



Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction

Edward J O'Brien¹, Joshua A Lerman¹, Roger L Chang, Daniel R Hyde and Bernhard Ø Palsson*

Department of Bioengineering, University of California San Diego, La Jolla, CA, USA

¹These authors contributed equally to this work.

* Corresponding author. Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, Mail Code 0412, PFBH Room 419, La Jolla, CA 92093-0412, USA. Tel.: +1 858 534 5668; Fax: +1 858 822 3120; E-mail: palsson@ucsd.edu

Received 22.4.13; accepted 5.9.13

Growth is a fundamental process of life. Growth requirements are well-characterized experimentally for many microbes; however, we lack a unified model for cellular growth. Such a model must be predictive of events at the molecular scale and capable of explaining the high-level behavior of the cell as a whole. Here, we construct an ME-Model for *Escherichia coli*—a genome-scale model that seamlessly integrates metabolic and gene product expression pathways. The model computes ~80% of the functional proteome (by mass), which is used by the cell to support growth under a given condition. Metabolism and gene expression are interdependent processes that affect and constrain each other. We formalize these constraints and apply the principle of growth optimization to enable the accurate prediction of multi-scale phenotypes, ranging from coarse-grained (growth rate, nutrient uptake, by-product secretion) to fine-grained (metabolic fluxes, gene expression levels). Our results unify many existing principles developed to describe bacterial growth.

Molecular Systems Biology 9: 693; published online 1 October 2013; doi:10.1038/msb.2013.52

Subject Categories: metabolic and regulatory networks; computational methods

Keywords: gene expression; genome-scale; metabolism; molecular efficiency; optimality

Introduction

The genotype–phenotype relationship is fundamental to biology. Historically, and still for most phenotypic traits, this relationship is described through qualitative arguments based on observations or through statistical correlations. Understanding the genotype–phenotype relationship demands vantage points at multiple scales, ranging from the molecular to the cellular. Reductionist approaches to biology have produced ‘parts lists’, and successfully identified key concepts (e.g., central dogma) and specific chemical interactions and transformations (e.g., metabolic reactions) fundamental to life. However, reductionist viewpoints, by definition, do not provide a coherent understanding of whole cell functions. For this reason, modeling whole biological systems (or subsystems) has received increased attention.

A number of modeling approaches have been developed to predict systems-level phenotypes. What distinguish these models from each other are the underlying assumptions they make, the input data they require, and the scope and precision of their predictions (Selinger *et al.*, 2003). The type of modeling formalism employed is influenced by all of these distinguishing characteristics (Machado *et al.*, 2011). Genome-scale optimality models of metabolism (termed as M-Models) have made much progress in recent years as they require only basic knowledge of reaction stoichiometry, are genome-scale in

scope, and have fairly accurate predictive power. Recently, M-Models have been extended to include the process of gene expression (termed as ME-Models) (Lerman *et al.*, 2012; Thiele *et al.*, 2012), opening up completely new vistas in the development of microbial systems biology. On the heels of these developments, a whole-cell model (WCM) of the human pathogen *Mycoplasma genitalium* appeared (Karr *et al.*, 2012). The WCM integrates many more cellular processes and can be used to simulate dynamic cellular states; however, it depends on detailed molecular measurements of an initial state (e.g., growth rate, biomass composition, and gene expression). While the model described by Karr *et al.* is a major advance toward whole-cell computation, many practical applications rely on the ability to compute optimal phenotypic states. The WCM does not have this ability owing to the disparate mathematical formalisms it employs. The WCM and genome-scale optimality models thus have different capabilities and will find use to predict and explain different biological phenomena.

Here, we construct an ME-Model for *E. coli* K-12 MG1655. The ME-Model is a microbial growth model that computes the optimal cellular state for growth in a given steady-state environment. It takes as input the availability of nutrients to the cell and produces experimentally testable predictions for: (1) the cell’s maximum growth rate (μ^*) in the specified

environment, (2) substrate uptake/by-product secretion rates at μ^* , (3) metabolic fluxes at μ^* , and (4) gene product expression levels at μ^* . The creation of this model required the development of a new modeling formalism and optimization procedure to couple gene expression with metabolism, which provided new insight into growth rate- and nutrient limitation-dependent changes in enzymatic efficiency. The model predicts three distinct regions of microbial growth, defined by the factors (nutrient and/or proteome) limiting growth. We show that proteomic constraints improve predictions of metabolism itself, rectifying dominant failure modes in M-Models. Finally, we compute gene expression changes as the cell transitions through and between the different growth regions. The ME-Model computes measurable coarse- and fine-grained cellular and molecular phenotypes, and provides unity in the field by reconciling a variety of principles related to cellular growth at various scales of complexity.

Results

Integration of genome-scale reaction networks of protein synthesis and metabolism

To create an ME-Model for *E. coli*, we started from two previous network reconstructions. The first reaction network includes all known metabolic pathways as of late 2011 (Orth *et al*, 2011) and is referred to as the M-Model throughout. The second accounts for reactions that describe gene expression and the synthesis of functional macromolecules in a mechanistically detailed manner (Thiele *et al*, 2009). The two reaction networks were integrated (see Materials and methods), and reactions and gene functions in both networks were updated to reflect gaps in knowledge that have been filled since their creation. We updated subunit stoichiometries for hundreds of multiprotein complexes and expanded the types of prosthetic groups, cofactors, and post-translational modifications required for catalytic activity (Materials and methods; Supplementary Table S1).

The scope and coverage of cellular processes in the integrated network is extensive. The integrated network mechanistically links the functions of 1541 unique protein-coding open reading frames (ORFs) and 109 RNA genes; it thus accounts for $\sim 35\%$ of the 4420 protein-coding ORFs, $\sim 65\%$ of the functionally well-annotated ORFs (Riley *et al*, 2006), and 53.7% of the non-coding RNA genes identified in *E. coli* K-12 (Keseler *et al*, 2013). In total, 1295 unique functional protein complexes are produced. Taken together, these complexes account for 80–90% of *E. coli*'s expressed proteome by mass (Supplementary Table S2).

The integrated reaction network covers and accurately predicts a large proportion of essential cellular functions. It includes 223 of the 302 (73.8%) genes classified as essential for cell growth under any condition (Kato and Hashimoto, 2007) (Supplementary Table S3A), and 166 of the 206 functions (80.6%) estimated as essential for a minimal organism (Gil *et al*, 2004) (Supplementary Table S3B). *In silico* prediction of gene essentiality in glucose M9 minimal media results in an accuracy of 88.8% (precision = 60.4%, recall = 75%, Supplementary Table S4). One of the dominant failure modes of essentiality predictions is due to the assumption that all tRNA and rRNA modifications are essential; removing these genes from predictions increases performance notably (accuracy = 92.3%, precision = 75.3%, recall = 75%, Supplementary Table S4). This accuracy is on par with previous approaches using the metabolic reaction network alone (accuracy = 91.2%, precision = 81%, recall = 68%) (Orth *et al*, 2011). Many of the key differences between the M-Model and the ME-Model essentiality predictions are due to the mechanistic treatment of cofactor and prosthetic group synthesis and utilization in the ME-Model. Specifically, for a protein complex to be functional in the ME-Model it has to contain the embedded prosthetic groups required for function; while this change in model structure results in some false predictions of essentiality compared with M-Models (which include all prosthetic groups in a biomass objective function that does not change across conditions), the essentiality predictions in the ME-Model can be directly related to the essential enzymes requiring the prosthetic group.

Growth demands and general constraints on molecular catalysis

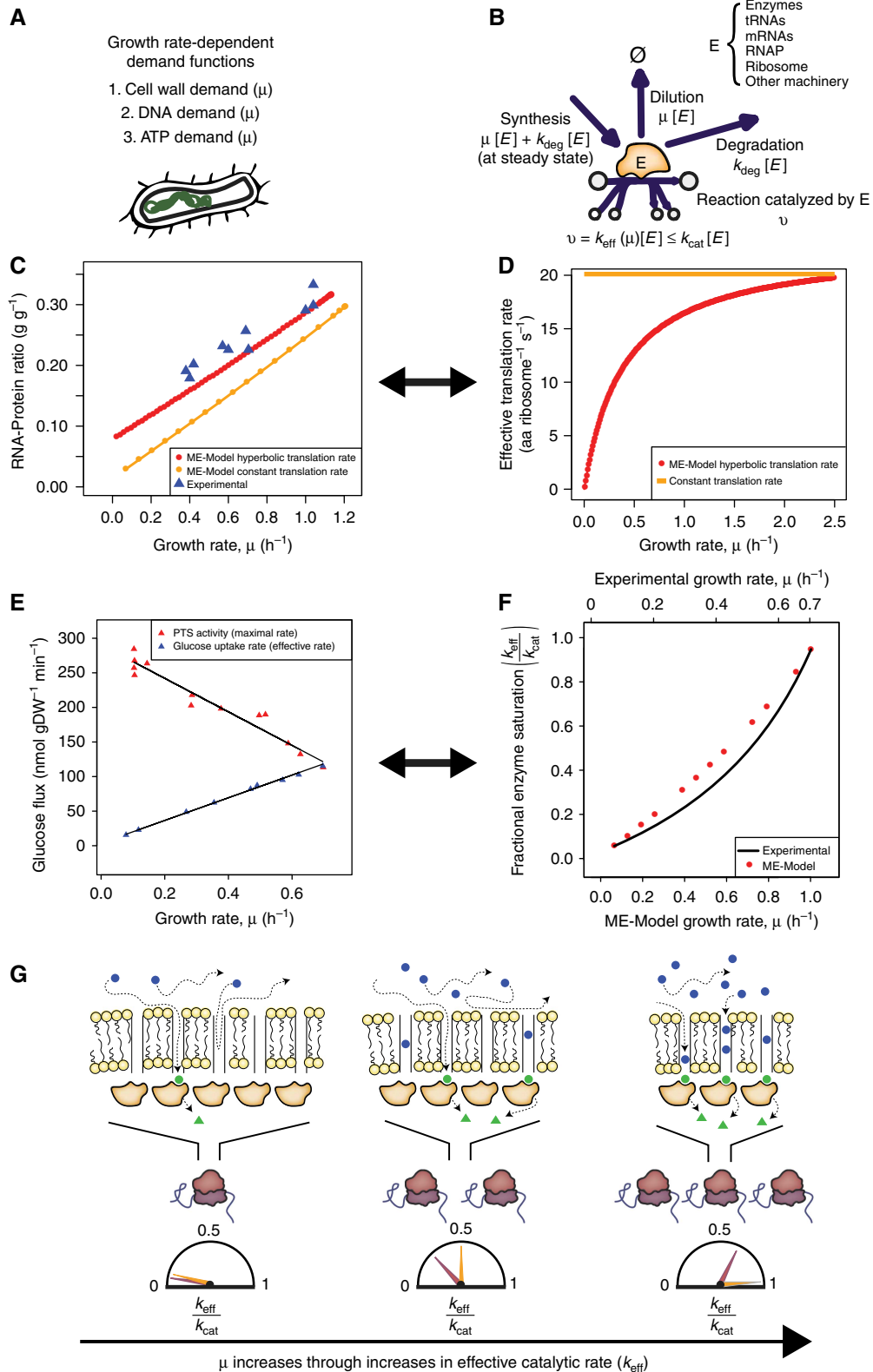
To compute functional states of the integrated network, growth demands are first imposed. Growth requires the replication of the organism's genome and synthesis of a new cell wall to contain the replicated DNA. In the ME-Model, growth rate-dependent DNA and cell wall demand functions formalize these requirements (Figure 1A; Supplementary information). We derived these demand functions from growth rate-dependent trends in cell size (Donachie and Robinson, 1987) and DNA content (Meyenburg and Hansen, 1987; Bremer and Dennis, 1996) (Supplementary information). In addition, as in previous models, we imposed growth-associated and non-growth-associated ATP utilization demands (Pirt, 1965) as the ostensible energy requirements (Neijssel *et al*, 1996; Zhuang *et al*, 2011).

One large improvement is that RNA and protein are not included as demand functions (as they are in M-Models; Feist

Figure 1 Growth demands and coupling constraints leading to growth rate-dependent changes in enzyme and ribosome efficiency. (A) Three growth rate-dependent demand functions derived from empirical observations determine the basic requirements for cell replication (detailed in Supplementary information). (B) Coupling constraints link gene expression to metabolism through the dependence of reaction fluxes on enzyme concentrations. (C, D) RNA:protein ratio predicted by the ME-Model with two different coupling constraint scenarios, one for variable translation rate versus growth rate (red lines) and one for constant translation rate (orange lines). Experimental data in (C) obtained from Scott *et al* (2010). (E) Phosphotransferase system (PTS) transient activity following a glucose pulse in a glucose-limited chemostat culture (red) and glucose uptake before the glucose pulse (blue) is plotted as a function of growth rate. The data shown were obtained from O'Brien *et al* (1980). Data from $\mu > 0.7 \text{ h}^{-1}$ were omitted. (F) Data from (E) are used to plot glucose uptake as a fraction of PTS activity. The resulting value is the fractional enzyme saturation (black line). The fractional enzyme saturation predicted by the ME-Model is plotted as a function of growth rate under carbon limitation (red dots). (G) The cartoon depicts changes in extra- (blue) and intra- (green) cellular substrate (circle) and product (triangle) concentrations and metabolic enzyme (orange) and ribosome (purple/maroon) levels as the concentration of a growth-limiting nutrient (and growth rate) increases. The dials show $k_{\text{eff}}/k_{\text{cat}}$, the effective catalytic rate over the maximum for metabolic enzymes (orange) and ribosomes (purple/maroon).

and Palsson, 2010); instead, expression of specific RNA and protein molecules are free variables determined during ME-Model simulations. 'Coupling constraints' (Thiele *et al*, 2010;

Lerman *et al*, 2012) relate the synthesis of RNA- and protein-based molecules to their catalytic functions in the cell (Figure 1B). The coupling constraints are based on parameters



that define the effective catalytic rate (k_{eff}) and degradation rate constant (k_{deg}) of molecular machines (Supplementary information).

A nutritional environment is then defined by setting constraints on the availability and uptake of nutrients. For a particular nutritional environment, there is a maximum growth rate at which the cell can no longer produce enough RNA and protein machinery to meet the demands of growth. The computed cellular state (biomass composition, substrate uptake and by-product secretion, metabolic flux, and gene expression) at this maximum growth rate is the predicted optimal response of the cell to the specified nutritional environment.

Derivation of constraints on molecular catalytic rates

Previous studies disagree as to if ribosomes translate with the same efficiency (amino acids per ribosome per second) across growth conditions (Young and Bremer, 1976; Scott *et al*, 2010). Here, we use the ME-Model and available data to determine an appropriate constraint for ribosomal efficiency as a function of growth rate. We find that if a constant translation rate of 20 amino acids per second is imposed as a constraint in the ME-Model, the model predicts a linear growth rate-dependent RNA-to-protein ratio (Figure 1C), consistent with the previous measurements (Scott *et al*, 2010); however, the predicted RNA content does not quantitatively match measured values. In particular, a constant translation rate results in no RNA production in the limit of no growth. We therefore hypothesized that ribosomal translation rate systematically varies with growth rate, and back-calculated a growth rate-dependent translation rate using measured growth rate-dependent RNA content (Supplementary information). Ultimately, we recovered a Michaelis–Menten-type rate law (Figure 1D) with a maximal rate (V_{max}) of ~ 20 amino acids per second, consistent with previous findings for maximal ribosomal speed (Bremer and Dennis, 1996); the rate law results in a quantitative match of RNA content compared with experimental data (Figure 1C, Pearson's $r = 0.96$). This rate law causes translation efficiency to increase under nutrient-rich conditions, which recent experimental evidence supports (Proshkin *et al*, 2010; Valgepea *et al*, 2013). Interestingly, when we applied the same Michaelis–Menten-type equations to constrain tRNA and mRNA catalytic rates, we recovered

maximal turnover rates highly consistent with previous estimates (Supplementary information).

The catalytic rates of metabolic enzymes are variable as well, and tend to decrease when nutrients are limited. Both metabolomics (Boer *et al*, 2010) and proteomics (Valgepea *et al*, 2013) data sets suggest a large-scale scaling of enzyme efficiencies under nutrient limitation. We approximate these changes in metabolic catalysis in the ME-Model with two minimal assumptions: (1) when the cell is nutrient-limited, protein content is maximized (at a given growth rate) and (2) this protein content specifically is metabolic enzymes not operating at their maximal catalytic rate (Valgepea *et al*, 2013) (i.e., $k_{\text{eff}}/k_{\text{cat}} < 1$, see Figure 1G and Supplementary information, Optimization procedure). These two assumptions allow us to predict average catalytic rates of metabolic enzymes under nutrient limitation. The nutrient limitation-dependent shape of our computed catalytic rates matches assays for glucose transporters under glucose limitation (O'Brien *et al*, 1980) (Figures 1E and F), LacZ under lactose limitation (Smith and Dean, 1972) (Supplementary Figure S1A), and the enzyme efficiency in a small-scale optimality model accounting for substrate concentrations with Michaelis–Menten kinetics (Molenaar *et al*, 2009) (Supplementary Figure S1B). However, because the current ME-Model simulation procedure assumes that k_{eff} decreases uniformly across metabolism, the model does not capture the importance of specific enzymes for particular nutrient limitations; recent data sets (Valgepea *et al*, 2013) and kinetic models (Kim *et al*, 2012) can help us understand and model these trends better at the genome-scale.

Growth regions under varying nutrient availability

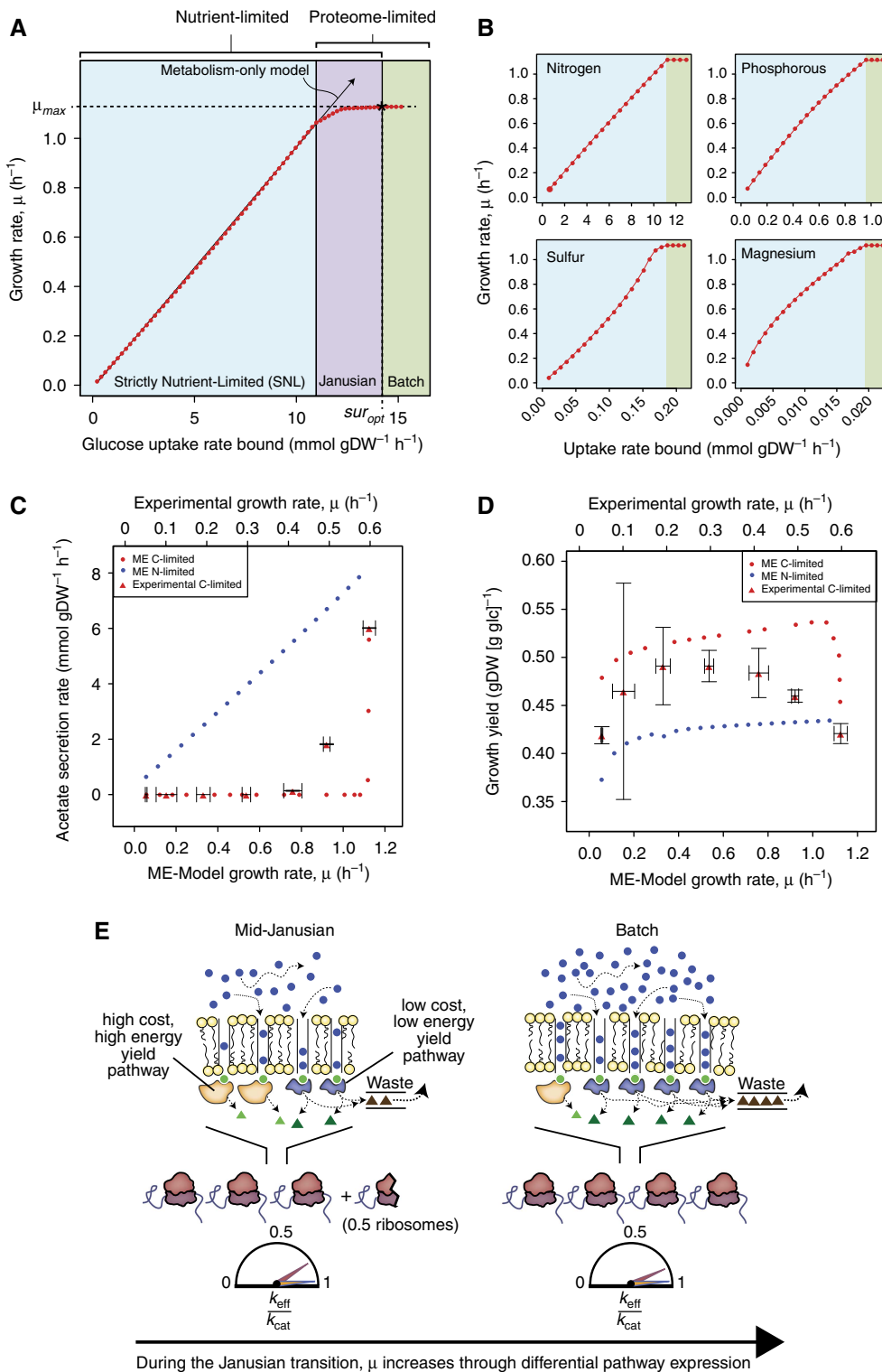
Upon derivation of the growth demands and molecular efficiencies, we investigate high-level model behavior to variable nutrient availability. Unlike previous genome-scale models (Orth *et al*, 2011; Thiele *et al*, 2012), growth rate in the ME-Model is a non-linear function of the substrate uptake rate bound (Figure 2A), and eventually reaches a maximum. This behavior is consistent with long-standing empirical models of microbial growth (Monod, 1949; Koch, 1997), in which growth is first nutrient-limited, but then limited by some intra-organismal bound.

Under nutrient-excess conditions, growth in the ME-Model is limited by internal constraints on protein production

Figure 2 Predicted growth, yield, and secretion. (A) Predicted growth rate is plotted as a function of the glucose uptake rate bound imposed in glucose minimal media. Three regions of growth are labeled Strictly Nutrient-Limited (SNL), Janusian, and Batch (i.e., excess of substrate) based on the dominant active constraints (nutrient and/or proteome limitation). The proteome-activity constraint inherent in the ME-Model results in a maximal growth rate and substrate uptake rate. The behavior of a genome-scale metabolic model (M-Model) is depicted with an arrow. (B) Predicted growth rates as a function of uptake of a limiting nutrient with glucose in excess. The shaded regions correspond to those as labeled in (A). (C) Experimental (triangle) and ME-Model-predicted (circle) acetate secretion in Nitrogen- (blue) and Carbon- (red) limited glucose minimal medium are plotted as a function of growth rate. Data were obtained from Zhuang *et al* (2011). The root-mean-square error (RMSE) between data and the ME-Model is 0.12 (for comparison, RMSE = 0.40 for the M-Model). (D) Experimental (triangle) and ME-Model-predicted (circle) carbon yield (gDW Biomass/g Glucose) in Carbon- (red) and Nitrogen- (blue) limited glucose minimal medium are plotted as a function of growth rate. Data were obtained from Zhuang *et al* (2011). RMSE between data and the ME-Model is 0.04 (for comparison, RMSE = 0.07 for the M-Model). (E) The cartoon depicts changes in extra- (blue) and intra- (green) cellular substrate (circle) and product (triangle) concentrations and metabolic enzyme (blue/orange) and ribosome (purple/maroon) levels during the Janusian region. Metabolic enzymes are saturated throughout the entire Janusian region. To increase the growth rate, the cell expresses metabolic pathways that have lower operating costs. (Pathways with the smaller blue proteins taken to be 0.25 the cost of the pathways with larger orange proteins.) A higher glucose uptake and turnover results, but energy yield is lower and some carbon is 'wasted' and secreted (brown triangles). The dials show $k_{\text{eff}}/k_{\text{cat}}$, the effective catalytic rate over the maximum for metabolic enzymes (blue/orange) and ribosomes (purple/maroon).

and catalysis—the cell is ‘proteome-limited’—resulting in a corresponding maximal growth rate (Figure 2A). This feature allows Batch culture growth to be simulated without specifying nutrient uptake bounds; instead, the ME-Model predicts a maximum batch growth rate and optimal substrate uptake rate.

Supporting the validity of the proteomic constraints limiting growth in Batch culture, optimal Batch growth rates, substrate uptake rates, and biomass yields correlate with experimental data for growth on different carbon sources (Supplementary Table S5). The ME-Model predicted substrate uptake and biomass yield closely matches laboratory evolved strains



(Pearson's $r = 0.89$ and $r = 0.91$, respectively) (Supplementary Table S5C, sensitivity analysis in Supplementary Table S6). Though less accurate, predicted growth rates by the ME-Model correlate with measured growth rates in batch culture better than standard M-Models, in which growth rate is maximized subject to a specified nutrient uptake, and the correlation increases when compared with laboratory evolved strains (M-Model Pearson's $r = 0.49$, ME-Model Pearson's $r = 0.61$) as opposed to wild-type strains (M-Model Pearson's $r = 0.30$, ME-Model Pearson's $r = 0.39$). Other methods that include various approximate constraints on the total flux through the metabolic network also show an increased performance in growth rate prediction, though all computational methods (Beg *et al*, 2007; Adadi *et al*, 2012) still correlate better with each other than with the experimental data (Supplementary Table S5B).

When the uptake of glucose is restricted below the amount required for optimal growth in batch culture, the cell's growth is carbon-limited. Growth rate linearly increases with glucose uptake when glucose availability is low. In this region (termed as the Strictly Nutrient-Limited (SNL) region in Figure 2A), the capabilities of the proteome are not fully utilized as the proteome could process more incoming glucose if it was available (Figures 1E–G). By varying the glucose availability, we find that a region exists in which the cell is both nutrient- and proteome-limited; we refer to this transition region as the Janusian region (Button, 1991). ME-Model computations thus reveal three distinct regions of microbial growth (Figure 2A; see Supplementary information, Optimization procedure, Computational definition, and identification of growth regions).

When the uptake of non-carbon sources is restricted below the amount required for optimal growth in batch culture, the cell's growth is limited by that nutrient. Unlike carbon-source limitation, we find the nutrient- and proteome-limited regions to be distinct (Figure 2B). However, in the SNL region, growth is sometimes non-linear as a function of uptake rate, due to changing biomass requirements (e.g., Sulfur and Magnesium).

Effect of proteome limitation on secretion phenotypes

To understand the proteome-limited growth regions in the ME-Model, we first investigate trends in secretion phenotypes and biomass yield. Under glucose limitation, different metabolic pathways are utilized in the Janusian region than in the SNL region, resulting in acetate secretion (Figure 2C, red). This metabolic switch, combined with growth rate-dependent ATP requirements, results in a concave biomass yield as a function of growth rate (Figure 2D, red). Both the biomass yield and secretion trends have repeatedly been experimentally observed (Zhuang *et al*, 2011).

The example of glucose limitation provides an illustrative example for the general behavior in the Janusian growth region. In the Janusian region, the cell increases its growth rate through differential expression of pathways, as illustrated in Figure 2E. Due to proteome limitations, the cell switches to pathways that require less protein mass but are lower in nutrient yield (defined as energy and/or biomass precursors produced per molecule of limiting nutrient consumed). This

behavior is in contrast to that in the SNL region, in which high-yield pathways are optimal (as in M-Models) and growth rate increases through changes in the effective catalytic rate of metabolic enzymes (Figure 1G). These results provide further support that 'overflow' metabolism can be understood in terms of proteomic constraints, as suggested with a small-scale model (Molenaar *et al*, 2009).

The ME-Model also predicts that acetate will be secreted at all growth rates when *E. coli* is Nitrogen (Ammonium)-limited (Figure 2C, blue). Experimentally, acetate is secreted under nitrogen limitation even at low growth rates (Hua *et al*, 2004). This secretion phenotype is explained by the ME-Model as follows: protein 'saved' by utilizing low-yield carbon metabolism is diverted to synthesize other enzymes that are not operating at their maximal catalytic capacity.

No Janusian region is observed under non-carbon limitation. In the ME-Model, this is likely due to reaction network topology—while there are many alternative pathways for energy, redox, and biomass precursor generation in carbon metabolism, non-carbon nutrient assimilation is often achieved using more linear pathways. As a result, there are fewer opportunities for trade-offs between uptake rate and biomass yield. However, perhaps including variable substrate affinities for alternative pathways would reveal Janusian regions corresponding to non-carbon limitations.

Central carbon fluxes reflect growth optimization subject to catalytic constraints

Further supporting the importance of proteomic constraints on metabolic phenotypes is the prediction of central carbon fluxes by the ME-Model. When glucose availability is varied, the ME-Model predicts changes in central carbon metabolism consistent with the changes from ^{13}C fluxomic data sets (Figure 3; Supplementary Figure S2, Pearson's $r = 0.93$, 0.90 , 0.86) (Nanchen *et al*, 2006; Schuetz *et al*, 2007, 2012). Importantly, the ME-Model predicts the dominant changes in pathway splits as the glucose availability is varied (Figure 3, insets).

Previous studies have evaluated the ability of M-Models together with assumed optimality principles to predict metabolic fluxes (Schuetz *et al*, 2007, 2012). These studies concluded that no single objective function applied to M-Models can accurately represent fluxomic data from all environmental conditions studied (Schuetz *et al*, 2007). Instead, metabolic fluxes can be understood as being Pareto optimal: multiple objectives are simultaneously optimized and their relative importance varies depending on the environmental condition (Schuetz *et al*, 2012). The three objectives needed to explain most of the variations in the data from Schuetz *et al* were (1) maximum ATP yield, (2) maximum biomass yield, and (3) minimum sum of absolute fluxes (which is a proxy for minimum enzyme investment). These three objectives formed a Pareto optimal surface that was valuable for interpreting fluxomic data; however, the surface was large and it was not possible to predict the importance of each of the objectives *a priori*.

By explicitly accounting for variable growth demands, enzyme expression, and constraints on enzymatic activity, the ME-Model eliminates the need for multiple objectives;

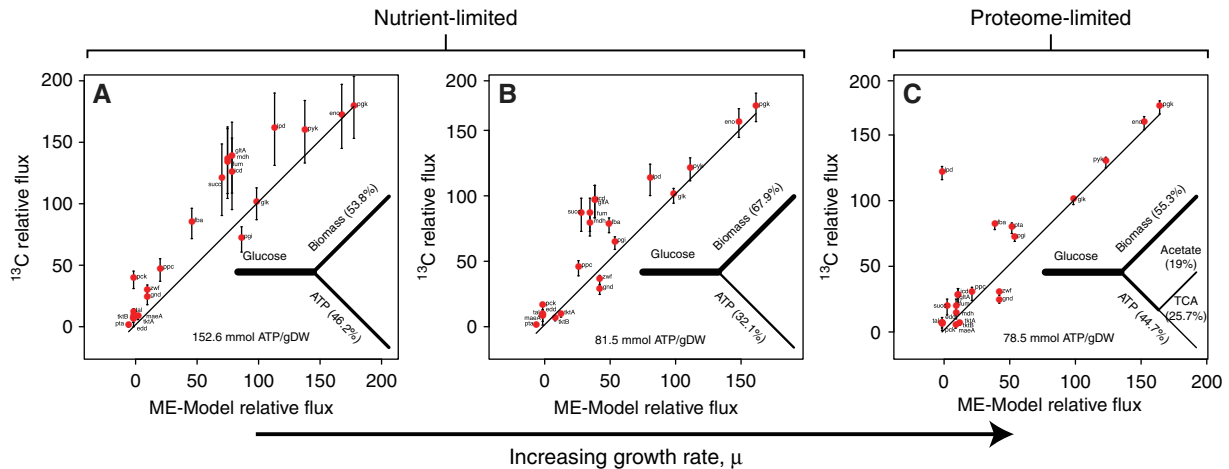


Figure 3 Central carbon metabolic flux patterns under glucose-limited and glucose-excess conditions. (A–C) Relative fluxes from ^{13}C experiments are plotted versus the fluxes predicted by the ME-Model. (A, B) Comparison of nutrient-limited model solutions with chemostat culture conditions and (C) comparison of the batch ME-Model solution with batch culture data. All simulations and experiments correspond to growth in glucose minimal media. Fluxes are normalized so that glucose uptake is 100. Insets show the main flux changes under increasing glucose concentrations. The only model parameter that is modulated is the glucose uptake rate bound. Data were obtained from Nanchen *et al* (2006) and Schuetz *et al* (2007). The ME-Model flux for the reaction 'pyk' is taken to include phosphoenolpyruvate (PEP) to pyruvate (PYR) conversion via the phosphotransferase system (PTS). Flux splits shown as insets were computed using the ME-Model. The percentages indicate the percent carbon (Glucose) converted to CO_2 (for branch labeled 'TCA'), acetate, and biomass. Both the TCA and acetate branches contribute to ATP production. The total mmol ATP per gDW biomass produced is indicated.

growth rate optimization alone is sufficient to predict the fluxes through central carbon metabolism (Figure 3; Supplementary Figure S2; Supplementary Table S7). The three original objectives chosen by Schuetz *et al* are biologically meaningful dimensions and required for interpreting fluxomic data when using an M-Model. In contrast, the ME-Model accounts for all three of these dimensions implicitly during growth rate maximization without adjusting any model parameters (see Supplementary information and Supplementary Table S7). Accordingly, ME-Models can determine, at least qualitatively, the importance and weighting of the objectives for growth in a given environment. Ultimately, the primary changes in flux through central carbon metabolism can be understood as responses to the same constraints causing the observed relationship in biomass yield (Figure 2D): at low growth rates under carbon limitation, the dominant changes are due to a changing ATP demand, and in the transition from carbon-limited to carbon-excess (proteome-limited) conditions, the primary changes are due to the switch to lower yield carbon catabolism (Figure 3, insets).

***In silico* gene expression profiling from nutrient-limited to batch growth conditions**

We now use the ME-Model to predict groups of proteins that change in expression under various degrees of glucose limitation. Under glucose limitation, the optimal proteome changes due to shifting growth demands and proteomic constraints. The groups of functionally related proteins that shift in our simulations match those previously reported experimentally (Vemuri *et al*, 2006; Nahku *et al*, 2010), but the model predictions of quantitative differential expression (at the level of single genes) are weak. We separate the analysis of

the SNL region (Figure 4; Supplementary Table S8A) from the Janusian region (Figure 5; Supplementary Table S8B), due to the different dominant constraints and phenotypic responses specific to each region.

In the SNL region, the expression of most proteins decreases as growth rate increases (Figure 4B, left side of tree, Supplementary Figure S3). The largest group of proteins includes those responsible for amino-acid and cell wall synthesis; the growth rate-dependent decrease in expression of these proteins is due to the combined effects of a decrease in cell wall and protein biomass (g/gDW) and an increase in the effective catalytic rate of enzymes (Figures 1E–G). Proteins involved in energy metabolism also decrease in expression with increasing growth rate due to changes in catalytic rate and growth rate-dependent demands. Surprisingly, the predicted expression levels of several accessory transcription proteins, including four stress-associated sigma factors (RpoS, RpoH, RpoE, and RpoN), are elevated at very low growth rates, reflecting an association with metabolic proteins needed for slow growth.

A smaller number of proteins show increases in their relative expression levels at higher growth rates (Figure 4B, right side of tree, Supplementary Figure S3). These proteins include those responsible for protein synthesis (ribosome, RNAP, and accessory proteins such as elongation factors) and proteins involved in RNA biosynthesis. The increase in expression of RNA biosynthetic machinery is necessary for *de novo* synthesis of ribonucleotides and to ensure flux through nucleotide salvage pathways (mainly to support an increase in rRNA biomass). Finally, the expression profile of the pentose phosphate pathway reflects the interplay between the increasing demand for ribonucleotide precursors and the decreasing demand for amino-acid precursors.

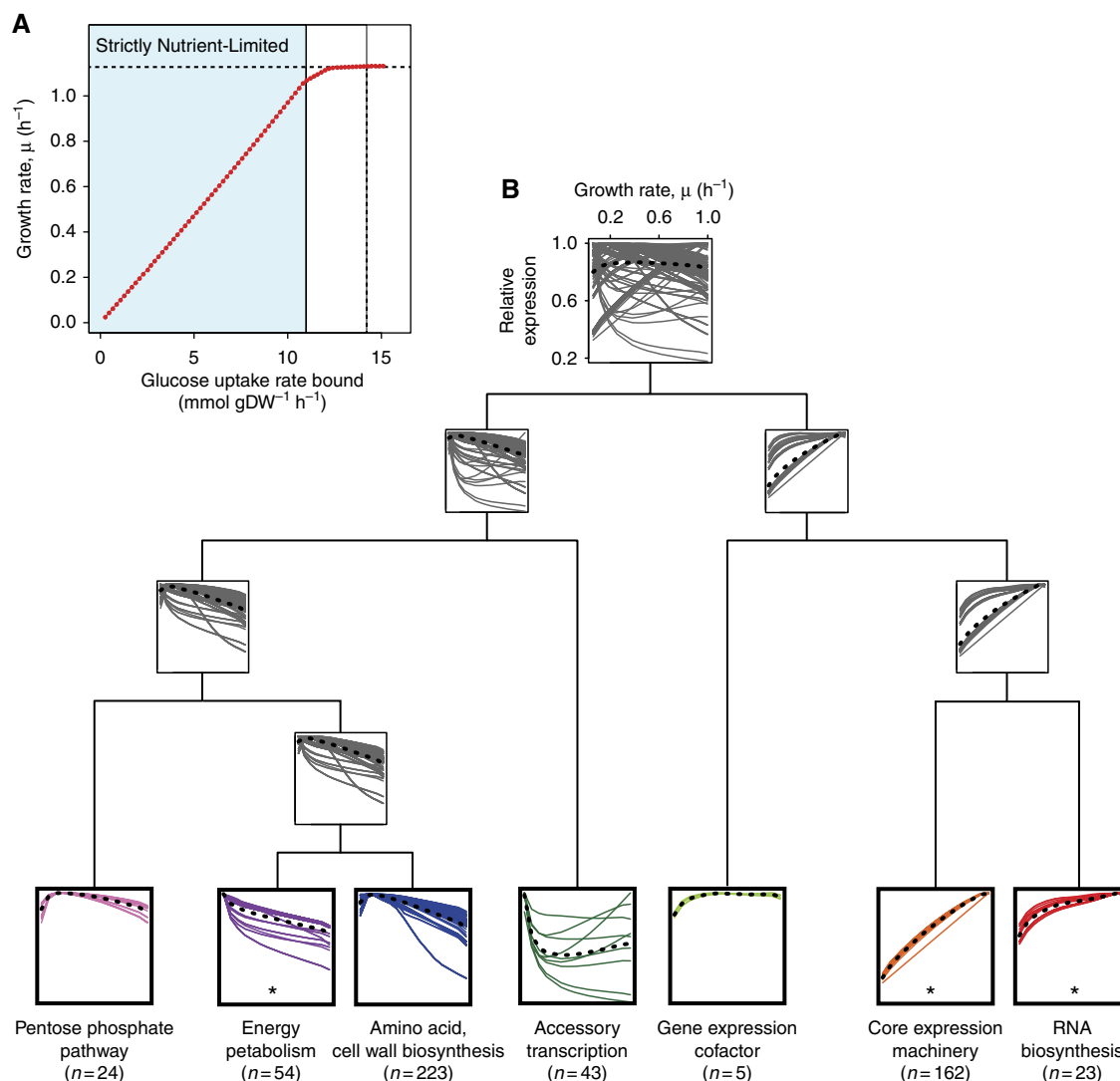


Figure 4 Growth rate-dependent gene expression under glucose limitation. **(A)** Gene expression changes predicted by the ME-Model to occur in the Strictly Nutrient-Limited (SNL) growth region indicated in light blue under glucose limitation in minimal media are analyzed. **(B)** ME-Model-computed relative gene–enzyme pair expression is plotted as a function of growth rate; the normalized *in silico* expression profiles are clustered hierarchically (see Materials and methods). Solid lines are expression profiles of individual gene–enzyme pairs and dotted black lines are the centroid of each cluster. Each leaf node is colored and qualitatively labeled by function. The number of genes in each leaf node is indicated and listed in Supplementary Table S8A. Asterisks indicate clusters with monotonic expression changes that significantly match the directionality observed in expression data (Wilcoxon signed-rank test, $P < 1 \times 10^{-4}$). Expression data were obtained from a previous study (Nahku *et al*, 2010), in which *E. coli* was cultivated in a chemostat at dilution rates 0.3 h^{-1} and $\sim 0.5 \text{ h}^{-1}$.

To validate our predicted expression changes, we compared gene clusters with expression data from *E. coli* grown at 0.3 h^{-1} and $\sim 0.5 \text{ h}^{-1}$ in a glucose-limited chemostat (Nahku *et al*, 2010). In this data set, genes in Energy Metabolism (purple), Core Expression Machinery (orange), and RNA Biosynthesis (red) all significantly change in the predicted direction (Wilcoxon signed-rank test, $P < 1 \times 10^{-4}$), supporting our predicted expression profiles. The other clusters showed no significant changes in the data set; these clusters are either small in size or do not change monotonically, hindering direct comparison with this data set. The ME-Model is not yet predictive of quantitative gene expression changes (at the level of single genes); the correlation over the entire data set is statistically

significant ($P < 0.005$), but weak (Pearson's $r = 0.14$). Our approach is at present limited to qualitative predictions of the direction of change of small groups of functionally related proteins.

In the Janusian region of growth (Figure 5), the cell transitions from carbon-limited to proteome-limited constraints, resulting in a distinct transcriptional response. At the beginning of this transition, the cell has reached a nutrient level where enzymes are saturated (Figure 1G); as growth rate increases, the total demand of anabolic processes increases, causing a global increase in the bulk of metabolism and gene expression machinery (Figure 5B). To meet these proteome demands, energy metabolism is altered to favor lower yield catabolic pathways that require less protein (so that the

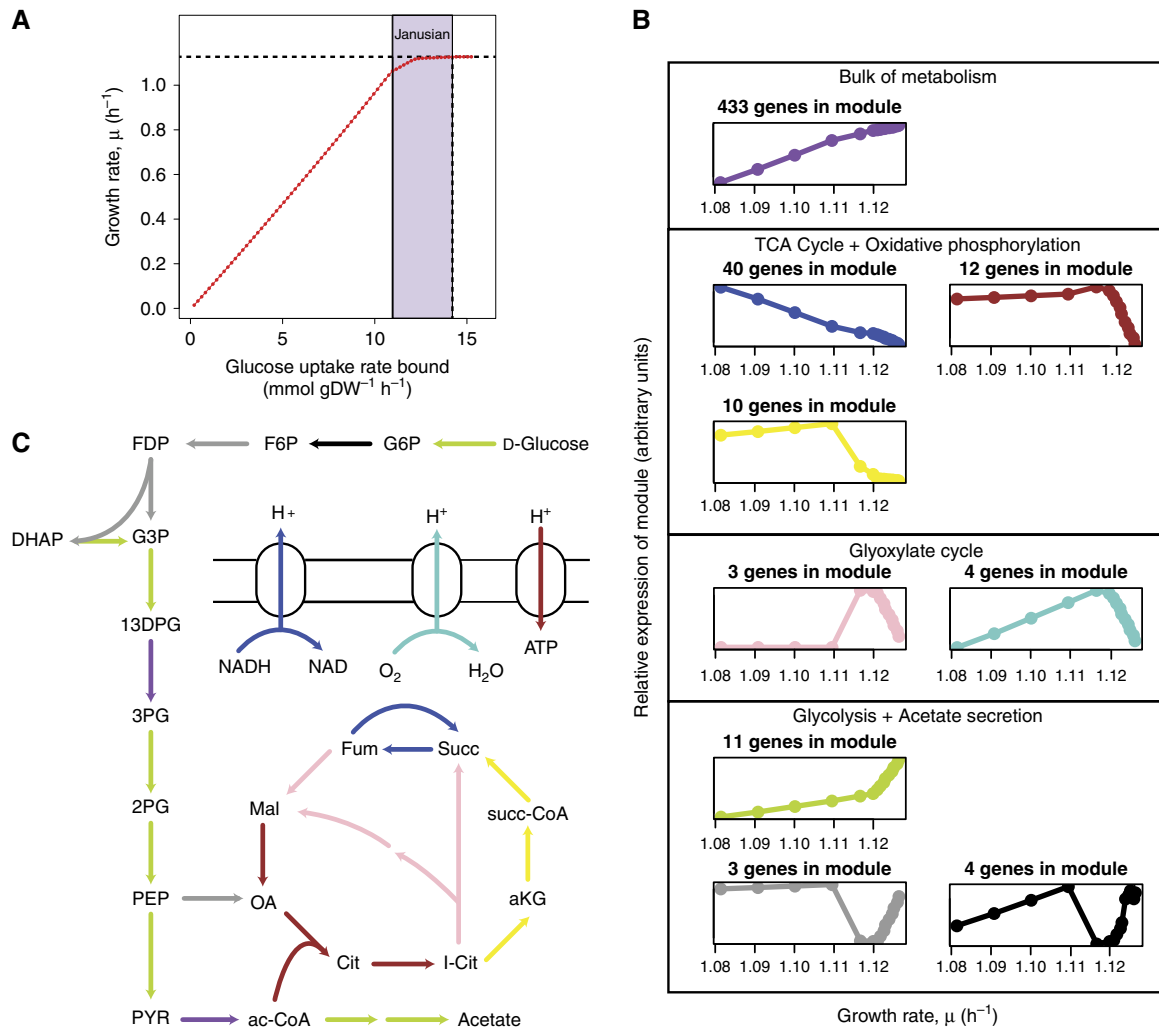


Figure 5 Gene expression during the Janusian region. **(A)** Gene expression changes predicted by the ME-Model to occur in the Janusian growth region indicated in purple under glucose limitation in minimal media are analyzed. **(B)** Simulated expression profiles for eleven clusters of genes (with $\beta = 25$ correlation similarity and average agglomeration) are analyzed. A freely available R package was used (Langfelder and Horvath, 2008). Eleven clusters resulted. Two small clusters were removed because they represented stochastic expression of alternative isozymes. The first principal component of the remaining nine clusters is displayed and grouped qualitatively by function. **(C)** Many of the expression modules correspond to genes of central carbon energy metabolism. Reactions are colored according to the module color in **(B)**.

protein can instead be used for anabolic processes); this is accomplished through a decrease in TCA Cycle and Oxidative Phosphorylation expression in favor of a transient increase in the Glyoxylate Cycle followed by a large increase in Glycolysis and acetate secretion (Figures 5B and C), consistent with previously observed changes in gene expression in the transition to glucose-excess environments (Vemuri *et al*, 2006).

The ME-Model predicts intricate expression changes as glucose availability changes by employing relatively simple constraints on molecular catalysis and biomass composition. This study is the first to attempt genome-scale prediction of gene expression levels under changing growth rate and/or nutrient limitation from optimality principles alone. Systematic consideration of transcriptional regulation and inclusion of missing constraints and parameters impacting optimality (e.g., kinetic constraints and parameters) are future endeavors

necessary to extend the predictive power to the level of single genes (see Discussion).

Discussion

The ME-Model is a microbial growth model that computes the optimal cellular state for growth in a given steady-state environment. It takes as input the availability of nutrients to the cell and produces experimentally testable predictions for: (1) the cell's maximum growth rate (μ^*) in the specified environment, (2) substrate uptake/by-product secretion rates at μ^* , (3) metabolic fluxes at μ^* , and (4) gene product expression levels at μ^* .

Important to the predictions of the ME-Model is the proper coupling between metabolism and gene product expression. Through comparison of model simulations with experimental

data, we derived two general classes of molecular efficiencies that vary based on the growth rate and the degree of nutrient limitation. For ribosomes (and tRNA and mRNA), we propose a growth rate-dependent Michaelis–Menten-type model for polymerization speed, which has preliminary experimental evidence (Proshkin *et al*, 2010), though we have not seen it previously proposed. We furthermore show that two simple assumptions allow us to approximate the effect of nutrient limitation on metabolic enzyme catalysis. While enzyme-specific trends in catalytic rates depend on the limiting nutrient (Bennett *et al*, 2009; Boer *et al*, 2010), our formulation is a first step toward modeling genome-scale effects of nutrient limitation and suggests that simple principles may underlie these trends. Both of these molecular efficiency variables are essential for genome-scale modeling of gene expression and warrant future studies to validate and refine them further. Paired proteomic and metabolomic data sets under nutrient-limited conditions will allow for a deeper understanding of nutrient limitation-dependent effective catalytic rates, and new data sets (Li *et al*, 2012) and models (Tuller *et al*, 2010) on the processes of gene expression can help to refine model parameters and determine their genome-scale effects.

The proteomic constraints inherent to the ME-Model result in qualitatively different growth predictions compared with previous genome-scale models. In the ME-Model, growth rate is not a simple linear function of substrate uptake bounds; instead, the ME-Model predicts a maximal growth rate and optimal substrate uptake rates, which better reflects empirical growth models and better predicts experimentally measured growth rates and substrate uptake rates. The ME-Model reveals three distinct growth regions, which we term SNL, Janusian, and Batch; while nutrient-limited (chemostat culture) and nutrient-excess (batch culture) conditions are commonplace, the Janusian region (where the cell is limited by both nutrient availability and proteome capacity) is rarely considered in microbiology. Interestingly, we observe the Janusian region to occur under carbon limitation but not under various non-carbon limitations. We take this to mean that Janusian regions may exist for non-carbon limitations, but the constraints that may cause them to arise are outside the scope of the current ME-Model.

The proteomic constraints in the ME-Model also improve predictions of by-product secretion and metabolic flux under both nutrient-excess and nutrient-limited conditions. By accounting for the metabolic cost of proteins and limitations of protein production capacity, the ME-Model accurately decouples substrate uptake, growth rate, and growth yield, allowing for important rate-yield trade-offs to be predicted. In particular, we show that seemingly inefficient metabolism in batch culture and under nitrogen limitation (both when carbon is in excess), can be explained and predicted through proteomic trade-offs. This capability rectifies the dominant failure mode in predicting metabolic flux previously reported for M-Models (Schuetz *et al*, 2012), and suggests that a single objective of growth rate (if the proper constraints are included) may be able to predict metabolic fluxes. This result shows that proteomic constraints are necessary to accurately predict metabolic responses—optimal growth and metabolic phenotypes cannot be fully understood without taking gene expression into account. From a practical standpoint, the

natural parsimony present in ME-Model simulations (Lerman *et al*, 2012) strongly reduces the optimal solution space, allowing for more precise predictions, an important feature in diverse applications. The effect of proteomic constraints on secretion phenotypes is of particular importance for applications in systems metabolic engineering, and will be necessary for simulating behavior in complex media and predicting nutrient preferences.

At the level of gene expression, the ME-Model predicts detailed behavior in each growth region. In the SNL and Janusian growth regions, gene modules have distinct nutrient limitation-dependent profiles. A number of the gene modules change in the correctly predicted direction compared with expression data from *E. coli* in a chemostat at different growth rates (Vemuri *et al*, 2006; Nahku *et al*, 2010), supporting our predicted expression profiles. By predicting optimal gene expression profiles, the ME-Model aids in understanding the factors shaping the evolution of gene expression patterns (e.g., proteomic constraints and changing biomass composition).

Modeling optimal transcriptional responses is complementary to the elucidation and modeling of specific regulatory mechanisms (Klumpp *et al*, 2009; Cho *et al*, 2011; Berthoumieux *et al*, 2013). It is tempting to relate the expression profiles predicted by the ME-Model to molecular mechanisms underlying the control gene expression *in vivo* (Klumpp *et al*, 2009; Berthoumieux *et al*, 2013; Gerosa *et al*, 2013). For example, constitutively expressed genes display growth rate-dependent expression trends (Klumpp and Hwa, 2008; Klumpp *et al*, 2009), which might provide the cell with an economical way of responding to global changes in metabolic efficiency (Valgepea *et al*, 2013). Also, PurR could be responsible for regulating the increase in expression of nucleotide biosynthesis genes at higher growth rates (as PurR is an autorepressor, this could be accomplished through mechanisms described in Klumpp *et al*, 2009). Finally, though the primary role of ArcA is to respond oxygen availability (Cho *et al*, 2006), it also represses many of the genes in the TCA cycle and Oxidative Phosphorylation that decrease during the glucose-limited to glucose-excess (Janusian) transition (Vemuri *et al*, 2006; Haverkorn van Rijsewijk *et al*, 2011). However, as regulatory mechanisms are not explicitly considered in the ME-Model, the relation between regulatory mechanisms and simulated expression profiles is indirect; while this comparison can assist in explaining and expanding upon the functional roles of cellular regulators, much further work is required to validate the resulting hypotheses.

As it is an optimality model, the ME-Model is particularly suited for studies related to adaptive laboratory evolution (ALE). Recently, it was reported that it is not possible to predict some changes that occur during ALE in Batch culture using an M-Model (Harcombe *et al*, 2013). This is because M-Models only take biomass yield optimization into account; these results are consistent with the rate-yield trade-offs present in the ME-Model under nutrient-excess conditions. In the ME-Model, a number of inherent factors can limit cellular growth (e.g., translation rate and metabolic catalysis); the ME-Model can thus provide alternative hypotheses for the mechanisms of growth increase and aid in understanding the results of ALE.

The ME-Model can simulate coarse- to fine-grained cellular and molecular phenotypes with an improved accuracy and

scope compared with previous genome-scale models. The ME-Model shows complex behavior as a result of linear constraints applied to an integrated network. The ME-Model thus shows that intricate and seemingly unintuitive phenotypes can be modeled at a genome-scale with simple enough assumptions to understand their underlying cause. Due to the richness of the model simulations, we primarily focused on *E. coli* growing in glucose minimal media at different growth rates by modulating the availability of glucose; there are therefore many future opportunities to investigate model predictions under many environmental and genetic conditions.

A whole-cell *E. coli* model has been desired for some time (Crick, 1973) as such a model would have profound impacts for basic microbiology, the study of microbial communities, antibiotic discovery, the elucidation of regulatory networks, and systems metabolic engineering. We hope the ME-Model will serve as a scaffold for continued model development toward these practical applications.

Materials and methods

Network reconstruction

The two primary reaction networks used to create the ME-Model were the most recent metabolic reconstruction (Orth *et al.*, 2011), and a network detailing the reactions of gene expression and functional enzyme synthesis (Thiele *et al.*, 2009). The gene expression reconstruction is formalized as a set of 'template reactions' that can be applied to different components (e.g., gene, peptide, and set of peptides) to generate balanced reactions. Merging the *E. coli* metabolic network reconstruction with the gene expression reconstruction required a conversion of the Boolean Gene-Protein-Reaction associations (GPRs) into protein complexes. We utilized EcoCyc's annotation to map gene sets to functional enzyme complexes. The content of the final reconstruction is detailed in Supplementary Tables S1, S9, and S10.

Coupling constraint formulation and imposition

Coupling constraints provide a mechanism for linking the flux values of one or more reactions in the ME-Model. For example, they were used to bound the number of proteins that may be translated from an mRNA before the mRNA decays or is transmitted to a daughter cell. They are also the mechanism through which we related enzyme abundance and activity. Often, the coupling constraints are a function of the organism's growth rate (μ). The coupling constraints are a set of inequality constraints appended to the stoichiometric matrix as additional rows. Assumptions and literature citations for all parameters used can be found in Supplementary information.

Optimization procedure

As the demand reactions and coupling constraints are functions of the organism's growth rate (μ), growth-rate optimization is not a linear program (LP) as in metabolic models, which rely on a linear biomass objective function. Instead, to optimize for growth rate, we solve a sequence of LPs to search for the maximum growth rate, μ^* , that still results in a feasible LP. This search for μ^* is accomplished through a binary search; the search procedure is slightly different depending on whether the cell is proteome-limited (Janusian and Batch growth modes) or SNL. Detailed traces of the execution of the optimization procedures can be found in Supplementary information.

Hierarchical clustering

For Figure 4B, relative fractional proteome mass was calculated for each gene-enzyme pair. If a gene is present in multiple enzyme

complexes, then it is represented twice, and all subunits of an enzyme complex are counted separately. To filter out the stochastic expression of alternative isozymes (to make the observed trends clear), we eliminated gene-enzyme pairs that were not expressed across all growth rates and filtered gene-enzyme pairs that changed in relative expression by >0.3 across more than one pair of consecutive growth rates. Hierarchical clustering was performed on the resulting expression profiles; we used a signed power ($\beta=6$) correlation similarity (as in Langfelder and Horvath, 2008) and average agglomeration.

File formats and accessibility

The model is freely available as part of the openCOBRA Project (<http://opencobra.sourceforge.net>).

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Thorsten Koch, Matthias Miltenberger, Ambros Gleixner, Michael Saunders, and Martina Ma for help solving linear programming problems. We thank Adam Feist, Harish Nagarajan, and Pep Charusanti for invigorating discussions. EJO and RLC were supported by NIH R01 GM057089. JAL was supported by NIH U01 GM102098. This work was supported in part by the US National Institute of Allergy and Infectious Diseases and the US Department of Health and Human Services through interagency agreement Y1-AI-8401-01. DRH is supported in part by Y1-AI-8401-01. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Author Contributions: JAL, EJO, and BØP conceived the study. JAL and EJO performed the computational analysis. JAL, EJO, and RLC performed model updates. RLC, DRH, and BØP supervised the study. JAL and EJO wrote the manuscript. All authors helped edit the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T (2012) Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput Biol* **8**: e1002575
- Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabasi AL, Oltvai ZN (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci USA* **104**: 12663–12668
- Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* **5**: 593–599
- Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, Ropers D, Geiselmann J (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol* **9**: 634
- Boer VM, Crutchfield CA, Bradley PH, Botstein D, Rabinowitz JD (2010) Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. *Mol Biol Cell* **21**: 198–211
- Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella*, Neidhardt FC (ed), pp 1553–1569. Washington D.C.: ASM Press

- Button DK (1991) Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the Michaelis constant. *Appl Environ Microbiol* **57**: 2033–2038
- Cho BK, Federowicz SA, Embree M, Park YS, Kim D, Pálsson BO (2011) The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res* **39**: 6456–6464
- Cho BK, Knight EM, Pálsson BO (2006) Transcriptional regulation of the fad regulon genes of *Escherichia coli* by ArcA. *Microbiology* **152**: 2207–2219
- Crick F (1973) Project K: 'The Complete Solution of *E. coli*'. In *Perspectives in Biology and Medicine*, Vol. 17, pp 67–70. Baltimore: Johns Hopkins University Press
- Donachie WD, Robinson AC (1987) Cell division: parameter values and the process. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*, Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umberger HE (eds), Vol. I and II, pp 1578–1593. Washington, D.C.: American Society for Microbiology
- Feist AM, Pálsson BO (2010) The biomass objective function. *Curr Opin Microbiol* **13**: 344–349
- Gerosa L, Kochanowski K, Heinemann M, Sauer U (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol* **9**: 658
- Gil R, Silva FJ, Pereto J, Moya A (2004) Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* **68**: 518–537
- table of contents
- Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ (2013) The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput Biol* **9**: e1003091
- Haverkorn van Rijsewijk BR, Nanchen A, Nallet S, Kleijn RJ, Sauer U (2011) Large-scale ¹³C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol Syst Biol* **7**: 477
- Hua Q, Yang C, Oshima T, Mori H, Shimizu K (2004) Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. *Appl Environ Microbiol* **70**: 2354–2366
- Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival BJ, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150**: 389–401
- Kato J, Hashimoto M (2007) Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol Syst Biol* **3**: 132
- Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V et al (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res* **41**: D605–D612
- Kim M, Zhang Z, Okano H, Yan D, Groisman A, Hwa T (2012) Need-based activation of ammonium uptake in *Escherichia coli*. *Mol Syst Biol* **8**: 616
- Klumpp S, Hwa T (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci USA* **105**: 20245–20250
- Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**: 1366–1375
- Koch AL (1997) Microbial physiology and ecology of slow growth. *Microbiol Mol Biol Rev* **61**: 305–318
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559
- Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, Schrimpe-Rutledge AC, Smith RD, Adkins JN, Zengler K, Pálsson BO (2012) In silico method for modelling metabolism and gene product expression at genome scale. *Nat Commun* **3**: 929
- Li GW, Oh E, Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**: 538–541
- Machado D, Costa R, Rocha M, Ferreira E, Tidor B, Rocha I (2011) Modeling formalisms in Systems Biology. *AMB Expr* **1**: 1–14
- Meyenburg KV, Hansen FG (1987) Regulation of chromosome replication. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*, Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umberger HE (eds), Vol. I and II, pp 1555–1577. Washington, D.C.: American Society for Microbiology
- Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* **5**: 323
- Monod J (1949) The growth of bacterial cultures. *Annu Rev Microbiol* **3**: 371–394
- Nahku R, Valgepea K, Lahtvee P-J, Erm S, Abner K, Adamberg K, Vilu R (2010) Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures. *J Biotechnol* **145**: 60–65
- Nanchen A, Schicker A, Sauer U (2006) Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl Environ Microbiol* **72**: 1164–1172
- Neijssel OM, Mattos MJT, Tempest DW (1996) Growth Yield and Energy Distribution. In *Escherichia coli and Salmonella*, Neidhardt FC (ed), pp 1683–1692. Washington D.C.: ASM Press
- Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Pálsson BO (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Mol Syst Biol* **7**: 535
- O'Brien RW, Neijssel OM, Tempest DW (1980) Glucose phosphoenolpyruvate phosphotransferase activity and glucose uptake rate of *Klebsiella aerogenes* growing in chemostat culture. *J Gen Microbiol* **116**: 305–314
- Pirt SJ (1965) The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci* **163**: 224–231
- Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**: 504–508
- Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett Gr, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot–2005. *Nucleic Acids Res* **34**: 1–9
- Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* **3**: 119
- Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. *Science* **336**: 601–604
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**: 1099–1102
- Selinger DW, Wright MA, Church GM (2003) On the complete determination of biological systems. *Trends Biotechnol* **21**: 251–254
- Smith RW, Dean AC (1972) Beta-galactosidase synthesis in *Klebsiella aerogenes* growing in continuous culture. *J Gen Microbiol* **72**: 37–47
- Thiele I, Fleming RM, Bordbar A, Schellenberger J, Pálsson BO (2010) Functional characterization of alternate optimal solutions of *Escherichia coli*'s transcriptional and translational machinery. *Biophys J* **98**: 2072–2081
- Thiele I, Fleming RM, Que R, Bordbar A, Diep D, Pálsson BO (2012) Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS ONE* **7**: e45635
- Thiele I, Jamshidi N, Fleming RM, Pálsson BO (2009) Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol* **5**: e1000312
- Tuller T, Carmi I, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**: 344–354
- Valgepea K, Adamberg K, Seiman A, Vilu R (2013) *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* **9**: 2344–2358

- Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA (2006) Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol* **72**: 3653–3661
- Young R, Bremer H (1976) Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. *Biochem J* **160**: 185
- Zhuang K, Vemuri GN, Mahadevan R (2011) Economics of membrane occupancy and respiration. *Mol Syst Biol* **7**: 500



Molecular Systems Biology is an open-access journal published by the *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported Licence. To view a copy of this licence visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.