Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity

Q. K. Beg*, A. Vazquez[†‡], J. Ernst[§], M. A. de Menezes[¶], Z. Bar-Joseph[§], A.-L. Barabási[‖], and Z. N. Oltvai*[,**]

*Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15261; [†]The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ 08540; [§]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15217; [¶]Instituto de Física, Universidade Federal Fluminense, 24210, Rio de Janeiro, Brazil; and [‖]Department of Physics and Center for Complex Networks Research, University of Notre Dame, Notre Dame, IN 46556

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved May 31, 2007 (received for review November 7, 2006)

The influence of the high intracellular concentration of macromolecules on cell physiology is increasingly appreciated, but its impact on system-level cellular functions remains poorly quantified. To assess its potential effect, here we develop a flux balance model of *Escherichia coli* cell metabolism that takes into account a systems-level constraint for the concentration of enzymes catalyzing the various metabolic reactions in the crowded cytoplasm. We demonstrate that the model's predictions for the relative maximum growth rate of wild-type and mutant *E. coli* cells in single substrate-limited media, and the sequence and mode of substrate uptake and utilization from a complex medium are in good agreement with subsequent experimental observations. These results suggest that molecular crowding represents a bound on the achievable functional states of a metabolic network, and they indicate that models incorporating this constraint can systematically identify alterations in cellular metabolism activated in response to environmental change.

flux balance analysis | metabolic networks | systems biology

An important aim of systems biology is the identification of the organizing principles and fundamental constraints that characterize the function of molecular interaction networks and the limits of an organism's phenotypic diversity (1–3). Flux balance-based modeling approaches, combining the constraints imposed by the metabolic network's structure with, e.g., mass- or energy-conservation principles (3–6), are especially successful in providing experimentally testable predictions on an organism's metabolic flux state and growth rate. A relative shortcoming of these approaches, however, is that they do not take into account the physical and spatial constraints resulting from the cell's unique intracellular milieu (7–9). For example, ≈20–30% of the *Escherichia coli* cytoplasm is occupied by macromolecules, many of them enzymes, whose cytoplasmic concentration cannot be further increased without drastically affecting protein folding, protein–protein association rates, biochemical reaction kinetics, and transport dynamics within a cell (9, 10). This suggests that constraint-based modeling approaches, such as flux balance analysis (FBA) (3, 11), could be improved if we take into account that the enzymes catalyzing each reaction compete for the available cytoplasmic space (12, 13), potentially limiting the attainable flux rates.

Current flux balance-based modeling approaches also have limited ability to predict substrate uptake from the environment. Extensive experimental data indicates that when grown in complex medium bacterial cells use the available substrates either preferentially or simultaneously depending on the growth condition (see, e.g., refs. 14–17). Efforts to model mixed-substrate growth have assumed specific kinetic expressions for substrate uptake and biomass growth rates, and their predictions are formulated in terms of known model parameters (15, 18). Similarly, FBA predictions are based on previous knowledge of the maximum uptake rates in the corresponding medium (the actual variables one aims to predict), and, in contrast to empirical evidence, FBA in itself predicts the simultaneous utilization of all carbon sources from a mixed-substrate growth medium. One way to overcome this deficiency is the superposition of regulatory mechanisms (in the form of mRNA expression signatures) on the FBA model, assessing which substrates are taken up and which are not (19). Yet regulatory mechanisms appear during the course of evolution because they result in a selective advantage for the cell. This selective advantage results from better use of the available resources within the metabolic constraints of the organism. Therefore, the metabolic constraint can be considered as the primary cause, whereas the regulatory processes represent the specific molecular mechanism developed to cope with this constraint. This fact opens the possibility for a FBA model that, after imposing the relevant constraints, predicts the selective advantage of implementing a regulatory mechanism. Here, we develop a modified FBA model that incorporates a solvent capacity constraint for the attainable enzyme concentrations within the crowded cytoplasm. Using this model, we predict the maximum growth rate of *E. coli* MG1655 wild-type and mutant strains on single carbon sources and for the dynamic patterns of substrate utilization from a mixed-substrate growth medium. We test the model predictions by using growth rate measurements and microarray and substrate concentration temporal profiles, and we obtain a good agreement between model predictions and experimental measurements. Taken together, these results suggest that macromolecular crowding indeed imposes a physiologically relevant constraint on bacterial metabolic activity and that incorporating this constraint allows for improved modeling of cell metabolism from system-level principles.

Results

FBA with Molecular Crowding. In the flux balance model of cellular metabolism a cell's metabolic network is mathematically represented by the stoichiometric matrix, S_{mi}, providing the stoichiometric coefficient of metabolite m ($m = 1, \ldots, M$) in reaction i ($i = 1, \ldots, N$) (3, 20), where M and N are the number of metabolites and reactions, respectively. The cell is assumed to be in a steady state, where the concentration of each intracellular metabolite (other than the metabolites that constitute the biomass) remains

Author contributions: Q.K.B. and A.V. contributed equally to this work; Q.K.B., A.V., M.A.d.M., A.-L.B., and Z.N.O. designed research; Q.K.B. and A.V. performed research; Q.K.B. and A.V. contributed new reagents/analytic tools; Q.K.B., A.V., J.E., and Z.B.-J. analyzed data; and Q.K.B., A.V., Z.B.-J., A.-L.B., and Z.N.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: FBA, flux balance analysis; FBAwMC, FBA with molecular crowding.

[‡]To whom correspondence may be addressed. E-mail: vazquez@ias.edu.

**To whom correspondence may be addressed at: University of Pittsburgh, S701 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261. E-mail: oltvai@pitt.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0609845104/DC1.

constant in time. Thus, the stationary reaction rates (fluxes) consuming and producing a metabolite should balance,

$$\sum_{i=1}^{N} S_{mi} f_i = 0, \qquad [1]$$

where f_i denotes the flux of reaction i. The study of the solution space defined by Eq. **1** together with maximum capacity constraints for the uptake rates of extracellular substrates constitutes the basis of FBA (3).

We extend this framework to consider the physical and spatial constraints resulting from the very high intracellular concentration of macromolecules (7–9). Given that the enzyme molecules have a finite molar volume v_i, we can fit only a finite number of them in a given volume V. Indeed, if n_i is the number of moles of the ith enzyme, then

$$\sum_{i=1}^{N} v_i n_i \leq V. \qquad [2]$$

Eq. **2** represents a constraint on the enzyme levels n_i, potentially affecting their maximum attainable values and relative abundance. Dividing by cell mass M, we can reformulate this constraint in terms of the enzyme concentrations $E_i = n_i/M$ (moles per unit mass), resulting in

$$\sum_{i=1}^{N} v_i E_i \leq \frac{1}{C}, \qquad [3]$$

where $C = M/V \approx 0.34$ g/ml is the *E. coli* cytoplasmatic density (21). Eq. **3** imposes a constraint on the maximum attainable enzyme concentrations and, therefore, we refer to it as the enzyme concentration constraint. This constraint is reflected in the metabolic fluxes as well. Indeed, an enzyme concentration E_i results in a flux $f_i = b_i E_i$ over reaction i, where the parameter b_i is determined by the reaction mechanism, kinetic parameters, and metabolite concentrations. Therefore, the enzyme concentration constraint (Eq. **3**) is reflected in the metabolic flux constraint

$$\sum_{i=1}^{N} a_i f_i \leq 1, \qquad [4]$$

where $a_i = Cv_i/b_i$, affecting the maximum attainable fluxes and the flux distribution among different metabolic reactions. From here on, we refer to this mathematical framework as "flux balance analysis with molecular crowding" (FBAwMC). Furthermore, because the coefficient a_i quantifies the contribution to the overall crowding by reaction i we refer to it as the "crowding coefficient of reaction i," or simply "crowding coefficient." Finally, we note that the enzyme concentration constraint is not the only additional constraint that could potentially restrict the metabolic capabilities of *E. coli* (for example, transporter capacities may be similarly limiting). Yet, our aim here is to test the predictive value of a model that assumes that the enzyme concentration constraint is indeed a main factor limiting the maximal metabolic capabilities of *E. coli*.

FBAwMC Predicts the Relative Maximum Growth of *E. coli* Growing on Single Carbon Sources. To examine the validity of macromolecular crowding as a constraint on a cell's metabolic activity, and to test the predictive capability of the FBAwMC framework, we first examined the phenotypic consequences of extracellular substrate availability during growth in single carbon-limited medium with oxygen being in abundance, focusing on the maximum growth rate. The FBAwMC contains as a free parameter the average crowding coefficient $\langle a \rangle$, and the model predictions for the maximum growth rate are proportional to $\langle a \rangle$. We first assumed that $\langle a \rangle$ is a constant independent of the substrates. In this case it is possible to make predictions for the maximum growth rate in different substrates in arbitrary units. To obtain the maximum growth rates in specific units we fit $\langle a \rangle$ to minimize the mean-square deviation between the predicted and measured growth rates, resulting in $\langle a \rangle = 0.0040 \pm 0.0005$ h·g/mmol, in which g is grams of dry weight. We have obtained an independent estimate of a_i for ≈100 *E. coli* enzymes as well [supporting information (SI) Datasets 1 and 2], resulting in values between 10^{-6} and 10^{-1} and most probable values between 10^{-5} and 10^{-2} (in units of h·g/mmol). The obtained $\langle a \rangle$ is, therefore, within the expected range.

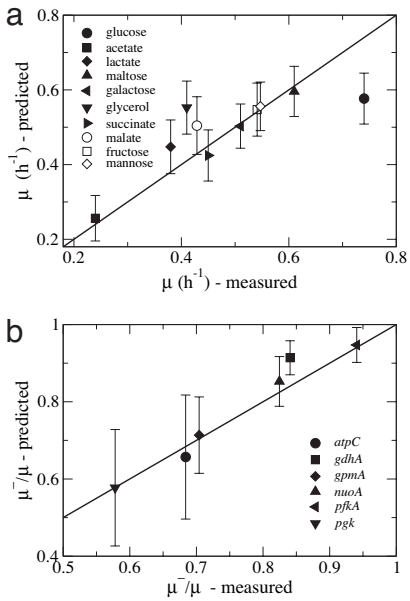

Fig. 1. Predicted and measured maximal growth rates comparison. (*a*) Comparison between the predicted (*y* axis) and measured (*x* axis) growth rates μ of *E. coli* MG1655 grown in M9 minimal medium with various carbon sources. For a perfect match between experiments and theory the symbols should fall on the black diagonal. The symbols indicate the carbon substrate identified in the key. The predicted growth rates were obtained by using $\langle a \rangle = 0.0040$ h·g/mmol (see *SI Text* sections S1 and S2). The error bars represent standard deviation over 1,000 sets of specific a_i parameters. (*b*) Same plot for single gene deletion *E. coli* mutants growing in glucose, the deleted genes being indicated in the key. The mutant growth rates μ^- are given relative to the predicted and measured maximal growth rate μ of wild-type *E. coli* cells growing in glucose-limited medium.

Using the reconstructed *E. coli* MG1655 metabolic network (22) (SI Dataset 1), we first tested the maximal growth rate of *E. coli* MG1655 cells in various single carbon-limited media and compared the results with the theoretically predicted growth rates (Fig. 1*a*). In most cases the line of perfect agreement falls within the standard deviation, implying an overall good agreement between the model predictions and the measured maximum growth rates. For glucose and glycerol, the line of perfect agreement is outside the standard deviation, indicating that our assumption of a substrate-independent $\langle a \rangle$ is not valid for these two substrates. *E. coli* is better adapted to growth on glucose, suggesting a smaller average crowding coefficient than in any of the other carbon sources. Indeed, the average crowding coefficient necessary to obtain a perfect agreement for glucose is smaller: $\langle a \rangle = 0.0031 \pm 0.0001$ h·g/mmol. In contrast, in some carbon-limited media *E. coli* reaches its predicted maximal growth rate only after a period of adaptive evolution (23, 24), suggesting a higher average crowding coefficient before metabolic adaptation. Indeed, the average crowding coefficient neces-

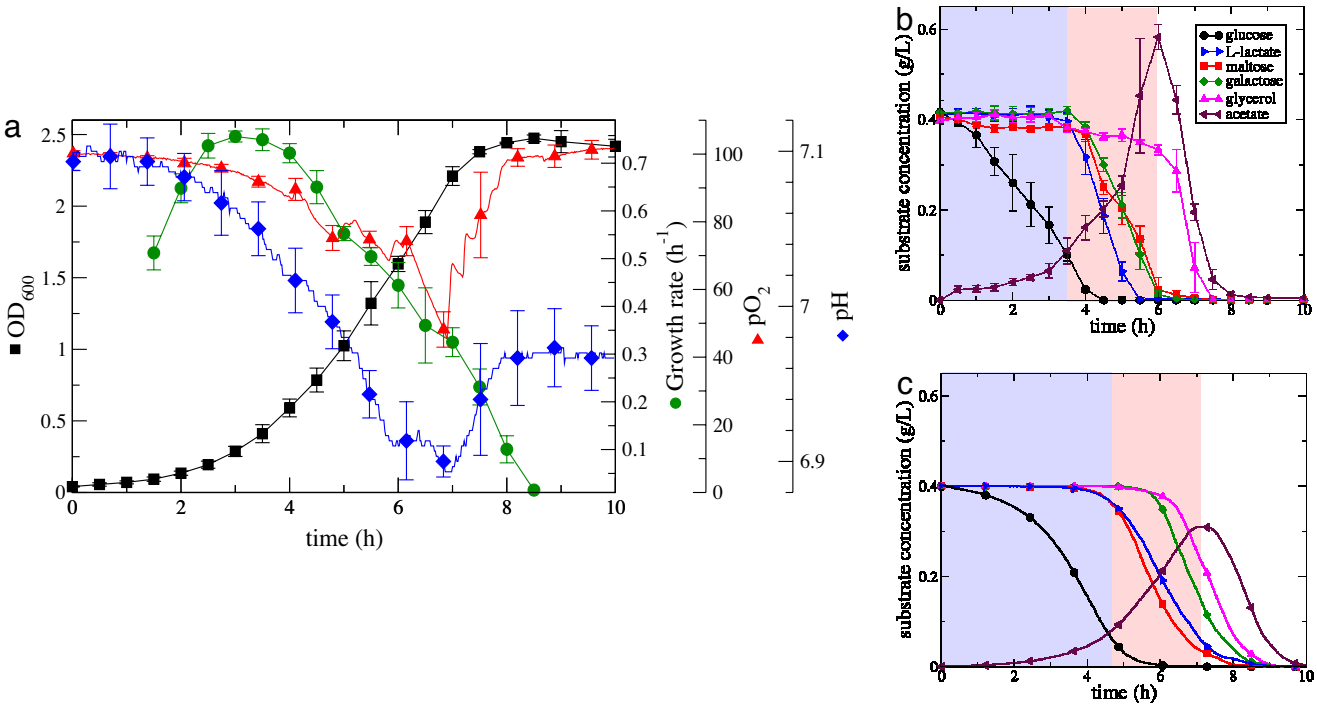

Fig. 2. *E. coli* growth profile and predicted vs. measured hierarchy of substrate utilization. (*a*) The absolute concentration (black curve) and maximal growth rates (green curve) of a batch culture of *E. coli* cells grown in M9 minimal medium containing an equal ratio of glucose, maltose, galactose, glycerol, and lactate are shown, together with the pH (blue curve) and oxygen concentration, pO_2 (red curve). (*b*) The measured concentration of the indicated carbon sources in the growth medium. The growth experiments were performed in triplicate (SI Fig. 8), and means and standard deviations are shown here. The three substrate utilization phases, phase 1 (exclusive glucose), phase 2 (mixed substrates), and phase 3 (glycerol and acetate), are indicated in light blue, purple, and white backgrounds, respectively. (*c*) Predicted substrate uptakes from the growth medium based on the FBAwMC model. The color coding for substrate utilization curves is identical in *b* and *c*, and the error bars represent the standard deviations of the data analyzed from the samples collected from three individual bioreactor runs (SI Fig. 8).

sary to obtain a perfect agreement for glycerol is larger: $\langle a \rangle = 0.0053 \pm 0.0001$ h·g/mmol.

The FBAwMC framework also allows us to predict the maximal growth rate of microbial strains with deleted metabolic enzymes, by simply removing the corresponding metabolic reaction from the FBAwMC model and recomputing the maximal growth rate. To test the power of this predictive capability we experimentally determined the maximal growth rate of several *E. coli* MG1655 single gene deletion mutants grown in glucose-limited medium. As shown in Fig. 1*b*, the agreement between predicted and measured maximal growth rates is remarkably good for various *E. coli* mutants, providing further evidence for the validity of our approach. It is worth noting that, as with FBA alone (25, 26), this analysis is not limited to single-enzyme mutants, but can be carried out for any combination of two or more enzyme deletions as well.

Substrate Hierarchy Utilization by *E. coli* Cells Growing in Mixed Substrates. Extensive experimental data indicate that when grown in complex media bacterial cells use the available substrates either preferentially or simultaneously (see, e.g., refs. 14, 15, and 17), depending on the growth conditions. To further assess the role of an enzyme concentration limit on cellular metabolism we next examined the substrate utilization of *E. coli* cells in a mixed carbon-limited medium, and we compared the results to the FBAwMC *E. coli* model-predicted substrate uptake and utilization (Fig. 2). We grew *E. coli* MG1655 for 12 h in a batch culture containing an equal concentration (0.04% each) of five different carbon sources (galactose, glucose, maltose, glycerol, and lactate) (Fig. 2*a* and SI Fig. 8). These substrates are taken up by *E. coli* through substrate-specific transport mechanisms and enter the central carbon metabolism through various substrate intermediates (Fig. 3 metabolic pathways). Note, that in single carbon-limited medium, maximum growth rates of *E. coli* in glucose (0.74 h^{-1}) was higher; whereas the experimentally measured maximal growth rates in glycerol (0.41 h^{-1}) and lactate (0.38 h^{-1}) were lower than the model predictions (Fig. 1*a*). In contrast, the maximal growth rates obtained on maltose (0.61 h^{-1}) and galactose (0.51 h^{-1}) were in good agreement with the FBAwMC-predicted values (Fig. 1*a*).

As typically seen in batch culture, initially *E. coli* cells showed minimal growth (lag phase) followed by rapid population expansion between 2 and 8 h (exponential growth phase) with no further growth afterward (stationary phase) (Fig. 2*a*). Parallel with this, the growth rate rapidly increased with the start of the logarithmic growth phase, reaching its maximum between 3 and 3.5 h. Thereafter the growth rate steadily declined, becoming negligible to zero after 8 h (Fig. 2*a*).

Of the five supplied carbon sources, in the first 3.5 h of growth only glucose was used (phase 1); it was depleted from the medium within the first 4 h (Fig. 2*b*). This "exclusive glucose use" phase coincided with the initial explosive growth and the maximal attained growth rate of the culture (Fig. 2*a*). At 3.5–4 h *E. coli* cells started to use all four remaining carbon sources, albeit at different rates. Galactose, lactate, and maltose were preferentially used during the next 2 h (phase 2), all three of them being depleted from the growth medium by the sixth hour (Fig. 2*b*). During this "mixed carbon utilization" phase lactate was used up at the fastest rate, followed by maltose and galactose. A small amount of glycerol was also taken up during this time interval, but its predominant utilization occurred only after 6 h, and it was completely depleted from the medium by 7.5 h (Fig. 2*b*). The concentration of acetate, a well known byproduct of rapid *E. coli* aerobic growth (27, 28), increased steadily, reaching its peak concentration in the growth medium at 6 h of growth. Thereafter, the process was reversed, and acetate, along with glycerol, was rapidly consumed and was depleted from

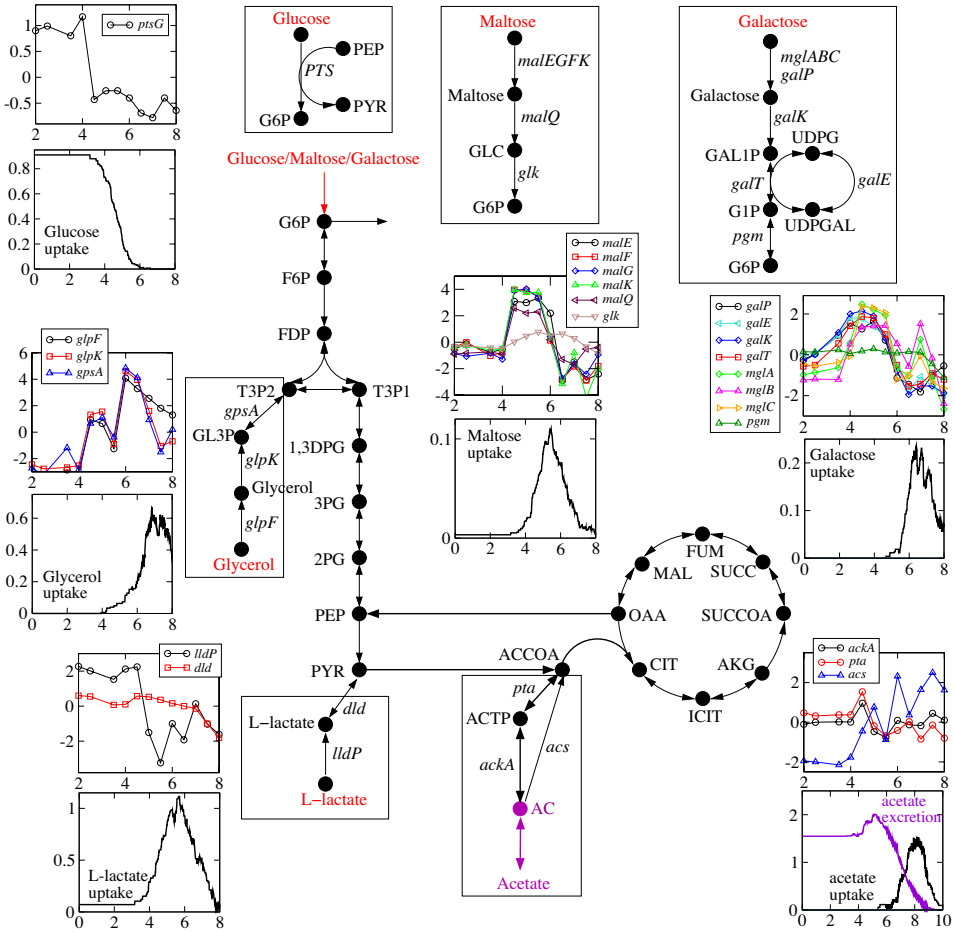

Fig. 3. Comparison between the gene expression profiles and predicted substrate uptake rates. The upper panel in each pair of graphs represents the measured relative gene expression profiles as a function of time (in hours). The lower panels represent the predicted substrate uptake profiles (mmol/min·g) also as a function of time (in hours). Of the carbon sources present in the original growth medium (red), the uptake and entry points of glucose, maltose, galactose, glycerol, and lactate into the *E. coli* glycolytic pathway and citric acid cycle are shown in the diagrammed metabolic pathways. Acetate (purple) is initially produced and later consumed by *E. coli* cells growing in batch culture. All other substrates are shown in black, and the genes encoding various enzymes catalyzing the transport and degradation of intermediary substrates are also italicized. The description of genes responsible for uptake and utilization of listed carbon sources, their biological roles, and description of substrate entry mechanisms are detailed in SI Table 1. mRNA expression profiles of genes encoding metabolic transporters and enzymes specifically involved in galactose, glucose, glycerol, lactate, acetate, and maltose metabolism are shown. The detailed profiles and the full microarray data are presented in the SI Dataset 3. Gene expression values (on *y* axis) in the time series microarray data are the calculated fold changes for each time point relative to the geometric mean of the hybridization intensity of all time points for each gene and are expressed as $\log_2$. The low values represent lower gene expression, and higher values represent higher gene expression.

the medium by 8 h (Fig. 2*b*), denoting a "late carbon utilization" phase in the culture (phase 3). Of note, as single carbon source, acetate provides a lower maximal growth rate (0.24 h^{-1}) than any of the five supplied carbon sources (Fig. 1*a*). Taken together, the sequential order of carbon substrates' uptake in the batch culture experiment only partially correlates with the maximal growth rate they individually provide: it appears earlier for lactate and later for maltose and glycerol.

Subsequently, we tested FBAwMC *E. coli* model on the mixed-substrate conditions. In contrast with FBA (3, 11), which predicts the simultaneous utilization of all carbon sources, we find a remarkably good correlation between the mode and sequence of FBAwMC-predicted and measured substrate uptake and consumption (Fig. 2 *b* and *c*). There are, however, two notable differences. First the FBAwMC predicts a lesser excretion of acetate. In turn the substrates are consumed faster *in vivo* (Fig. 2*b*), because a larger fraction of the carbon source is diverted toward the excretion of acetate. As a consequence the different phases of substrate consumption are shifted to the right (longer times) for the model predictions. The second major discrepancy is the delayed consumption of galactose in the model predictions (Fig. 2*c*). Yet, overall FBAwMC correctly predicts the existence of the three experimentally observed phases of substrate consumption: initial consumption of glucose, intermediate mixed-substrate consumption, and late consumption of glycerol and acetate.

As surrogate markers of cellular metabolism, during the batch culture experiments we also traced the changes in pH and oxygen concentrations in the growth medium. We observed a steady decline in pH during the first 6 h, followed by a slight increase then decrease between 6 and 7 h, and finally an increase between 7 and 8 h (Fig. 2*a* and SI Fig. 8). There was also an accelerating decline in the dissolved oxygen concentration (pO_2) in the medium during the first 7 h, followed by a rapid stepwise increase during the next 30 min. However, the decline phase (indicating aerobic respiration in an increasingly acidic environment because of acetate excretion) was consistently interrupted by rapid upswings in pO_2 concentration (Fig. 2*a* and SI Fig. 8). These spikes indicate brief pauses in aerobic metabolism likely due to switches in predominant substrate use. Indeed, the first of these spikes, at ≈4 h, correlates with the depletion of glucose and initiation of mixed-substrate utilization;

the second, at ≈5 h, with the depletion of lactate and increased utilization of maltose; and the third, at ≈6 h, with the start of joint glycerol and acetate utilization. Similarly, the first rapid increase in oxygen concentration at ≈7 h correlates with the near-depletion of glycerol, followed by a final increase after 30 min corresponding with the depletion of acetate from the medium (Fig. 2*b*).

The Mode and Sequence of Substrate Utilization Correlate with the Expression of Genes Participating in the Uptake Modules. We also prepared mRNA from samples obtained at 30-min intervals between 2 and 8 h and processed them for microarray analysis presented as SI Dataset 3. At the level of substrate uptake pathways (Fig. 3) it is evident that the expression of *ptsG*, the gene encoding the glucose transporter PtsG/Crr, was at a high level from the first time point up until the depletion of glucose from the growth medium, and *ptsG* expression was rapidly turned off afterward. Similarly, the expression of the gene (*lldP*) encoding the lactate transporter, LldP, was high during the first 4.5 h of growth and was turned off rapidly thereafter, in agreement with the earlier than predicted utilization of lactate. In contrast, the expression of gene products responsible for maltose and galactose uptake and utilization were turned on much later and peaked at 4.5–5.5 h, corresponding with the period of their maximal uptake. The expression of gene products responsible for glycerol uptake and utilization peaked in two waves, the smaller one between 4.5 and 5 h and the larger one at 6.5–7 h, the latter corresponding to maximal glycerol consumption from the medium (Fig. 2*b*). Finally, the expression of *acs*, whose gene product catalyzes acetate uptake toward the citric acid cycle, peaked between 6 and 8 h of growth, corresponding to the maximal uptake of previously secreted acetate from the growth medium. We note that all these changes are in good agreement with the FBAwMC model-predicted uptake of the corresponding substrates (Fig. 3, black tracings).

To assess the quality of the microarray profiles and to identify genes with expression patterns that are similar to those of genes encoding enzymes of the uptake pathways we used TimeSearcher (29). We find that most genes displaying expression patterns similar to those of the query genes are colocalized with them in the same operon (SI Figs. 10–15). For example, for the maltose uptake module genes (*malEFGK*, *malQ*, and *glk*), TimeSearcher identified several other genes (*lamB*, *malM*, *malP*, *malS*, and *malZ*) with similar expression profiles. These genes are part of various operons within the maltose regulon (30), although not all of them directly participate in maltose uptake. Similarly, for glycerol metabolism several related glycerol utilization genes (*glpA*, *glpB*, *glpC*, *glpD*, *glpQ*, and *glpT*) displayed expression patterns that were similar to those of the three genes responsible for glycerol uptake (*glpF*, *glpK*, and *gpsA*). The products of these genes are part of the pathway for glycerol catabolism after its uptake.

Activation of Stress Programs upon Switching Metabolic Phases. To assess the global state of *E. coli* transcriptome during the various metabolic phases of the time course experiment, we used three different data analysis methods to analyze the full microarray data. These methods included hierarchical clustering with optimal leaf ordering (31, 32) (Fig. 4), principal component analysis (PCA) (33) (SI Fig. 9), and a probabilistic clustering method based on hidden Markov models (HMMs) (34) (SI Figs. 19–26). It is evident that during the exclusive glucose utilization phase there are similar expression profiles in all samples collected between 2 and 3.5 h, followed by transition in the transcriptome state at the beginning of mixed utilization phase after 4 h (Fig. 4). Within the latter phase, the up-regulation of, e.g., the genes of the maltose regulon is clearly evident (Fig. 4).

Samples obtained during the mixed-substrate utilization phase (5 and 5.5 h) and the late carbon utilization phase (6.5 h) display similar global expression profiles (Fig. 4), interrupted by a significant alteration in the expression profile at 6 h that denotes the

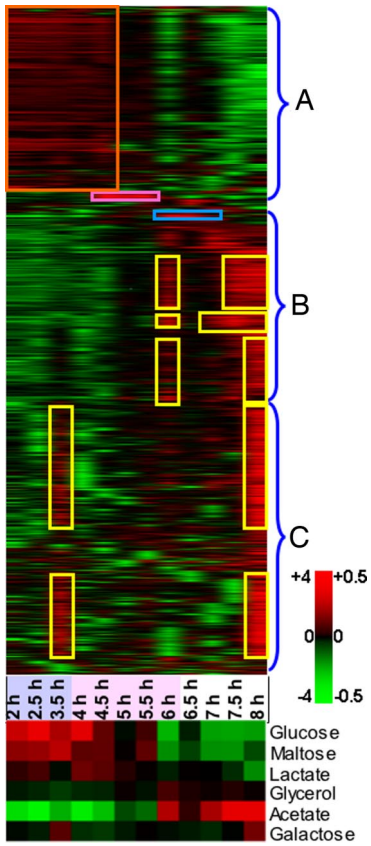

Fig. 4. Analysis of microarray expression data. (*Upper*) Hierarchical clustering with optimal leaf ordering (31) identifies three major expression modes. Relative gene expression values from the highest (red) to the lowest (green) are shown, as indicated by the left side of the color scale bar (+4 to −4). Expression mode A: genes that are up-regulated until 4.5 h (red box). Expression mode B: genes with peak expression at 6 h and after 7.5 h. Expression mode C: genes with peaks at 3.5 h and after 7.5 h. See the GO analysis of these three expression modes in SI Figs. 16–18. The purple and blue boxes indicate up-regulation of maltose and glycerol regulons, respectively. The temporal order of the three phases of substrate utilization is shown in light blue, purple, and white shades (as in Fig. 2). (*Lower*) Matrix comparing the overall correlation of expression profiles at the given time points with that of obtained in mid-logarithmic batch cultures of the indicated single carbon-limited media, as indicated by the right side of the color scale bar [+0.5 (red–high) and −0.5 (green–low)].

switch from acetate secretion to acetate utilization (Fig. 3). Interestingly, the transcriptome at 6 h displays substantial similarity to that characterizing *E. coli* cells at the (near) exhaustion of all substrates from the medium (7–8 h), a phase that is characterized by generic stress response (Fig. 4 and SI Fig. 27). Similarly, many of the genes up-regulated at the end of the last phase are also up-regulated to a lesser extent at 3.5 h, the stage of switching from exclusive glucose utilization to a mixed-substrate utilization phase.

To further characterize the time-point-specific expression profiles, we also prepared mRNA samples from individual mid-logarithmic batch culture *E. coli* cells (OD_{600} ≈ 0.2) grown in glucose-, maltose-, glycerol-, acetate-, lactate, or galactose-limited medium, processed them for microarray analysis (presented as SI Datasets 4 and 5), and compared the obtained transcriptome profiles with those of the individual time points (Fig. 4 *Lower*) from mixed-substrate experiment. It is evident that the transcriptome profiles during the glucose-only and mixed-substrate utilization phases display the highest correlation to that of glucose- and maltose-limited cultures, especially at the initial time points, whereas the late carbon utilization profiles are most similar to that of glycerol- and, especially, acetate-limited cultures. Highly notable

is the transcriptome profile of galactose-limited cultures, which shows some similarity to that of cells at the stage of switching from exclusive glucose utilization to a mixed-substrate-utilization phase (3.5 h), and an even higher similarity to the transcriptome profiles of cells when all carbon sources are depleted (8 h). Thus, *E. coli* cells display a partial adaptation/stress response at each major metabolic transition, followed by a generic stress response (SI Fig. 27) and implementation of a foraging program (35) at complete exhaustion of all extracellular substrates that seems to be most primed for acetate and galactose catabolism.

Discussion

A key aim of systems biology is the identification of the organizing principles and fundamental constraints that characterize the function of molecular interaction networks, including those that define cellular metabolism. In the present work we have focused on the identification of principles that define the growth and substrate utilization mode of bacterial cells in complex environments. Our experimental results indicate the occurrence of three major metabolic phases during the growth of *E. coli* on one type of mixed-substrate medium. Glucose, which by itself provides the highest growth rate, is preferentially used by *E. coli*, followed by simultaneous utilization of maltose, L-lactate, and galactose. Glycerol and (secreted) acetate are used at a third and final stage of growth. In addition, global mRNA expression data indicate that the organism-level integration of cellular functions in part involves the appearance of partial stress response by *E. coli* at the boundaries of major metabolic phases, and, as previously shown (35), the activation of a foraging program upon exhaustion of substrates from the growth medium (Fig. 4).

The simulation results show that the FBAwMC model introduced here successfully captures all main features of the examined metabolic activities. First, there is a significant correlation between *in vivo* relative maximal growth rates of *E. coli* in different carbon-limited media and the *in silico* predictions of the FBAwMC (Fig. 1). Second, the FBAwMC model predicts remarkably well the existence of three metabolic phases and hierarchical mode (i.e., single- or mixed-substrate utilization) of substrate utilization in mixed-substrate growth medium (Figs. 2–4). In essence, our modeling approach indicates that when *E. coli* cells grow in conditions of substrate abundance their growth rate is determined by the solvent capacity of the cytoplasm; vice versa, the solvent capacity should be saturated at the maximal growth rate. Therefore, when growing in a mixture of abundant carbon sources *E. coli* cells should preferentially consume the carbon source resulting in the highest growth rate. At solvent capacity saturation, the synthesis of metabolic enzymes for the utilization of a second, less efficient, carbon source can take place only at the expenses of degrading metabolic enzymes involved in the consumption of the more efficient carbon source. However, this would result in a growth rate reduction and, therefore, cells preferentially using the more efficient carbon source would outgrow those that allow the simultaneous utilization of other carbon sources.

We observe, however, two discrepancies of the FBAwMC model predictions: (*i*) a higher than predicted amount of secreted acetate in the growth medium, and (*ii*) a somewhat earlier uptake and consumption of various substrates from the medium compared with that predicted by the model. The first discrepancy is likely rooted on the contribution of other cell components apart from metabolic enzymes. With increasing growth rate higher concentrations of ribosomal proteins, mRNA, and DNA are required in addition to metabolic enzymes (36). This observation indicates that the FBAwMC model may underestimate the impact of macromolecular crowding and the resulting excretion of acetate. The second discrepancy is quite likely a consequence of the first one, as acetate secretion is generally correlated with an increased carbon source uptake rate (27).

Taken together, our results show that *in silico* models incorporating flux balance and other physicochemical constraints can capture increasingly well the metabolic activity of bacterial cells, and that the maximum enzyme concentration is a key constraint shaping the hierarchy of substrate utilization in mixed-substrate growth conditions. Yet, while the metabolic capabilities of a cell are limited by such constraints, in reality any change in metabolic activity is controlled by regulatory mechanisms evolved in the context of these constraints. Therefore, constrained optimization approaches are also expected to help us better understand and/or uncover regulatory mechanisms acting in *E. coli* and other organisms.

Materials and Methods

Mathematical Framework. The FBAwMC modeling framework has been established, as described in *Results* and as detailed in *SI Text*, S1 and S2.

Growth Experiments, Carbon Substrate, and Microarray Analyses. The *E. coli* K12 strain MG1655 (F^- λ^- *ilvG rfb50 rph1*) was used throughout the work. Isogenic *E. coli* mutants (*pgk*, *atpC*, *gpmA*, *nuoA*, *gdhA*, and *pfkA*) were obtained from F. Blattner (University of Wisconsin, Madison) (37). The experimental details of the growth rate measurements, substrate concentration assays and microarray analyses are detailed in *SI Text*, S3–S12.

We thank N. Gerry and M. Lenburg (Boston University, Boston, MA) for their help with microarray experiments, and an anonymous reviewer for comments on the manuscript. Research at the University of Notre Dame and at the University of Pittsburgh was supported by the National Institutes of Health Grant U01 AI070499.

1. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ (2004) *Curr Opin Struct Biol* 14:300–306.
2. Barabási A-L, Oltvai ZN (2004) *Nat Rev Genet* 5:101–113.
3. Price ND, Reed JL, Palsson BO (2004) *Nat Rev Microbiol* 2:886–897.
4. Segre D, Vitkup D, Church GM (2002) *Proc Natl Acad Sci USA* 99:15112–15117.
5. Beard DA, Liang SD, Qian H (2002) *Biophys J* 83:79–86.
6. Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V (2006) *Biophys J* 90:1453–1461.
7. Ellis RJ (2001) *Trends Biochem Sci* 26:597–604.
8. Hall D, Minton AP (2003) *Biochim Biophys Acta* 1649:127–139.
9. Minton AP (2005) *J Pharm Sci* 94:1668–1675.
10. Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S (1999) *J Bacteriol* 181:197–203.
11. Reed JL, Palsson BO (2004) *Genome Res* 14:1797–1805.
12. Brown GC (1991) *J Theor Biol* 153:195–203.
13. Heinrich R, Schuster S (1996) *The Regulation of Cellular Systems* (Chapman & Hall, New York).
14. Harder W, Dijkhuizen L (1982) *Philos Trans R Soc London B* 297:459–480.
15. Egli T (1995) *Adv Microb Ecol* 14:305–386.
16. Raamsdonk LM, Diderich JA, Kuiper A, van Gaalen M, Kruckeberg AL, Berden JA, Van Dam K (2001) *Yeast* 18:1023–1033.
17. Baev MV, Baev D, Radek AJ, Campbell JW (2006) *Appl Microbiol Biotechnol* 71:323–328.
18. Zinn M, Witholt B, Egli T (2004) *J Biotechnol* 113:263–279.
19. Covert MW, Palsson BO (2002) *J Biol Chem* 277:28058–28064.
20. Schilling CH, Palsson BO (1998) *Proc Natl Acad Sci USA* 95:4193–4198.
21. Zimmerman SB, Trach SO (1991) *J Mol Biol* 222:599–620.
22. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) *Genome Biol* 4:R54.
23. Fong SS, Palsson BO (2004) *Nat Genet* 36:1056–1058.
24. Ibarra RU, Edwards JS, Palsson BO (2002) *Nature* 420:186–189.
25. Thiele I, Vo TD, Price ND, Palsson BO (2005) *J Bacteriol* 187:5818–5830.
26. Deutscher D, Meilijson I, Kupiec M, Ruppin E (2006) *Nat Genet* 38:993–998.
27. El-Mansi EM, Holms WH (1989) *J Gen Microbiol* 135:2875–2883.
28. Reiling HE, Laurila H, Fiechter A (1985) *J Biotechnol* 2:191–206.
29. Hochheiser H, Baehrecke E, Mount S, Shneiderman B (2003) in *Proceedings 2003 International Conference on Multimedia and Expo* (IEEE, Piscataway, NJ), Vol 3, pp III-453–III-456.
30. Balázsi G, Barabási A-L, Oltvai ZN (2005) *Proc Natl Acad Sci USA* 102:7841–7846.
31. Bar-Joseph Z, Gifford DK, Jaakkola TS (2001) *Bioinformatics* 17:S22–S29.
32. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) *Proc Natl Acad Sci USA* 95:14863–14868.
33. Peterson LE (2003) *Comput Methods Programs Biomed* 70:107–119.
34. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z (2007) *Mol Syst Biol* 3:74.
35. Liu M, Durfee T, Cabrera JE, Zhao K, Jin DJ, Blattner FR (2005) *J Biol Chem* 280:15921–15927.
36. Neidhardt FC, Ingraham JL, Schaechter M (1990) *Physiology of the Bacterial Cell: A Molecular Approach* (Sinauer, Sunderland, MA).
37. Kang Y, Durfee T, Glasner JD, Qiu Y, Frisch D, Winterberg KM, Blattner FR (2004) *J Bacteriol* 186:4921–4930.