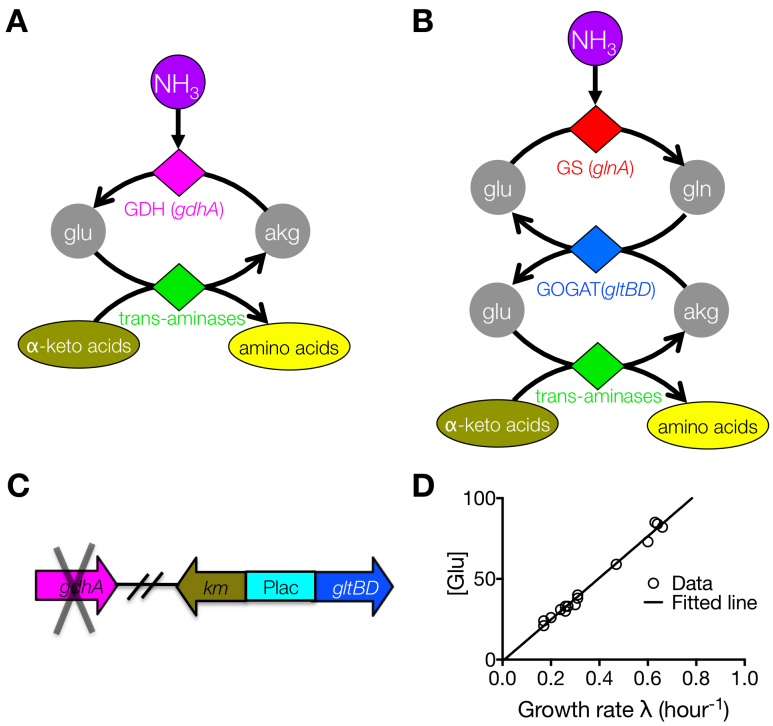


Figure 2.3. A-limitation by titrating ammonia assimilation.

To impose A-limitation, we constructed the strain NQ393 whose capacity to assimilate ammonium as the sole nitrogen source can be varied in graded manners.

(A) and (B) illustrate the two known pathways for the assimilation of ammonium in E. coli. In pathway A, ammonium is fixed onto alpha-ketoglutarate (akg) via the enzyme glutamate dehydrogenase (GDH, purple diamond, encoded by gdhA) to form glutamate (glu), which subsequently trans-aminates (green diamond) one of many alpha-keto acids (light blue oval) to form amino acids (yellow oval), regenerating akg in the process. In pathway B, the overall process is the same except that GDH is replaced by two enzymes, glutamine synthetase (GS, red diamond, encoded by glnA) and glutamate synthase (GOGAT, blue diamond, encoded by the gltBD operon). In this pathway, ammonium is first assimilated into glutamine (gln) and then passed on to glu. Note that among the biosynthetic pathways of the 20 amino acids, only the tryptophan pathway does not involve transamination reaction.

(C) In strain NQ393 the gene gdhA is deleted and the promoter of gltBD is replaced by the Plac promoter, so that pathway A is broken and ammonium assimilation must proceed by pathway B. Strain NQ393 was grown on glucose minimal medium, supplemented with various concentrations of IPTG (30-100 μM) to titrate the expression of GOGAT. See Experimental Procedures for details of strain construction.

(D) The intracellular glu pool concentration ([Glu]) increases linearly with the growth rate. The data are for a strain that has the native lac promoter replaced with the glnA promoter but otherwise is identical to NQ393. It suggests that the growth reduction of this strain and also of NQ393 is due to limitation in glu, which presumably directly affects amino acid synthesis via trans-amination (A). The method for measuring the glu pool was described in Okano et al. (Okano et al., 2010).

2.3 Quantitative proteomic mass spectrometry

Proteomic mass spectrometry is a powerful tool for quantifying changes in global protein expression patterns (Aebersold and Mann, 2003; Bantscheff et al., 2007; Han et al., 2008; Ong and Mann, 2005). The cellular content of each protein is directly measured, whereas inferences of protein levels based on mRNA transcripts obtained from the more commonly used transcriptomic approaches are complicated by complex growth-rate dependent relations linking the mRNA and protein levels (Klumpp et al., 2009; Maier et al., 2009; 2011; Taniguchi et al., 2010; Vogel et al., 2010). 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 ^{15}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.

2.3.1 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}N-labeled purified ribosomes. The observed relative levels, measured by ratios of the labeled to the unlabeled ribosomal proteins (or $^{15}N/^{14}N$), agree extremely well with the expected values over a range of about two orders of magnitude (Figure 2.4A). 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 Figure 2.4B. The effective precision of relative protein quantification is ±18%, based on analysis of the 1:1 sample (Figure 2.5). Thus, subtle changes in proteome composition that are much less than 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 (Figure 2.6A) and also with quantitation of LacZ using a β-galactosidase assay (Figure 2.6B).

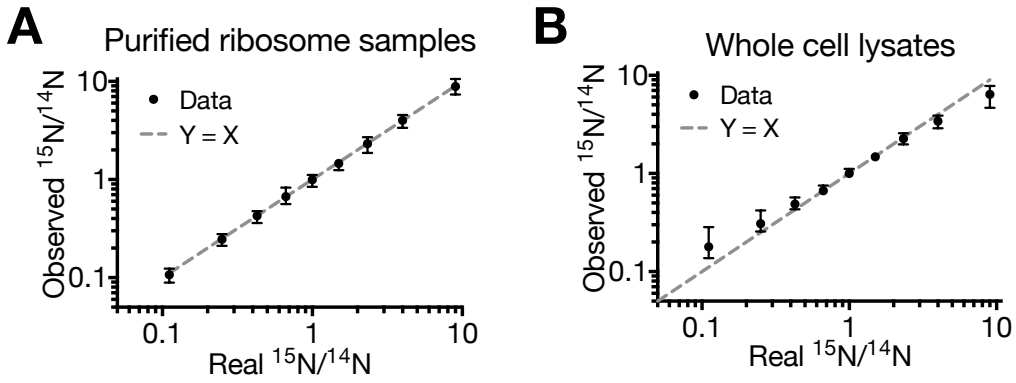


Figure 2.4. The accuracy of quantitative protein mass spectrometry.

(A) Observed values versus real values for ratios of ^{15}N ribosomal proteins to ^{14}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}N proteins to ^{14}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.

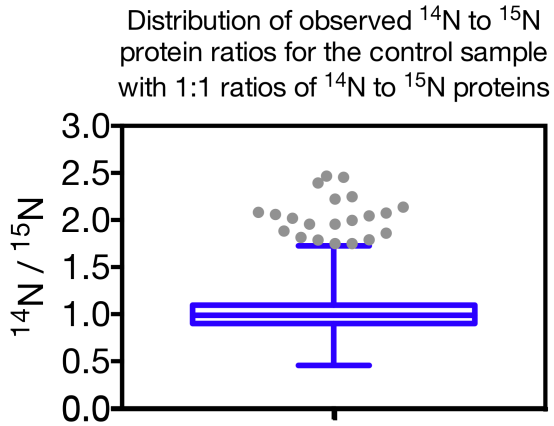


Figure 2.5. The precision o quantitative protein mass spectrometry.

To measure the precision of the relative protein quantification method using mass spectrometry, we focus on the control sample which consists of 1:1 mixture of ^{15}N and ^{14}N samples, where each protein has an expected ^{15}N to ^{14}N ratio of 1. Above is the distribution of the observed ^{14}N over ^{15}N ratios for 638 proteins in the control sample. The box includes data points between first quartile (0.906) and the third quartile (1.114), with the line in the center of the box representing the median value (0.998). The upper and low hinges represent the maximum and minimum data points, excluding 21 outliers. The outliers are shown in gray points and are defined as points that are at least 3×(3rd quartile – 1st quartile) away from either the 1st quartile or the 3rd quartile. The standard deviation for all the data points except the outliers is calculated to be 0.179, or about 18%, which is taken as the precision of the method.

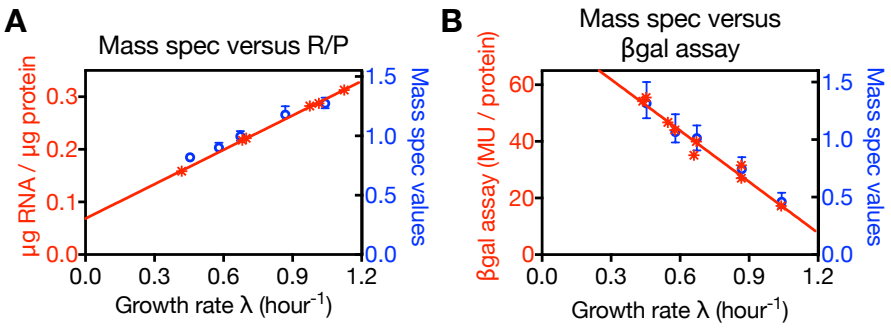


Figure 2.6. Comparison of the relative mass spectrometry method to traditional biochemical methods.

(A) Comparison between mass spectrometry data and the measurements of the total RNA to total protein ratio (R/P). The ratio between total RNA and total protein is well established as a good proxy (with a constant converting factor) for ribosome content, and has a linear relation with growth rate for nutrient-limited growth (Maaloe, 1979; Schaechter et al., 1958; Scott et al. 2010). The red dots are R/P data for cells grown on various carbon sources. The blue circles are the relative change of ribosomal proteins under C-limitation as detected by mass spectrometry. The mass spectrometry values for the ribosomal proteins were taken as the medians of the 52 ribosomal proteins detected. The error associated with each value was taken as the quartiles.

(B) Comparison between mass spectrometry data (in blue) and the β-galactosidase assay data (in red), both under C-limitation (lactose-limited growth).

2.3.2 Data sets and protein coverage

For the C-, A-, and R- limitations, a series of cultures was prepared with varying growth rates. For the C-limitation series, controlled inducible expression of the *lacY* gene gave doubling times from 40-92 minutes (5 conditions), for the A-limitation series, controlled expression of GOGAT gave doubling times from 43-91 minutes (5 conditions), and for the R-limitation series, inhibition of protein synthesis with chloramphenicol gave doubling times from 42-147 minutes (4 conditions), as detailed in Table 2.1. Samples from each of the fourteen cultures were collected, and the relative protein levels were determined using mass spectrometry, as described in the Experimental Procedures. For C-, A-, and R- limitations, the number 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 data sets and a total of 1053 unique proteins in any dataset. The current protein coverage represents ~80% of the total proteome mass, due to a highly non-uniform distribution of protein abundance (Figure 2.7). For data analysis, the combined datasets were represented as a matrix of 1053 proteins across the 14 growth conditions (Table 2.2), graphically shown in Figure 2.8.

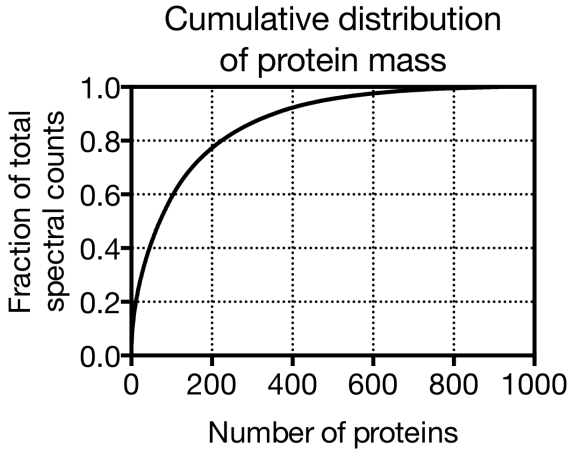


Figure 2.7. Estimation of coverage of total protein mass by mass spectrometry.

To estimate the fraction of total protein mass covered by mass spectrometry, we rely on two pieces of information: 1) the highly non-uniform distribution of individual protein mass as given by the method of spectral counting (shown above); and 2) the absolute protein quantitation results from the 2D gel study by Pedersen et al. (Pedersen et al., 1978). The plot shows the cumulative distribution of protein mass detected in the standard condition (i.e., WT cells growing in glucose minimal medium), with the proteins ranked from high to low according to their masses as measured by spectral counts. The 2D gel study found that in glucose minimal medium the most abundant 190 proteins account for about 60% of the total protein mass. Recent 2D gel absolute protein quantitation study (private communication with Scott, et al.) found a similar number for the same medium, with the top 190 proteins accounting for about 64% of the total protein mass. Here the plot shows that top 190 proteins comprise 76% of the total spectral counts. Therefore, the total proteome mass detected by the mass spectrometry is estimated to be between 60%/76%=79% and 64%/76%=84%. We thus take 80% as the estimated value for the coverage of total protein mass by mass spectrometry.

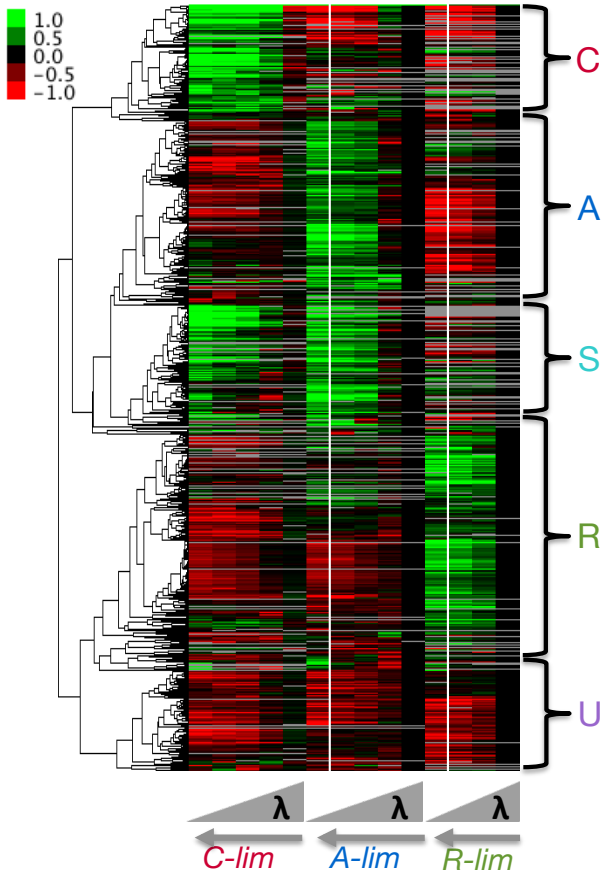


Figure 2.8. The expression matrix and clustering results.

The matrix is composed of 1053 proteins (rows) and 14 conditions (columns); see Table 2.2. 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 log2-transformed, with expression values at the standard condition as zero (see Experimental Procedures), 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 Experimental Procedures), with the 5 major clusters shown on the right side of the matrix.

2.3.3 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 (Experimental Procedures), and the resulting dendrogram is shown on the expression matrix in Figure 2.8. 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 stress, and that the response of the proteome to the environment might be amenable to a quantitative coarse-graining analysis.

2.4 Quantitative characterization of the proteome responses

2.4.1 Linear growth-rate dependence of the relative protein levels

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). Inspection of the expression data in Figure 2.8 suggested that many proteins exhibited a similar linear trend, and the coefficient of determination (R^2) for the expression of each protein was calculated for each limitation series. The cumulative distribution of R^2 for each mode of growth limitation is shown in Figure 2.9. Approximately half of the proteins in C-limitation and A-limitation data sets and about two thirds of the proteins in R-limitation data set exhibit $R^2 > 0.7$. The slope and R^2 of each linear fit are given in Table 2.2. As a comparison, the plot of the cumulative distribution of R^2 for randomly permuted expression data of the A-limitation is shown, where less than one tenth of the proteins have $R^2 > 0.7$. The widespread linear dependence on growth rate further supported a coarse-grained analysis of the proteome response to growth limitations.

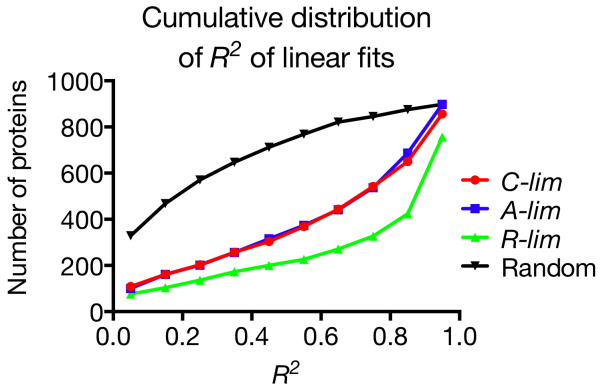


Figure 2.9. Cumulative distribution of R squared values (R^2) of linear fits.

For each of the three limitation data sets, a line was fit for each protein and its R^2 value was calculated (See Table 2.2 for the parameters of fits). The red symbols and line show the cumulative distribution of R^2 for C-limitation, while the blue and green data are for A- and R-limitation, respectively. The black symbols and line are for the A-limitation data with the expression values for each protein randomly permuted.

2.4.2 Strategy for a coarse-grained characterization of the proteome

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 (↑), while a positive slope corresponds to a lower expression level at slower growth rate, referred to as the "downward" response (↓).

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↑A↓R↓, C↑A↑R↓, C↓A↑R↓, C↓A↑R↑, C↓A↓R↑, C↑A↓R↑, C↑A↑R↑, and C↓A↓R↓, 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↑A↓R↓ group consists of proteins that have upward response under C-limitation and downward

responses under both the A- and R- limitation.

The collective behavior of a protein group can be approximated by coarse graining, effectively summing the mass of proteins in the same group. The summed mass of proteins for a protein group, from one of the 8 groups defined above and normalized by total protein mass, is referred to as the "proteome fraction". Due to the almost constant proportionality between the total protein mass and dry mass per cell, and the constancy of the cell's density at different growth rates (Woldringh et al., 1981), proteome fraction is a proxy for protein concentration at all growth rates explored in this paper (Klumpp et al., 2009) in contrast to other measures such as protein copy number per cell which do not track concentration due to the huge changes in cell size across growth conditions (Milo, 2013; Schaechter et al., 1958). To carry out the coarse-graining approach on our data, the proteins first need to be placed into one of the 8 groups and then the proteome fraction needs to be estimated for each of the groups. The grouping is readily accomplished using the relative protein expression data. However, determining the proteome fraction requires an estimate of the absolute protein level for each protein.

2.4.3 Defining the protein groups based on the relative expression data

To partition proteins into one of the 8 groups, we calculated the slope of a linear fit to each gene under each limitation. The results of linear-fit of the relative protein expression data are given in Table 2.2, where positive slope corresponds to downward response and negative slope to upward response. Due to the sensitivity limitations of the method, proteins exhibiting < 25% change under each of the growth limitations were partitioned into a separate, growth-rate independent group (or I-group). The membership

of proteins in the resulting 9 groups is given in Table 2.2. The corresponding expression matrix (Figure 2.10) is seen to capture similar global expression patterns revealed by the clustering method at a qualitative level (Table 2.3).

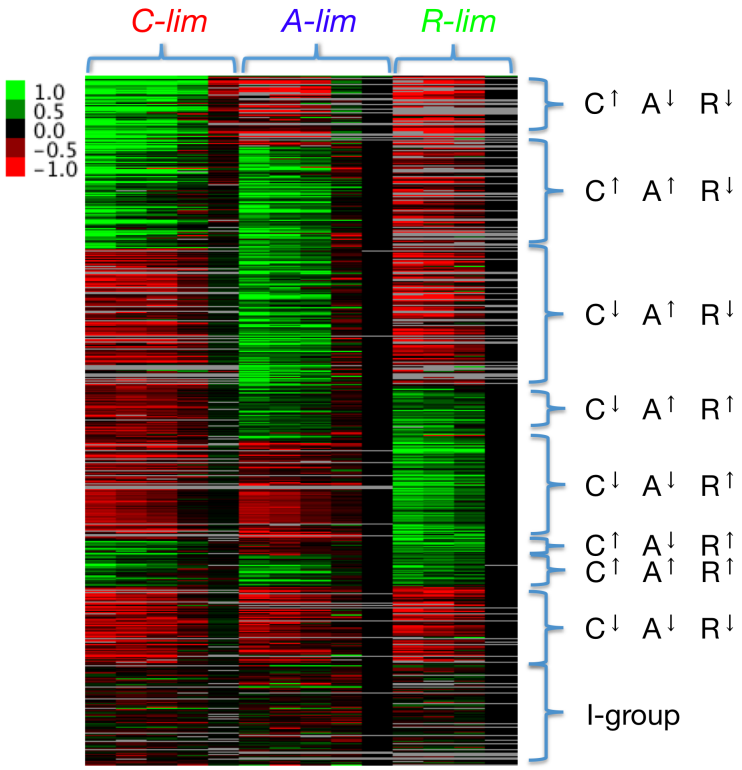


Figure 2.10. Grouping proteins into 9 groups.

In the expression matrix, the first five columns are for C-limitation, the next five columns for A-limitation, and the last four columns for R-limitation. Within each limitation, the growth rate increases from left to right. Red color indicates negative values, green color indicates positive values, and black indicates zero values. Gray indicates missing entries. The right side of the expression matrix shows the 9 groups. From top to bottom, the groups are C↑A↓R↓, C↑A↑R↓, C↓A↑R↓, C↓A↑R↑, C↓A↓R↑, C↑A↓R↑, C↑A↑R↑, C↓A↓R↓, and I-group, where the upward arrow denotes expression values going up (specific response) as growth rate goes down in a limitation and down arrow means the opposite (general response). If a protein is missing (i.e., having no values) under a limitation, we treated its response as general response.

2.4.4 Absolute quantitation estimates for protein groups from spectral counting

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;

Schmidt et al., 2011; Silva et al., 2006; Vogel and Marcotte, 2008), the method of spectral counting takes the number of peptides recorded for each protein as proxy for the absolute abundance of the protein (Malmstrom 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, as shown in Figure 2.11A. The comparison of spectral counting data for ribosomal proteins with the estimation based on biochemical measurements is in good agreement (Figure 2.11B).

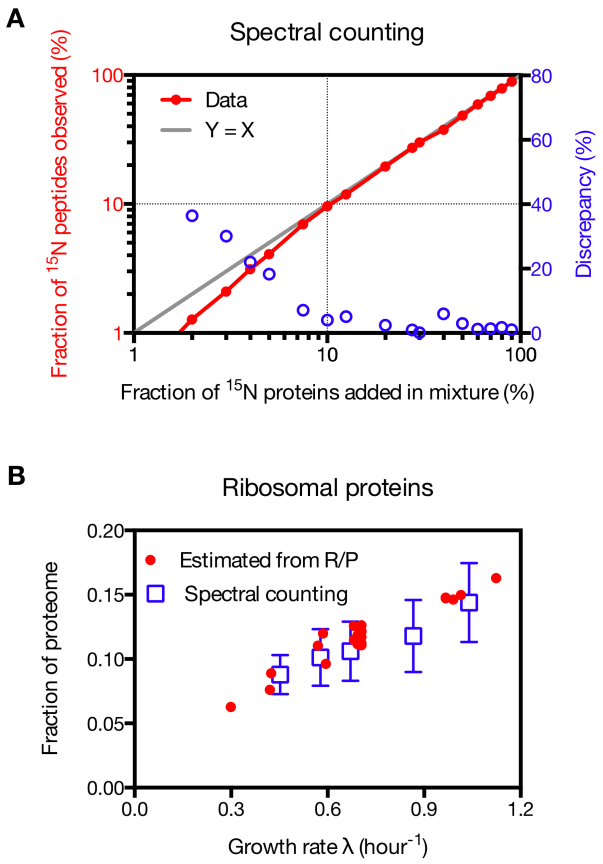


Figure 2.11. Absolute protein quantitation with spectral counting.

(A) Spectral counting data from the whole cell series. ^{15}N-labeled cell sample was mixed with unlabeled cell sample at different proportions. The estimated fractions of the ^{15}N-labeled proteins based on spectral counting are plotted against the real fractions (red symbols and line). Discrepancy between the estimated value from spectral counting and the expected value is defined as the absolute value of the difference between the two values divided by the expected value (blue symbols). The discrepancy quickly goes down as the fraction goes up, with around 20% for 5% of expected fraction and less than 10% for 7.5% of expected fraction.

*(B) Comparison of spectral counting data with the R/P data for the proteome fraction of ribosomal proteins. The red dots are estimated from R/P measurements (Experimental Procedures), with the formula: fraction of proteome = 0.52*R/P (See Eq. [S1] in (Scott et al., 2010). The blue squares and error bars are the mean values and standard deviations of spectral counting data from triplicate mass spectrometry runs.*

2.4.5 Absolute abundance of the coarse-grained proteome sectors

The proteome fractions for the 9 protein groups defined in Table 2.2 were

determined for each of the 3 series of growth limitations by applying the spectral counting method (Figure 2.12). It is clear from Figure 2.12 that some groups occupy significant fractions of the proteome while others are minor constituents. The top three groups are C↑A↓R↓, C↓A↑R↓, and C↓A↓R↑. 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 (Figure 2.13A-C). The C↓A↓R↓ group includes proteins that are uninduced by any of the three applied limitations, and is referred to as the U-sector (Figure 2.13D). Another significant protein sector is the C↑A↑R↓ 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 Figure 2.13E. The three remaining groups (i.e., C↑A↑R↑, C↑A↓R↑, and C↓A↑R↑ groups) are small, with most of the data at or below 5% of the proteome, below the accuracy of the spectral counting method (Figure 2.11A). The three small groups and the I-group were placed together into the O-sector (Figure 2.13F). 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 Figure 2.13, with complete data for all fractions shown in Table 2.4.

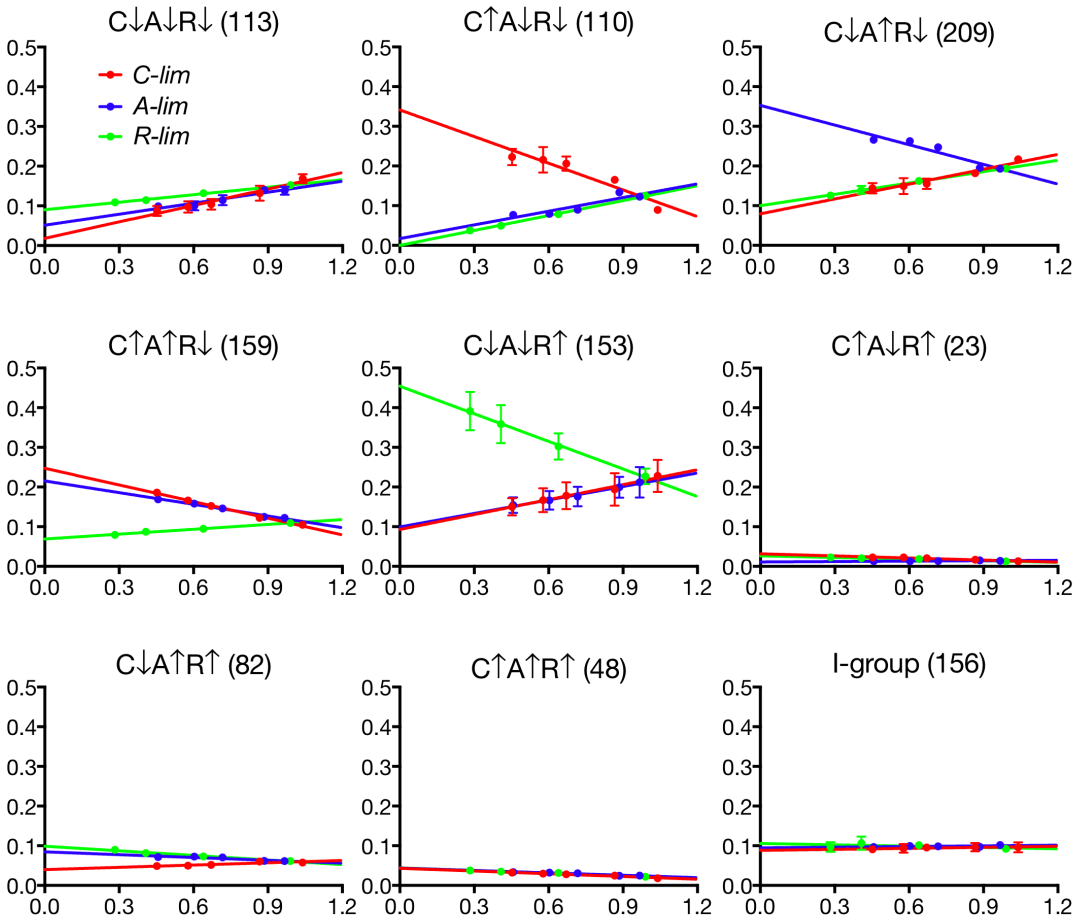


Figure 2.12. Coarse-grained results for the 9 protein groups.

The Y-axis of each of the plots is fraction of proteome and the X-axis is the growth rate (in units of per hour). The red symbols and lines are for C-limitation, blue for A-limitation, and green for R-limitation. The lines are the best linear fits to the data represented by symbols of the same colors. The title of a group indicates the types of response the group has to the three limitations, with an upward arrow (↑) for a line with negative slope and downward arrow (↓) for a line with positive slope. The number in the title indicates the number of proteins in the group.

The variation of the abundance for the triplicate runs is much larger for the C↓A↓R↑ group (or R-sector in Figure 2.13) than other sectors. This reflects the coarse-graining method we used for estimating the absolute abundance for proteome sectors. The method assumes a diverse representation of proteins with broad distributions of efficiencies in various steps of the experimental flow. The R-sector includes most of the r-proteins which together form one complex, ribosome. Similar behaviors of this large group of proteins in terms of noise could cause the observed large variation for the R-sector.

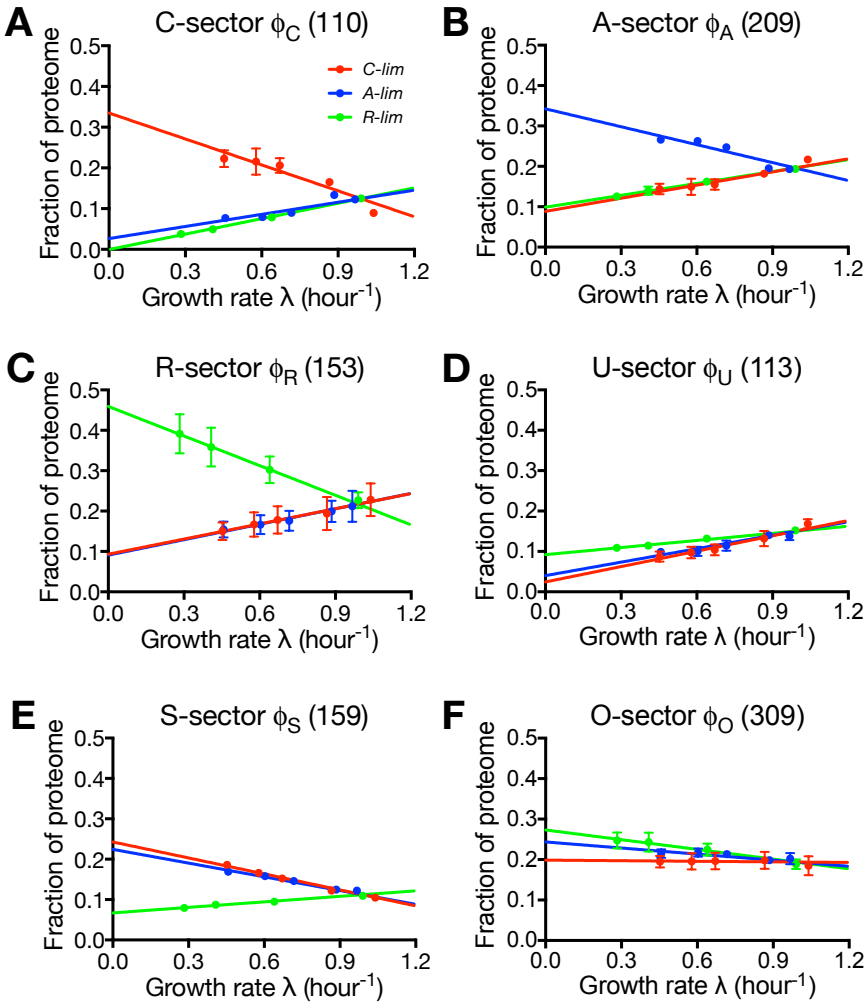


Figure 2.13. The coarse-grained proteome sectors.

Coarse-grained responses of the C-, A-, R-, U-, S-, and O- sectors to the three modes of growth limitation. As indicated in panel 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. 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 Table 2.4 for the data on proteome fraction and Table 2.5 for parameters of the fitted lines.

2.4.6 Linear responses of proteome sectors

The proteome fraction of a sector s under the *l*-mode of growth limitation is

denoted by $\phi_{\sigma,l}$ with σ being one of the 6 sectors (C,A,R,U,S,O) and *l* being C-, A-, or R-limitations. The data for each sector merge at the growth rate of ~1.0 hour^{-1}, because the growth rate dependence of the proteome fraction was quantified relative to the proteome fractions under the "glucose standard condition", which is the condition of WT cells growing on glucose minimal medium. These points define the set of reference proteome fractions ϕ_σ^* under the standard growth rate λ^*. The growth rate dependences of the proteome sectors $\phi_{\sigma,l}(\lambda)$ shown in Figure 2.13 are well described by a linear relation:

$$\phi_{\sigma,l}(\lambda) = \phi_{\sigma,l,0} + \frac{\phi_\sigma^* - \phi_{\sigma,l,0}}{\lambda^*}\lambda \,. \qquad [2.1]$$

where $\phi_{\sigma,l,0}$ is the y-intercept of the line. The least squares fits to Eq. [2.1] are shown as the lines of corresponding colors in Figure 2.13, and the set of 4×6=24 fit parameters that describe the data are given in Table 2.5.

Closer scrutiny of the data and the fits suggests additional simplicity in the structure of the responses. In particular, the downward responses in Figure 2.13 (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 will be addressed below.

2.5 Biological functions of the proteome sectors

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. Briefly, individual GO terms that accounted for a small mass fraction of a sector were removed, as were those terms (such as "metabolic process") that are too general for distinguishing among the sector functions. GO terms that were not enriched in a sector compared to the entire proteome were also removed. The sector functions were identified based on the remaining GO terms accounting for a significant fraction of a sector with minimal overlap with other sectors, as described in Chapter 3.

The results of the abundance-based GO analysis are summarized in Figure 2.14, with each bar graph describing the major proteome composition for each sector. The list of genes corresponding to each GO term and each sector is provided in Table 2.6. With the exception of the O-sector, between 1 and 3 terms accounted for more than 60% of the proteome in each sector, providing a simple interpretation of their biological functions and the nature of the sector responses. 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 counteract the applied growth limitation by increasing the abundance of "translational" proteins (Scott et al., 2010).

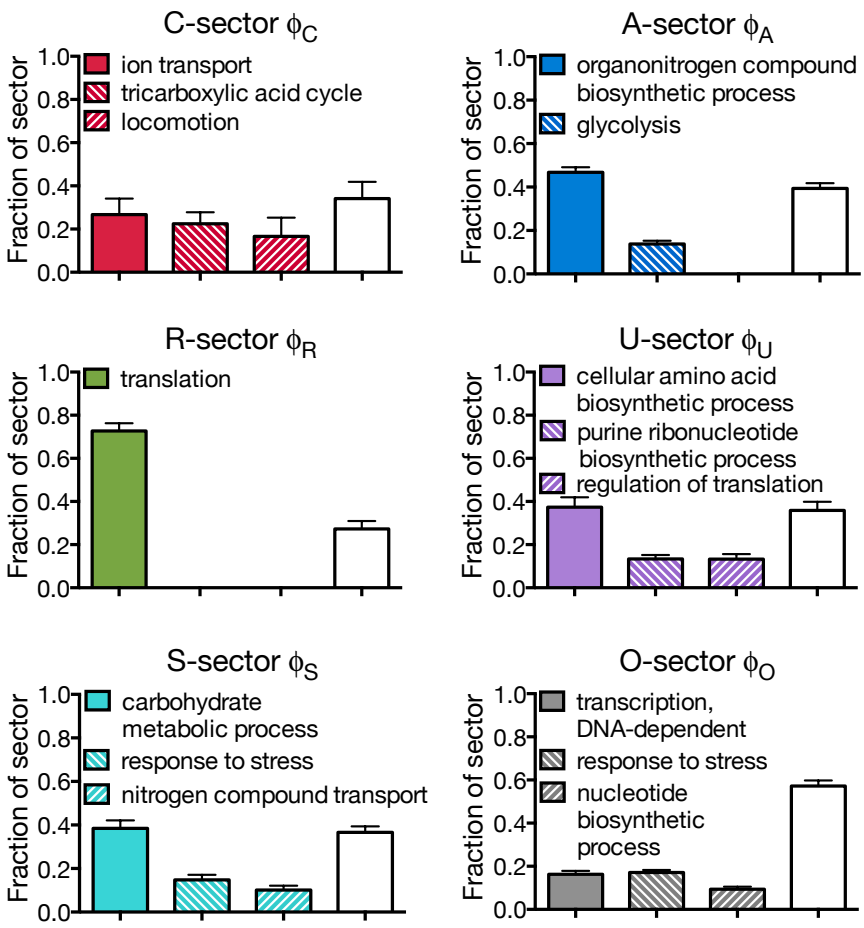


Figure 2.14. Composition of proteome sectors.

Each bar graph shows the results of the abundance-based Gene Ontology analysis for each of the 6 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 Chapter 3 for details of the method.

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 (Table 2.6). Again, similar to the 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 (Figure 2.3). Interestingly, "glycolysis" 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 cysteine, methionine, and tryptophan biosynthesis (Table 2.6). 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 (Figure 2.3), 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 any of the growth limitations. The third term is "regulation of translation", where the proteins are mostly related to synthesis of tRNAs with the exception of one ribosomal protein (Table 2.6). In summary, the U-sector is composed of enzymes for making a diverse group of building blocks including some amino acids, purine ribonucleotides, and some tRNAs.

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 "nitrogen compound 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- and A- limitations. This notion is best illustrated by the term 'nitrogen compound transport", consisting mostly of transporters for peptides and amino acids which can clearly be used to counteract both C- and A- limitations (Table 2.6). The O-sector is more diverse, with the top three terms describing about 40% of the sector. The terms are "transcription, DNA-dependent", "response to stress", and "nucleotide biosynthetic process", reflecting some house-keeping activities of proteins in this sector.

In summary, the GO analysis reveals that the R-sector consists mostly of the translational machinery, the C-sector engages in carbon scavenging, 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, there coarse-grained enzyme functionalities correspond to steady fluxes of biomass. As illustrated in Figure 2.15, 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. The O-sector apparently corresponds to house-keeping functions that are carried out independently from any of the applied limitations.

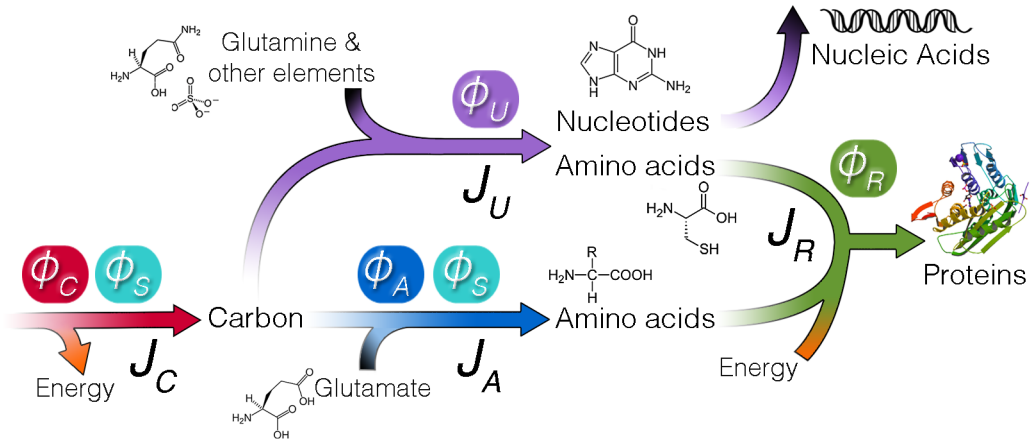


Figure 2.15. Association between metabolic fluxes and proteome sectors.

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 ϕs on top of the fluxes represent the corresponding proteome fractions that carry the fluxes. Note that the S-sector proteins contribute to both J_C and J_A.

2.6 A proteome-based flux model

As summarized in Figure 2.15, 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 Figure 2.13, the proteome is partitioned into six sectors, ϕ_σ, each of which is comprised of a growth-rate independent component, $\phi_{\sigma,0}$, and a growth-rate dependent component, $\Delta\phi_\sigma(\lambda)$, i.e.,

$$\phi_\sigma(\lambda) = \phi_{\sigma,0} + \Delta\phi_\sigma(\lambda). \qquad [2.2]$$

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

$$J_\sigma = k_\sigma \cdot \Delta\phi_\sigma \quad [2.3]$$

where k_σ is a coarse-grained kinetic coefficient describing the efficiency of the metabolic sector σ. As shown below, the model comprising of Eqs. [2.2-2.3] can quantitatively account for all of the observations summarized in Figure 2.13.

General Response: The "downward" general responses in Figure 2.13 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 (ϕ_R) is given by

$$J_R = k_R \cdot \Delta\phi_R , \quad [2.4]$$

where k_R is the corresponding enzyme kinetic parameter (given by the peptide elongation rate. 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 and Palsson, 1994), the growth rate dependent proteome fraction for the R-sector is given by

$$\Delta\phi_R(\lambda) = \lambda / \nu_R , \quad [2.5]$$

where $\nu_R = k_R c_R$ is an effective rate constant for the R-sector. Upon applying the C- or A- limitation, the peptide elongation is not affected, ν_R is constant and Eq. [2.5] describes a linear relation between ϕ_R and λ, which is the "general response". Note that this model explicitly predicts identical general responses for the R-sector under C- and A-limitations (Eq. [2.S1] of Table 2.7), in good agreement with the data of Figure 2.13.

Similarly, for the U-sector:

$$\Delta\phi_U(\lambda) = \lambda / \nu_U . \quad [2.6]$$

The downward lines of the U-sector in Figure 2.13 is produced by Eq. [2.6] as long as none of the growth limitations affects the value of ν_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 (Eq. [2.S2] of Table 2.7), which is consistent with the data in Figure 2.13.

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 (Figure 2.15). 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. Then S-sector proteins is 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 (Figure 2.16). Applying the same linear relation between proteome fractions $\phi_{\tilde{C}}, \phi_{\tilde{A}}$ and with the fluxes, i.e.,

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

we obtain in the simplest case

$$\begin{aligned} \Delta\phi_C(\lambda) &= (1-f)\cdot\lambda/\nu_C, \\ \Delta\phi_A(\lambda) &= (1-f)\cdot\lambda/\nu_A, \\ \Delta\phi_S(\lambda) &= f\cdot[\lambda/\nu_C + \lambda/\nu_A], \end{aligned} \qquad [2.7]$$

with f being the fraction of $\tilde{C}$- and $\tilde{A}$- sector proteins that are in common (Figure 2.16). These relations describe the general responses of the C-, A- and S- sectors (Eqs. [2.S3-2.S5] of Table 2.7), for growth limitations that do not affect ν_C or ν_A. Finally, we assume the existence of a growth-rate independent sector, and identify it with the O-sector, i.e., $\phi_O(\lambda) = \phi_{O,0}$ (Eq. [2.S6] of Table 2.7) with

$$\Delta\phi_O(\lambda) = 0\,. \qquad [2.8]$$

Specific response: A striking result of this flux model is that the "specific" upward responses of the C-, A-, R-, and S-sectors in Figure 2.13 can also be produced by Eqs. [2.5-2.8], without introducing any additional parameters. For example, under R-

limitation the value of ν_R changes in response to the limitation, and consequently, the growth-rate dependence of ϕ_R can no longer be obtained from Eq. [2.5]. However, $\Delta\phi_R(\lambda)$ can be obtained from the important constraint $\sum_\sigma \phi_\sigma(\lambda) = 1$, or equivalently

$$\sum_\sigma \Delta\phi_\sigma(\lambda) = \phi_{\max}, \qquad [2.9]$$

where $\phi_{\max} \equiv 1 - \sum_\sigma \phi_{\sigma,0}$. Since under R-limitation only ν_R is reduced, all other sectors still follow the general responses. Using Eqs. [2.5-2.9], the expression for the specific response of the R-sector becomes:

$$\Delta\phi_R(\lambda) = \phi_{\max} - \lambda/\kappa_R, \qquad [2.10]$$

with $\kappa_R^{-1} \equiv \nu_C^{-1} + \nu_A^{-1} + \nu_U^{-1}$ (Eq. [2.S7] of Table 2.7). Note that both parameters appearing in Eq. [2.10] 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 ν s, with no additional parameters; see Eqs. [2.S8-2.S10] of Table 2.7 with derivation given in the following.

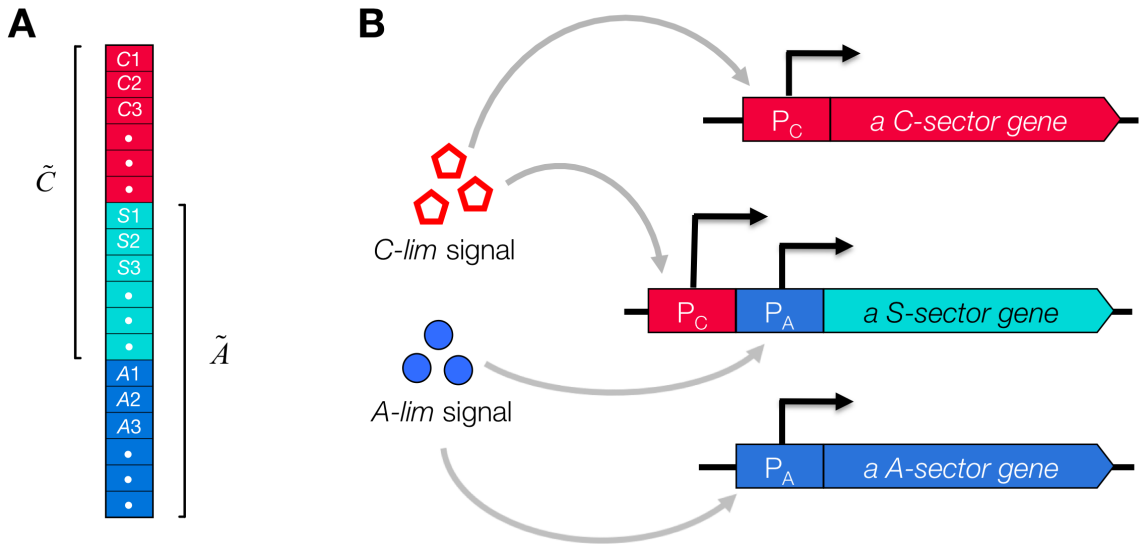


Figure 2.16. S-sector proteins respond to both C- and A- limitations.

(A) Illustration of two hypothetical lists of proteins, $\tilde{C}$ and $\tilde{A}$, responding to only C-limitation and only A-limitation, respectively. While the C- and A- sector proteins belong only to the $\tilde{C}$ and $\tilde{A}$ respectively, the "multi-purpose" S-sector proteins belong to both lists.

(B) An illustrative mechanism generating the expression pattern of an S-sector protein: the corresponding gene is expressed by the activation of either the promoter P_c which responds to signals for C- limitation or the promoter P_A which responds to signals for A- limitation.

We first hypothesize that there are two lists of proteins, $\tilde{C}$ and $\tilde{A}$, responding to only C-limitation and only A-limitation, respectively (Figure 2.16A). While the C- and A- sector proteins belong only to the $\tilde{C}$ and $\tilde{A}$ respectively, the "multi-purpose" S-sector proteins belong to both lists. Figure 2.16B illustrates a mechanism generating the expression pattern of an S-sector protein: the corresponding gene is expressed by the activation of either the promoter P_c which responds to signals for C- limitation or the promoter P_A which responds to signals for A- limitation. The growth-rate dependent components of $\phi_{\tilde{C}}$ and $\phi_{\tilde{A}}$ are denoted as $\Delta\phi_{\tilde{C}}$ and $\Delta\phi_{\tilde{A}}$, respectively. Similar to the R- and U- sectors (Eqs [2.5-2.6]), we have for $\tilde{C}$ proteins

$$\Delta\phi_{\tilde{C}}(\lambda) = \lambda / \nu_C \,, \qquad [2.11]$$

and

$$\Delta\phi_{\tilde{A}}(\lambda) = \lambda / \nu_A \,, \qquad [2.12]$$

for $\tilde{A}$ proteins, with ν_C and ν_A being the respective rate constants. We assume that a constant fraction (f_C) of $\Delta\phi_{\tilde{C}}$ belongs to $\Delta\phi_S$, the growth-rate dependent component of the S-sector. Similarly, we assume that a constant fraction (f_A) of $\Delta\phi_{\tilde{A}}$ also belongs to $\Delta\phi_S$. We then have

$$\Delta\phi_S(\lambda) = f_A \cdot \lambda / \nu_A + f_C \cdot \lambda / \nu_C \,. \qquad [2.13]$$

The remaining parts of $\Delta\phi_{\tilde{C}}$ and $\Delta\phi_{\tilde{A}}$ are respectively $\Delta\phi_C$ and $\Delta\phi_A$, i.e.,

$$\Delta\phi_C = (1 - f_C) \cdot \lambda / \nu_C \,, \qquad [2.14]$$

and

$$\Delta\phi_A = (1 - f_A) \cdot \lambda / \nu_A \,. \qquad [2.15]$$

Eqs. [2.13-2.15] describe the general responses of the S-, C-, and A- sectors. To derive the specific responses of the sectors, we use the constraint given by Eq. [2.9]. For example, under C-limitation where only ν_C is changed, A-, R-, and U- sectors still follow the general responses. Using Eqs. [2.5-2.6], [v], and [2.9], we have

$$\Delta\phi_S + \Delta\phi_C = \phi_{\max} - \lambda \cdot (\nu_R^{-1} + \nu_U^{-1} + (1 - f_A) \cdot \nu_A^{-1}) \,. \qquad [2.16]$$

Using Eqs. [2.13-2.14], we have

$$(1 - f_C) \cdot \Delta\phi_S - f_C \cdot \Delta\phi_C = (1 - f_C) \cdot f_A \cdot \lambda / \nu_A \,. \qquad [2.17]$$

Solving Eqs. [2.16] and [2.17] for $\Delta\phi_S$ and $\Delta\phi_C$ gives

$$\begin{aligned} \Delta\phi_S(\lambda) &= f_C \cdot \phi_{\max} - \lambda \cdot (f_C \cdot (\nu_R^{-1} + \nu_U^{-1}) + (f_C - f_A) \cdot \nu_A^{-1}) \\ \Delta\phi_C(\lambda) &= (1 - f_C) \cdot (\phi_{\max} - \lambda / \kappa_C) \end{aligned} \,, \qquad [2.18]$$

where $\kappa_C^{-1} \equiv \nu_A^{-1} + \nu_R^{-1} + \nu_U^{-1}$.

Similarly, under A-limitation, we obtain

$$\begin{aligned}\Delta\phi_S(\lambda) &= f_A \cdot \phi_{\max} - \lambda \cdot (f_A \cdot (\nu_R^{-1} + \nu_U^{-1}) + (f_A - f_C) \cdot \nu_C^{-1}) \\ \Delta\phi_A(\lambda) &= (1 - f_A) \cdot (\phi_{\max} - \lambda / \kappa_A)\end{aligned}, \qquad [2.19]$$

with $\kappa_A^{-1} \equiv \nu_C^{-1} + \nu_R^{-1} + \nu_U^{-1}$.

Eqs. [2.18-2.19] describe the specific responses of the three sectors. Inspired by the similar specific responses of the S-sector to both C- and A- limitations (see the two upward lines in Figure 2.13E), we simply used $f = f_A = f_B$ in Eqs. [2.18-2.19], yielding simpler equations for specific responses of the C-, A-, and S-sector (Eqs. [2.S8-2.S10] of Table 2.7). Similarly, Eqs. [2.S3-2.S5] of Table 2.7 are the results of applying this simplification to Eqs. [2.13-2.15]. This simplification still allows good quantitative description of the data (Figure 2.17; Table 2.8).

In summary, the linear equations in Table 2.7 describe the prediction of the simple flux model (Eqs. [2.2-2.3]) 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 Figure 2.17; Table 2.8) is comparable to the 24-parameter fit for each individual response (Figure 2.13; Table 2.5).

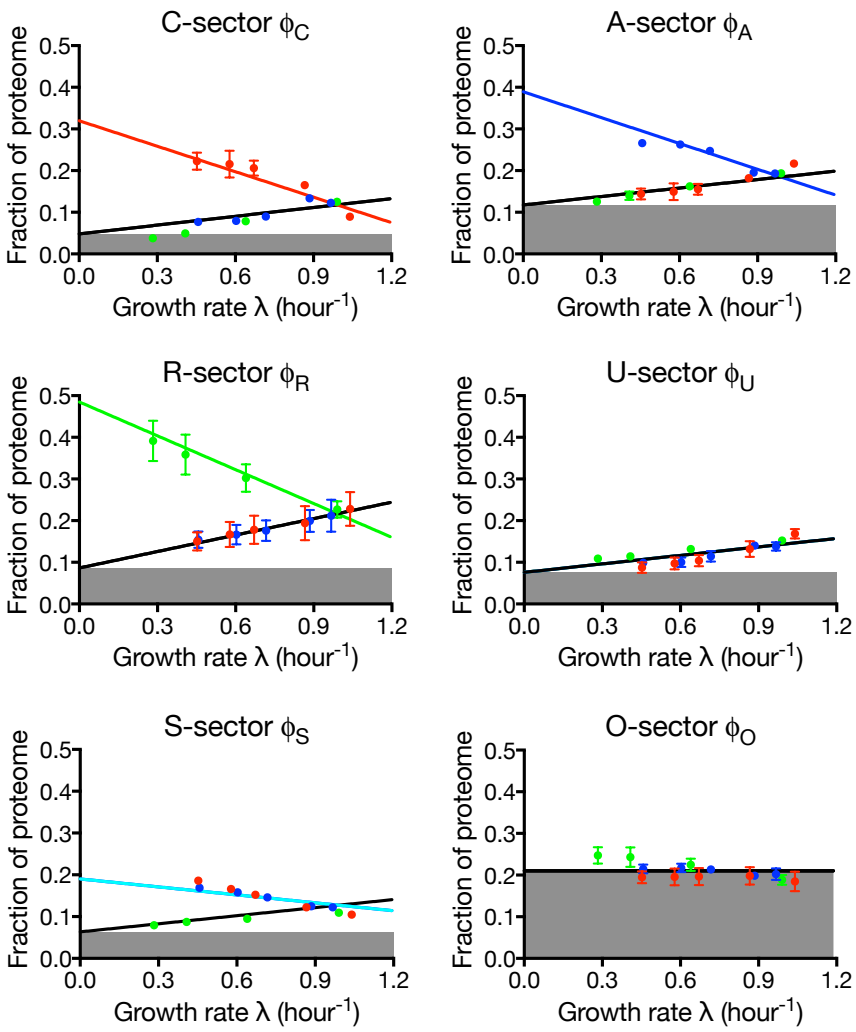


Figure 2.17. Performance of the proteome-based flux model.

The data points are identical to those in Figure 2.13. The lines here are the result of a global fit to the predictions of the flux-based proteome model (Table 2.7). The growth-rate independent component of each sector ($\phi_{\sigma,0}$) is represented as the height of the filled area in the corresponding plot. See Table 2.8 for parameters of the fitted lines.

2.7 Simple responses of proteome

The straightforward meanings of the remaining 10 parameters are illustrated by the cartoon in Figure 2.18. The top pie chart in Figure 2.18 represents the proteome fractions for the sectors under the glucose standard condition, with the growth-rate

independent fraction of the proteome, ϕ_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, $\phi_{\max}$.

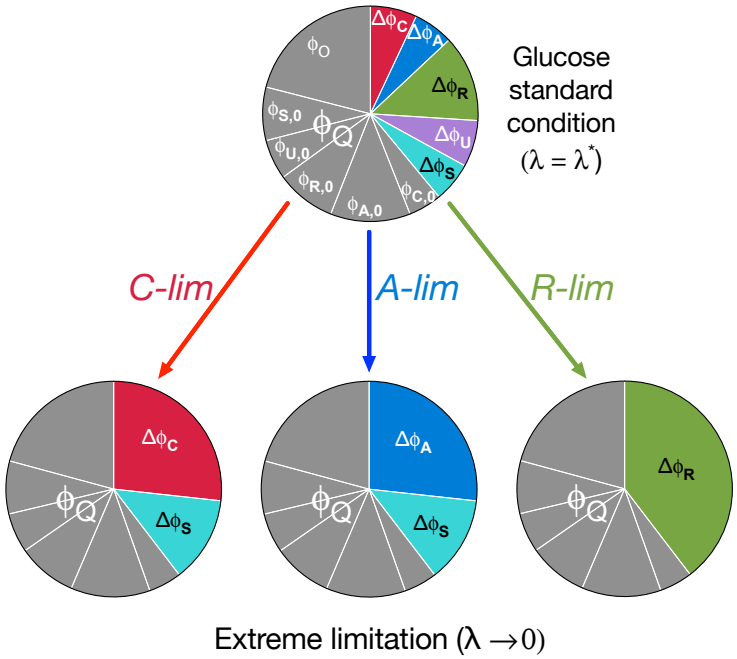


Figure 2.18. Representation of the proteome responses under extreme growth limitations and interpretation of the model parameters.

The growth-rate independent component of the protein is represented as ϕ_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 Table 2.8 for their values. The growth-rate dependent part of a sector σ is labeled as $\Delta\phi_\sigma$, distinguished by the different colors. The colored wedges in the top pie chart shows the sizes of these sectors, $\Delta\phi_\sigma(\lambda^*)$, under the glucose standard condition (with growth rate λ^*). Their values are: $\Delta\phi_C = 0.07$, $\Delta\phi_A = 0.06$, $\Delta\phi_R = 0.13$, $\Delta\phi_U = 0.07$, and $\Delta\phi_S = 0.06$. The pie charts at the bottom show the sizes of these sectors under the three modes of growth limitations in the extreme limit $\lambda \to 0$. Theses sizes are governed by two parameters, $\phi_{\max} = 1 - \phi_Q$ and $f \approx 0.32$.

The bottom three pie charts in Figure 2.18 describe the responses of the proteome

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

Among the 10 parameters of the model, the 4 values of ν_σs are dependent on the growth medium, while the $\phi_{\sigma,0}$ 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 Figure 2.13) were obtained using glucose minimal medium as the standard condition (with the ν_σs taking on the values ν_σ^*), with each mode of growth limitation corresponding to varying one of the ν_σs away from ν_σ^*. The proteome flux model also makes explicit predictions on the response of the proteome under combinatorial modes of growth limitation, corresponding to varying multiple ν_σs. The effect of varying multiple ν_σs 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 (Table 2.9). Compared to the standard condition with glucose

minimal medium, the glycerol minimal medium should differ by only the value of ν_C, which is fixed by the growth rate for the glycerol standard condition (Table 2.8). Using this new value of ν_C, together with the values of the other 9 parameters obtained from the glucose data, the model describes the new data remarkably well (Figure 2.19; Table 2.10). Thus, the model can describe experiments in different standard conditions, an important benchmark for its ability to capture proteome responses to combinatorial limitations.

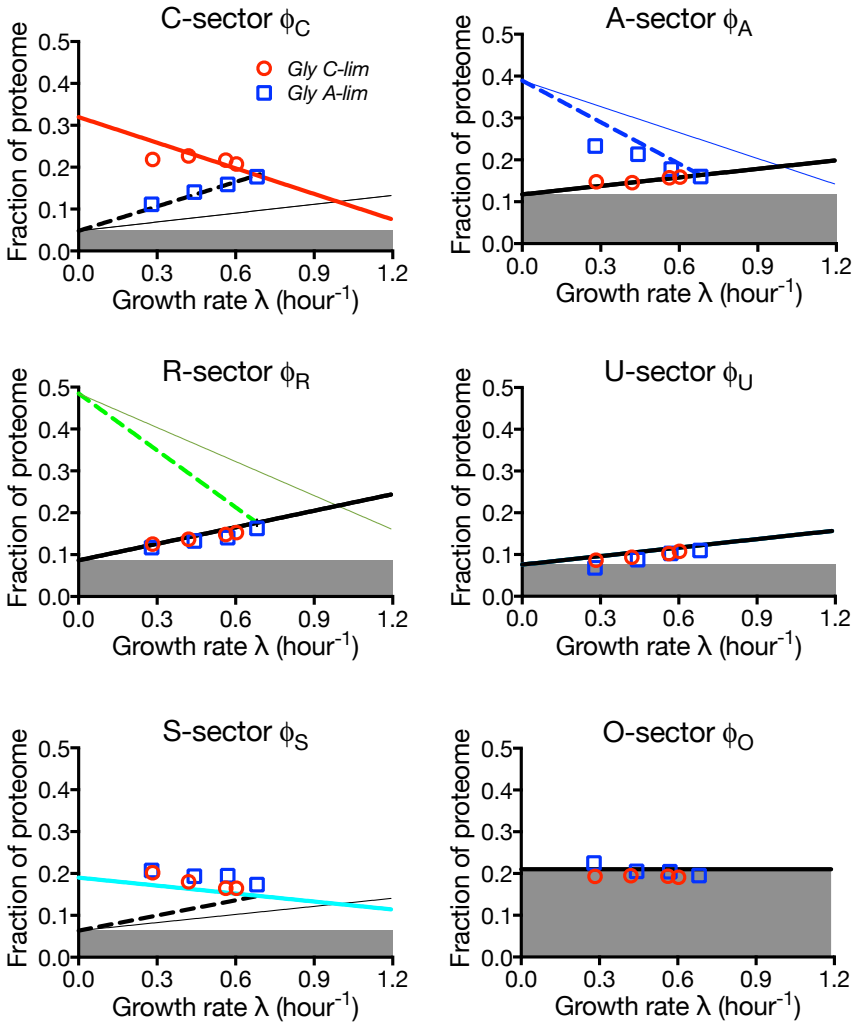


Figure 2.19. Proteome fractions under growth limitations with respect to the glycerol standard condition.

Proteome fractions ϕ_σ for C- and A- limitation under glycerol standard condition are shown as the red and blue circles respectively for each of the 6-sectors; see Table 2.10 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 ν_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 ν_C. See Eqs. [2.S3], [2.S9], [2.S7], and [2.S5] describing the dashed lines for the C-, A-, R-, and S- sectors, respectively. For comparison, the four respective proteome responses under glucose standard condition are also shown as thin solid lines. All solid lines are from Figure 2.17. Note that the new value of ν_C, is determined from the growth rate of cells in glycerol standard condition (Table 2.8). Thus, all predictions for the glycerol standard condition were generated with no adjustable parameters.

2.8 The economic allocation of proteome resources

Understanding the fundamental principles governing gene expression across the proteome is a major goal of molecular systems biology. However, the enormous intricacy of regulatory processes makes this goal difficult to realize using purely bottom-up modeling approaches. Similarly, measuring the allocation of resources across the proteome is irreducibly dependent on global absolute abundance data, which makes proteome management difficult to study using global relative quantitation data in isolation.

Performing a binary classification of relative protein responses to three basic modes of growth limitations for ~1,000 proteins in *E. coli*, measured using quantitative mass spectrometry, we identified six distinct sectors of the proteome. An 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.

During balanced exponential growth, a constant flux of matter from environmental nutrients flows through the metabolic network of the cell to form biomass. In contrast to bottom-up description of the metabolic network as an object of incredible complexity (e.g., the KEGG map), our results highlight an enzymatic network that is simply coarse-grained according to the functional grouping of the proteome sectors. The mass fractions of the various proteome sectors increase or decrease in a strikingly linear fashion with the growth rate, according to the nature of the applied limitation. For example, R-sector, consisting mostly of ribosomal proteins, increases in response to the limitation of protein translation.

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, e.g. ppGpp directs ribosome synthesis in accordance to the level of amino acid depletion (Ross et al., 2013), cAMP-Crp coordinates catabolic protein expression in accordance to the availability of alpha-keto acids (You et al., 2013) and the PII/NtrBC system determines the degree of nitrogen assimilation in accordance to the availability of glutamine (Reitzer, 2003). However, many mysteries remain, e.g., the coherent response of the expression of proteins in the anabolic sector is well beyond what is known to be controlled by nitrogen regulatory system, and yet the enzymes for amino acid synthesis and nucleic acid synthesis are clearly distinguished in their responses. Also, a substantial number of proteins are in the S-sector which respond 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.

Despite the lack of regulatory information, the simple character of coordinated sector behavior shows that the physiological rationale of proteome management follows basic economic principles of supply and demand that can be understood independent of the underlying mechanisms. A phenomenological model that stipulates flux matching in the flow of material between the sectors of the coarse-grained reaction network, along

with the mutual constraint of a finite proteome, is sufficient to capture the observed sector behavior quantitatively with a few parameters. These governing principles imply that the flux through each sector σ is carried by a mass fraction ϕ_σ whose size is determined by the cost of supplying flux through the given sector under the given mode of growth limitation, with the cost given by $\nu_\sigma^{-1} \equiv \frac{\partial \phi_\sigma}{\partial \lambda}$. In this way, proteome management is analogous to the economic concept of "division of labor", with finite capital allocated according to an effective pricing system given by the ν_σ 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, $\frac{\partial \phi_C}{\partial \lambda}$, is increased while the price to carry flux through the other sectors remains the same. This requires an increased investment of capital, e.g., proteome fraction, to carry the requisite flux J_C. The predictions of the model in two series of test conditions with a different standard condition are in close agreement with measurements, demonstrating the robustness of these conclusions.

While the growth-rate dependent components ($\Delta\phi_\sigma$ s) closely follow economic principles, much of the growth-rate independent component (ϕ_Q) comprises of offsets of the identified proteome sectors, i.e., $\phi_{\sigma 0}$ s as shown in Figure 2.13. It was shown previously that abundance in the growth-rate independent sector directly diminishes the growth rate. 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 (Dennis and Bremer, 1974; Kjeldgaard et al., 1958; Koch and Deppe, 1971). Note that these offsets are the remaining components of proteome sectors extrapolated linearly to zero growth. This work examined cells kept at moderate growth rates; it is unclear when or if the observed linear relations cease to hold. Other competing considerations may well arise at very slow growth, or in starvation conditions, adding to the principles of proteome management revealed by this work.

The quantitative approach described in this work, perhaps among the first to capture noisy -omics data by a simple mathematical model, relies on several key ingredients. As the majority of observed protein expression levels vary within a range of two fold in the observed growth rate range, effective binary classification depends crucially on the highly precise relative quantitation method used to measure individual protein level changes. Measuring the mass fraction of each sector would not have been possible without a method for absolute protein concentration. To overcome the unreliability and noise-dominated signal of individual peptide spectral counts, we developed a coarse-grained spectral counting method that enabled reproducible measurements of absolute abundance for the proteome sectors. In the case of the ribosome sector, the measurement of the R-sector absolute abundance using quantitative mass spectrometry and using a classic biochemical assay is in close agreement. Early works on proteome partition were, by necessity, based upon extrapolations from measurements on a representative set of genes. Here, leveraging a high coverage qMS platform, the rules governing proteome management could be determined using data from a vast majority (~80%) of the expressed proteins in *E. coli*, corresponding to ~1,000 genes. The phenomenological coarse-graining modeling approach presented here is

readily extensible to dissecting global gene expression responses to other perturbations, and suggests a general path forward in the quantitative analysis of multivariate -omics data.

Tables

Table 2.1. Strains and growth conditions.

Three strains were used in this study: the wild type NCM3722, NQ381, and NQ393. The latter two strains are based on NCM3722. The C-limitation was carried out by titrating the lactose uptake for the strain NQ381 growing on lactose minimal medium. Four growth rates were obtained for four different 3MBA levels. The fifth growth condition in the C-limitation series was WT NCM3722 growing on lactose minimal medium. The lactose minimal media were prepared with 1 mM IPTG. NQ393 was used for the A-limitation, with four growth rates corresponding to four different IPTG levels in the glucose minimal medium. Similarly, WT NCM3722 growing on glucose minimal medium was the fifth growth condition in the A-limitation series. WT NCM3722 was used for the R-limitation, with four growth rates corresponding to four chloramphenicol levels in the glucose minimal medium. The fastest growth condition in both the A- and R-limitation series is the condition of WT cells growing on glucose minimal medium. We refer to this growth condition as the "glucose standard condition", from which cell growth was A-limited or R-limited.

Growth limitations	Medium	Strains, inducers/antibiotic amounts, and doubling times					
C-limitation	Lactose minimal medium	Titratable LacY NQ381 (*attB*::$P_{Llac\text{-}O1}$-*xylR*, *lacY*::*km-Pu-lacY*)					WT NCM3722
		3MBA (uM)	0	25	50	500	
		Dbl (min)	92	72	62	48	40
A-limitation	Glucose minimal medium	Titratable GOGAT NQ393 (*attB::Sp-lacIQ-tetR*, Δ*lacY*, Δ*gdhA,* $P_{Llac\text{-}O1}$-*gltBD*)					WT NCM3722
		IPTG (uM)	30	40	50	100	
		Dbl (min)	91	69	58	47	43
R-limitation	Glucose minimal medium	WT NCM3722 (wild type)					
		Chloramphenicol (uM)	8	4	2	0	
		Dbl (min)	147	102	65	42	

Table 2.2. Relative protein expression data, parameters of linear fits, and membership in proteome sectors.

See Supplemental file "Table2_2.xlsx"

Table 2.3. Comparison between the sectors and the clusters.

The numbers inside the parentheses indicate the number of proteins included in the corresponding clusters or sectors. With the exception of the O-sector, each sector is overwhelmingly present in one of the 5 hierarchical clusters, with the bold number indicating the number of overlapping proteins between the corresponding sector and hierarchical cluster.

Sector\Cluster	C-cluster (160)	A-cluster (266)	R-cluster (325)	U-cluster (137)	S-cluster (165)
C-sector (110)	**80**	2	6	4	17
A-sector (209)	3	**150**	19	11	24
R-sector (153)	0	0	**144**	8	0
U-sector (113)	9	6	19	**79**	0
S-sector (159)	33	34	5	0	**86**
O-sector (309)	35	61	132	35	38

Table 2.4. Proteome fractions for the 6 sectors under the three limitations.

Proteome fraction data for the triplicate runs (the 4th-6th columns) and their means (the 7th column) are listed for each sector under each limitation. The means and the corresponding standard deviations are shown in Figure 2.13, with same color scheme for each of the three growth limitations. See Experimental Procedures for how the proteome fraction data were obtained.

		Doubling time (min)	Replicate 1	Replicate 2	Replicate 3	Mean
C-sector	*C-lim* ●	92	0.203	0.221	0.244	0.223
		72	0.185	0.213	0.249	0.216
		62	0.190	0.203	0.226	0.206
		48	0.156	0.166	0.174	0.165
		40	0.086	0.086	0.097	0.089
	A-lim ●	91	0.073	0.080	0.077	0.077
		69	0.076	0.083	0.080	0.080
		58	0.090	0.089	0.090	0.090
		47	0.134	0.136	0.131	0.134
		43	0.119	0.128	0.121	0.123
	R-lim ●	147	0.042	0.036	0.035	0.038
		102	0.051	0.048	0.050	0.049
		65	0.079	0.079	0.077	0.078
		42	0.126	0.122	0.127	0.125
A-sector	*C-lim* ●	92	0.155	0.148	0.129	0.144
		72	0.166	0.156	0.127	0.150
		62	0.165	0.160	0.140	0.155
		48	0.187	0.182	0.176	0.182
		40	0.219	0.218	0.213	0.217
	A-lim ●	91	0.263	0.270	0.265	0.266
		69	0.259	0.265	0.264	0.263
		58	0.245	0.251	0.246	0.247
		47	0.197	0.199	0.190	0.195
		43	0.199	0.196	0.184	0.193
	R-lim ●	147	0.136	0.120	0.120	0.125
		102	0.150	0.138	0.131	0.140
		65	0.170	0.161	0.156	0.162
		42	0.195	0.192	0.193	0.193

Table 2.4. Continued.

		Doubling time (min)	Replicate 1	Replicate 2	Replicate 3	Mean
R-sector	*C-lim* ●	92	0.129	0.149	0.173	0.150
		72	0.141	0.159	0.200	0.167
		62	0.148	0.172	0.215	0.178
		48	0.160	0.183	0.240	0.194
		40	0.189	0.226	0.270	0.228
	A-lim ●	91	0.135	0.152	0.175	0.154
		69	0.145	0.164	0.191	0.167
		58	0.154	0.172	0.203	0.176
		47	0.175	0.197	0.227	0.199
		43	0.178	0.204	0.253	0.212
	R-lim ●	147	0.336	0.414	0.425	0.391
		102	0.303	0.382	0.391	0.358
		65	0.264	0.318	0.325	0.302
		42	0.205	0.232	0.243	0.227
U-sector	*C-lim* ●	92	0.098	0.091	0.073	0.087
		72	0.109	0.102	0.081	0.097
		62	0.114	0.109	0.089	0.104
		48	0.138	0.146	0.111	0.132
		40	0.179	0.171	0.157	0.169
	A-lim ●	91	0.099	0.103	0.094	0.099
		69	0.111	0.101	0.088	0.100
		58	0.125	0.118	0.100	0.114
		47	0.147	0.144	0.130	0.140
		43	0.147	0.140	0.127	0.138
	R-lim ●	147	0.117	0.101	0.109	0.109
		102	0.121	0.114	0.107	0.114
		65	0.137	0.131	0.127	0.132
		42	0.153	0.158	0.146	0.152
S-sector	*C-lim* ●	92	0.181	0.185	0.193	0.186
		72	0.167	0.164	0.166	0.166
		62	0.151	0.153	0.152	0.152
		48	0.127	0.121	0.120	0.122
		40	0.108	0.106	0.101	0.105
	A-lim ●	91	0.170	0.170	0.167	0.169
		69	0.157	0.160	0.158	0.158
		58	0.147	0.150	0.141	0.146
		47	0.125	0.123	0.127	0.125
		43	0.123	0.123	0.122	0.122
	R-lim ●	147	0.082	0.083	0.073	0.079
		102	0.093	0.082	0.086	0.087
		65	0.099	0.089	0.096	0.095
		42	0.112	0.108	0.108	0.109

Table 2.4. Continued.

		Doubling time (min)	Replicate 1	Replicate 2	Replicate 3	Mean
O-sector	*C-lim* ●	92	0.210	0.193	0.181	0.195
		72	0.215	0.197	0.174	0.195
		62	0.217	0.197	0.176	0.196
		48	0.220	0.196	0.179	0.198
		40	0.208	0.185	0.161	0.185
	A-lim ●	91	0.227	0.209	0.208	0.215
		69	0.229	0.213	0.210	0.217
		58	0.218	0.209	0.213	0.213
		47	0.208	0.196	0.192	0.198
		43	0.216	0.202	0.189	0.202
	R-lim ●	147	0.270	0.239	0.234	0.247
		102	0.270	0.229	0.230	0.243
		65	0.242	0.217	0.215	0.225
		42	0.202	0.185	0.179	0.189

Table 2.5. Parameters describing the linear growth-rate dependence of the 6 proteome sectors under the three growth limitations (Eq. [2.1])

For a sector σ, 4 parameters are required to describe the responses to the three growth limitations, with 3 for the Y-intercepts ($\phi_{\sigma,l,0}$) and 1 for the proteome fraction at the glucose standard condition (ϕ_σ^*). The fitted lines are shown in Figure 2.13. The R^2 of the fit measures the quality the overall fit (i.e., the 6×3=18 lines) with respect to the mean proteome fraction data (the last column in Table 2.4). This value of R^2 is also useful for later comparison with the quality of fit by the flux model (Table 2.8). See Experimental Procedures for the definition of R^2.

$\phi_{\sigma,l,0}$	*C-lim ($l = C$)*	*A-lim ($l = A$)*	*R-lim ($l = R$)*	ϕ_σ^*
C-sector ($\sigma = C$)	0.33±0.01	0.03±0.01	-0.00±0.01	0.12±0.00
A-sector ($\sigma = A$)	0.09±0.01	0.34±0.01	0.10±0.01	0.20±0.00
R-sector ($\sigma = R$)	0.09±0.01	0.09±0.01	0.46±0.01	0.22±0.00
U-sector ($\sigma = U$)	0.02±0.01	0.04±0.01	0.09±0.01	0.15±0.00
S-sector ($\sigma = S$)	0.24±0.01	0.22±0.01	0.07±0.01	0.11±0.00
O-sector ($\sigma = O$)	0.20±0.01	0.24±0.01	0.27±0.01	0.19±0.00
R^2 of the fit	0.99			

Table 2.6. Lists of genes associated with each of the GO terms identified by the abundance-based GO analysis.

See Supplemental file "Table2_6.xlsx".

Table 2.7. Flux model equations describing responses of the six sectors to the three growth limitations.

The table lists equations describing all 18 responses, 6 sectors to 3 growth limitations. As developed in the text, the equations are the results of the proteome-based flux model. The equations contain 16 parameters, four of which (κ_C, κ_A, κ_S, and κ_R) are expressed as functions of rest of the parameters (Eqs. 2.S8b], [2.S9b], [2.S10b], and [2.S7b]), reducing the number of free parameters to 12. Due to the definition $\phi_{\max} \equiv 1 - \sum_{\sigma} \phi_{\sigma,0}$, the number of free parameters is further reduced to 11. For a given condition, Eqs. [2.5-2.9] yields an expression of growth rate λ as a function of the effective rate constants and ϕ_{max}, i.e.,

$$\lambda = \frac{\phi_{\max}}{\nu_C^{-1} + \nu_A^{-1} + \nu_R^{-1} + \nu_U^{-1}}, \qquad [2.S11]$$

which further eliminates one parameter if the growth rate of the condition is given.

	C-lim	***A-lim***	***R-lim***
C	$\phi_C(\lambda) = \phi_{C,0} + (1-f)\cdot(\phi_{\max} - \lambda/\kappa_C)$ [2.S8a] $\kappa_C^{-1} = \nu_A^{-1} + \nu_R^{-1} + \nu_U^{-1}$ [2.S8b]	$\phi_C(\lambda) = \phi_{C,0} + (1-f)\cdot\lambda/\nu_C$ [2.S3]	$\phi_C(\lambda) = \phi_{C,0} + (1-f)\cdot\lambda/\nu_C$ [2.S3]
A	$\phi_A(\lambda) = \phi_{A,0} + (1-f)\cdot\lambda/\nu_A$ [2.S4]	$\phi_A(\lambda) = \phi_{A,0} + (1-f)\cdot(\phi_{\max} - \lambda/\kappa_A)$ [2.S9a] $\kappa_A^{-1} = \nu_C^{-1} + \nu_R^{-1} + \nu_U^{-1}$ [2.S9b]	$\phi_A(\lambda) = \phi_{A,0} + (1-f)\cdot\lambda/\nu_A$ [2.S4]
R	$\phi_R(\lambda) = \phi_{R,0} + \lambda/\nu_R$ [2.S1]	$\phi_R(\lambda) = \phi_{R,0} + \lambda/\nu_R$ [2.S1]	$\phi_R(\lambda) = \phi_{R,0} + \phi_{\max} - \lambda/\kappa_R$ [2.S7a] $\kappa_R^{-1} = \nu_C^{-1} + \nu_A^{-1} + \nu_U^{-1}$ [2.S7b]
U	$\phi_U(\lambda) = \phi_{U,0} + \lambda/\nu_U$ [2.S2]	$\phi_U(\lambda) = \phi_{U,0} + \lambda/\nu_U$ [2.S2]	$\phi_U(\lambda) = \phi_{U,0} + \lambda/\nu_U$ [2.S2]
S	$\phi_S(\lambda) = \phi_{S,0} + f\cdot(\phi_{\max} - \lambda/\kappa_S)$ [2.S10a] $\kappa_S^{-1} = \nu_R^{-1} + \nu_U^{-1}$ [2.S10b]	$\phi_S(\lambda) = \phi_{S,0} + f\cdot(\phi_{\max} - \lambda/\kappa_S)$ [2.S10a] $\kappa_S^{-1} = \nu_R^{-1} + \nu_U^{-1}$ [2.S10b]	$\phi_S(\lambda) = \phi_{S,0} + f\cdot\lambda\cdot(1/\nu_C + 1/\nu_A)$ [2.S5]
O	$\phi_O(\lambda) = \phi_{O,0}$ [2.S6]	$\phi_O(\lambda) = \phi_{O,0}$ [2.S6]	$\phi_O(\lambda) = \phi_{O,0}$ [2.S6]

Table 2.8. Parameters of the flux model.

List here are 12 parameters, including 6 growth-rate independent components of the sectors, 4 effective rate constants, the constant f, and ϕ_{max}. Only 10 of them are free parameters due to two relations among the parameters. The first one is the definition $\phi_{\max} \equiv 1 - \sum_{\sigma} \phi_{\sigma,0}$. The second relation is Eq. [2.S11] given that the growth rate λ is known for a condition. For the glucose standard condition, $\lambda = \lambda^*$ (corresponding to a doubling time of 42 min). The parameter values were determined by fitting the 10-parameter flux model (Table 2.7) to the proteome responses data with respect to the glucose standard condition (Table 2.4). The results of the fit are shown as lines in Figure 2.17. The quality of the fit is measured by the value of R^2. See Experimental Procedures for its definition. For the glycerol standard condition, all parameters except ν_C are expected to have these same values. The new ν_C value (indicated in the table with "glycerol" next to it) was determined by Eq. [2.S11], using the growth rate $\lambda^\dagger$ (corresponding to a doubling time of 61 min) of the glycerol standard condition and parameter values from this table. This new value of ν_C, together with the values of other parameters listed in this table, are used for the model (Table 2.7) to give the thick (both solid and dashed) lines in Figure 2.19.

Parameters		Determined values
6 growth-rate independent components	$\phi_{C,0}$	0.05±0.01
	$\phi_{A,0}$	0.12±0.01
	$\phi_{R,0}$	0.09±0.01
	$\phi_{U,0}$	0.08±0.01
	$\phi_{S,0}$	0.06±0.01
	$\phi_{O,0}$	0.21±0.00
4 effective rate constants	ν_C^{-1}	0.10±0.02 0.28±0.04 (glycerol)
	ν_A^{-1}	0.10±0.02
	ν_R^{-1}	0.13±0.02
	ν_U^{-1}	0.07±0.02
	f	0.32±0.04
	$\phi_{\max}$	0.40±0.02
R^2 of the fit		0.94

Table 2.9. Strains and growth conditions for the C- and A- limitations in the glycerol minimal medium.

The glycerol C-limitation was carried out by titrating the glycerol uptake for the strain NQ399 (You et al., 2013) growing on glycerol minimal medium. Three growth rates were obtained for three different 3MBA levels. The fourth growth condition in the glycerol C-limitation series was NCM3722 growing on glycerol minimal medium. The four glycerol C-limitation conditions all contained 1 mM IPTG. Strain NQ393 was used for the glycerol A-limitation, with the four growth rates corresponding to four different IPTG levels in the glycerol minimal medium.

Growth limitations	Medium	Strains, inducers/antibiotic amounts, and doubling times					
Glycerol C-limitation	Glycerol minimal medium	NQ399 (*attB*::$P_{Llac-O1}$-*xylR*, *km-Pu-glpFK*)					NCM3722 (*wild type*)
		3MBA (uM)	25	100	500		
		Dbl (min)	147	99	74		69
Glycerol A-limitation	Glycerol minimal medium	NQ393 (*attB::Sp-lacIQ-tetR*, Δ*lacY*, Δ*gdhA,* $P_{Llac-O1}$-*gltBD*)					
		IPTG (uM)	20	30	40	75	
		Dbl (min)	149	94	73	61	

Table 2.10. Proteome fraction data for the 6 sectors under the growth limitations in glycerol medium.

Proteome fraction data for the 6 proteome sectors are listed for each of the two growth limitations in glycerol medium. See Experimental procedures for how the proteome fraction data were obtained.

	Doubling time (min)	C-sector	A-sector	R-sector	U-sector	S-sector	O-sector
Gly C-lim	147	0.219	0.148	0.125	0.086	0.202	0.193
	99	0.227	0.146	0.137	0.094	0.180	0.195
	74	0.217	0.157	0.148	0.102	0.165	0.194
	69	0.208	0.159	0.153	0.108	0.165	0.191
Gly A-lim	149	0.111	0.233	0.117	0.068	0.207	0.225
	94	0.141	0.213	0.133	0.088	0.194	0.205
	73	0.159	0.178	0.141	0.103	0.195	0.204
	61	0.177	0.160	0.163	0.110	0.173	0.195

2.9 Experimental procedures used in Chapter 2

2.9.1 Growth of bacterial culture

Growth medium: All growth media used in this study were based on the MOPS-buffered minimal medium used by Cayley et al. (Cayley et al., 1989) with slight modifications. The base medium contains 40 mM MOPS and 4 mM tricine (adjusted to pH 7.4 with KOH), 0.1 M NaCl, 10 mM NH_4Cl, 1.32 mM KH_2PO_4, 0.523 mM $MgCl_2$, 0.276 Na_2SO_4, 0.1 mM $FeSO_4$, and the trace micronutrients described in Neidhardt et al. (Neidhardt et al., 1974). For ^{15}N-labeled media, $^{15}NH_4Cl$ was used in place of $^{14}NH_4Cl$.

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 β-D-1-thiogalactopyranoside (IPTG) and various concentrations (0-500 μM) 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 μM) were added to the glucose minimal medium. Various concentrations of chloramphenicol (0-8 μM) 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 μM) of 3MBA. The same glycerol minimal medium with no 3MBA and various amounts of IPTG was used for the A-limitation on glycerol.

Growth measurements: All batch culture growth was performed in a 37°C water bath shaker shaking at 250 rpm. The culture volume was at most 10 ml in 25 mm × 150 mm test tubes. Each growth experiment was carried out in three steps: “seed culture” in

LB broth, “pre-culture” and “experimental culture” in identical minimal medium. For seed culture, one colony from fresh LB agar plate was inoculated into liquid LB and cultured at 37°C with shaking. After 4-5 hrs, cells were centrifuged and washed once with desired minimal medium. Cells were then diluted into the minimal medium and cultured in 37°C water bath shaker overnight (pre-culture). The overnight pre-culture was allowed to grow for at least 3 doublings. Cells from the overnight pre-culture was then diluted to OD600 = 0.005-0.025 in identical pre-warmed minimal medium, and cultured in 37°C water bath shaker (experimental culture). 200 µl cell culture was collected in a Starna Sub-Micro Cuvette (Starna Cells, Atascadero, CA) for OD600 measurement using a Thermal GENESYSTM 20 Spectrophotometer around every half doubling of growth. About 5-7 OD600 data points within the range of ~0.05 and ~0.5 (Above OD600=~0.6 the spectrophotometer was determined to be slightly nonlinear.) were used for calculating growth rate.

2.9.2 Strain construction

The strains used in this study are derived from *E. coli* K12 strain NCM3722 (Lyons et al., 2011; Soupene et al., 2003) and summarized in Table 2.1 and Table 2.9.

Construction of titratable lacY (NQ381) and titratable glpFK (NQ399) strains: DNA fragment containing the P*u* promoter (-1 bp to -178 bp relative to the transcriptional start site) was amplified by PCR from a P*u* promoter containing plasmid pEZ9, then inserted into the *SalI* and *BamHI* sites of plasmid pKD13, producing plasmid pKDP*u*. Using this plasmid as a template, the region containing the km gene and P*u* promoter was PCR amplified and integrated into the chromosome of *E. coli* strain NQ351 between the

lacZ and *lacY* (from *lacZ* stop codon to *lacY* start codon), and in front of *glpF* (-1 bp to -252 bp relative to the translational start point of glpF) respectively, by using the λ Red system (Datsenko and Wanner, 2000). Because the activation of P*u* promoter needs the XylR protein, we constructed a strain NQ386 in which a synthetic lac promoter $P_{Llac\text{-}O1}$ (Lutz and Bugard, 1997) (a promoter that is repressed by LacI but does not need Crp-cAMP for activation) driving *xylR* (*xylR* gene was cloned from pEZ6 (de Lorenzo et al., 1991)) was inserted at the *attB* site. The *kan*-P*u*-*lacY* and *kan*-P*u*-*glpFK* constructs in NQ351 were transferred into strain NQ386 containing P*Llac*-*O1*-*xylR* by P1 transduction (Thomason et al, 2007), resulting in strains NQ381 and NQ399, respectively.

Construction of titratable GOGAT strain (NQ393): Using the λ Red system, we replaced the promoter (+123 bp to -176 bp) of *gltBDF* operon by the synthetic *lac* promoter $P_{Llac\text{-}O1}$ (a promoter that is repressed by LacI but does not need Crp-cAMP for activation) together with selection maker *kan* gene. The resulting *kan*-$P_{Llac\text{-}O1}$-*gltBDF* construct was transferred to strain NCM3722 by P1 transduction. The km gene was then eliminated by using plasmid pCP20 (Cherepanov and Wackernagel, 1995). A *sp*-*lacIQ*-*tetR* cassette providing constitutive expression of *lacI* to tightly repress $P_{Llac\text{-}O1}$ activity was inserted at the *attB* site by P1 transduction. Lactose permease encoded by *lacY* can concentrate intracellular IPTG and will narrow the titration range, we inactivated *lacY* by P1 transduction using strain JW0334-1 from CGSC (*E. coli* Genetic Stock Center, Yale University) as Δ*lacY* donor following by *kan* gene elimination. The *gdhA* gene was knocked out by P1 transduction using strain JW1750-2 from CGSC as Δ*gdhA* donor following by *kan* gene elimination to obtain the final strain NQ393.

2.9.3 Total protein and total RNA Measurements, and β-Galactosidase Assay

Total protein quantitation: The Biuret method was used for total protein quantitation (Herbert et al., 1971). Briefly, 1.8 ml of cell culture at around OD600=0.5 during the exponential phase was collected by centrifugation. The cell pellet was washed with water and re-suspended in 0.2 ml water and fast frozen on dry ice. The cell pellet was then thawed in water bath at RT. 0.1 ml 3M NaOH was added to the cell pellet and samples were incubated at 100°C heat block for 5 min to hydrolyze proteins. Samples were then cooled in water bath at RT for 5 min. The biuret reactions are carried out by adding 0.1 ml 1.6% $CuSO_4$ to above samples with thorough mixing at RT for 5 min. Samples were then centrifuged and the absorbance at 555 nm was measured by a spectrophotometer. Same biuret reaction was also applied to a series of BSA standards to get a standard curve. Protein amounts in the above samples were determined by the BSA standard curve.

Total RNA quantitation: The RNA quantitation method is based on the method used by Benthin et al. (Benthin et al, 1991) with modifications. Briefly, 1.5 ml of cell culture at around OD600=0.5 during the exponential phase was collected by centrifugation and the cell pellet was fast frozen on dry ice. The cell pellet was thawed and washed twice with 0.6 ml cold 0.1 M $HClO_4$, then digested with 0.3 ml 0.3 M KOH for 60 min at 37°C with constant shaking. The cell extracts were then neutralized with 0.1 ml 3 M $HClO_4$ and centrifuged at 13,000 rpm for 5 min. The supernatant was collected and the precipitate was washed twice with 0.55 ml 0.5 M $HClO_4$. A final volume of 1.5 ml of supernatant was then centrifuged and the supernatant was measured for its absorbance at 260 nm on a Bio-Rad spectrophotometer. The RNA concentration (μg/ml/

OD600) was given by OD260 x 31/OD600, where we have used the converting factor of 31 between the OD260 and RNA concentration. The converting factor of 31 is based on the molar extinction coefficient is 10.5 $mmole^{-1}cm^{-1}$ and the average molecular weight of an *E. coli* RNA nucleotide residue is 324.

β-Galactosidase Assay: Samples (0.2 ml cell culture) were collected, fast frozen on dry ice and stored at -80°C prior to β-Galactosidase assay. Four samples were collected for each culture during exponential growth (for OD600 = 0.1~0.5). For each sample collected, b-Galactosidase activity was measured at 37°C by the traditional Miller method (Miller, 1972). The activities obtained (in unit of U/ml=OD420/min/ml) were plotted against the respective OD600, and the resulting slope from linear regression is taken to be the "LacZ expression level" (in unit of U/ml OD600, or "Miller Unit").

2.9.4 ^{15}N-labeled proteomic mass spectrometry

Sample preparation: 1.8 ml of cell culture at OD600=0.4~0.5 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}N reference cell sample (or labeled cell sample) was mixed with each of the ^{14}N cell samples (or non-labeled cell samples), which contained the same amount of proteins. Each aliquot of the ^{15}N samples contained about 100 μg of proteins. Each of the ^{14}N cell samples also contained about 100 μg proteins. For each mode of growth limitation, a ^{15}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 hour. The protein precipitates were sped down by centifugation at 16000 g for 10 min at 4^{o}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 µl 100 mM NH4HCO3 with 5% acetonitrile (ACN). Then 8 µl of 50 mM dithiothreitol (DTT) was added to reduce the disulfide bonds before the samples were incubated at 65^{o}C for 10 min. Cysteine residues were modified by the addition of 8 µl of 100 mM iodoacetamide (IAA) followed by incubation at 30^{o}C for 30 min in the dark. The proteolytic digestion was carried out by the addition of 8 µl of 0.1 µg/µl trypsin (Sigma-Aldrich, St. Louis, MO) with incubation overnight at 37^{o}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 µL sample buffer (5% ACN and 0.1% formic acid).

Mass spectrometry setttings: 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 µL) were injected using an autosampler. The samples were first loaded onto a Nano cHiPLC Trap column 200µm x 0.5mm ChromXP C18-CL 3µm 120Å (Eksigent) at a flow rate of 2 µL/min for 10 minutes. The peptides were then separated on a Nano cHiPLC column

75µm x 15cm ChromXP C18-CL 3µm 120Å (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 400-1250 and accumulation time 0.5 seconds. MS2 settings: mass range of m/z 100-1800, accumulation time 0.05 seconds, high sensitivity mode, charge state 2 to 5, selecting anything over 100 cps, maximal number of candidate/cycle 50, and excluding former targets for 12 seconds 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 ± 0.1 Da, MS/MS tolerance ± 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 the Mascot search results. For our ^{15}N and ^{14}N mixture samples, only the ^{14}N spectra were counted. As described in the main text, 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}N spectra in the sample.

2.9.5 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 less than 3. As described in the main text, the sizes of the three final expression matrices are 856 × 5, 898 × 5, and 756 × 4, respectively for the C-, A-, and R- limitation.

Scaling of the expression matrices: Because different ^{15}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 un-limited 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 1053 × 14 expression matrix, with 14 (=5+5+4) for the total number of growth conditions and 1053 for the total number of unique proteins. The pair-wise distance (d) used for clustering was defined as $d = 1 - \rho$, where ρ 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 log2-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 to measure how the quality of fit. Assuming a data set has values y_i, and the predicted values f_i based on the fit, R^2 is defined as $R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \overline{y})^2}$, where $\overline{y}$ 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.

Determination of the resolution of the relative protein expression method: To decide whether a protein expression level change under a growth limitation is significant, we need to determine the resolution of our method. In our case, multiple (4 or 5) data points were obtained for the level of a protein under a growth limitation. The "fold change" for a protein under a particular mode of growth limitation is determined in the following way: First, the expression level data is fit to a linear function parameterized by the growth rate. Next, obtain the expression levels for the slowest and fastest growth rates (for this particular growth limitation) using the fitted line. Finally, calculate the ratio between the expression levels at the slowest and fastest growth rates. The ratio is taken as the fold change for this protein under the growth limitation. The change of the protein level is taken as significant only if the value of this fold change exceeds the resolution of the method.

To determine the resolution of the method, we draw 4 or 5 random data values

from a normal distribution with mean of 1 and standard deviation of 0.179 (which is the precision of the method; see Figure 2.5), and then calculate the fold change as defined above. The procedure is repeated many times and the standard deviation of the values of fold change is defined the resolution of the method, which was determined to be 0.24 while 5 data values were drawn (as in the case of C- and A- limitations), and 0.25 while 4 data values were drawn (as in the case of R-limitation). We thus took 25% as the resolution of our method.

2.10 Acknowledgement

With permission from the coauthors, chapters 2-3 of this dissertation contain work from a manuscript that has been submitted for publication with the working title, "Metabolic partitioning of the bacterial proteome as a principle of global gene regulation" 2014, Hui, S; Silverman, J; Chen, S; Erickson, D; Wang, J; Hwa, T; Williamson, J.

Chapter 3

An abundance-based functional analysis of proteome sectors

3.1 Introduction

In this chapter, I describe our method for identifying the biological functions of each of the proteome sectors. The usual analysis of Gene Ontology (GO) terms identifies a list of terms that are enriched for a given set of genes compared to a background list of genes, e.g., the genome. Functions for the set of genes are then inferred from the GO terms identified. This approach is not well suited for our case due to the fact that proteome sectors occupy a mass fraction of the proteome, and individual proteins have vastly different abundance inside the cell. To take this fact into consideration, we carry out an abundance-based analysis, aiming to identify for each sector a list of GO terms that best reflect the functions of the sector. We refer to such a list as "best-representing list".

3.2 Data files

3.2.1 Ontology and gene association files

The ontology file and the gene association file of *E. coli* were downloaded from the Gene Ontology project website (http://geneontology.org). The "data-version" of the

ontology file is “2013-07-17” and the “date” is “16:07:2013 13:38”. The gene association file has a “submission date” of “6/5/2013” and a “GOC validation date” of “6/14/2013”. The ontology file contains information for the hierarchical relations between GO terms. In the gene association file, a gene is associated with GO terms that lie at the bottom of the hierarchy. The two files together provide full correspondence between genes and GO terms. For our purpose, we consider only the “biological process” GO terms.

3.2.2 Abundance data file

We use the spectral counting data for the glucose standard condition, which was obtained by merging the spectral counting data of the triplicate runs of the R-limitation sample with no chloramphenicol. As becomes clear later, we found it convenient to introduce the mathematical concept of set for representing spectra. We denote all the spectra in the data set as members in a set S0, so that the number of members in the set (or |S0|) is the spectral counts (e.g., the number of spectra) in the data set. Note that a spectrum in the data set is a recording of a peptide by the mass spectrometer. The same peptide can occur multiple times and each occurrence is counted as one spectrum.

3.3 Filtering out GO terms

We first filter out GO terms that are not enriched in any of the sectors, and are therefore not justified for inclusion in the best-representing list. For a proteome sector *i*, a GO term *t* can be assigned with a value of “fraction of sector” ($\omega_{i,t}$), which is defined as the following,

$$\omega_{i,t} \equiv \frac{|S_t \cap S_i|}{|S_i|},$$

where S_i is a set of spectra that belong to a proteome sector i, and S_t is a set of spectra that are associated with the GO term t. $\omega_{i,t}$ represents how much of the abundance of the sector i the GO term t can account for.

Similarly, for the whole proteome, the GO term t can be assigned with a value of "fraction of proteome" (Ω_t):

$$\Omega_t \equiv \frac{|S_t|}{|S_0|}.$$

We first make a scatter plot of $\omega_{i,t}$ versus Ω_t for all GO terms and for all sectors (Figure 3.1).

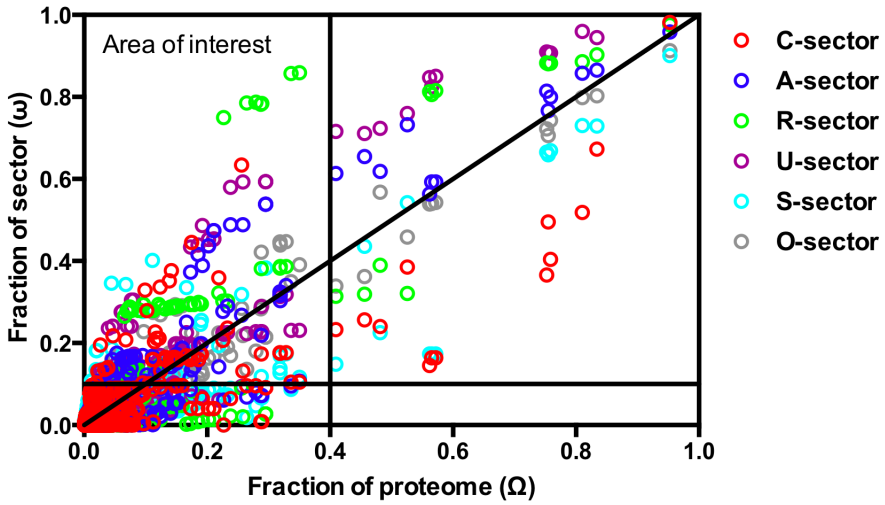


Figure 3.1. Scatter plot of fraction of sector versus fraction of proteome for all GO terms.

In Figure 3.1, we applied three criteria to define the "area of interest" indicated in the figure, outside of which the GO terms have been filtered out, i.e., they are not

qualified to be in the best-representing list and will not be considered in following analysis. The three criteria correspond to the three black lines in Fig. N1: 1) $\omega_{i,t} > 0.1$; 2) $\Omega_t < 0.4$; and 3) $\omega_{i,t} > \Omega_t$. The first criterion filters out GO terms that are small for a sector, i.e., accounting for less than 10% of the sector. The second criterion removes GO terms that are too general (see the list in Table 3.1). The third criterion is a measure of "enrichment", e.g., terms above the line are more enriched in a sector, compared to their distribution in the whole proteome.

Table 3.1. The list of general GO terms.

Fraction of proteome	**GO name**	**GO ID**
0.95	biological_process	GO:0008150
0.83	cellular process	GO:0009987
0.81	metabolic process	GO:0008152
0.76	cellular metabolic process	GO:0044237
0.75	primary metabolic process	GO:0044238
0.75	organic substance metabolic process	GO:0071704
0.57	biosynthetic process	GO:0009058
0.57	organic substance biosynthetic process	GO:1901576
0.56	cellular biosynthetic process	GO:0044249
0.53	single-organism metabolic process	GO:0044710
0.48	nitrogen compound metabolic process	GO:0006807
0.46	small molecule metabolic process	GO:0044281
0.41	organonitrogen compound metabolic process	GO:1901564

The three filters reduce the number of GO terms from 1584 to less than 100 for each of the sectors (See the row of "After the three filters" in Table 3.2).

Table 3.2. Number of GO terms for each sector after applying the three filters.

	Sector	**C**	**A**	**R**	**U**	**S**	**O**
Number of remaining GO terms	**After the three filters**	38	58	82	75	42	29
	After the 4th filter	26	44	42	50	33	27

Next focusing on the remaining GO terms for a sector *i*, we continue filtering out

a term t_1 if there is another term t_2 for which t_1 is a parent term and has $\omega_{i,t_1} = \omega_{i,t_2}$. t_1 is a parent term to t_2 if t_2 has a relation of "is_a" with t_1 according to the gene ontology file. This fourth filter is reasonable because t_2 is more "specific" than t_1 while both account for the same fraction of sector. Table 3.3 lists for the C-sector the terms that are filtered out in this way. For example, the GO term "taxis" is filtered out because the term "taxis" is a parent term to the term "chemotaxis" which accounts for the same fraction of the C-sector. This fourth filter was applied to all of the sectors and further reduced the number of remaining GO terms (See the row of "After the 4th filter" in Table 3.2). We denote the remaining GO terms for a sector i as set T_i.

Table 3.3. List of the GO terms that are filtered out from the C-sector by applying the 4th filter.

	Terms filtered out		**Corresponding "specific" terms**	
Fraction of sector	**GO name**	**GO ID**	**GO name**	**GO ID**
0.21	biological regulation	GO:0065007	regulation of biological process	GO:0050789
0.17	nucleoside metabolic process	GO:0009116	purine nucleoside metabolic process	GO:0042278
0.17	glycosyl compound metabolic process	GO:1901657	nucleoside metabolic process	GO:0009116
0.17	purine-containing compound metabolic process	GO:0072521	purine nucleoside metabolic process	GO:0042278
0.16	purine nucleotide metabolic process	GO:0006163	purine ribonucleotide metabolic process	GO:0009150
0.16	ribonucleoside metabolic process	GO:0009119	purine ribonucleoside metabolic process	GO:0046128
0.16	nucleoside triphosphate metabolic process	GO:0009141	purine nucleoside triphosphate metabolic process	GO:0009144
0.16	nucleoside triphosphate metabolic process	GO:0009141	ribonucleoside triphosphate metabolic process	GO:0009199
0.16	purine nucleoside triphosphate metabolic process	GO:0009144	purine ribonucleoside triphosphate metabolic process	GO:0009205
0.16	ribonucleotide metabolic process	GO:0009259	purine ribonucleotide metabolic process	GO:0009150
0.16	ribonucleoside triphosphate metabolic process	GO:0009199	purine ribonucleoside triphosphate metabolic process	GO:0009205
0.16	ribose phosphate metabolic process	GO:0019693	ribonucleotide metabolic process	GO:0009259
0.13	taxis	GO:0042330	chemotaxis	GO:0006935

3.4 Idnetification of the best-representing list of GO terms

Our task now is to identify from T_i a list of GO terms l (denoted as set T_l) that best represents the sector. Note that $T_l \subseteq T_i$ and thus $|T_l|$ (denoted as k) can be any integer from 1 to n, where $n \equiv |T_i|$. It is clear that k should not be too big for the best-representing list or else the GO terms in the list will overlap with one another.

3.4.1 Lists with small degree of overlapping

Two GO terms t_1 and t_2 overlap with each other in a sector i if $\left|(S_i \cap S_{t_1}) \cap (S_i \cap S_{t_2})\right| > 0$. To quantify the extent of term overlap, we introduce a measure called the "degree of overlapping" ($O_{i,l}$) for a list l and a sector i. To calculate $O_{i,l}$, we first calculate for a GO term t in the list T_l (i.e., $t \in T_l$) the following quantity:

$$O_{i,l,t} \equiv \frac{\left|(S_t \cap S_i) \cap \left(\bigcup_{j \neq t}^{k} S_j\right)\right|}{|S_t \cap S_i|},$$

where $j \in T_l$ and $k \equiv |T_l|$). We then take the minimal value of $O_{i,l,t}$ for $t \in T_l$ to be the value of $O_{i,l}$. According to this definition, $O_{i,l} = 0$ for lists with single terms and for lists with non-overlapping terms. At the other extreme where one term of a list accounts for a subset of spectra that another term in the same list accounts for, the degree of overlapping is at its maximum, or $O_{i,l} = 1$. Note that as k increases, the values of O of the lists tend to get bigger. Figure 3.2 shows for the C-sector the minimal value of $O_{C,l}$ for all l's plotted against the size of lists k. For $k < 4$, there exist lists with zero overlapping, or $O_{C,l} = 0$, while for larger k the minimal value of $O_{C,l}$ quickly goes up. By choosing a cutoff value

for O, we can decide a maximum value for k. For example, a rather large (or loose) cutoff value of 0.3 for $O_{C,l}$ already allows us to only consider lists with $k < 5$.

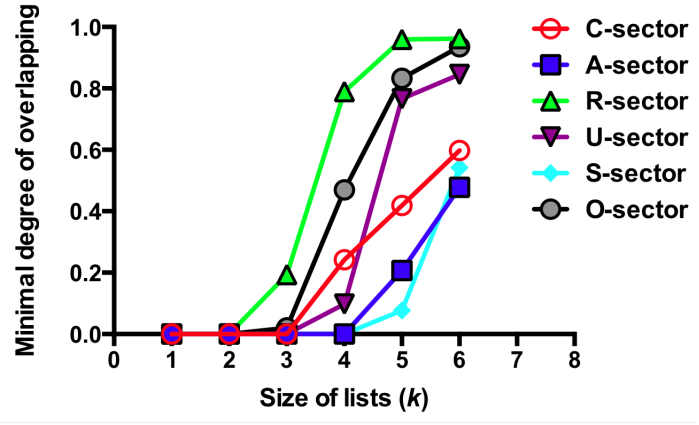


Figure 3.2. The degree of overlapping for a list quickly goes up as the size of the list increases.

To decide the cutoff value for $O_{C,l}$, we focus on the region with $O_{C,l} \leq 0.3$ (Figure 3.3). The histogram reveals a clear cutoff for $O_{C,l}$ at 0.05.

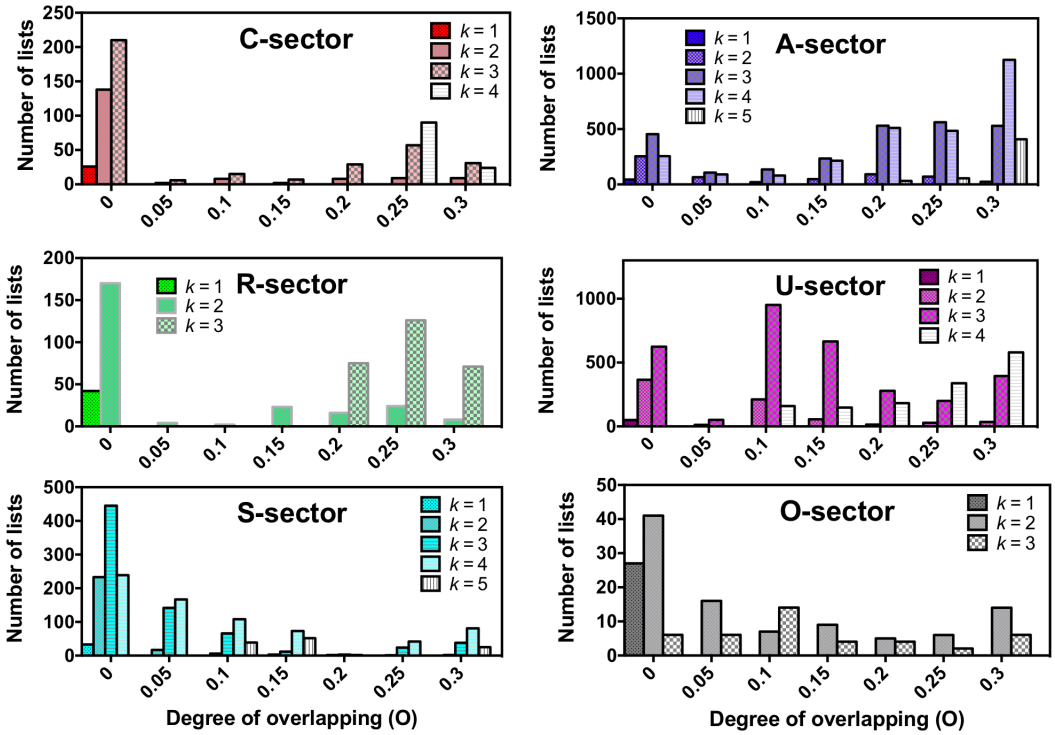


Figure 3.3. Distribution of the degree of overlapping for lists of GO terms.

3.4.2 Lists with large fraction

The concept of "fraction of sector" can be extended to a list of GO terms ($\omega_{i,l}$),

$$\omega_{i,l} \equiv \frac{\left| \left(\bigcup_{j=1}^{k} S_j \right) \cap S_i \right|}{|S_i|},$$

where $j \in T_l$ and $k \equiv |T_l|$).

It is clear that the best-representing list should have a large value of ω. We choose $\omega_{C,l} > 0.6$, i.e., the list l has to account for at least 60% of the C-sector.

In summary, by requiring $O_{C,l} > 0.05$ and $\omega_{C,l} > 0.6$, we reached 75 lists for the C-sector, from which we continue identifying the best-representing one for the sector. We

carried out the same procedure to the other sectors (Figure 3.3) and Table 3.4 shows the number of lists that satisfy the respective cutoff values for O and ω.

Table 3.4. Cutoff values for the degree of overlapping and for the fraction of sector, and the number of remaining lists.

Sector	**C**	**A**	**R**	**U**	**S**	**O**
Cutoff value for O	0.05	0.05	0.05	0.05	0.05	0.05
Cutoff value for ω	0.6	0.6	0.6	0.6	0.6	0.4
Number of lists	75	176	7	173	116	32

3.4.3 Lists with large gene coverage

There is another attribute for a list GO terms and it is what we refer to as "gene coverage". A GO term t is associated with a number of genes in the genome, which is denoted as the set G_t. The set of genes included in a proteome sector i is denoted as set G_i. The "gene coverage" $\chi_{i,t}$ for the GO term in the proteome sector is defined as follows,

$$\chi_{i,t} \equiv \frac{|G_t \cap G_i|}{|G_t|}.$$

The definition can be generalized to a list of GO terms and is denoted as $\chi_{i,l}$ for the list l and for the sector i,

$$\chi_{i,l} \equiv \frac{\left|\left(\bigcup_{j=1}^{k} G_j\right) \cap G_i\right|}{\left|\bigcup_{j=1}^{k} G_j\right|},$$

where $j \in T_l$ and $k \equiv |T_l|$).

Figure 3.4 shows the histogram of χ for each of the sectors, with the total number of lists in each histogram given in Table 3.4.

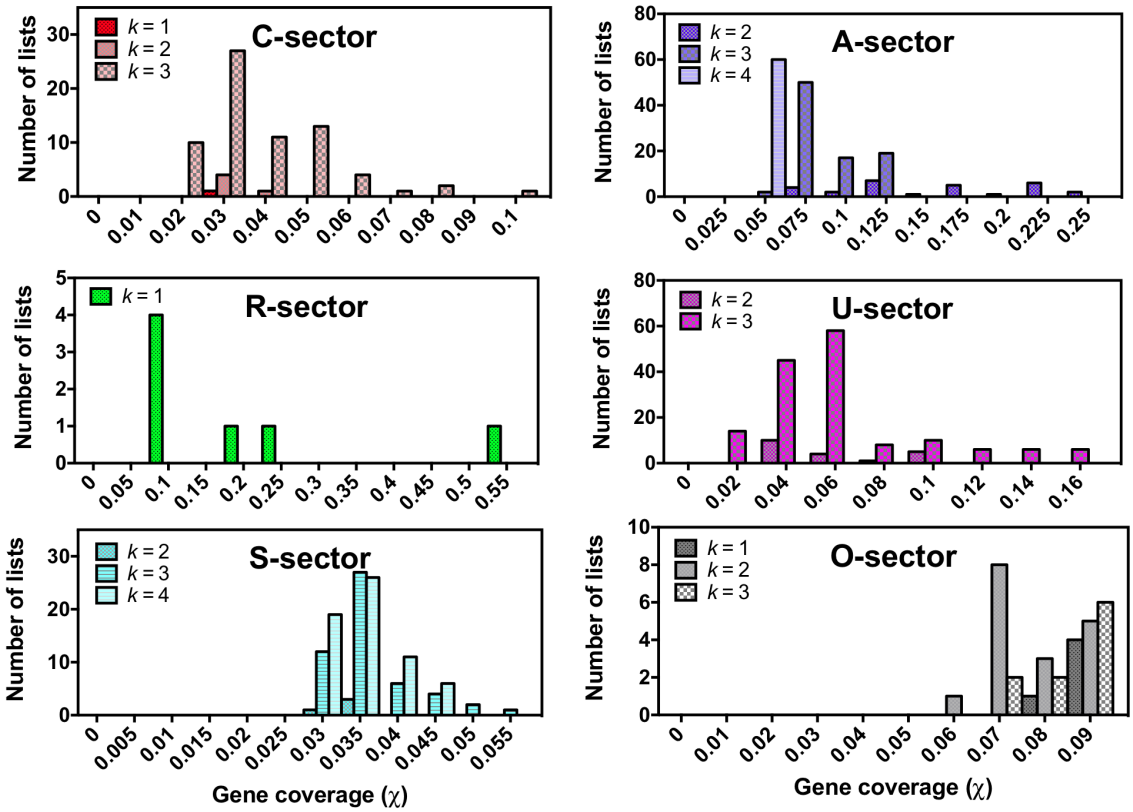


Figure 3.4. Distribution of the gene coverage for lists of GO terms.

3.5 Top lists of GO terms for each sector.

In Figure 3.4, the histograms for the C-, A-, R-, and U- sectors have "long tails", which means that the lists on the right side of the distribution are clearly better than the other lists. Table 3.5 shows a few lists with top χ values for each of the four sectors. The best-representing list (highlighted in light orange color in Table 3.5) for each sector is then manually picked from these small number of choices.

Table 3.5. Top lists and the picked best-representing lists for the C-, A-, R-, and U- sectors.

	ω	O	χ	GO name			ω of individual GO terms		
C	0.75	0.002	0.10	tricarboxylic acid cycle	ion transport	locomotion	0.21	0.33	0.22
	0.67	0.003	0.08	tricarboxylic acid cycle	ion transport	chemotaxis	0.21	0.33	0.13
	0.67	0.003	0.08	tricarboxylic acid cycle	ion transport	response to external stimulus	0.21	0.33	0.14
A	0.65	0.000	0.22	glucose catabolic process	cellular amino acid metabolic process		0.16	0.49	
	0.64	0.000	0.22	glucose catabolic process	small molecule biosynthetic process		0.16	0.47	
	0.60	0.000	0.25	glucose catabolic process	carboxylic acid biosynthetic process		0.16	0.44	
	0.65	0.000	0.17	glucose catabolic process	organonitrogen compound biosynthetic process		0.16	0.49	
	0.63	0.000	0.22	Glycolysis	cellular amino acid metabolic process		0.14	0.49	
	0.62	0.000	0.22	Glycolysis	small molecule biosynthetic process		0.14	0.47	
	0.63	0.000	0.17	glycolysis	organonitrogen compound biosynthetic process		0.14	0.49	
	0.65	0.000	0.18	cellular amino acid metabolic process	carbohydrate catabolic process		0.49	0.17	
	0.65	0.000	0.22	cellular amino acid metabolic process	hexose catabolic process		0.49	0.17	
	0.64	0.000	0.18	carbohydrate catabolic process	small molecule biosynthetic process		0.17	0.47	
	0.60	0.000	0.19	carbohydrate catabolic process	carboxylic acid biosynthetic process		0.17	0.44	
	0.64	0.000	0.22	hexose catabolic process	small molecule biosynthetic process		0.17	0.47	
	0.60	0.000	0.24	hexose catabolic process	carboxylic acid biosynthetic process		0.17	0.44	
	0.65	0.000	0.17	hexose catabolic process	organonitrogen compound biosynthetic process		0.17	0.49	

Table 3.5. Continued.

	ω	O	χ	GO name			ω of individual GO terms		
R	0.75	0.000	0.57	translation					
U	0.67	0.003	0.17	regulation of translation	purine ribonucleoside monophosphate biosynthetic process	alpha-amino acid biosynthetic process	0.12	0.12	0.43
	0.68	0.002	0.16	purine ribonucleoside monophosphate biosynthetic process	posttranscriptional regulation of gene expression	alpha-amino acid biosynthetic process	0.12	0.12	0.43
	0.69	0.003	0.16	regulation of translation	purine ribonucleotide biosynthetic process	alpha-amino acid biosynthetic process	0.12	0.13	0.43
	0.69	0.002	0.15	purine ribonucleotide biosynthetic process	posttranscriptional regulation of gene expression	alpha-amino acid biosynthetic process	0.13	0.12	0.43
	0.69	0.003	0.15	regulation of translation	purine-containing compound biosynthetic process	alpha-amino acid biosynthetic process	0.12	0.14	0.43
	0.69	0.002	0.15	posttranscriptional regulation of gene expression	purine-containing compound biosynthetic process	alpha-amino acid biosynthetic process	0.12	0.14	0.43
	0.68	0.003	0.15	regulation of translation	cellular amino acid biosynthetic process	purine ribonucleoside monophosphate biosynthetic process	0.12	0.44	0.12
	0.68	0.002	0.15	cellular amino acid biosynthetic process	purine ribonucleoside monophosphate biosynthetic process	posttranscriptional regulation of gene expression	0.44	0.12	0.12
	0.69	0.003	0.14	regulation of translation	cellular amino acid biosynthetic process	purine ribonucleotide biosynthetic process	0.12	0.44	0.13
	0.70	0.002	0.14	cellular amino acid biosynthetic process	purine ribonucleotide biosynthetic process	posttranscriptional regulation of gene expression	0.44	0.13	0.12
	0.70	0.003	0.14	regulation of translation	cellular amino acid biosynthetic process	purine-containing compound biosynthetic process	0.12	0.44	0.14
	0.70	0.002	0.14	cellular amino acid biosynthetic process	posttranscriptional regulation of gene expression	purine-containing compound biosynthetic process	0.44	0.12	0.14
	0.69	0.003	0.13	regulation of translation	purine ribonucleoside monophosphate biosynthetic process	carboxylic acid biosynthetic process	0.12	0.12	0.45
	0.69	0.002	0.13	purine ribonucleoside monophosphate biosynthetic process	posttranscriptional regulation of gene expression	carboxylic acid biosynthetic process	0.12	0.12	0.45
	0.65	0.000	0.12	purine ribonucleoside monophosphate biosynthetic process	negative regulation of cellular process	alpha-amino acid biosynthetic process	0.12	0.10	0.43

The histogram for the S-sector (Figure 3.4) does not show a clear tail. We list in Table 3.6 all the lists with $\chi_{S,l} > 0.04$. The best-representing list is again highlighted in light orange color.

Table 3.6. Top lists and the picked best-representing list for the S-sector.

ω	O	χ	GO name				ω of individual GO terms			
0.63	0.020	0.056	carbohydrate metabolic process	response to chemical stimulus	nitrogen compound transport		0.40	0.13	0.10	
0.61	0.020	0.050	carboxylic acid metabolic process	response to chemical stimulus	nitrogen compound transport		0.38	0.13	0.10	
0.64	0.038	0.050	carbohydrate metabolic process	response to stress	nitrogen compound transport		0.40	0.13	0.10	
0.61	0.022	0.047	glucose metabolic process	glyoxylate cycle	response to stress	organic substance transport	0.15	0.18	0.13	0.14
0.60	0.008	0.047	electron transport chain	response to chemical stimulus	nitrogen compound transport	organic substance catabolic process	0.13	0.13	0.10	0.25
0.62	0.022	0.046	response to stress	carboxylic acid metabolic process	nitrogen compound transport		0.13	0.38	0.10	
0.61	0.019	0.046	carbohydrate metabolic process	cellular response to stimulus	nitrogen compound transport		0.40	0.11	0.10	
0.61	0.022	0.045	monosaccharide metabolic process	glyoxylate cycle	response to stress	organic substance transport	0.15	0.18	0.13	0.14
0.67	0.038	0.044	carbohydrate metabolic process	response to stress	organic substance transport		0.40	0.13	0.14	
0.61	0.016	0.044	response to stress	electron transport chain	nitrogen compound transport	organic substance catabolic process	0.13	0.13	0.10	0.25
0.62	0.022	0.044	glyoxylate cycle	response to stress	dicarboxylic acid metabolic process	organic substance transport	0.18	0.13	0.16	0.14
0.62	0.016	0.044	tricarboxylic acid cycle	response to stress	organic substance transport		0.34	0.13	0.14	
0.61	0.022	0.043	glucose metabolic process	response to stress	monocarboxylic acid metabolic process	organic substance transport	0.15	0.13	0.19	0.14
0.61	0.028	0.042	glyoxylate cycle	response to stress	single-organism carbohydrate metabolic process	organic substance transport	0.18	0.13	0.16	0.14
0.61	0.022	0.042	monosaccharide metabolic process	response to stress	monocarboxylic acid metabolic process	organic substance transport	0.15	0.13	0.19	0.14
0.66	0.022	0.041	response to stress	carboxylic acid metabolic process	organic substance transport		0.13	0.38	0.14	
0.62	0.044	0.041	glucose metabolic process	glyoxylate cycle	response to stress	single-organism transport	0.15	0.18	0.13	0.16
0.65	0.019	0.041	carbohydrate metabolic process	cellular response to stimulus	organic substance transport		0.40	0.11	0.14	
0.62	0.022	0.040	response to stress	monocarboxylic acid metabolic process	dicarboxylic acid metabolic process	organic substance transport	0.13	0.19	0.16	0.14
0.62	0.044	0.040	monosaccharide metabolic process	glyoxylate cycle	response to stress	single-organism transport	0.15	0.18	0.13	0.16

The histogram for the O-sector (Figure 3.4) shows that all lists have similar values of χ. We list all the lists of GO terms with $\chi_{O,l} > 0.08$ in Table 3.7. Again, the best-representing list is highlighted in light orange color.

Table 3.7. Top lists and the picked best-representing list for the O-sector.

ω	O	χ	GO name			ω of individual GO terms		
0.44	0.020	0.095	response to stress	nucleotide biosynthetic process	RNA biosynthetic process	0.17	0.10	0.17
0.43	0.020	0.094	transcription, DNA-dependent	response to stress	nucleotide biosynthetic process	0.16	0.17	0.10
0.45	0.027	0.093	response to stress	RNA biosynthetic process	carbohydrate derivative biosynthetic process	0.17	0.17	0.12
0.46	0.020	0.093	response to stress	cellular nitrogen compound biosynthetic process		0.17	0.29	
0.44	0.027	0.093	transcription, DNA-dependent	response to stress	carbohydrate derivative biosynthetic process	0.16	0.17	0.12
0.44	0.027	0.093	response to stress	RNA biosynthetic process	organophosphate biosynthetic process	0.17	0.17	0.11
0.44	0.027	0.092	transcription, DNA-dependent	response to stress	organophosphate biosynthetic process	0.16	0.17	0.11
0.44	0.020	0.092	response to stress	nucleobase-containing compound biosynthetic process		0.17	0.27	
0.46	0.020	0.092	response to stress	heterocycle biosynthetic process		0.17	0.29	
0.46	0.020	0.091	response to stress	organic cyclic compound biosynthetic process		0.17	0.29	
0.45	0.020	0.090	response to stress	aromatic compound biosynthetic process		0.17	0.28	
0.42	0.000	0.089	nucleobase-containing compound metabolic process					
0.45	0.000	0.088	heterocycle metabolic process					
0.44	0.000	0.086	cellular nitrogen compound metabolic process					
0.45	0.000	0.086	organic cyclic compound metabolic process					
0.44	0.000	0.084	cellular aromatic compound metabolic process					
0.45	0.020	0.080	ion transport	response to stress	RNA biosynthetic process	0.11	0.17	0.17

See Table 2.6 of Chapter 2 for lists of genes for the best-representing GO term lists.

Acknowledgement

With permission from the coauthors, chapters 2-3 of this dissertation contain work from a manuscript that has been submitted for publication with the working title, "Metabolic partitioning of the bacterial proteome as a principle of global gene regulation" 2014, Hui, S; Silverman, J; Chen, S; Erickson, D; Wang, J; Hwa, T; Williamson, J.

Chapter 4

The overflow metabolism in *E. coli*

4.1 Introduction

In this chapter, I describe a quantitative, physiological study of the overflow metabolism, where the cell utilizes the less efficient fermentation pathway for energy production and excretes acetate as a result, even when oxygen is available. By taking the constraint of allocating proteome resources into consideration, we are able to provide understanding of this long-standing mysterious phenomenon.

4.2 The acetate line

In previous studies, a dependence of acetate excretion on dilution rates in carbon-limited chemostats has been observed for various organisms (el-Mansi and Holms, 1989; Holms, 1996; Meyer et al, 1984; Nanchen et al., 2006; Renilla et al, 2011; Valgepea et al., 2010) (Figure 4.1).

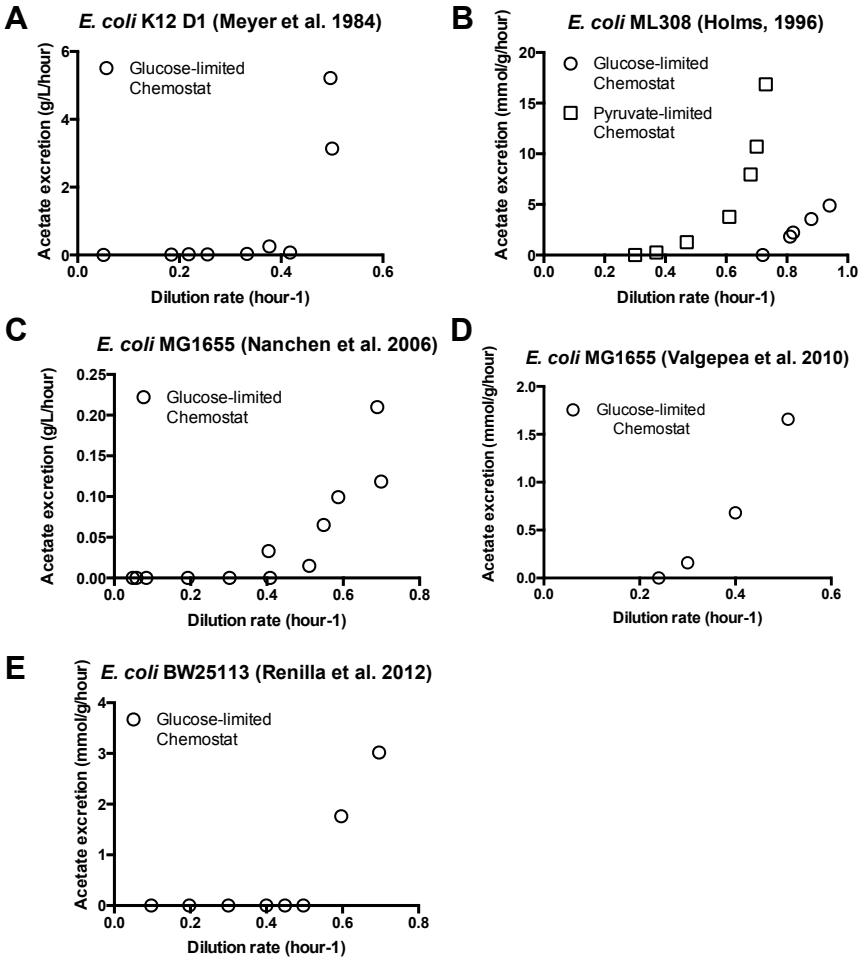


Figure 4.1. Acetate excretion data from previous chemostat studies.

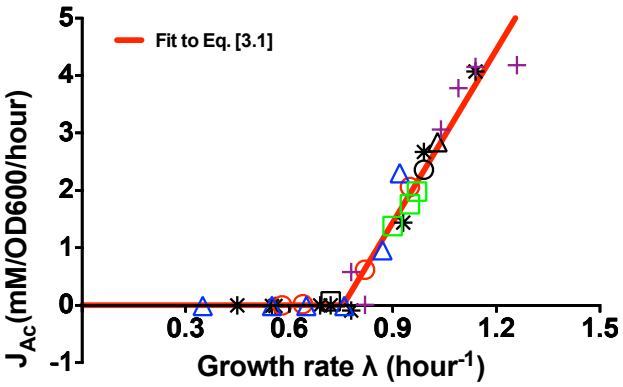
(A) Glucose-limited Chemostat data based on the figure 1 of the reference (Meyer et al., 1984). (B) Glucose-limited and pyruvate–limited Chemostat data from the table 7 of the reference (Holms, 1996). (C) Glucose-limited Chemostat data based on the the figure 3 of the reference (Nanchen et al., 2006). Only data with dilution rate less than the apparent washout dilution rate are plotted here. (D) Glucose-limited Chemostat data from the table 1 of the reference (Valgepea et al., 2010). (E) Glucose-limited Chemostat data based on the figure 4 of the reference (Renilla et al., 2012).

In order to gain a better understanding of this growth rate dependence, we measured acetate excretion and growth rates of a wild-type *E. coli K-12* strain (Table 4.1) grown in minimal medium with a variety of glycolytic carbon sources (Table 4.2, black stars in Figure 4.2). Strikingly, we found that the rate of acetate excretion (in units of mM/OD600/hour), J_{ac}, exhibits a simple linear dependence on the growth rate λ, above a

characteristic growth rate ($\lambda_0 \approx 0.76$/hr, or 55 min per doubling) at which acetate excretion disappears. This relation can be expressed mathematically as

$$J_{ac} = \begin{cases} s \cdot (\lambda - \lambda_0) & for \quad \lambda \geq \lambda_0 \\ 0 & for \quad \lambda < \lambda_0 \end{cases} \qquad [4.1]$$

where s≈10mM/OD600. This functional growth rate dependence is referred to as the "acetate line" (the red line in Figure 4.2). Exceptions to this rule are found for pyruvate and lactate, however these substrates are considered gluconeogenic, as the glycolysis pathway is reversed for growth on these substrates (Figure 4.3). Acetate excretion was observed for growth on gluconic acid, which is metabolized via the pentose phosphate pathway or the Entner-Doudoroff pathway (Figure 4.3). We did not observe acetate excretion for carbons metabolized via the tricarboxylic acid (TCA) cycle (Figure 4.3).



Series	Strain	Description	Carbon	Supplement	Symbol
Various carbons	NCM3722	WT	Various		✳
Reducing glucose uptake	NCM3722	WT	Glucose		○
	NQ1243	Titratable PtsG	Glucose	Various inducer levels	○
Reducing lactose uptake	NCM3722	WT	Lactose		△
	NQ381	Titratable LacY	Lactose	Various inducer levels	△
Increasing glycerol uptake	NCM3722	WT	Glycerol		□
	NQ636 NQ638 NQ640	glpK mutants	Glycerol		□
Reducing precursor demand	NCM3722	WT	Various	7 amino acids	+

Figure 4.2. The acetate line.

Acetate excretion rate as a function of growth rate under carbon limitation for various carbons, uptake titration with Pu-ptsG and Pu-lacY, glpK mutants and with the addition of non-metabolizable amino acids to the growth medium. Acetate excretion rates under different modes of carbon limitation exhibit the same functional dependence on growth rate.

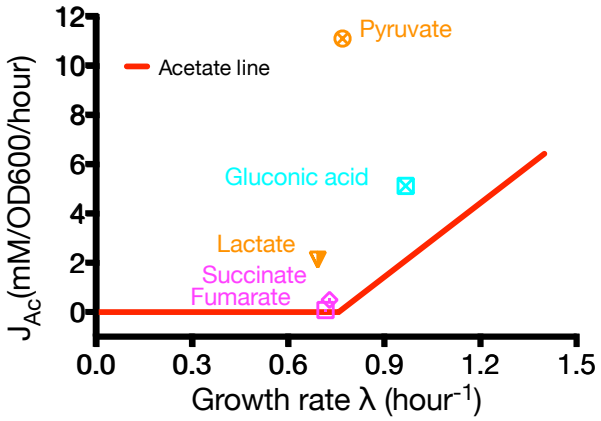


Figure 4.3. Acetate excretion for cells growing on media with non-glycolytic carbons.

E. coli WT NCM3722 was grown on minimal medium with each of the five non-glycolytic carbons, which include two gluconeogenic carbons (i.e., pyruvate and lactate), one carbon metabolized via the pentose phosphate pathway (i.e., gluconic acid), and two carbons metabolized via the TCA pathway (i.e., succinate and fumarate).

We next examined acetate excretion in strains with titratable glucose and lactose uptake systems (Table 4.1, Figure 4.4, Figure 2.2), grown in saturating amounts of glucose and lactose respectively. In both cases (Table 4.3, Table 4.4, red circles and blue triangles in Figure 4.2), acetate excretion follows the same acetate line as defined in Eq. [4.1]. These results suggest that acetate overflow is an innate response that depends on the degree of carbon influx and not specifically on the nature of carbon sources.

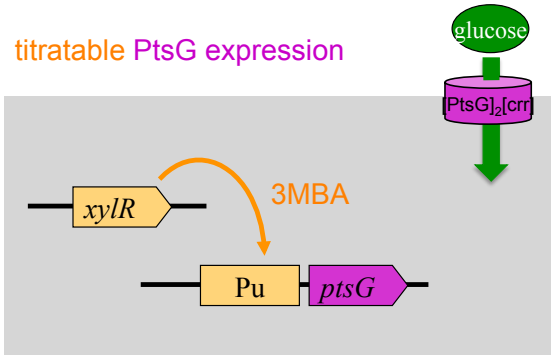


Figure 4.4. Reducing glucose uptake by titrating the expression of PtsG, a subunit of the glucose PTS peamease.

The glucose PTS peamease consists of two subunits, PtsG and Crr. The strain NQ1243 was constructed by replacing the ptsG promoter with a titratable Pu promoter from Pseudomonas putida. The expression of the Pu is activated by the regulator XylR upon induction by 3-methylbenzyl alcohol (3MBA). Strain NQ1243 was grown in glucose minimal medium, supplemented with various 3MBA levels (0-800 μM) to stimulate XylR and titrate the expression of PtsG.

A vivid demonstration of this effect is seen by the behavior of cells growing on glycerol: Wild type *E. coli* grow on glycerol minimal medium at a rate that is below λ_0, and do not excrete acetate (Table 4.2, black square in Figure 4.2) in accordance with Eq. [4.1]. Three isogenic strains expressing different mutant forms of Glycerol Kinase (glpK*, detailed in Figure 4.5) grew at rates faster than λ_0 on glycerol minimal medium. These mutant strains excrete acetate (Table 4.5, green squares in Figure 4.2), with rates dictated by their respective growth rates according to Eq. [4.1].

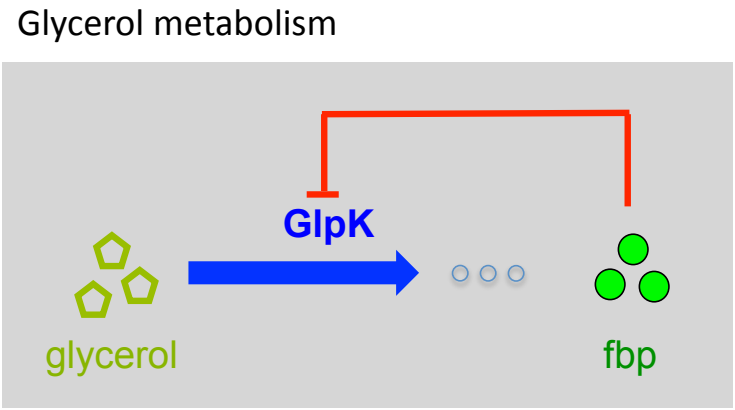


Figure 4.5. Mutants of glycerol kinase (glpK)*

The glpK muants have weaker repression of GlpK by the metabolite fbp, leading to higher catabolism of glycerol (Applebee et al., 2011).

Instead of changing the carbon influx, we also characterized the effect of reducing the metabolic demand for carbon on acetate excretion. This is done by supplementing minimal medium with different carbon sources by amino acids which do not support growth as sole carbon sources (Table 4.6). The addition of these non-metabolizable amino acids resulted in significantly enhanced growth rates, presumably by relieving the demand for carbon precursors needed to synthesize these amino acids. The corresponding acetate excretion rates again fall on the acetate line (Table 4.7, purple crosses in Figure 4.2). Data for other amino acids and carbon combinations is shown in Figure 4.6.

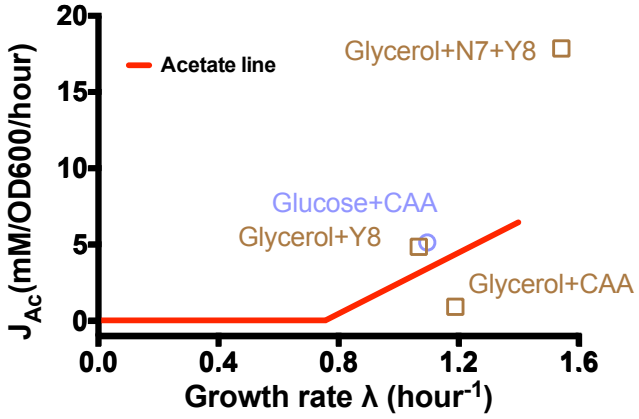


Figure 4.6. Acetate excretion for cells growing on media supplemented with various groups of amino acids.

E. coli WT NCM3722 was grown on glucose or glycerol medium, supplemented with different groups of amino acids. The group "Y8" refers to the mixture of 8 amino acids that have degradation pathways: Alanine (0.8 mM), Arginine (5.2 mM), Aspartate (0.4 mM), Glutamic acid (0.6 mM), Glycine (0.8 mM), Proline (0.4 mM), Serine (10.0 mM), and Threonine (0.4 mM). "N7" refers to the mixture of 7 amino acids that do not have degradation pathways: Histidine (0.2 mM), Isoleucine (0.4 mM), Leucine (0.8 mM), Lysine (0.4 mM), Methionine (0.2 mM), Phenylalanine (0.4 mM), and Valine (0.6 mM). The concentrations of amino acids are based on (Neidhardt et al., 1974). "CAA" is casamino acids.

4.3 A qualitative picture

4.3.1 Two energy-generating pathways

The simple and robust relation described by Eq. [4.1] is reminiscent of other linear dependences of protein expression levels on the growth rate established in recent studies (Scott et al., 2010; You et al., 2013) and this work (Chapter 2). In this case, acetate excretion is a measure of the carbon flux directed towards energy generation (i.e., ATP molecules) by fermentation, mediated not only by glycolytic enzymes but also enzymes in the oxidative phosphorylation system (Figure 4.7). Let the proteome fraction of such enzymes used by the cells for fermentation (a certain portion of each of the

enzymes shown in Figure 4.7) be ϕ_f.

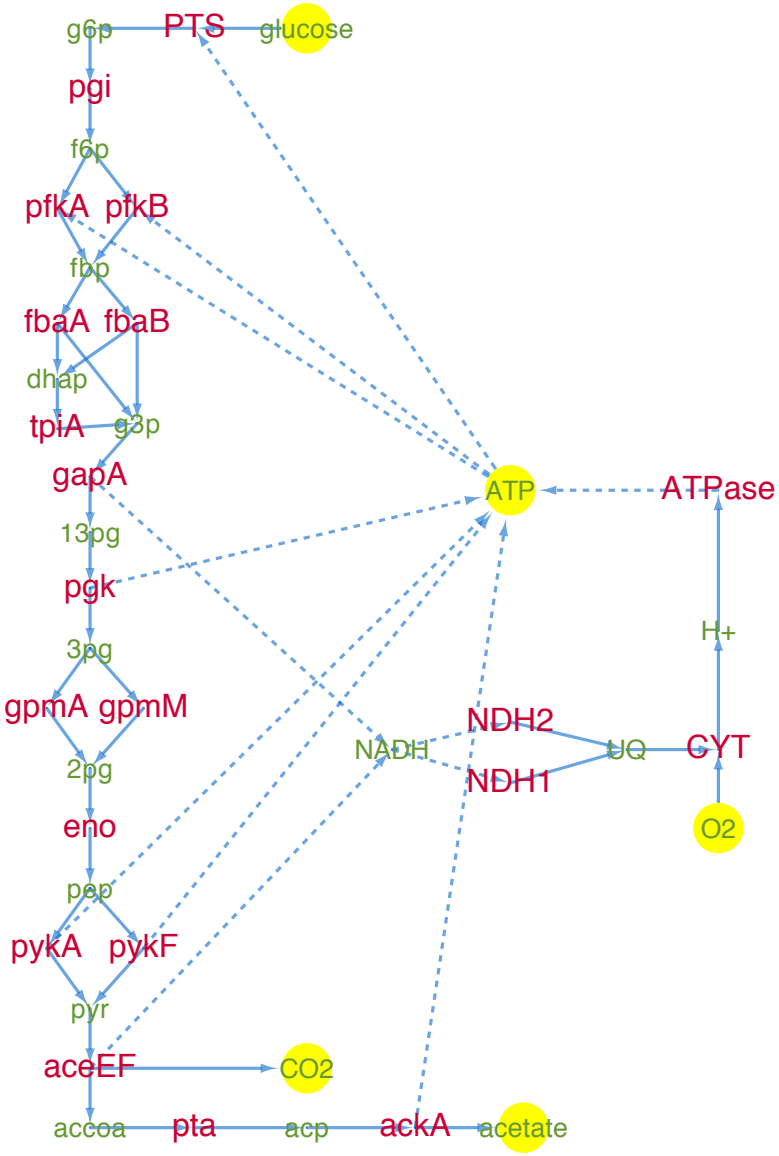


Figure 4.7. The fermentation pathway.

The pathway is based on EcoCyc (Keseler et al., 2012). Green labels indicate metabolites and red labels are for enzymes catalyzing the biochemical reactions, which are represented as blue connecting lines. Highlighted in yellow are the end metabolites of the pathway.

Similarly, respiration is carried out by portions of glycolytic enzymes, TCA enzymes, and also enzymes in the oxidative phosphorylation system (Figure 4.8) and let

their proteome fraction be ϕ_r.

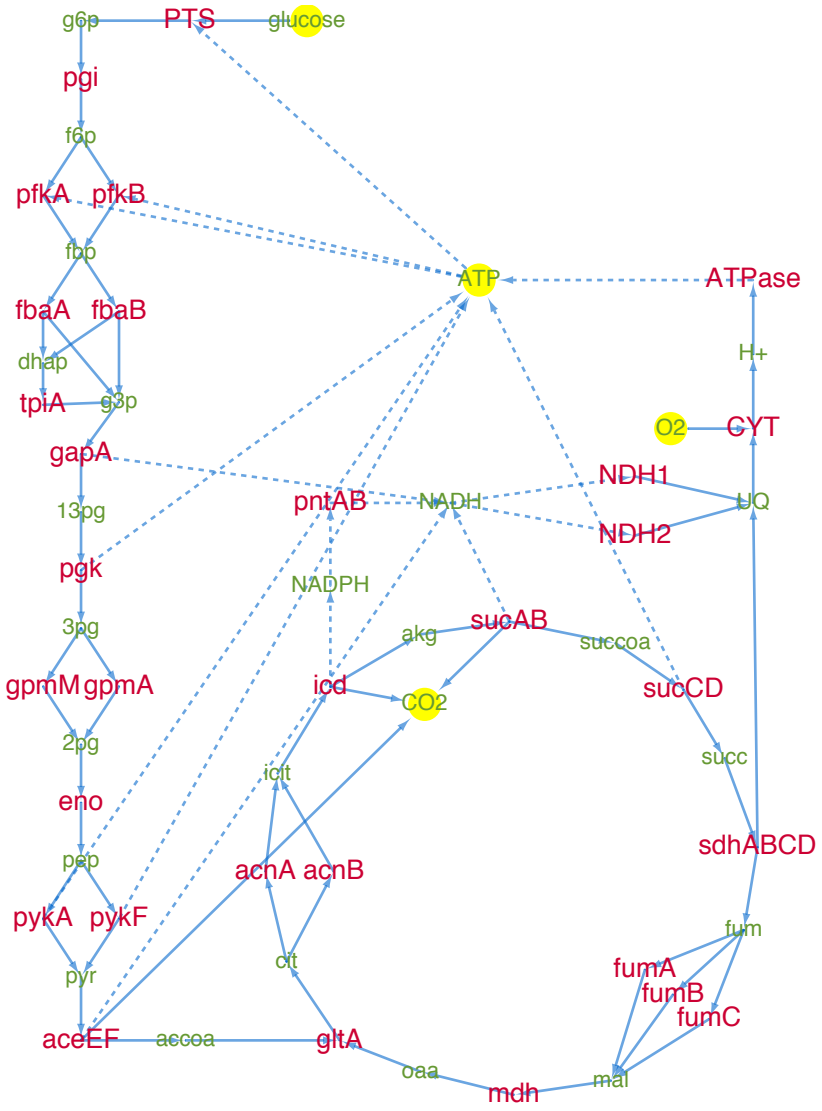


Figure 4.8. The respiration pathway.

The pathway is based on EcoCyc (Keseler et al., 2012). Green labels are for metabolites and red for enzymes catalyzing the biochemical reactions, which are represented as blue connecting lines. Highlighted in yellow are the end metabolites of the pathway.

4.3.2 A picture with energy balance, carbon balance, and proteome constraint

Both the fermentation and respiratory proteins draw carbon away from biomass synthesis, with fluxes $J_{C,f}$ and $J_{C,r}$, respectively, to generate energy fluxes $J_{E,f}$ and $J_{E,r}$.

Biomass synthesis itself is accomplished by the remaining part of the proteome, denoted by ϕ_x, which includes catabolism, anabolism, ribosomes as investigated previously. While biomass synthesis is not the subject of this study, it couples to energy generation since the proteome fractions must add up to unity, i.e.,

$$\phi_f + \phi_r + \phi_x = 1 \cdot \quad [4.2]$$

The cell must choose how much of the proteome to allocate to each energy pathway such that the total energy flux generated satisfies the energy demand for cell growth (denoted by $J_E(\lambda)$), i.e.,

$$J_{E,f} + J_{E,r} = J_E(\lambda) \cdot \quad [4.3]$$

At the same time, not too much carbon should be diverted from the total influx $J_{C,in}$ to meet the demand for biomass synthesis (flux denoted by $J_{C,bm}(\lambda)$), i.e.,

$$J_{C,in} = J_{C,f} + J_{C,r} + J_{C,bm}(\lambda) \cdot \quad [4.4]$$

To a large extent, this allocation depends on the efficiencies of the energy pathways. There are two very different efficiencies: It is well known that respiration has higher "carbon-efficiency" – the energy flux generated per carbon is larger for respiration than fermentation (Unden and Dunnwald, 2008). On the other hand, if respiration has higher protein cost, or lower "proteome-efficiency" than fermentation, i.e., if the energy flux generated per proteome fraction devoted to a pathway is much lower for respiration than fermentation (Figure 4.9A), then we have a scenario that may qualitatively explain the observed disappearance of acetate flux at slow growth rates. This is depicted in Figure 4.9: When carbon is readily available (large $J_{C,in}$) and the cell has the potential to grow rapidly, it is advantageous to generate energy by the more proteome-efficient fermentation pathway, so that more of the proteome can be directed towards biosynthesis.

Conversely, when carbon availability is poor (small $J_{C,in}$), it is more advantageous to generate energy by the more carbon-efficient respiration pathway.

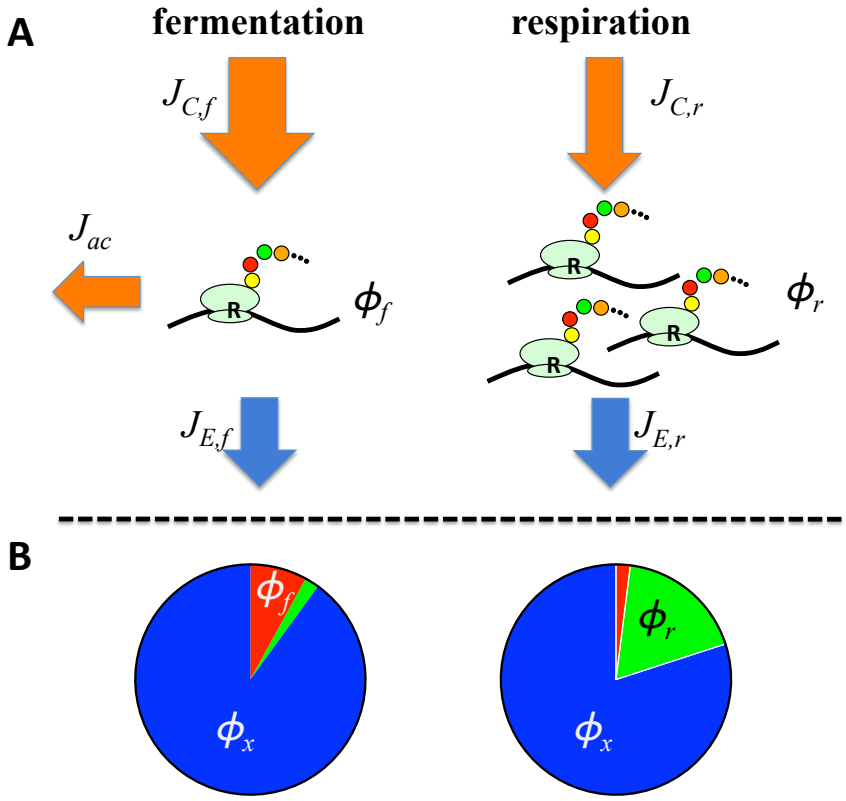


Figure 4.9. Model illustration.

A) Illustration of the properties of fermentation and respiration pathways for energy production. In order to produce the same flux of ATP, fermentation consumes a much larger glucose flux than respiration and a large part of the carbon flux is excreted as acetate. However, at the same time, in order to generate the same flux of energy, fermentation requires a smaller amount of protein investment as compared to respiration.

B) Due to the smaller protein investment required for fermentation, utilization of this pathways frees up a larger fraction of the proteome for the production of biomass enabling a faster growth rate (right). However, this is only possible with a sufficient carbon supply to meet the demands of fermentation in addition to biomass production. Therefore, with limited carbon supply, the cell resorts to respiratory pathways, which have higher protein cost, but require less carbon.

4.3.3 Perturbing proteome availability

In the picture formulated above, acetate excretion is the result of the coordination

of energy demand with carbon influx given constrained proteomic resources. Therefore, we expect perturbations in proteome availability to have significant effects on the rates of acetate excretion. Constraints in proteome availability can be generated by expressing large quantities of useless proteins (Scott et al., 2010), and can be quantified by introducing another protein sector (z) of proteome fraction ϕ_z, with

$$\phi_f + \phi_r + \phi_x + \phi_z = 1 \cdot \quad [4.5]$$

This additional protein load results in a smaller proteome sector available for energy production and therefore should lead to higher acetate excretion rates at a given growth rate.

Experimentally, we implemented the expression of large amounts of the harmless lacZ protein using a plasmid system. For a number of glycolytic carbon sources, we characterized acetate excretion rate for varying degrees of LacZ expression. We observed it to decrease for increasing expression levels of LacZ, which appears to be contradictory to our expectation. However, useless protein expression also reduces cell growth. Plotting acetate excretion against growth rate for varying degrees of LacZ overexpression surprisingly leads to a simple proportionality relation for each carbon source (Table 4.8, Figure 4.10a). For a more systematic analysis, acetate excretion is plotted together with the corresponding LacZ levels and growth rates in a three dimensional graph (Figure 4.10a).

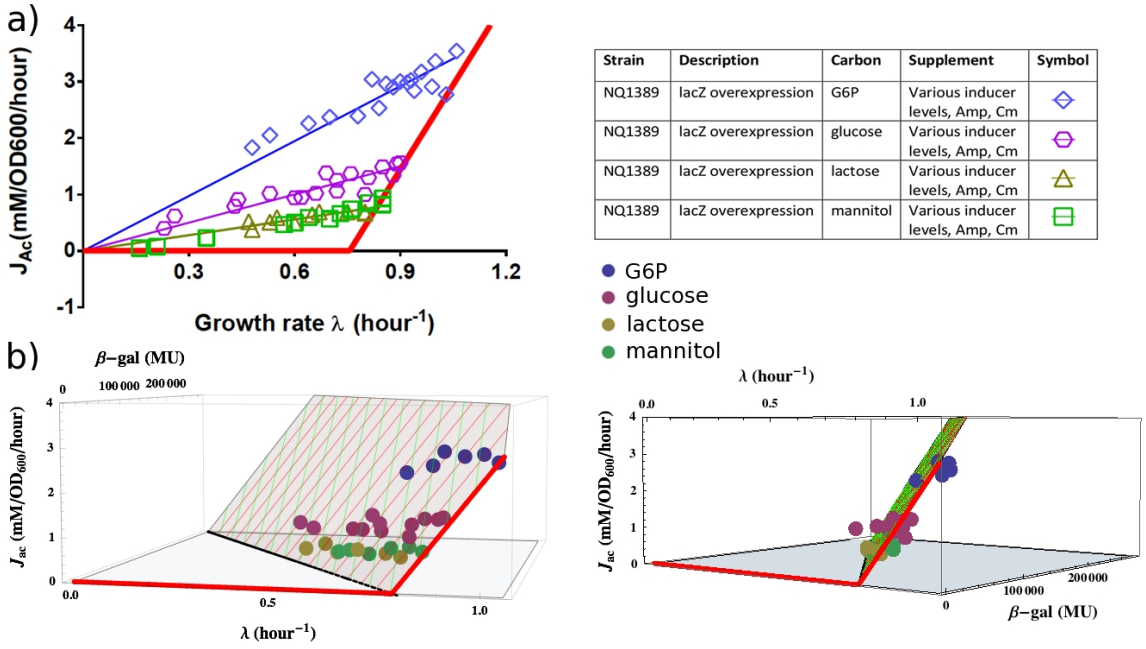


Figure 4.10. Protein overexpression.

a) Acetate excretion rates as a function of growth rate for increasing lacZ expression levels. Fits assuming direct proportionality yield satisfactory agreement with the data. b) Acetate excretion rates, beta-galactosidase activity and corresponding growth rates for overexpression of large quantities of lacZ for G6P, glucose, lactose and mannitol. The plane is obtained from anchoring in the standard acetate line and a one-parameter model fit. For a constant growth rate, acetate excretion increases with increasing lacZ expression level (green lines). On the other hand, for a constant expression level of lacZ, the standard acetate line (red) is shifted parallel (red lines).

Remarkably, we found that the different data points lie on a single plane anchored by the acetate line (thick red line in Figure 4.10b). On this plane, acetate excretion increases linearly with LacZ overexpression at each fixed growth rate (green lines in Figure 4.10b). On the other hand, for each fixed LacZ level, the plane produces a series of parallel shifts of the standard acetate line (thin red lines in Figure 4.10b). These lines are still described by Eq. [4.1], but with a reduction of the threshold growth rate, λ_0, linear with increasing LacZ level (Z) (black line), i.e.

$$\lambda_0(Z) = \lambda_0 \cdot \left(1 - Z/Z_{\max}\right), \qquad [4.6]$$

with $Z_{\max} \approx 2.9 \cdot 10^5$ Miller Units.

We also tested the effect of an opposite shift, to see whether a decrease in acetate excretion is possible as expected by our model. Using a mutant (Δ*flhD*), which does not express the large amount of motility proteins that are not important for growth in well-shaken batch culture, we found that acetate excretion indeed disappears at a faster growth rate in this mutant (Table 4.9, Figure 4.11). The above observations are consistent with the qualitative expectations based on the proteome partition model (Figure 4.9) and furthermore, the simplicity of these results (i.e., the linearity) suggests simple systematic dependences underlying the coordination of acetate overflow with growth rate.

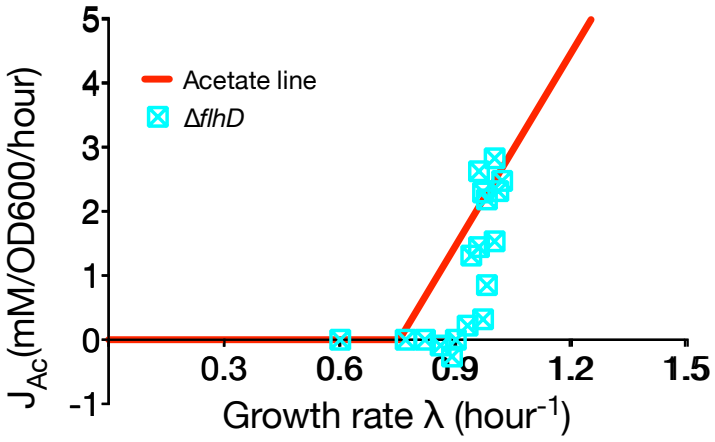


Figure 4.11. Acetate excretion for the flhD deletion strain.

E. coli NQ1389 strain was grown on glucose minimal medium. Various concentrations of 3MBA were used to titrate the glucose uptake.

4.4 A quantitative model

4.4.1 Quantitative assumptions

The concepts represented by Eqs. [4.2-4.4] are transformed into a quantitative and predictive model by making several quantitative assumptions. First, we assume simple proportionalities of energy fluxes to proteome sectors, i.e.,

$$\begin{aligned} J_{E,f} &= \varepsilon_f \cdot \phi_f \\ J_{E,r} &= \varepsilon_r \cdot \phi_r \end{aligned}, \qquad [4.7]$$

where ε_f and ε_r are the proteome efficiencies of the two energy pathways.

The carbon efficiencies for the two pathways are denoted as e_f and e_r, and relate the energy fluxes to the carbon fluxes as

$$\begin{aligned} J_{E,f} &= e_f \cdot J_{C,f} \\ J_{E,r} &= e_r \cdot J_{C,r} \end{aligned}. \qquad [4.8]$$

Note that for a given carbon source, the values for e_f and e_r are known from biochemical stoichiometry.

Eqs. [4.7-4.8] give the expression of carbon fluxes in terms of proteome fractions,

$$\begin{aligned} J_{C,f} &= \frac{\varepsilon_f}{e_f} \cdot \phi_f \\ J_{C,r} &= \frac{\varepsilon_r}{e_r} \cdot \phi_r \end{aligned}. \qquad [4.9]$$

We then assume a linear growth-rate dependence of the remaining proteome sector ϕ_x responsible for biomass synthesis under carbon-limitation,

$$\phi_x = \phi_0 + b\lambda, \qquad [4.10]$$

which is consistent with previous findings (You et al., 2013) and results in Chapter 2, and furthermore, simple direct proportionalities of biomass carbon demand to growth rate,

$$J_{C,bm}(\lambda) = \beta\lambda\,, \qquad [4.11]$$

where the parameter β is known from the published flux data (Fischer and Sauer, 2003; Perrenoud and Sauer, 2005; Teixeira de Mattos and Neijssel, 1997; Usui et al., 2012).

Finally, for the energy demand of the cell, we assume a linear functional dependence on growth rate given by

$$J_E(\lambda) = \sigma\lambda\,. \qquad [4.12]$$

With these assumptions, Eqs. [4.2-4.4] are then transformed into the following equations, respectively,

$$\phi_f + \phi_r = 1 - \phi_0 - b\cdot\lambda\,, \qquad [4.13]$$

$$\varepsilon_f\cdot\phi_f + \varepsilon_r\cdot\phi_r = \sigma\cdot\lambda\,, \qquad [4.14]$$

and

$$\frac{\varepsilon_f}{e_f}\cdot\phi_f + \frac{\varepsilon_r}{e_r}\cdot\phi_r = J_{C,in} - \beta\cdot\lambda\,. \qquad [4.15]$$

In the following sections, we explore the consequences of the simple model, as represented by Eqs. [4.13-4.15].

4.4.2 Prediction of the acetate line

This model can be solved to yield simple solutions for proteome fractions and growth rate, with ϕ_f and ϕ_r given respectively by

$$\phi_f = \frac{(b\cdot\varepsilon_r + \sigma)\cdot\lambda - \phi_{E,\max}\cdot\varepsilon_r}{\varepsilon_f - \varepsilon_r}\,, \qquad [4.16]$$

and

$$\phi_r = \frac{-(b\cdot\varepsilon_f + \sigma)\cdot\lambda + \phi_{E,\max}\cdot\varepsilon_f}{\varepsilon_f - \varepsilon_r}\,, \qquad [4.17]$$

where the expression for λ is given by

$$\lambda = \frac{(\frac{1}{\varepsilon_r} - \frac{1}{\varepsilon_f}) \cdot J_{C,in} + (\frac{1}{e_f} - \frac{1}{e_r}) \cdot \phi_{E,\max}}{(\frac{1}{\varepsilon_r} - \frac{1}{\varepsilon_f}) \cdot \beta + (\frac{1}{e_f} - \frac{1}{e_r}) \cdot b + (\frac{1}{\varepsilon_r \cdot e_f} - \frac{1}{\varepsilon_f \cdot e_r}) \cdot \sigma}, \quad [4.18]$$

where $\phi_{E,\max} \equiv 1 - \phi_0$.

The carbon fluxes through fermentation ($J_{C,f}$) and respiration ($J_{C,r}$) pathways can be obtained from Eq. [4.16] and Eq. [4.17], respectively as,

$$J_{C,f} = \frac{\varepsilon_f}{e_f} \cdot \frac{(b \cdot \varepsilon_r + \sigma) \cdot \lambda - \phi_{E,\max} \cdot \varepsilon_r}{\varepsilon_f - \varepsilon_r}, \quad [4.19]$$

and

$$J_{C,r} = \frac{\varepsilon_r}{e_r} \cdot \frac{-(b \cdot \varepsilon_f + \sigma) \cdot \lambda + \phi_{E,\max} \cdot \varepsilon_f}{\varepsilon_f - \varepsilon_r}. \quad [4.20]$$

Using the stoichiometry relation $J_{ac} = \frac{1}{3} \cdot J_{C,f}$ (Keseler et al., 2012), acetate excretion rate is simply given by,

$$J_{ac} = \frac{1}{3} \cdot \frac{\varepsilon_f}{e_f} \cdot \frac{(b \cdot \varepsilon_r + \sigma) \cdot \lambda - \phi_{E,\max} \cdot \varepsilon_r}{\varepsilon_f - \varepsilon_r}, \quad [4.21]$$

which as a function of growth rate indeed takes the linear form we observe experimentally, i.e., Eq. [4.1]. The model yields the dependence of the acetate line on the available proteome fraction and the energy generation efficiencies ε_f and ε_r: $\lambda_{ac,0} = \frac{\phi_{E,\max}}{b + \sigma/\varepsilon_r}$ and $s_{ac} = \frac{1}{3} \cdot \frac{\varepsilon_f}{e_f} \cdot \frac{b + \sigma/\varepsilon_r}{\varepsilon_f/\varepsilon_r - 1}$. Notably, according to this simple model, acetate excretion only occurs under the condition $\varepsilon_f > \varepsilon_r$, and therefore, the energy generation efficiency of fermentation ε_f must be higher than that of respiration ε_r.

4.4.3 Prediction of the linear rise of respiration

Furthermore, the model also yields a corresponding linear increase of respiration with decreasing growth rate (see Eq. [4.20]), which can be tested by determining the CO2

flux through the pathway (called respiratory CO2 flux J_{rCO2}). Using the stoichiometry relation $J_{rCO2} = J_{C,r}$, from Eq. [4.20] the growth rate dependence of J_{rCO2} is simply given by

$$J_{rCO2} = s_{rCO2} \cdot (-\lambda + \lambda_{rCO2,0}), \qquad [4.22]$$

where $s_{rCO2} = \frac{\varepsilon_r}{e_r} \cdot \frac{b + \sigma/\varepsilon_f}{1 - \varepsilon_r/\varepsilon_f}$ and $\lambda_{rCO2,0} = \frac{\phi_{E,\max}}{b + \sigma/\varepsilon_f}$.

Experimentally, J_{rCO2} was obtained by measuring the total CO2 production, acetate excretion, and growth rates for a series of points along the standard acetate line using the titratable lactose uptake system (Figure 2.2). Remarkably, J_{rCO2} exhibits a simple linear increase with decreasing growth rate (Table 4.10, Figure 4.12), in accordance with our expectations from the model. See Chapter 5 for details on the experimental procedures and analysis of CO2 fluxes.

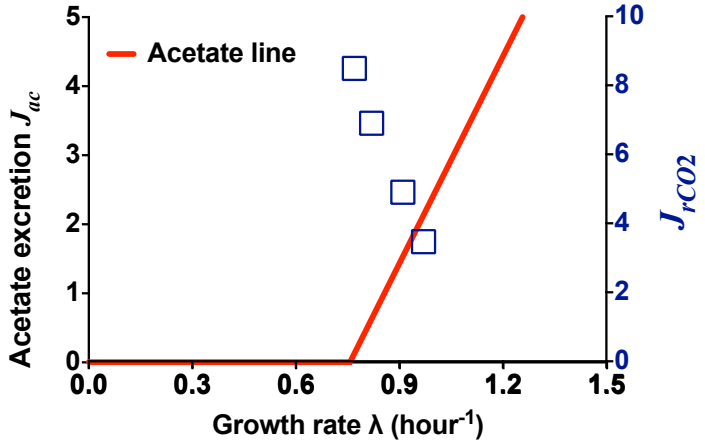


Figure 4.12. The respiration CO2 flux (J_{rCO2}) for cells with varying lactose uptake.

E. coli WT NCM3722 and NQ381 strain were grown on lactose minimal medium in a bioreactor. 3MBA was used to titrate the lactose uptake of NQ381.

4.4.4 Protein overexpression

In the framework of the model, protein overexpression reduces the "available" proteome fraction for energy production, $\phi_{E,\max}$. Hence, from the model expression for Eq. [4.21], we see that this reduces the threshold growth rate $\lambda_{ac,0}$ of the acetate line, but leaves the slope s_{ac} unchanged. We therefore expect parallel shifts of the standard acetate excretion line for a constant level of protein overexpression. This prediction is in accordance with our experimental results (red lines in Figure 4.10b).

4.4.5 Inhibiting peptide elongation

In order to further test the quantitative model established above, we performed additional experimental perturbations. Translational limitation by sub-lethal doses of the antibiotic chloramphenicol (Cm) inhibits peptide elongation and makes the cell respond by allocating a larger proteome fraction to ribosomes. In the language of the model, chlorampenicol results in an increase of the biomass sector ϕ_x for an identical growth rate. Because the need for ribosomes is proportional to growth rate, the size of the increase in ϕ_x due to chloramphenicol is also proportional to growth rate and therefore corresponds to an increased parameter b. The model then predicts increased acetate excretion, with modified acetate lines terminating at reduced thresholds $\lambda_{ac,0}$ and steeper slopes s_{ac} such that the lines all extrapolate to the same vertical intercept $J_{ac,0}$, i.e.,

$$J_{ac,0} = s_{ac} \cdot \lambda_{ac,0} = \frac{1}{3} \cdot \frac{\varepsilon_f}{e_f} \cdot \frac{\phi_{E,\max}}{\varepsilon_f / \varepsilon_r - 1}. \qquad [4.23]$$

This expectation is in good agreement with the data presented in Figure 4.13 (Table 4.11), as demonstrated by the one-parameter model fits.

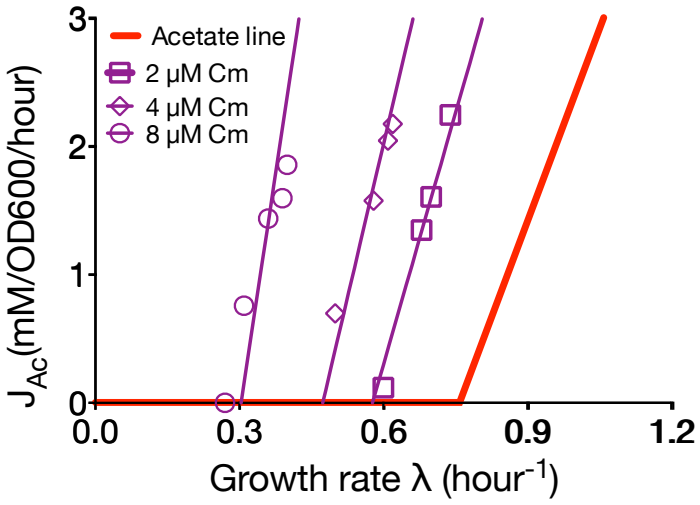


Figure 4.13. Acetate excretion for cells grown in media with chloramphenicol.

E. coli NQ381 was grown on lactose minimal medium, with 2, 4, and 8 μM chloramphenicol. For each concentration of chloramphenicol, various concentrations of 3MBA were used to titrate lactose uptake. Each of the linear fits for the three concentrations of chloramphenicol was one-parameters fit with the Y-axis fixed at $J_{ac,0}$.

4.4.6 Draining energy

Instead of directly perturbing proteome allocation, we also investigated the effect of altering the cell's energy balance. Intuitively, one might expect that an increased energy demand should lead to a larger fraction of carbon uptake being used for respiration and a corresponding decrease in acetate excretion, in order to produce more energy from the available carbon. However, in the framework of our model, we anticipate the opposite response: At a given growth rate, the increased energy demand forces the cell to shift to the more proteome-efficient energy production pathway, i.e., fermentation, resulting in increased acetate excretion. More quantitatively an energy leakage flux can be added to the right hand side of Eq. [4.14] and (assuming the same growth rate dependence of the biomass sector $\phi_x = \phi_0 + b\lambda$) the model then predicts a parallel shift of

the acetate line to higher excretion rates (Eq. [4.21]).

The effect of energy dissipation was tested by expressing a mutant lactose transporter (LacYVal^{177}) known to leak protons across the inner membrane (Brooker, 1991), which directly affects the energy budget of the cell. By titrating carbon uptake while keeping a constant expression level of LacYVal^{177}, we obtained the modified acetate line (blue squares in Figure 4.14,

Table 4.12) as anticipated, i.e., a simple parallel shift to higher acetate excretion rates. In contrast, the control strain with the same expression of wildtype LacY (magenta triangles in Figure 4.14, Table 4.12) closely follows the standard acetate line.

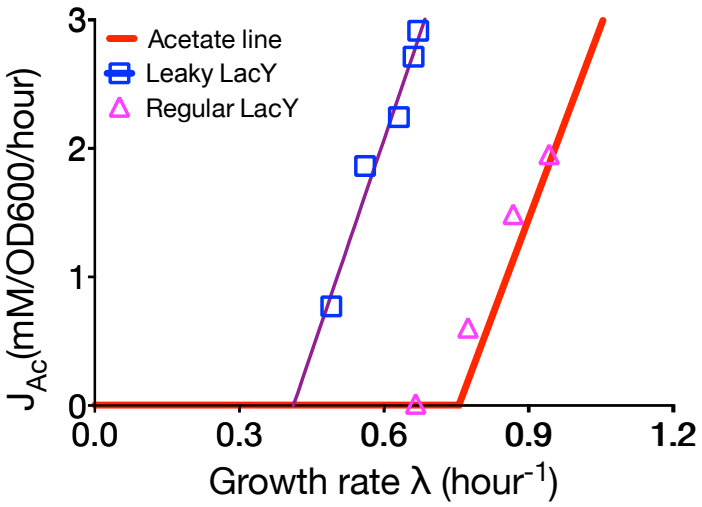


Figure 4.14. Acetate excretion for cells with leaky LacY.

E. coli NQ1313 was grown on glucose minimal medium. Various concentrations of 3MBA were used to titrate the glucose uptake. The best linear fit (purple line) to the data has a slope of 11.0 (compare to the slope of 10.0 of the acetate line). For control, the strain NQ1314 was used.

4.4.7 Experimental determination of proteome efficiencies

Furthermore, our measurement of the CO2 production rate, in combination with growth, carbon uptake and acetate excretion rates allowed us to partition proteins measured by mass spectrometry into different proteome sectors, according to the sector's contribution to the flux. Thereby, we were able to determine the proteome fractions of the different model sectors (see Chapter 5). The energy production fluxes of fermentation and respiration are plotted as a function of their determined proteome fractions in Figure 4.15 (Table 4.10). The slopes of these curves directly represent the energy production efficiencies defined in our model and we obtain $\varepsilon_f \approx 770$ mM/OD600/hour, $\varepsilon_r \approx 380$ mM/OD600/hour. This directly validates our key hypothesis that $\varepsilon_f > \varepsilon_r$.

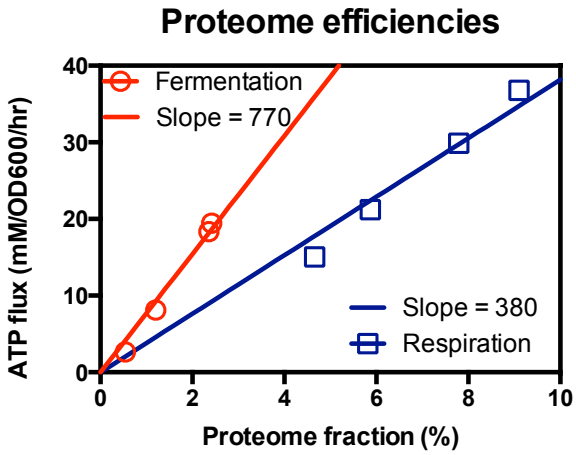


Figure 4.15. The proteome efficiencies of the two energy pathways.

4.5 Summary

We observed a robust linear relation between acetate excretion rate and growth

rate when the status of carbon supply was varied, e.g., by using various carbon sources, titrating the expression of carbon uptake proteins, etc. Furthermore, distinct but still linear relations were obtained when the proteome availability was perturbed (e.g., by expressing useless proteins), and when energy demand was increased (e.g., by introducing a futile proton cycle).

These linear relations under different modes of growth limitation can be quantitatively described by the extension of a coarse-grained resource allocation model (Scott et al., 2010) explicitly including proteome fraction for energy generation. The model invokes a single unknown parameter that describes the difference in protein cost for fermentation and respiration. Using flux analysis and protein mass spectrometry measurements (Chapter 2), we were able to determine directly the protein costs of the two energy-generating pathways and validated quantitatively the key hypothesis of the model.

These results establish that acetate overflow in *E. coli* results from the tradeoff between efficient utilization of carbon resources by respiration and efficient utilization of proteome resources by fermentation, and suggest optimal allocation of carbon and proteome resources as a possible driving force for the occurrence of overflow metabolism in other organisms.

4.6 Tables

Table 4.1. Strains used in this study.

All the strains used are derived from the E. coli K12 strain NCM3722.

Strain	Genotype	Description
NCM3722	wild-type *E. coli* K12 strain	parent strain for all strains used here.
NQ381	*attB*::$P_{Llac-O1}$-*xylR*, *lacY*::*km-Pu-lacY*	titratable LacY
NQ636	*glpK g184t*	Glpk mutant
NQ638	*glpK a218t*	Glpk mutant
NQ640	*glpK g692a*	Glpk mutant
NQ1243	*ycaD::FRT:Ptet:xylR PptsG::kan:Pu:ptsG*	Titratable PtsG
NQ1313	*ycaD::FRT:Ptet:xylR PptsG::kan:Pu:ptsG; Ptet-leaky_lacY on pZA31*	Leaky LacY
NQ1314	*ycaD::FRT:Ptet:xylR PptsG::kan:Pu:ptsG; Ptet-_lacY on pZA31*	LacY control
NQ1388	*ycaD::FRT:Ptet:xylR Pu:ptsG; ΔflhD-kan*	*flhD* deletion
NQ1389	*Ptet-tetR on pZA31; Ptetstab-lacZ on pZE1*	LacZ over-expressing strain

Table 4.2. The acetate excretion for cells grown in various carbon sources.

E. coli NCM3722 strain was grown in minimal medium supplied with each of the above carbon sources. These data are shown as black stars (✳) in Figure 4.2, except that glucose data point is shown as black circle (○), lactose as black triangle (△), and glycerol as black square (□).

Carbon source	Growth rate λ ($hour^{-1}$)	Acetate excretion J_{Ac} (mM/OD600/hour)
0.2% (w/v) glucose	1.01	2.32
0.2% (w/v) lactose	1.03	2.84
0.2% (w/v) glycerol	0.72	0.07
0.2% (w/v) arabinose	0.99	2.67
0.2% (w/v) sorbitol	0.55	-0.01
0.2% (w/v) galactose	0.56	-0.01
0.2% (w/v) melibiose	0.69	0.00
0.2% (w/v) maltose	0.78	-0.09
0.2% (w/v) fructose	0.72	0.00
0.2% (w/v) mannitol	0.93	1.44
0.2% (w/v) mannose	0.45	0.00
20 mM Glucose 6-phosphate disodium	1.14	4.07

Table 4.3. The acetate excretion for cells with reduced glucose uptake by titrating the expression of PtsG.

E. coli NQ1243 (Titratable ptsG; Figure 4.4) strain was grown in minimal medium supplied with 0.2% glucose (w/v) and each of the 3MBA levels. 3MBA was used to titrate the expression of PtsG. These data are shown as red circles (○) in Figure 4.2.

3MBA (μM)	Growth rate λ (h^{-1})	Acetate excretion J_{Ac} (mM/OD600/hour)
0	0.58	0.00
20	0.64	0.02
300	0.82	0.62
800	0.95	2.06

Table 4.4. The acetate excretion for cells with reduced lactose uptake by titrating the expression of LacY.

E. coli NQ381 (Titratable lacY; Figure 2.2) strain was grown in minimal medium supplied with 0.2% lactose (w/v) and each of the 3MBA levels. 3MBA was used to titrate the expression of LacY. These data are shown as blue triangles (△) in Figure 4.2.

3MBA (μM)	Growth rate λ (h^{-1})	Acetate excretion J_{Ac} (mM/OD600/hour)
0	0.35	-0.01
25	0.55	-0.01
50	0.65	-0.01
100	0.76	-0.01
400	0.87	0.96
800	0.92	2.30

Table 4.5. The acetate excretion for cells with increased glycerol uptake.

E. coli NQ636, NQ638, and NQ640 (Figure 4.5) strains were grown in minimal medium supplied with 0.2% glycerol (w/v). These data are shown as green squares (□) in Figure 4.2.

Strain	Growth rate λ (h^{-1})	Acetate excretion J_{Ac} (mM/OD600/hour)
NQ636	0.97	1.98
NQ638	0.90	1.38
NQ640	0.95	1.76

Table 4.6. The existence of degradation pathways for amino acids according to EcoCyc.

"Y" indicates there exists at least one degradation pathways for an amino acid, while "N" indicates there is no degradation pathway for an amino acid. Among the 20 amino acids, 8 have no degradation pathways. The 7 non-degradable amino acids we used are highlighted in red. We did not use Tyrosine due to its low solubility.

Amino acid	**Existence of degradation pathway (Y/N)**
Alanine	Y
Arginine	Y
Asparagine	Y
Aspartic acid	Y
Cysteine	Y
Glutamic acid	Y
Glutamine	Y
Glycine	Y
Histidine	N
Isoleucine	N
Leucine	N
Lysine	N
Methionine	N
Phenylalanine	N
Proline	Y
Serine	Y
Threonine	Y
Tryptophan	Y
Tyrosine	N
Valine	N

Table 4.7. The acetate excretion for cells growing in media supplemented with 7 amino acids.

E. coli NCM3722 strain was grown in minimal media with various carbon sources, supplemented with 7 non-degradable amino acids with concentration specified in the reference (Neidhardt et al., 1974): Histidine (0.2 mM), Isoleucine (0.4 mM), Leucine (0.8 mM), Lysine (0.4 mM), Methionine (0.2 mM), Phenylalanine (0.4 mM), and Valine (0.6 mM). These data are shown as purple crosses (+) in Figure 4.2.

Carbon source	**Growth rate λ (hour^{-1})**	**Acetate excretion J_{Ac} (mM/OD600/hour)**
0.2% (w/v) glucose	1.09	3.78
0.2% (w/v) lactose	1.16	4.15
0.2% (w/v) mannitol	1.04	3.06
0.2% (w/v) maltose	0.78	0.58
0.2% (w/v) glycerol	0.8	-0.01
20 mM Glucose 6-phosphate disodium	1.26	4.18

Table 4.8. The acetate excretion for cells with LacZ overexpression.

E. coli NQ1389 strain was grown on minimal media with various carbon sources and concentrations of cTc. cTc was used to induce the expression of LacZ protein. These data are shown in Figure 4.10.

Carbon	**cTc (µg/ml)**	**Growth rate λ (hour-1)**	**Acetate excretion rate J_{ac} (mM/OD600/hour)**	**LacZ activity (Miller Unit)**
Glucose-6-Phosphate	0	0.93	3.02	
Glucose-6-Phosphate	2.5	0.82	3.04	
Glucose-6-Phosphate	3.75	0.70	2.37	
Glucose-6-Phosphate	5	0.64	2.26	
Glucose-6-Phosphate	6.25	0.53	2.05	
Glucose-6-Phosphate	7.5	0.48	1.83	
Glucose-6-Phosphate	0	1.03	2.77	21320
Glucose-6-Phosphate	1.25	0.99	2.91	35161
Glucose-6-Phosphate	2.5	0.94	2.84	44172
Glucose-6-Phosphate	3.75	0.88	2.90	64339
Glucose-6-Phosphate	5	0.84	2.53	85293
Glucose-6-Phosphate	6.25	0.78	2.39	74023
Glucose-6-Phosphate	0	1.06	3.54	
Glucose-6-Phosphate	1.25	1.00	3.36	
Glucose-6-Phosphate	2.5	0.96	3.17	
Glucose-6-Phosphate	3.75	0.92	2.98	
Glucose-6-Phosphate	5	0.90	3.00	
Glucose-6-Phosphate	6.25	0.86	2.96	
Glucose	0	0.88	1.34	
Glucose	2.5	0.76	1.37	
Glucose	5	0.60	0.94	
Glucose	10	0.43	0.79	
Glucose	15	0.26	0.62	
Glucose	20	0.23	0.40	
Glucose	0	0.90	1.56	23495
Glucose	1.25	0.85	1.49	32749
Glucose	2.5	0.80	1.00	52745
Glucose	3.75	0.72	1.06	67055
Glucose	5	0.66	1.02	83115
Glucose	6.25	0.62	0.95	102929
Glucose	0	0.89	1.54	17636
Glucose	2.5	0.81	1.30	46593
Glucose	5	0.72	1.25	61979
Glucose	7.5	0.69	1.38	77435
Glucose	10	0.53	1.02	84828
Glucose	12.5	0.44	0.91	143783

Table 4.8. Continued.

Carbon	cTc (µg/ml)	Growth rate λ (hour-1)	Acetate excretion rate J_{ac} (mM/OD600/hour)	LacZ activity (Miller Unit)
Lactose	0	0.80	0.69	
Lactose	2.5	0.65	0.64	
Lactose	3.75	0.55	0.59	
Lactose	5	0.47	0.52	
Lactose	6.25	0.39	0.57	
Lactose	7.5	0.33	0.68	
Lactose	0	0.80	0.67	24133
Lactose	1.25	0.75	0.69	36266
Lactose	2.5	0.67	0.69	50908
Lactose	3.75	0.60	0.49	76551
Lactose	5	0.53	0.51	115731
Lactose	6.25	0.48	0.38	117256
Mannitol	0	0.85	0.93	
Mannitol	2.5	0.73	0.66	
Mannitol	5	0.57	0.47	
Mannitol	10	0.35	0.23	
Mannitol	15	0.21	0.08	
Mannitol	20	0.16	0.05	
Mannitol	0	0.85	0.82	19353
Mannitol	1.25	0.81	0.85	34446
Mannitol	2.5	0.76	0.74	50693
Mannitol	3.75	0.70	0.57	56365
Mannitol	5	0.64	0.59	68975
Mannitol	6.25	0.60	0.50	78773

Table 4.9. The acetate excretion for cells with flhD deletion.

E. coli NQ1388 strain was grown on glucose minimal medium with various 3MBA concentraions. 3MBA was used to titrate the expression of PtsG, thus to control the glucose uptake. These data are shown in Figure 4.11.

3MBA (μM)	Growth rate λ (h^{-1})	Acetate excretion J_{Ac} (mM/OD600/hour)
800	1.00	1.53
400	0.97	0.32
200	0.90	0.00
100	0.82	0.00
50	0.77	0.00
0	0.60	0.00
1000	1.00	2.82
900	0.97	2.29
800	0.98	2.18
700	0.96	2.62
600	0.96	1.44
500	0.94	1.31
600	1.02	2.47
500	1.01	2.31
400	0.98	0.85
300	0.93	0.22
200	0.89	-0.26
100	0.86	-0.09

Table 4.10. Flux and proteome fraction data.

Strain	3MBA (μM)	λ	J_{ac}	J_{rCO2}	$J_{ATP,f}$	$J_{ATP,r}$	J_{ATP}	ϕ_f	ϕ_r	$\phi_f + \phi_r$
NCM3722	0	0.97	3.24	3.47	19.44	15.05	34.49	2.42%	4.66%	7.08%
NQ381	800	0.91	3.06	4.90	18.36	21.22	39.58	2.36%	5.87%	8.23%
NQ381	500	0.82	1.35	6.90	8.1	29.88	37.98	1.20%	7.79%	9.00%
NQ381	400	0.77	0.44	8.49	2.64	36.78	39.42	0.54%	9.10%	9.65%

Table 4.11. The acetate excretion for cells with chloramphenicol.

E. coli NQ381 cells were grown on lactose minimal medium with various 3MBA concentrations. 3MBA was used to induce the expression of LacY. These data are shown in Figure 4.13.

Chloramphenicol (μM)	3MBA (μM)	Growth rate λ (h^{-1})	Acetate excretion J_{Ac} (mM/OD600/hour)
2	50	0.49	-0.02
2	100	0.55	-0.01
2	200	0.60	0.12
2	400	0.68	1.35
2	600	0.70	1.61
2	800	0.74	2.25
4	100	0.45	-0.17
4	200	0.50	0.70
4	400	0.58	1.58
4	600	0.61	2.05
4	800	0.62	2.18
8	100	0.27	0.00
8	200	0.31	0.76
8	400	0.36	1.44
8	600	0.39	1.60
8	800	0.40	1.86

Table 4.12. The acetate excretion for cells with leaky lacY.

E. coli NQ1313 strain was grown on lactose minimal medium with various concentrations of 3MBA. These data are shown in Figure 4.14.

Strain	3MBA (μM)	Growth rate λ (h^{-1})	Acetate excretion J_{Ac} (mM/OD600/hour)
NQ1313	0	0.40	-0.13
NQ1313	50	0.49	0.77
NQ1313	100	0.56	1.86
NQ1313	200	0.63	2.24
NQ1313	400	0.67	2.91
NQ1313	800	0.66	2.71
Control	0	0.52	-0.08
Control	50	0.66	0.01
Control	100	0.70	-0.05
Control	200	0.77	0.60
Control	400	0.87	1.49
Control	800	0.94	1.95

Table 4.13. Primers used in this chapter.

Primer	Plasmid/Construct	Use/ Digestion Sites
ptetstab-F	pZE1 P*tetstab-lacZ*	Forward amplification P*tetstab*, *XhoI*
ptetstab-R	pZE1 P*tetstab-lacZ*	Reverse amplification P*tetstab*, *BamHI*
ptet-F	pZE1 P*tetstab-lacZ*, pZA31-lacY, pZA31-*lacY Val*^{177}	P*tet* forward amplification, *XhoI*
lacY-R	pZA31-*lacY*, pZA31-*lacYVal*^{177}	Reverse amplification of *lacY*, *BamHI*
lacYfusion-F	pZA31-*lacYVal*^{177}	Forward amplification of DNA fragment for fusion PCR with lacY-R
lacYfusion-R	pZA31-*lacYVal*^{177}	Reverse amplification of DNA fragment for fusion PCR with ptet-F

4.7 Experimental procedures for Chapter 4

4.7.1 Growth of bacterial culture

All growth media used in this study were based on the MOPS-buffered minimal medium used by Cayley et al. (Cayley et al., 1989) with slight modifications. See Chapter 2 for the composition of the base medium. The concentrations of carbon sources are indicated in the relevant tables. Also see Chapter 2 for the protocol for growth measurements in water bath shaker.

4.7.2 Growth in bioreactor and CO2 measurement

To measure CO2 production from the bacterial growth, cells were grown in a Multifors bioreactor (Infors HT, Switzerland). Around 400 ml of medium was used in a 750 ml vessel, which has an inlet for compressed air and out outlet for the exhaust gas. The vessel is otherwise closed except during brief period of sample collection. Samples of the cell culture (for reading OD600, assaying lactose and acetate, etc) can be taken by using a syringe connected to the vessel.

The air flow rate to the inlet was controlled by a mass flow controller (Cole-Parmer, Model #: 32907-67) and maintained at 400 ml/min. The outlet was connected to a BlueInOne Cell sensor unit (BlueSens, Germany) for measuring CO2 concentration. The stir rate in the growth vessel was set as 800 rpm and temperature was maintained at 37^{o}C.

4.7.3 Strain construction

Construction of Pu-ptsG strain (NQ1243): The Δ*ptsG468*::Φ(*kan:*P*u*) allele in which *ptsG* promoter is replaced by P*u* promoter was made as follows. The region containing the *kan* gene and P*u* promoter was PCR amplified by primers SDY158 and SDY159 from NQ381 and integrated at *ptsG* locus, resulting in the replacement of 342 base before the open reading frame with a tandem array of *kan* gene and P*u* promoter by using the λ Red system (Datsenko and Wanner, 2000), which results in the replacement of *ptsG* promoter with P*u* promoter.

Construction of lacZ overexpression strain (NQ1389): The *lacZ* structural gene region was amplified from MG1655 chromosomal DNA with upstream and downstream primers including the digestion sites *XhoI* and *BamHI* respectively. The PCR products were gel purified, digested with *XhoI* and *BamHI*, then inserted into the same sites immediately downstream of P*tet* in the pZE11 plasmid (Lutz and Bujard, 1997), yielding pZE11-*lacZ*. To improve the stability of P*tet* with respect to homologous recombination, we later replaced the promoter sequence with the following modified Ptet promoter sequence (the underlined bases are changed as compared to the original Ptet):

CTCGAGTCCCTATCAGTGATAGCTCTTGACAGATCTATCAATGATAGAGATAC
TGAGCACATATGCAGCAGGACGCACTGACCGAATTCATTAAAGAGGAGAAA
GGTACC

To do this, we first synthesized a single DNA fragment (CTCTTGACAGATCTATCAA TGATAGAGATACTGAGCACATATGCAGCAGGACGCACTGAC) that served as template for PCR amplification of the modified stable P*tet* (P*tetstab*) using primers ptetstab-F and ptetstab-R. The products were purified, digested with *XhoI* and *KpnI* and subsitituted for P*tet* in pZE11-*lacZ*. This yielded pZE1 P*tetstab-lacZ*. We then transformed this plasmid into NCM3722 in combination with the auto-regulated TetR plasmid pZA31 P*tet-tetR* (Klumpp et al., 2009), creating a stable, titratable system capable of high levels of lacZ expression. The primers are listed in Table 4.13.

Construction of lacY/leaky-lacY strains (NQ1313, NQ1314): The P*tet-lacY* region was amplified from the pZE12 P*tet-lacY* plasmid (Kuhlman et al., 2007) with upstream and downstream primers including the digestion sites *XhoI* and *BamHI* respectively using the primers ptet-F and lacY-R. The resulting DNA fragment was substituted for P*tet-gfp* in pZA31-*gfp* (Levine et al., 2007), yielding pZA31-*lacY*. The same procedure was employed to generate the *leak- lacY* mutant, but fusion PCR was used to introduce a point mutation Val^{177} into the *lacY* sequence (Brooker, 1991). For this, two overlapping parts of the P*tet-lacY* region were PCR amplified with the primers ptet-F, lacYfusion-R and lacYfusion-F, lacY-R, where the point substitution leading to the Val^{177} mutation from was included in the primers lacYfusion-F and lacYfusion-R. These two overlapping DNA fragments were fused together by PCR using primers ptet-F and lacY-R. The resulting

P*tet-lacY* fragment that carries the desired mutation was inserted into pZA31, yielding pZA31-*lacYVal*^{177}. The resulting plasmids were transformed into NQ1243. The primers are listed in Table 4.13.

Construction of flhD deletion strain (NQ1388): The *flhD* deletion allele in strain JW1881-1 (*E. coli* Genetic Stock Center, Yale Univ.), in which a *kan* gene is substituted for the *flhD* gene, was transferred to NQ1243 (deleted of *kan*) by phage P1 vir mediated transduction.

4.7.4 Lactose assay, acetate assay, and β-Galactosidase assay with plate reader

Lactose assay and acetate assay: 500 μL samples were taken at 3 or 4 different times during exponential growth (typically at OD600 between 0.1 and 0.5) by centrifugation at maximum (15000 rpm) for 1 min. The supernatant was transferred on to ice.

To assay lactose, ~10 μL of the supernatant was first digested by β-galactosidase (Sigma-Aldrich) in Z-buffer at 37°C for 20 min. The released glucose was then assayed enzymatically by the kit commercially available (Glucose Assay Kit, GAHK20; Sigma-Aldrich). As a control, the sample was treated in the same way without β-galactosidase. Little glucose was detected in the control.

About 100 μL of supernatant was used to measure acetate concentration using the Acetate Assay kit (Catalog #: 10148261035, R-Biopharm).

β-Galactosidase Assay with plate reader: The assay was performed following a similar protocol as detailed in a previous study (Scott et al., 2010). 100 μL samples were taken for at least 4 different times during exponential growth (typically at OD600

between 0.1 and 0.6). LacZ samples were immediately added to 0.9 mL of freshly prepared Z-buffer (in 1 L: 8.52g Na2HPO4, 5.5g NaH2PO4 H2O, 0.75g KCl and 0.25g MgSO4 7H2O, with 0.004%(w/v) SDS and 40mM β-mercaptoethanol) with 100 µL chloroform. Cells were disrupted by vortexing and stored at 4 ^{o}C until all samples were collected. After all samples were collected, they were briefly vortexed a second time. After 5-10 minutes at room temperature to settle the chloroform, the lysates were diluted (typically 1:4) in Z-buffer. 200 µL was then added to a 96-well plate (Sarstedt). Immediately prior to reading in GENiosPro (Tecan) plate reader, 40 µL of of 4 mg/mL ortho-Nitrophenyl-β-galactoside (Sigma) in 0.1 M phosphate buffer (pH=7.0) was added to each well. The plate reader was set to read absorbance at a wavelength of 420 nm every minute for 60 to 120 minutes at 28 ^{o}C.

The slope of the plot of OD420 vs. time (in minutes) for all replicates was used to calculate the β-galactosidase activity (Units/mL) in the original sample via the following conversion:

$$\beta\text{-galactosidase activity (U/mL)} = 1000 \cdot \left[\frac{OD_{420}}{\min} \right] \cdot (\text{fold dilution}) \cdot 2.66.$$

The factor 2.66 converts the plate-reader data to the activity obtained using the original assay protocol by Miller (Miller, 1972) and is specific to the path-length through the sample (i.e. 240 µL in a 96-well plate). The slope of the plot of activity (U/mL) versus the sample OD600 yields the activity in Miller units (U/mL/ OD600). The enzyme activity was expressed in µg of β-galactosidase (1082 U/mL/µg of β-galactosidase) calibrated with pure enzyme (Sigma), and normalized to total protein content.

4.7.5 Proteomic mass spectrometry

See the Experimental procedure section in Chapter 2.

4.8 Acknowledgement

With permission from the coauthors, chapters 4-5 of this dissertation contain work from a manuscript that has been submitted for publication with the working title, “Coordination of carbon supply and proteomic resources for energy metabolism results in acetate overflow” 2014, Basan, M; Hui, S; Shen, Y; Zhang, Z; Williamson, J; Hwa, T. The dissertation author and Markus Basan are the co-first authors of the work.

Chapter 5

Measurements and analysis of metabolic fluxes and protein mass

5.1 Introduction

In this chapter, I describe the experimental procedures and analysis for determining the metabolic fluxes and protein mass of the two energy pathways, which are necessary for testing the model in Chapter 4.

5.2 Flux measurements and analysis

5.2.1 Flux balance

The carbon flux through respiration ($J_{C,r}$) can be deduced using carbon balance, as shown in Eq. [4.4]: The toal carbon influx ($J_{C,in}$) is equal to the sum of three carbon fluxes, the carbon flux through fermentation ($J_{C,f}$), the carbon flux going to biomass production ($J_{C,bm}$), and $J_{C,r}$. (For convenience, we refer 'carbon flux' as 'single carbon flux'. For example, for one unit of glucose influx, $J_{C,f}$ is equal to 6 because one glucose molecule contains 6 carbon atoms.) $J_{C,f}$ is given by the acetate flux (J_{ac}) as $J_{C,f} = 3 \cdot J_{ac}$ (Keseler et al., 2012). Using Eq. [4.11], $J_{C,bm}$ is readily obtained by just measuring growth rate. And finally since $J_{C,in}$ can be directly measured, $J_{C,r}$ can be calculated according to Eq. [4.4]. In practice, however, the resulting $J_{C,r}$ would have large

uncertainty because $J_{C,bm}$ is much larger than $J_{C,r}$. A small noise for $J_{C,bm}$ would translate into a huge noise for $J_{C,r}$. To avoid this problem, we turn to the flux balance equation for CO2,

$$J_{CO2,f} + J_{CO2,r} + J_{CO2,bm} = J_{CO2}, \quad [5.1]$$

where $J_{CO2,f}$, $J_{CO2,r}$, and $J_{CO2,bm}$ refer to the CO2 fluxes produced by the fermentation, respiration, and biomass production, respectively, and J_{CO2} is the total CO2 production rate by the cell. Note that in the notation J_{rCO2} in Chapter 4 is identical to $J_{CO2,r}$ here. As we will show below, the CO2 fluxes by the three pathways have similar magnitude and therefore noises in measurements or estimates of $J_{CO2,f}$, $J_{CO2,bm}$, and J_{CO2} do not result in magnified noise in the estimate of $J_{C,r}$.

5.2.2 Measurement of total CO2 production rate

To measure the CO2 production rate, we have a bioreactor setup. The system has an air-inlet and an outlet for exhaust gas. A constant rate of the air-flow (f) to the inlet is maintained by a mass flow controller. The exhaust gas line is connected to a CO2 sensor and the concentration of CO2 at time t, $c(t)$, in the exhaust gas is constantly measured. The OD600 of the bacterial culture (with volume V) is measured between every half-doubling time and one doubling time and the OD600 at the time t, $OD(t)$, can be deduced from the growth curve. At steady state, the CO2 production rate is equal to the CO2 flow, or

$$OD(t) \cdot V \cdot J_{CO2} = k \cdot f \cdot (c(t) - c_0), \quad [5.2]$$

where $k = 39.3$ μmol/ml is a constant converting gas volume and amount of gas under the condition of 1 atm and 37°C. Re-arranging Eq. [5.2] gives

$$c(t) = J_{CO2} \cdot \frac{V}{k \cdot f} \cdot OD(t) + c_0 . \quad [5.3]$$

The slope (s) of the $c(t)$ versus $OD(t)$ plot can be obtained by fitting the data with a line (Figure 5.1). Subsequently, J_{CO2} is given by

$$J_{CO2} = \frac{k \cdot f}{V} \cdot s . \quad [5.4]$$

Using a culture of NCM3722 grown on lactose minimal medium, where f = 400 ml/min, $V = 410$ ml, and $s = 0.0046$ OD^{-1} (Figure 5.1), we have $J_{CO2} = 10.57$ mM/OD/hr.

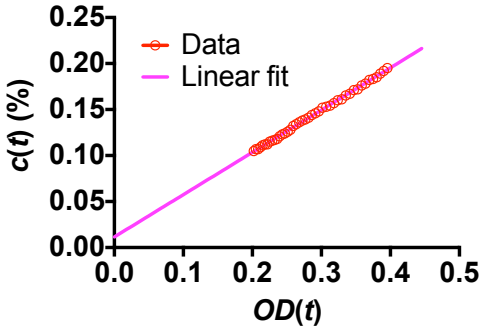


Figure 5.1. CO2 concentration versus OD600.

The concentration (%) of CO2 in the exhaust gas is plotted a function of the OD600 of the cell culture.

5.2.3 CO2 production associated with biomass production

We assume that

$$J_{CO2,bm} = \lambda \cdot \beta_{CO2} , \quad [5.5]$$

where β_{CO2} (in unit of mmol/g) represents the amount of CO2 produced for making 1 g of cell dry weight.

In cases where we can deduce $J_{CO2,f}$ and $J_{CO2,r}$ from measured flux distribution, we can use Eq. [5.1] to get $J_{CO2,bm}$ if J_{CO2} has also been measured. With given growth rate

λ, we can obtain β_{CO2} from Eq. [4.5].

In the reference(Fischer and Sauer, 2003), we get $J_{CO2,f} = 7.2$ mmol/g/hr, $J_{CO2,r} = 6.3$ mmol/g/hr, and $J_{CO2} = 19.6$ mmol/g/hr. We thus have $J_{CO2,bm} = 6.1$ mmol/g/hr (based on Eq. [5.1]), which leads to $\beta_{CO2} = 7.18$ mmol/g given that $\lambda = 0.85$ /hr (based on Eq. [5.5]).

Another way to estimate β_{CO2} is to start from biomass composition and deduce from biochemical stoichiometry the amount of CO2 involved for making each of the biomass constituent.

Using the latest *E. coli* reconstruction model iAF1260 (Feist et al., 2007), we ran a Flux Balance Analysis (FBA) to maximize biomass production rate under the constraint of limited glucose uptake rate. From the resulting CO2 production flux, we obtained $\beta_{CO2} = 6.32$ mmol/g. This number is the theoretical minimum for a given biomass composition. To see how sensitive this number depends on biomass composition, we next vary the composition to see how β_{CO2} changes accordingly.

In the original iAF1260 model, protein and RNA account for about 56% and 21% of the dry weight, respectively, while the rest includes about 9% lipid, 2.5% murein, 3% DNA, 2.5% glycogen, 3% outer membrane materials, and 3% soluble metabolites. Assuming this rest 23% is constant, we vary the mass fraction of protein from 40% to 70% (meaning that RNA mass fraction is varied from 37% to 7%). See Figure 5.2 for the corresponding values of β_{CO2}.

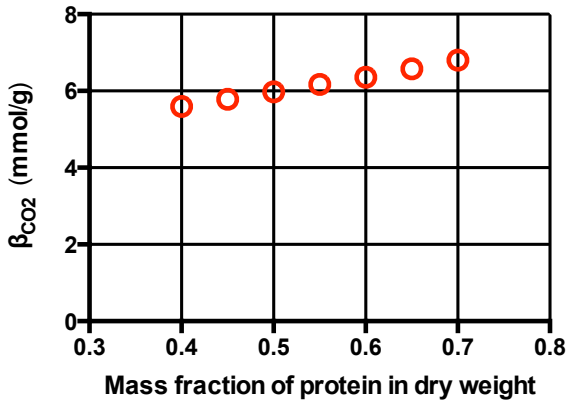


Figure 5.2. β_{CO2} as a function of mass fraction of protein.

We also vary the total mass fraction of protein and RNA, from 70% to 90%, assuming the relative composition of the rest mass fraction is fixed. See Figure 5.3 for values of β_{CO2}.

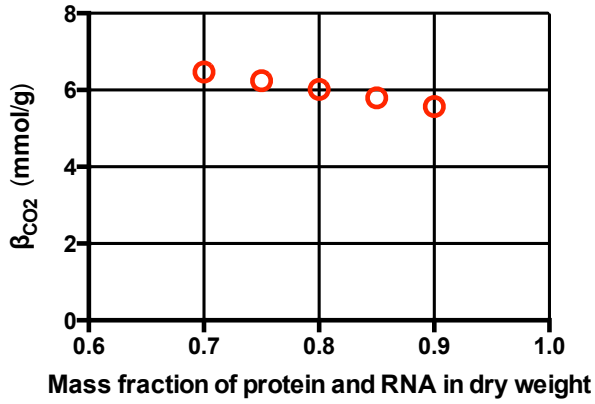


Figure 5.3. β_{CO2} as a function of mass fraction of protein and RNA.

The results above show that β_{CO2} does not have strong dependence on biomass composition, with ~10% deviation from the value of $\beta_{CO2} = 6.32$ even for the extreme biomass compositions.

Note that the value of $\beta_{CO2} = 6.32$ is a theoretical minimum. We thus take the above estimate from the reference (Fischer and Sauer, 2003), i.e., $\beta_{CO2} = 7.18$ with 10% of error.

5.2.4 CO2 flux by respiration

Since $J_{CO2,f} = J_{ac}$, we now have

$$J_{CO2,r} = J_{CO2} - J_{ac} - \beta_{CO2} \cdot \lambda . \quad [5.6]$$

For cultures with different lactose uptake rates (using WT NCM3722 and the strain NQ381), we have measured total CO2 production rate (J_{CO2}), acetate excretion rate (J_{ac}), and growth rate (λ). The results for the CO2 flux by respiration $J_{CO2,r}$ (or equivalently, J_{rCO2}) deduced based on Eq. [5.6] are shown in Figure 4.12.

5.2.5 Energy fluxes

Once one type of metabolic flux is known for a pathway, it is simple biochemical stoichiometry that determines the other fluxes through the pathway. In table below we list the stoichiometry among the various fluxes through the two energy pathways.

Table 5.1. The stoichiometry for various metabolic fluxes for the two energy pathways.

$J_{ATP,sub}$ refers to the ATP flux produced by the substrate phosphorylation, while J_{ATP} refers to the total ATP flux assuming some converting factors between the redox molecules and the ATP molecule, i.e., NADH = 2 ATP, NADPH = 2 ATP, and FADH2 = 1 ATP (Unden and Dunnwald, 2008).

	J_{6C}	J_C	J_{ac}	J_{CO2}	J_{NADH}	J_{NADPH}	J_{FADH2}	$J_{ATP,sub}$	J_{ATP}
Fermentation	1	6	2	2	4	0	0	4	12
Respiration	1	6	0	6	8	2	2	4	26

It is clear from Table 5.1 that we have the carbon efficiencies as $e_f = \frac{J_{ATP,f}}{J_{C,f}} = 2$ and $e_r = \frac{J_{ATP,r}}{J_{C,r}} = 13/3$.

5.3 Estimate of protein mass of the energy pathways

As demonstrated in Chapter 4, the proteome efficiencies of the energy pathways (i.e., ϕ_f, ϕ_r) are crucial parameters in our understanding of the acetate overflow problem. To experimentally determine their values, we need to estimate the protein mass for each of the energy pathways. In this chapter, I describe our approach for achieving this goal.

In Chapter 2, we have shown that spectral counting data can be used for estimating absolute protein abundance. This method, however, is limited for obtaining coarse-grained abundance for a large group of proteins whose summed proteome fraction is above ~10% (Figure 2.11). As we will show later in this chapter, the proteome fraction of either of the two energy pathways is well below this threshold. Thus, we need to develop a new method of absolute protein quantitation.

5.3.1 Absolute protein quantitation using mass spectrometry

As demonstrated in Chapter 2, the relative protein quantitation using mass spectrometry is accurate and precise (Figure 2.4, Figure 2.5). To infer the absolute protein abundance from relative protein abundance, we need a standard for which the absolute protein abundance is known. This usually means a practically daunting task of purifying all the proteins of interest. Recent work by Li et al. (Li et al., 2014), however, shows that ribosome profiling can reliably quantify absolute protein abundance. The

work has quantified a WT *E. coli* MG1655 strain (denoted as EQ383 by us) grown on three different conditions. Below in Table 5.2 lists the absolute protein abundance for proteins in the two energy pathways for the strain grown on the glucose minimal medium (Li et al., 2014).

Table 5.2. Absolute protein abundance for the proteins in the two energy pathways, as determined by the reference (Li et al., 2014) for EQ383 grown on glucose minimal medium.

Note	Gene	Protein abundance
Lactose degradation	lacY	0.000%
	lacZ	0.001%
	glk	0.013%
	pgm	0.051%
	galM	0.027%
	galK	0.007%
	galT	0.004%
	galE	0.019%
Glycolysis	pgi	0.149%
	pfkA	0.069%
	pfkB	0.029%
	fbaA	0.382%
	fbaB	0.073%
	tpiA	0.148%
	gapA	1.467%
	pgk	0.351%
	gpmA	0.220%
	gpmI	0.111%
	eno	0.678%
	pykF	0.219%
	pykA	0.058%
	aceE	0.407%
	aceF	0.241%
	lpdA	0.374%
	pta	0.066%
	ackA	0.036%
TCA	gltA	0.286%
	acnA	0.059%
	acnB	0.503%
	icd	1.244%
	sucA	0.156%
	sucB	0.175%
	lpdA	0.374%
	sucC	0.185%

Table 5.2. Continued.

Note	Gene	Protein abundance
TCA	sucD	0.133%
	sdhA	0.154%
	sdhB	0.075%
	sdhC	0.036%
	sdhD	0.023%
	fumA	0.083%
	fumB	0.001%
	fumC	0.015%
	mdh	0.365%
	ppc	0.308%
NADH:ubiquinone oxidoreductase I (NDH-1)	nuoA	0.012%
	nuoB	0.028%
	nuoC	0.053%
	nuoE	0.021%
	nuoF	0.036%
	nuoG	0.082%
	nuoH	0.016%
	nuoI	0.013%
	nuoJ	0.007%
	nuoK	0.005%
	nuoL	0.024%
	nuoM	0.023%
	nuoN	0.025%
NADH:ubiquinone oxidoreductase II (NDH-2)	ndh	0.029%
cytochrome *bo* oxidase (CyoABCD)	cyoA	0.083%
	cyoB	0.113%
	cyoC	0.043%
	cyoD	0.037%
cytochrome *bd*-I oxidase (CydABX)	cydA	0.025%
	cydB	0.013%
	cydX	0.004%
cytochrome *bd*-II oxidase (AppCB)	appC	0.001%
	appB	0.000%

Table 5.2. Continued.

Note	Gene	Protein abundance
F_1 complex of ATP synthase	atpA	0.369%
	atpC	0.038%
	atpD	0.289%
	atpG	0.069%
	atpH	0.036%
F_0 complex of ATP synthase	atpB	0.074%
	atpE	0.136%
	atpF	0.054%

Thanks to Gene-Wei Li for kindly providing the strain. We grew the strain EQ383 on the same glucose minimal medium as used in Li et al. (Li et al., 2014). The collected sample from the culture was spiked with the same ^{15}N reference sample as our samples of various lactose uptake rates. Table 5.3 lists the relative protein levels as determined by mass spectrometry. The absolute protein abundance in our samples can then be inferred from the relative protein abundance data.

Table 5.3. Relative protein abundance data as determined by mass spectrometry.

To have a series with varying lactose uptake rates, WT NCM3722 and NQ381 with three different concentrations of 3MBA were used. The numbers inside the parenthesis in the header row of the table indicate the concentrations of 3MBA. To infer absolute protein abundance, EQ383 was also included as one of the mass spectrometry samples.

Gene	EQ383	NQ381 (400)	NQ381 (500)	NQ381 (800)	NCM3722
lacZ		1.05	1.09	1.10	0.96
glk	0.67	0.87	1.08	0.92	0.87
galM	0.31	0.89	0.84	0.87	0.88
galK		0.96	1.08	1.09	0.92
galE		0.73	0.75	0.99	1.11
pgi	1.63	0.95	0.97	1.08	1.16
pfkA	1.13	1.00	1.02	1.02	1.03
fbaA	1.31	1.08	1.19	1.20	1.34
tpiA	1.08	0.88	1.13	0.98	1.34
gapA	1.27	1.03	1.12	1.15	1.28
pgk	1.16	0.99	1.02	1.05	1.19
gpmA	1.52	1.07	1.05	1.06	1.12
gpmI	1.44	1.04	1.08	1.26	1.42
eno	1.10	1.03	1.14	1.21	1.35
pykF	1.42	0.96	1.00	1.17	1.37
pykA		1.17	1.09	0.97	0.97
aceE	1.21	0.91	0.96	1.01	1.22
aceF	1.12	0.97	1.11	1.13	1.15
lpdA	0.68	1.16	1.12	1.02	0.96
pta	0.73	1.08	1.11	1.09	1.06
gltA	0.45	1.39	1.26	0.96	0.75
acnB	0.69	1.26	1.16	0.91	0.78
icd	0.89	1.11	1.11	0.93	0.99
sucA	0.50	1.07	1.17	0.94	0.82
sucB	0.50	1.08	1.11	1.00	0.86
lpdA	0.68	1.16	1.12	1.02	0.96
sucC	0.43	1.23	1.16	0.89	0.71
sucD	0.45	1.23	1.19	0.96	0.73
sdhA	0.54	1.15	1.05	0.99	0.83
sdhB	0.52	1.10	1.06	0.92	0.78
fumA	0.45	1.28	1.11	0.90	0.70

Table 5.3. Continued.

Gene	EQ383	NQ381 (400)	NQ381 (500)	NQ381 (800)	NCM3722
mdh	1.05	1.29	1.29	1.18	1.12
ppc	1.83	0.93	1.07	1.07	1.14
nuoC	0.76	1.10	1.07	0.93	0.81
atpA	0.81	1.12	1.11	1.04	0.99
atpD	0.86	1.18	1.12	1.10	0.97
atpG	0.83	1.21	1.11	0.96	0.85
atpH	0.89	1.01	1.15	1.21	0.98
atpF	1.02	1.11	1.11	1.12	1.12

Only about half of the proteins have reliable data. This limited protein coverage, however, does not prevent us from obtaining useful estimate. For the 34 proteins that are present in the mass spectrometry data for EQ383, the total protein abundance is 9.46%, which represents 88% of the total protein abundance for the 74 proteins listed in the table. Among the remaining 40 proteins, LacZ could have significant protein mass, as reflected by 0.8-0.9% of proteome fraction by spectral counting data. To obtain accurate estimate for this abundant protein, we have assayed the b-galactosidase activity of one of the conditions (i.e., NCM3722 cells growing on lactose minimal medium). With a Miller Unit of 4700 and a conversion factor of 1.7% proteome fraction for 10,000 MU (You et al. 2013), we got 0.78% of proteome fraction for LacZ in this condition, which leads to proteome fraction of LacZ for other conditions given the relative protein abundance in the table. For the remaining 39 proteins, we used spectral counting data. The results are summarized in Table 5.4.

Table 5.4. Absolute protein abundance for the proteins in the two energy pathways.

Gene	NQ381 (400)	NQ381 (500)	NQ381 (800)	NCM3722
lacY	0.000%	0.000%	0.000%	0.000%
lacZ	0.851%	0.885%	0.893%	0.780%
glk	0.017%	0.021%	0.018%	0.017%
pgm	0.030%	0.000%	0.000%	0.000%
galM	0.077%	0.072%	0.075%	0.076%
galK	0.179%	0.195%	0.224%	0.128%
galT	0.030%	0.033%	0.032%	0.064%
galE	0.149%	0.163%	0.289%	0.257%
pgi	0.087%	0.089%	0.099%	0.106%
pfkA	0.061%	0.062%	0.062%	0.063%
pfkB	0.000%	0.000%	0.000%	0.000%
fbaA	0.316%	0.348%	0.351%	0.390%
fbaB	0.000%	0.000%	0.000%	0.000%
tpiA	0.120%	0.155%	0.134%	0.183%
gapA	1.191%	1.292%	1.328%	1.474%
pgk	0.300%	0.308%	0.318%	0.361%
gpmA	0.154%	0.153%	0.153%	0.162%
gpmI	0.080%	0.083%	0.097%	0.109%
eno	0.634%	0.702%	0.747%	0.832%
pykF	0.148%	0.155%	0.181%	0.212%
pykA	0.179%	0.130%	0.064%	0.064%
aceE	0.305%	0.322%	0.340%	0.410%
aceF	0.210%	0.238%	0.243%	0.247%
lpdA	0.635%	0.614%	0.561%	0.524%
pta	0.097%	0.100%	0.098%	0.095%
ackA	0.060%	0.065%	0.064%	0.000%
gltA	0.885%	0.805%	0.615%	0.477%
acnA	0.030%	0.000%	0.000%	0.000%
acnB	0.919%	0.844%	0.662%	0.567%
icd	1.553%	1.555%	1.309%	1.393%
sucA	0.331%	0.362%	0.290%	0.254%
sucB	0.378%	0.389%	0.353%	0.301%
lpdA	0.635%	0.614%	0.561%	0.524%
sucC	0.523%	0.495%	0.376%	0.301%
sucD	0.364%	0.352%	0.284%	0.217%

Table 5.4. Continued.

Gene	NQ381 (400)	NQ381 (500)	NQ381 (800)	NCM3722
sdhA	0.331%	0.302%	0.285%	0.238%
sdhB	0.160%	0.154%	0.133%	0.114%
sdhC	0.000%	0.000%	0.000%	0.000%
sdhD	0.000%	0.000%	0.000%	0.000%
fumA	0.236%	0.205%	0.165%	0.128%
fumB	0.000%	0.000%	0.000%	0.000%
fumC	0.000%	0.000%	0.000%	0.000%
mdh	0.450%	0.449%	0.410%	0.390%
ppc	0.156%	0.181%	0.180%	0.192%
nuoA	0.000%	0.000%	0.000%	0.000%
nuoB	0.000%	0.000%	0.000%	0.000%
nuoC	0.077%	0.074%	0.065%	0.057%
nuoE	0.000%	0.000%	0.000%	0.000%
nuoF	0.060%	0.065%	0.032%	0.000%
nuoG	0.060%	0.130%	0.032%	0.000%
nuoH	0.000%	0.000%	0.000%	0.000%
nuoI	0.000%	0.033%	0.000%	0.000%
nuoJ	0.000%	0.000%	0.000%	0.000%
nuoK	0.000%	0.000%	0.000%	0.000%
nuoL	0.000%	0.000%	0.000%	0.000%
nuoM	0.000%	0.000%	0.000%	0.000%
nuoN	0.000%	0.000%	0.000%	0.000%
ndh	0.000%	0.000%	0.032%	0.000%
cyoA	0.000%	0.033%	0.032%	0.032%
cyoB	0.000%	0.000%	0.000%	0.000%
cyoC	0.000%	0.000%	0.000%	0.000%
cyoD	0.000%	0.000%	0.000%	0.000%
cydA	0.000%	0.000%	0.000%	0.000%
cydB	0.000%	0.000%	0.000%	0.000%
cydX	0.000%	0.000%	0.000%	0.000%
appC	0.000%	0.000%	0.000%	0.000%
appB	0.000%	0.000%	0.000%	0.000%
atpA	0.514%	0.508%	0.474%	0.451%
atpC	0.000%	0.000%	0.000%	0.000%
atpD	0.395%	0.378%	0.371%	0.326%
atpG	0.101%	0.093%	0.080%	0.071%

Table 5.4. Continued.

Gene	NQ381 (400)	NQ381 (500)	NQ381 (800)	NCM3722
atpH	0.041%	0.047%	0.049%	0.040%
atpB	0.000%	0.000%	0.000%	0.000%
atpE	0.030%	0.000%	0.032%	0.000%
atpF	0.059%	0.059%	0.059%	0.059%

5.3.2 Partitioning mass of individual enzymes based on fluxes

Most of the proteins in Table 5.2 are shared among the two energy pathways and the biomass production pathway. To determine the protein mass for each of the energy pathways, it is necessary to partition the share proteins according to the fluxes they carry. For example, for a protein p whose mass is ϕ_p, if the carbon fluxes it carries are the fermentation flux ($J_{p,f}$), respiration flux ($J_{p,r}$), and biomass flux ($J_{p,bm}$), then protein p contributes

$$\phi_{p,f} = \phi_p \cdot \frac{J_{p,f}}{J_p} \quad [5.7]$$

to the fermentation pathway, and

$$\phi_{p,r} = \phi_p \cdot \frac{J_{p,r}}{J_p} \quad [5.8]$$

to the respiration pathway, where $J_p = J_{p,f} + J_{p,r} + J_{p,bm}$.

The total proteome fractions for the fermentation and respiration pathways are then simply

$$\phi_f = \sum_p \phi_{p,f} \quad [5.9]$$

and

$$\phi_r = \sum_p \phi_{p,r}, \quad [5.10]$$

respectively.

Within each of the energy pathways or the biomass pathway, the stoichiometry between the fluxes of different proteins is fixed. The stoichiometry is shown in Table 5.5 for the three pathways. With the stoichiometry $S_{p,i}$, for protein p and pathway i, given that the carbon flux feeding to the pathway is $J_{C,i}$, the flux for the protein is simply

$$J_{p,i} = S_{p,i} \cdot J_{C,i}. \qquad [5.11]$$

Eqs. [5.7-5.11], together with Table 5.4 (for providing values for ϕ_p) and Table 5.5 (for providing values for $S_{C,i}$) enable the estimates for the protein mass (i.e., ϕ_f, ϕ_r) of the two energy pathways. The results are shown in Figure 4.15.

Table 5.5. Stoichiometry for the two energy pathways and the biomass production pathway. The stoichiometry for the biomass pathway is based on (Neidhardt et al., 1990).

Gene	Biomass	Fermentation	Respiration
lacY	1	1	1
lacZ	1	1	1
glk	1	1	1
pgm	1	1	1
galM	1	1	1
galK	1	1	1
galT	1	1	1
galE	1	1	1
pgi	1.628	2	2
pfkA	1.61	2	2
pfkB	1.61	2	2
fbaA	1.61	2	2
fbaB	1.61	2	2
tpiA	1.61	2	2
gapA	3.188	4	4
pgk	3.188	4	4
gpmA	2.808	4	4
gpmI	2.808	4	4
eno	2.808	4	4
pykF	1.948	4	4
pykA	1.948	4	4
aceE	1.228	4	4
aceF	1.228	4	4
lpdA	1.228	4	4
pta	0	4	0
ackA	0	4	0
gltA	0.274	0	4
acnA	0.274	0	4
acnB	0.274	0	4
icd	0.274	0	4
sucA	0	0	4
sucB	0	0	4
lpdA	0	0	4
sucC	0	0	4
sucD	0	0	4

Table 5.5. Continued.

Gene	Biomass	Fermentation	Respiration
sdhA	0	0	4
sdhB	0	0	4
sdhC	0	0	4
sdhD	0	0	4
fumA	0	0	4
fumB	0	0	4
fumC	0	0	4
mdh	0	0	4
ppc	0.728	0	0
nuoA	1.678	8	22
nuoB	1.678	8	22
nuoC	1.678	8	22
nuoE	1.678	8	22
nuoF	1.678	8	22
nuoG	1.678	8	22
nuoH	1.678	8	22
nuoI	1.678	8	22
nuoJ	1.678	8	22
nuoK	1.678	8	22
nuoL	1.678	8	22
nuoM	1.678	8	22
nuoN	1.678	8	22
ndh	1.678	8	22
cyoA	1.678	8	22
cyoB	1.678	8	22
cyoC	1.678	8	22
cyoD	1.678	8	22
cydA	1.678	8	22
cydB	1.678	8	22
cydX	1.678	8	22
appC	1.678	8	22
appB	1.678	8	22
atpA	1.678	8	22
atpC	1.678	8	22
atpD	1.678	8	22
atpG	1.678	8	22
atpH	1.678	8	22

Table 5.5. Continued.

Gene	Biomass	Fermentation	Respiration
atpB	1.678	8	22
atpE	1.678	8	22
atpF	1.678	8	22

5.4 Acknowledgement

With permission from the coauthors, Chapters 4-5 of this dissertation contain work from a manuscript that has been submitted for publication with the working title, "Coordination of carbon supply and proteomic resources for energy metabolism results in acetate overflow" 2014, Basan, M; Hui, S; Shen, Y; Zhang, Z; Williamson, J; Hwa, T. The dissertation author and Markus Basan are the co-first authors of the work.

References

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198–207.

Applebee, M.K., Joyce, A.R., Conrad, T.M., Pettigrew, D.W., and Palsson, B.O. (2011). Functional and Metabolic Effects of Adaptive Glycerol Kinase (GLPK) Mutants in Escherichia coli. Journal of Biological Chemistry *286*, 23150–23159.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. Analytical and Bioanalytical Chemistry *389*, 1017–1031.

Bennett, P., and Maaloe, O. (1974). The effects of fusidic acid on growth, ribosome synthesis and RNA metabolism in Escherichia coli. J Mol Biol *90*, 541.

Benthin, S., Nielsen, J., and Villadsen, J. (1991). A simple and reliable method for the determination of cellular RNA content. Biotechnology Techniques *5*, 39–42.

Beynon, R.J., Doherty, M.K., Pratt, J.M., and Gaskell, S.J. (2005). Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat Meth *2*, 587–589.

Brauer, M., Huttenhower, C., Airoldi, E., Rosenstein, R., Matese, J., Gresham, D., Boer, V., Troyanskaya, O., and Botstein, D. (2008). Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast. Molecular Biology of the Cell *19*, 352.

Bremer, H., and Dennis, P.P. (2009). 5.2.3 Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal 1–95.

Brooker, R. (1991). An analysis of lactose permease“ sugar specificity” mutations which also affect the coupling between proton and lactose transport. I. Val177 and Val177/Asn319 permeases facilitate proton uniport and sugar uniport. Journal of Biological Chemistry *266*, 4131.

Cayley, S., Record, M.T., and Lewis, B.A. (1989). Accumulation of 3-(N-morpholino)propanesulfonate by osmotically stressed Escherichia coli K-12. Journal of Bacteriology *171*, 3597–3602.

Chen, S.S., Sperling, E., Silverman, J.M., Davis, J.H., and Williamson, J.R. (2012). Measuring the dynamics of E. coli ribosome biogenesis using pulse-labeling and quantitative mass spectrometry. Molecular BioSystems *8*, 3325–3334.

Cherepanov, P.P., and Wackernagel, W. (1995). Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9–14.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences *97*, 6640–6645.

De Deken, R.H. (1966). The Crabtree effect: a regulatory system in yeast. Journal of General Microbiology *44*, 149–156.

de Lorenzo, V., Herrero, M., Metzke, M., and Timmis, K.N. (1991). An upstream XylR- and IHF-induced nucleoprotein complex regulates the sigma 54-dependent Pu promoter of TOL plasmid. Embo J *10*, 1159–1167.

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

el-Mansi, E.M., and Holms, W.H. (1989). Control of carbon flux to acetate excretion during growth of Escherichia coli in batch and continuous cultures. Journal of General Microbiology *135*, 2875–2883.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., and Palsson, B.Ø. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol *3*, 121.

Fischer, E., and Sauer, U. (2003). A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry Escherichia coli. Journal of Biological Chemistry *278*, 46446.

Han, X., Aslanian, A., and Yates, J.R., III (2008). Mass spectrometry for proteomics. Current Opinion in Chemical Biology *12*, 483–490.

Herbert, D., Phipps, P.J., and Strange, R.E. (1971). Chemical analysis of microbial cells. In Methods in Microbiology, J.R. Norris, and D.V. Ribbons, eds. (Elsevier), pp. 209–342.

Holms, H. (1996). Flux analysis and control of the central metabolic pathways in Escherichia coli. FEMS Microbiol Rev *19*, 85–116.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and 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, F.U., Kerner, M.J., and Frishman, D. (2008). Protein abundance profiling of the Escherichia coli cytosol. BMC Genomics 2008 9:102 *9*, 102.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2012). EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Research *41*, D605–D612.

Kjeldgaard, N.O., Maaloe, O., and SCHAECHTER, M. (1958). The transition between different physiological states during balanced growth of Salmonella typhimurium. Journal of General Microbiology *19*, 607–616.

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. Cell *139*, 1366–1375.

Koch, A., and Deppe, C. (1971). In vivo assay of protein synthesizing capacity of Escherichia coli from slowly growing chemostat cultures. J Mol Biol *55*, 549.

Kuhlman, T., Zhang, Z., Saier, M.H., and Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of Escherichia coli. Proc Natl Acad Sci USA *104*, 6043–6048.

Levine, E., Zhang, Z., Kuhlman, T., and Hwa, T. (2007). Quantitative Characteristics of Gene Regulation by Small RNA. Plos Biol *5*, e229.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. Cell *157*, 624–635.

Lovell, M.C. (2004). Economics with Calculus (World Scientific Publishing Co.).

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2006). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol *25*, 117–124.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Research *25*, 1203–1210.

Lyons, E., Freeling, M., Kustu, S., and Inwood, W. (2011). Using genomic sequencing for classical genetics in E. coli K12. PLoS ONE *6*, e16717.

Maaloe, O. (1979). Regulation of the protein synthesizing machinery - ribosomes, tRNA, factors, and so on. Biological Regulation and Development Volume 1 - Gene Expression 1–56.

Maier, T., Güell, M., and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. FEBS Lett *583*, 3966–3973.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R., and Serrano,

L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol Syst Biol *7*, 511.

Malmström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E.W., and Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. Nature *460*, 762–765.

Mankiw, N.G. (2011). Principles of Economics (Gengage Learning).

Meyer, H., Leist, C., and FIECHTER, A. (1984). Acetate formation in continuous culture of Escherichia coli K 12 D 1 on defined and complex media. Journal of Biotechnology *1*, 355–358.

Miller, J.H. (1972). Experiments in Molecular Genetics (Cold Spring Harbor Laboratory Press).

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. Bioessays *35*, 1050–1055.

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

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

Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. Journal of Bacteriology *119*, 736–747.

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1990). Physiology of the Bacterial Cell: A Molecular Approach (Sunderland, Massachusetts: Sinauer Associates Inc).

Oda, Y., Huang, K., Cross, F.R., Cowburn, D., and Chait, B.T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci USA *96*, 6591–6596.

Okano, H., Hwa, T., Lenz, P., and Yan, D. (2010). Reversible Adenylylation of Glutamine Synthetase Is Dynamically Counterbalanced during Steady-State Growth of Escherichia coli. J Mol Biol *404*, 522–536.

Ong, S., and Mann, M. (2005). Mass spectrometry–based proteomics turns quantitative. Nat Chem Biol *1*, 252–262.

Orth, J.D., Fleming, R.M.T., and Palsson, B.Ø. (2010). 10.2.1 Reconstruction and Use of Microbial Metabolic Networks_ the Core Escherichia coli Metabolic Model as an Educational Guide. EcoSal 1–73.

Pedersen, S., Bloch, P., Reeh, S., and 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.

Perrenoud, A., and Sauer, U. (2005). Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli. Journal of Bacteriology *187*, 3171–3179.

Reitzer, L. (2003). Nitrogen assimilation and global regulation in Escherichia coli. Annual Reviews in Microbiology *57*, 155–176.

Renilla, S., Bernal, V., Fuhrer, T., Castaño-Cerezo, S., Pastor, J.M., Iborra, J.L., Sauer, U., and Cánovas, M. (2011). Acetate scavenging activity in Escherichia coli: interplay of acetyl–CoA synthetase and the PEP–glyoxylate cycle in chemostat cultures. Applied Microbiology and Biotechnology *93*, 2109–2124.

Ross, W., Vrentas, C.E., Sanchez-Vazquez, P., Gaal, T., and Gourse, R.L. (2013). The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. Molecular Cell *50*, 420–429.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

Sauer, U. (2006). Metabolic networks in motion: 13C-based flux analysis. Mol Syst Biol *2*, 62.

Schaechter, M., Maaloe, O., and 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., and Aebersold, R. (2011). Absolute quantification of microbial proteomes at different states by directed mass spectrometry. Mol Syst Biol *7*, 510.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099–1102.

Silva, J.C., Gorenstein, M.V., Li, G.-Z., Vissers, J.P.C., and Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics *5*, 144–156.

Soupene, E., van Heeswijk, W.C., Plumbridge, J., Stewart, V., Bertenthal, D., Lee, H., Prasad, G., Paliy, O., Charernnoppakul, P., and Kustu, S. (2003). Physiological studies of Escherichia coli strain MG1655: growth defects and apparent cross-regulation of gene expression. Journal of Bacteriology *185*, 5611–5626.

Sperling, E., Bunner, A.E., Sykes, M.T., and Williamson, J.R. (2008). Quantitative Analysis of Isotope Distributions In Proteomic Mass Spectrometry Using Least-Squares Fourier Transform Convolution. Anal Chem *80*, 4906–4917.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science *329*, 533–538.

Teixeira de Mattos, M.J., and Neijssel, O.M. (1997). Bioenergetic consequences of microbial adaptation to low-nutrient environments. Journal of Biotechnology *59*, 117–126.

Thomason, L.C., Costantino, N., and Court, D.L. (2007). E. coli Genome Manipulation by P1 Transduction. In Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Unden, G., and Dunnwald, P. (2008). 3.2.2 The Aerobic and Anaerobic Respiratory Chain of Escherichia coli and Salmonella enterica: Enzymes and Energetics. EcoSal 1–65.

Usui, Y., Hirasawa, T., Furusawa, C., Shirai, T., Yamamoto, N., Mori, H., and Shimizu, H. (2012). Investigating the effects of perturbations to pgi and eno gene expression on central carbon metabolism in Escherichia coli using 13C metabolic flux analysis. Microb Cell Fact *11*, 87.

Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.-J., Arike, L., and Vilu, R. (2010). Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase. BMC Systems Biology 2010 4:166 *4*, 166.

Vander Heiden, M., Cantley, L., and Thompson, C. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science *324*, 1029.

Varma, A., and Palsson, B.O. (1994). Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Bio/Technology *12*.

Vogel, C., and Marcotte, E.M. (2008). Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. Nat Protoc *3*, 1444–1451.

Vogel, C., Abreu, R. de S., Ko, D., Le, S.-Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol Syst Biol *6*, 400.

Weinberg, R.A. (2013). The Biology of Cancer (New York: Garland Science).

Woldringh, C.L., Binnerts, J.S., and Mans, A. (1981). Variation in Escherichia coli buoyant density measured in Percoll gradients. Journal of Bacteriology *148*, 58–63.

Wolfe, A.J. (2005). The acetate switch. Microbiol Mol Biol Rev *69*, 12–50.

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

Yuan, J., Bennett, B., and Rabinowitz, J. (2008). Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat Protoc *3*, 1328–1340.