In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data

Jeremy S. Edwards[1,2], Rafael U. Ibarra[1], and Bernhard O. Palsson[1]*

[1]*Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412.* [2]*Current address: Department of Chemical Engineering, University of Delaware, Newark, DE 19716. *Corresponding author (palsson@ucsd.edu).*

A significant goal in the post-genome era is to relate the annotated genome sequence to the physiological functions of a cell. Working from the annotated genome sequence, as well as biochemical and physiological information, it is possible to reconstruct complete metabolic networks. Furthermore, computational methods have been developed to interpret and predict the optimal performance of a metabolic network under a range of growth conditions. We have tested the hypothesis that *Escherichia coli* uses its metabolism to grow at a maximal rate using the *E. coli* MG1655 metabolic reconstruction. Based on this hypothesis, we formulated experiments that describe the quantitative relationship between a primary carbon source (acetate or succinate) uptake rate, oxygen uptake rate, and maximal cellular growth rate. We found that the experimental data were consistent with the stated hypothesis, namely that the *E. coli* metabolic network is optimized to maximize growth under the experimental conditions considered. This study thus demonstrates how the combination of *in silico* and experimental biology can be used to obtain a quantitative genotype–phenotype relationship for metabolism in bacterial cells.

The complete genome sequence for numerous microbes is currently available[1]. The genome annotation, along with biochemical and strain-specific information, provides the information needed to reconstruct complete metabolic networks of these microbes[2–6]. The reconstructed metabolic networks can be used to analyze, interpret, and predict the metabolic flux distribution of the reconstructed metabolic network[7–11].

The interpretation and prediction of metabolic flux distributions requires mathematical modeling and computer simulation, and there exists a long history of quantitative metabolic modeling[12]. Currently, several well-developed mathematical approaches exist for the analysis of cellular metabolism and its regulation[13–21]. Most of these methods require detailed kinetic and concentration information about enzymes and various cofactors. Even though information about cellular components is growing rapidly, the application of many mathematical modeling methods is hampered by a lack of the required kinetic and enzyme concentration data. The human red blood cell remains as a notable exception[22–24].

To deal with the lack of kinetic information, an alternative approach has been used to study feasible and optimal metabolic flux distributions: flux balance analysis (FBA)[25–30]. FBA can be used to analyze the capabilities of a reconstructed metabolic network solely on the basis of the systemic mass-balance and reaction capacity constraints[31] (Fig. 1). As a result of the incomplete set of constraints on the metabolic network (that is, kinetic constant constraints and gene expression constraints are not considered), FBA does not yield a unique solution for the flux distribution. Rather, FBA provides a solution space that contains all the possible steady-state flux distributions that satisfy the applied constraints. Subject to the imposed

constraints, optimal metabolic flux distributions can be determined from the set of all allowable flux distributions using linear programming (LP)[30,32]. To identify an "optimal" solution, it is necessary to state an objective function. For studies involving microbial cells, maximal cellular growth has been defined as the objective function (see Experimental Protocol and refs 8, 26, 27, 33). Although such optimal solutions can be calculated, we still do not know if these calculated flux distributions represent actual cell behavior. If the predicted and actual metabolic fluxes agree, the data would support the hypothesis that microbial cells use their metabolic networks to achieve their objective: maximal cellular growth rate. Here we address this hypothesis using *E. coli* as a model organism.

Results

Previously, we have reconstructed the *E. coli* metabolic network for qualitative and quantitative *in silico* simulations[8]. We have also developed an approach (phenotype phase plane (PhPP) analysis) to map the optimal utilization of a reconstructed metabolic network as a function of two nutrient uptake rates (see Experimental Protocol and Supplementary Appendix 1 in the Web Extras page of *Nature Biotechnology* Online)[31]. In the present study, we used the *in silico E. coli* metabolic reconstruction and PhPP analysis to design experiments that address the stated hypotheses about optimal cellular growth rates.

Phenotype phase plane analysis. The metabolic network can be examined using FBA. All the metabolic flux vectors (or metabolic phenotypes) attainable by a reconstructed metabolic network are mathematically confined to a region known as the flux cone, and the optimal solution to the linear programming problem will then lie on

RESEARCH ARTICLES

Metabolic reconstruction

Metabolic pathway analysis and phenotype phase plane analysis

Flux balance analysis – linear programming

Figure 1. From genome sequence to metabolic characteristics. The metabolic network can be reconstructed from the annotated genome sequence[2,5,51]. The global properties of the metabolic reconstruction can be studied to determine the feasible steady-state metabolic flux distributions, and this process can be performed with a whole-cell pathway analysis[34,35]. However, this process is computationally intense; thus we can map feasible steady-state metabolic flux distributions using an alternative approach known as phenotype phase plane (PhPP) analysis[7,31,36]. PhPP analysis consists of calculating the optimal solution using linear programming as a function of two fluxes in the metabolic network (by setting the α and β value in equation 1). This process constructs the surface revealed in the figure. Finally, linear programming can be applied to calculate the value of the objective function for specific values of the uptake fluxes; additionally, the optimal value of all other fluxes is calculated[26,29,30]. The results from this process can be compared to experimental data to evaluate the suitability of this modeling framework. Furthermore, this modeling approach can be used to guide the metabolic engineering of industrial microorganisms.

and in Supplementary Appendix 1 in the Web Extras page of *Nature Biotechnology* Online.

Acetate. Optimal growth performance on acetate was investigated *in silico*, and the predictions generated were compared to experimental data. The *in silico* study started with a PhPP analysis, with the acetate and oxygen uptake rates defined as the axes of the two-dimensional projection of the flux cone representing the capabilities of *E. coli* metabolism (Fig. 2A). The flux cone is the region of all admissible steady-state metabolic flux distributions (for a complete description of the flux cone see ref. 35). Furthermore, a three-dimensional projection of the flux cone with the growth rate defined as the third dimension was utilized (Fig. 2B). The *in silico* analysis of the acetate-oxygen PhPP has been described [31]. Briefly, the acetate-oxygen PhPP consists of two phases (Fig. 2A) separated by a line that defines the optimal (with respect to cellular growth) relation between the acetate and oxygen uptake rate, and this line is referred to as the line of optimality (LO).

The PhPP was used to analyze and interpret the operation of the metabolic network. For example, under oxygen limitations the characteristics of the metabolic network may be defined by region 2 of the PhPP (Fig. 2A, B), where the acetate uptake rate exceeds the optimal relation to the oxygen uptake rate. From Figure 2B, it can be seen that if the metabolic network were operating within region 2, the optimal capability to support growth would be increased by reducing the acetate uptake rate to a point on the LO. A similar interpretation can be made for points within region 1, with oxygen and acetate switching roles. Hence, metabolic flux vectors defining a point in region 1 or region 2 would indicate inefficient utilization of the available resources. Thus, the *in silico* PhPP analysis led to the conclusion that if the regulation of the *E. coli* metabolic network has evolved to operate optimally to support growth with acetate as the sole carbon source, the relation between the acetate and oxygen uptake rate and the growth rate should be defined by the LO (Fig. 2A, B).

The relation between the acetate and oxygen uptake rates and the growth rate was experimentally examined by cultivating *E. coli* MG1655 on acetate minimal medium. The acetate uptake rate was experimentally controlled by changing the acetate concentration in the minimal medium. The uptake rates of acetate and oxygen and the growth rate were measured, and the experimental points were plotted on the PhPP (Figs 2 and 3). The calculated optimal relation between the acetate and oxygen uptake rate was then compared to the experimental data (Fig. 2A). Comparison of the experimental data to the *in silico* predictions indicated a 14% difference between the slope (0.91) of the linear regression line for the experimental data and the slope (1.04) of the *in silico*-defined LO for aerobic growth on acetate minimal medium.

The measured and calculated growth rates were plotted as the third dimension above the acetate–oxygen PhPP (Fig. 2B). The color-coded surface represents the three-dimensional projection of

an edge or vertex of the flux cone[34, 35]. Linear programming was used to search the flux cone for a solution that maximizes an objective (here the objective was defined as the growth flux). However, the optimal flux distribution is only meaningful when interpreted in terms of the flux constraints (α and β, see Experimental Protocol) on the transport fluxes. Therefore, phenotype phase planes[7,31,36] have been developed to define the range of optimal flux vectors, and how the optimal flux vector is dependent on the transport fluxes (or the α and β values described in the Experimental Protocol). The methodology for defining PhPPs has been described[31]. The construction of PhPPs is briefly described in the Experimental Protocol

A

B

Figure 2. *In silico* predictions of growth and metabolic functions and comparisons to experimental data. (A) The acetate uptake rate (AUR) (mmol/g DW/h) versus oxygen uptake rate (OUR)(mmol/g DW/h) phenotype phase plane. Red line is the *in silico*-defined line of optimality (LO). The slope of this line is indicated in the figure. The experimental data points are displayed on the figure. The error bars represent a single standard deviation, and the error bars are displayed for both the AUR and the OUR measurements. A linear regression was performed (the dashed line) on the data points to define the experimentally reconstructed LO. The R^2 value for the curve fit is 0.92. Regions 1 and 2 represent nonoptimal metabolic phenotypes. (B) The three-dimensional rendering of the phase surface. The x- and y-axis represent the same variables as in (A). The third dimension represents the cellular growth rate. The z-axis values are color-coded with the optimal growth rate value quantitatively indicated on the corresponding legend. The LO in three dimensions is indicated. The parametric equation of LO in three dimensions is indicated in the text. The black lines define the surface of the metabolic capabilities in the three-dimensional projection of the flux cone and represent constant values of the AUR or OUR. The quantitative effect on cellular growth potential of increasing the AUR (without proportional increase in the OUR) can be visualized. The data points (in blue) are also plotted on the three-dimensional figure and error bars have been omitted.

the flux cone. In other words, the color-coded surface defines the solution space, and all feasible steady-state metabolic flux distributions are confined within the surface. The LO on the two-dimensional phase plane (Fig. 2A) is a projection of the edge on the three-dimensional surface onto the x,y-plane (acetate uptake rate, oxygen uptake rate). The experimental data were plotted in the three-dimensional space (Fig. 2B). To quantitatively visualize the proximity of the data points to the LO in three dimensions, the *in silico* predictions and the experimental data were projected onto each plane formed by the basis vectors.

The projection of the three-dimensional LO and the experimental data points onto the (x,y) (x,z), and (y,z) planes is indicated in Figures 2A, 3A, and 3B, respectively, where the quality of the linear regression is indicated by the correlation coefficient, and the data are compared to the *in silico* predictions. (Note: x-axis is acetate uptake rate; y-axis is oxygen uptake rate; z-axis is growth rate.) The predicted and the observed metabolic fluxes (substrate and oxygen uptake rates and growth rate) for each point were directly compared and the *in silico* predictions and had an overall average error of 5.8% (data available as Supplementary Table 4 in the Web Extras page of *Nature Biotechnology* Online). At this point, we should note that the information used to reconstruct the metabolic network was obtained

independently from the present experiments[8]. The calculated PhPP represents a priori interpretation and prediction of the data obtained in the present study.

Succinate. The succinate–oxygen PhPP (Fig. 4) was more complicated than the acetate–oxygen PhPP. The succinate–oxygen PhPP (Fig. 4A) had four distinct regions of qualitatively distinct optimal metabolic network utilization. Regions 1 and 4 of the succinate–oxygen PhPP were analogous to regions 1 and 2 of the acetate–oxygen PhPP. For example, it can be seen from Figure 4C that the maximal growth flux for a flux vector in region 4 can be increased if the succinate uptake is reduced to a point defined by the region 3, 4 demarcation. Furthermore, from the PhPP analysis, region 3 is defined as a single substrate-limited region. The single substrate-limited region indicates that the succinate uptake rate has little effect on the maximal growth flux in region 3, whereas the oxygen uptake rate has a positive effect on the growth rate.

Region 2 is defined as a dual substrate-limited region, because in region 2 the metabolic network can support an increased growth rate if the succinate or oxygen uptake rate is increased. The *in silico* analysis shows that the cellular growth rate can be increased by operating the metabolic network off of the LO in region 2, by implementing a partially aerobic metabolism and the secretion of a metabolic by-product. The optimal metabolic by-product was calculated to be acetate. The production of a reduced metabolic by-product in region 2, however, reduces the overall biomass yield. Therefore, it is possible to surmise from the PhPP analysis that, if the regulation of the metabolic network were to evolve toward optimal growth with succinate as the sole carbon source, the metabolic network would operate with a flux vector along the LO. However, the growth rate can be increased by moving the flux vector into region 2; thus we expect that the network should only operate in region 2 when oxygen is limited and succinate is plentiful if the stated hypothesis is true.

We carried out *E. coli* growth experiments on succinate minimal M9 media to critically test the hypothesis given in the above *in silico* analysis. Multiple batch cultures were grown at various succinate concentrations

A

B
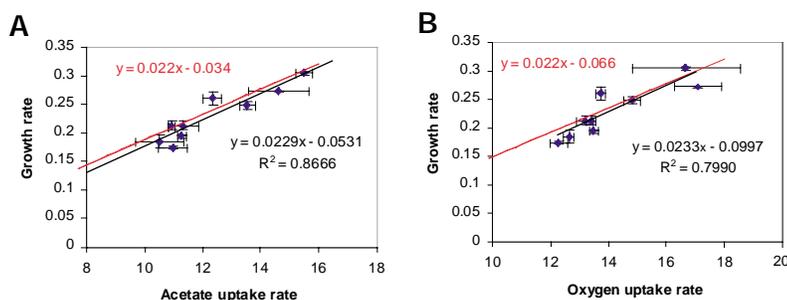
Figure 3. Line of optimality (LO) projected onto each pair of basis vectors. (A) Acetate uptake rate versus growth rate. (B) Oxygen uptake rate versus growth rate. The *in silico*-defined LO is indicated in red with the corresponding regression equation. The data points have also been projected on the respective basis vectors, and a linear regression was performed in the two-dimensional plane to experimentally define the LO

RESEARCH ARTICLES

A

B

C
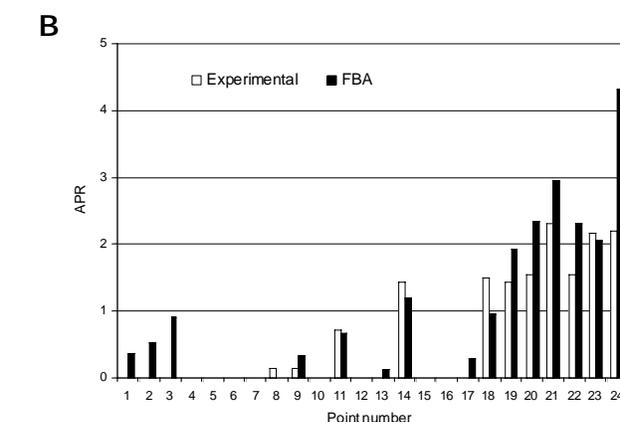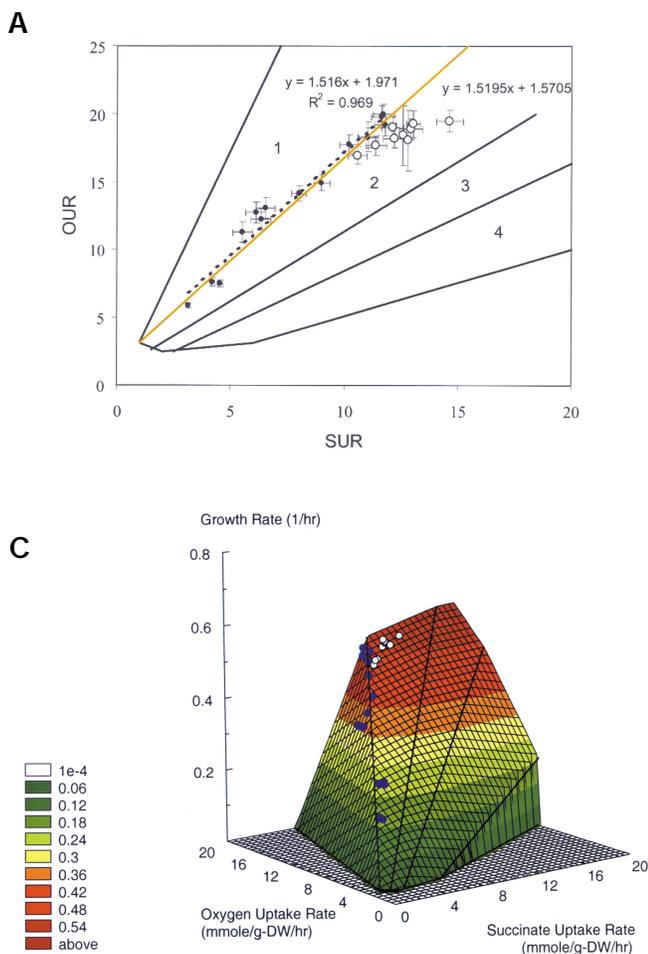
Figure 4. *In silico* predictions of growth and metabolic functions and comparisons to experimental data. (A) The succinate uptake rate (SUR) (mmol/g DW/h) versus oxygen uptake rate (OUR) (mmol/g DW/h) phenotype phase plane. Red line is the *in silico*-defined line of optimality (LO). The experimental data points are displayed on the figure. The error bars are displayed for both the SUR and OUR measurements and represent a single standard deviation. Cultivations for which acetate was produced above a threshold of 0.3 mmol/g DW/h are indicated by open circles; filled circles identify either no acetate production or production below the threshold. The black dotted line represents the linear regression of the data points with no acetate production. (B) The measured acetate production vs. the *in silico* predictions for each point illustrated in (A). The data points are rank-ordered by the SUR. (C) Three-dimensional phenotype phase surface analysis. The x- and y-axis represent the same variables as in (A). The third dimension represents the cellular growth rate. The z-axis values are color-coded with the corresponding legend in the figure. The demarcation lines separating the colored regions represent constant OURs and AURs, and the quantitative effect of moving away from the LO can be visualized. The data points are plotted in this three-dimensional figure with the exception of the error bars.

and temperatures to span a range of succinate uptake rates. The aeration and agitation were held constant to maintain a consistent maximal oxygen diffusion rate in all the cultures. The succinate and oxygen uptake rates and the growth rate were measured separately for each independent growth experiment. The experimental data were then directly compared to the *in silico* predictions (Fig. 4).

The experimental data points were consistent with the stated hypothesis: the flux vector consistently identified points along the LO for oxygen uptake rates below a critical value (~18.8 mmol/g DW/h). (Note: g DW is grams dry weight.) Furthermore, the cultures that identified points along the LO produced little or no acetate as a metabolic by-product (as predicted by the *in silico* analysis; see Fig. 4B). As hypothesized, the experimental data indicate horizontal movement of the flux vector within region 2 for the experimental systems that are oxygen limited but have plentiful succinate. The breakpoint in the experimental data was determined to correspond to a maximal oxygen uptake rate of 18.8 ± 0.5 mmol/g DW/h. Flux vectors within regions 3 or 4 were never observed. Acetate production was measured for the cultures identified in region 2, and quantitatively compared to the *in silico* predictions in Figure 4B.

The optimal growth rate surface was constructed over the succinate–oxygen PhPP, and the measured flux vectors fell near the edge of the fluxcone that corresponded to the LO (Fig. 4C). The flux vectors also identified a locus of points on the phase surface in region 2 with a constant oxygen uptake rate equal to the maximal oxygen uptake limit of the system. To quantitatively test the predictive capability of the *in silico* analysis and the *in silico* derived-hypothesis, we employed a piecewise linear model to describe our hypothesis and the experimentally observed flux vectors. The piecewise linear

model is defined as follows: we identified the locus of points defined by the flux vector for a range of succinate uptake rates and an oxygen uptake limit. Below the oxygen uptake limit, the locus of points lies along the LO, and above the oxygen uptake limit the locus of points lies along the phase surface with a constant oxygen uptake rate (the oxygen uptake limit). On the basis of the piecewise linear model, we used the succinate uptake rate to predict the oxygen uptake rate and the growth rate, and considered the other two permutations as well. From this analysis an overall average error between the *in silico* predictions and the experimental data was 10.7% (data available as Supplementary Table 5 in the Web Extras page of *Nature Biotechnology* Online).

Discussion

We have studied the metabolic behavior and capabilities of *E. coli* MG1655 *in silico* and formulated experiments that directly test the optimal growth of *E. coli*. For the considered growth conditions, quantitative predictions of the substrate and oxygen uptake rates, by-product secretion rates, and cellular growth rates were obtained. Two metabolic fluxes corresponding to the substrate uptake rate (acetate/succinate and oxygen uptake) and the growth flux were chosen to define the three-dimensional phenotype phase surface. Under the examined growth conditions, the hypothesis that *E. coli* optimizes its growth rate subject to systemic capacity and stoichiometric constraints was consistent with the experimental data. Thus, for the growth conditions considered, it was possible to use an *in silico* metabolic reconstruction to quantitatively interpret metabolic physiology.

The *in silico* approach utilized is a departure from traditional approaches to the detailed modeling of physicochemical systems. Traditional approaches call for the statement of fundamental princi-

ples (such as the diffusion equation and chemical potential differences), upon which a detailed mathematical model is based. In contrast, the approach we take here is to impose known systemic constraints on the function of the entire metabolic network and to study the feasible metabolic behavior within these systemic constraints. Using this approach, we are able to simulate whole-cell metabolic flux in a steady state, but we are unable to determine the metabolic concentrations or track the system dynamics. Despite the limitations of the approach, our results indicate that the growth of *E. coli* in acetate and succinate minimal media is consistent with the optimal utilization of the metabolic network subject to the stoichiometric and capacity constraints. In the future as more detailed and quantitative biochemical information becomes available, we will be able to state additional unambiguous constraints on cellular metabolism to further restrict the allowable behavior of the integrated function of biological networks. However, in spite of the impressive advances in genome-scale experimental approaches and the rapid development of bioinformatics, it is still uncertain whether in the foreseeable future we will have all of the necessary information to construct detailed kinetic models based on fundamental principles.

Although this study represents an important step that experimentally verifies *in silico* numerical predictions generated on the basis of genomic, biochemical, and strain-specific data, we expect that further experimental work will be needed to examine the generality of the optimal growth behavior of *E. coli* and the computational approach used. Here we have measured the exchange fluxes, and further experiments are needed to determine the consistency between the optimal and actual intracellular fluxes. "Genome-scale" measurements of intracellular metabolic fluxes are currently difficult to obtain; nonetheless, to fully address the predictive capability of the *in silico* analysis, intracellular measurements will ultimately be required[37,38]. Despite the existing lack of intracellular flux measurements, other measurements indicative of the intracellular fluxes will provide essential clues for deducing the predictive capability of *in silico* analysis and in addition will reveal a greater understanding of how the metabolic network operates. Advances in genomic technologies, such as DNA microarrays[39,40], GeneChips[41], and proteomics[42–44] will undoubtedly be useful, making it possible to assess predicted shifts in pathway utilization[45].

The work described here should be considered a step toward the integrative analysis of bioinformatic databases to predict and understand cellular function on the basis of the underlying genetic content. More extensive studies are required to assess the generality of the results and conclusions presented.

Experimental protocol

Strains and media. *Escherichia coli* MG1655 (American Type Culture Collection No. 47076) was used for all of the experiments. The *E. coli* MG1655 annotated genome sequence and the biochemical literature[46] were used to construct the *in silico E. coli* strain as described[8]. The experiments were carried out in M9 minimal media[47] with the addition of acetate or succinate as the carbon source. Cellular growth rate was varied by changing the conditions, both the carbon source concentration (0.05–4 g/L) and the temperature (27.5–37°C).

Batch cultivation. Batch reactors were set up at two different volume scales. 1 L cultures were carried out in 1.5 L Erlenmeyer flasks with aeration, and the large-volume batch cultures were used to continuously monitor the oxygen uptake rate (OUR) online with an off gas analyzer. To carry out 200 ml cultures, 500 ml Erlenmeyer flasks were used, and for the 200 ml cultures the OUR was monitored online polarographically and by measuring the mass transfer coefficient for oxygen ($k_l a$). The temperature was controlled using a circulating waterbath (Haake, Berlin, Germany). All measurements and data analysis were restricted to the exponential phase of growth. The biomass (see Supplementary Table 3 in the Web Extras page of *Nature Biotechnology* Online) and the concentration of the substrate (acetate or succinate) in the medium were monitored throughout the experiment.

Analytical procedures. Cellular growth was monitored by measuring the

optical density (OD) at 600 nm and 420 nm and by cell counts (Coulter Electronics Inc., Hialeah, FL). OD to cellular dry weight correlations were determined by two different measurements: (1) spun-down cells were dried at 75°C to a constant weight, and (2) 15–25 ml (taken throughout the culture) samples were filtered and dried to a constant weight. The concentration of metabolites in the culture medium was determined by high-performance liquid chromatography (HPLC; Rainin Instruments Co. Inc., Woburn, MA). An Aminex HPX-87H ion exchange carbohydrate-organic acid column (Bio-Rad Laboratories, Hercules, CA) (at 66°C) was used with degassed 5 mM H_2SO_4 as the mobile phase and UV detection. The dissolved oxygen in the culture was monitored using a polarographic dissolved oxygen probe (Cole-Parmer Instruments Co., Vernon Hills, IL). Oxygen consumption was measured by three different methodologies: (1) passing the effluent gas through a 1440C Servomex oxygen analyzer (Servomex Co., Inc. Norwood, MA); (2) calculating from the dissolved oxygen reading and $k_l a$ measurements; and (3) in a respirometer chamber in a separate 50 ml flask. All three methods used for measuring the OUR gave similar and reproducible results.

Flux balance analysis. The metabolic capabilities of the defined *E. coli* MG1655 metabolic genotype were assessed computationally using FBA as described in the literature[26,29,30] (an online FBA primer is available as Supplementary Appendix 1 in the Web Extras page of *Nature Biotechnology* Online; also see http://gcrg.ucsd.edu). FBA can be used to analyze metabolic systems by defining a series of constraints on the metabolic network. The stoichiometric constraints and the capacity constraints were considered herein. The FBA equation that defines the stoichiometric (or mass balance constraints) is $\mathbf{S} \bullet \mathbf{v} = \mathbf{0}$, where $\mathbf{S}$ is the stoichiometric matrix and the vector $\mathbf{v}$ defines the metabolic fluxes (includes internal and external fluxes). Capacity constraints were placed on the value of each flux in the metabolic network. The capacity constraints were utilized to set the uptake rate for the transport reactions and define the reversibility of each metabolic reaction. Capacity constraints were defined for transport reactions. The phenotype phase planes were constructed by varying the capacity constraints on the carbon source and oxygen uptake rate. Transport fluxes for metabolites not available in the media were always restricted to zero. Additionally, the transport of inorganic phosphate, ammonia, carbon dioxide, sulfate, potassium, and sodium were unconstrained, and metabolic by-products were always allowed to leave the metabolic system. The capacity constraints were also used to define the reaction reversibility; for example, α_i for the internal fluxes was set to zero for all irreversible fluxes, and all reversible fluxes were unbounded.

The determination of a particular metabolic flux distribution was formulated as a linear programming problem, in which the solution that maximizes an objective function was identified. The linear programming problem was formulated as shown below:

Minimize Z subject to: $\mathbf{S} \bullet \mathbf{v} = \mathbf{0}$
$$\alpha_i \le v_i \le \beta_i$$

where

$$Z = \sum c_i \cdot v_i = \langle \mathbf{c} \bullet \mathbf{v} \rangle \qquad (1)$$

The vector $\mathbf{c}$ was used to select a linear combination of metabolic fluxes to include in the objective function[33]. Herein, we have defined cellular growth as the objective function; therefore, $\mathbf{c}$ was defined as the unit vector in the direction of the growth flux, and the growth flux was defined in terms of the biosynthetic requirements as follows:

$$\sum_{all\ m} d_m \cdot X_m \xrightarrow{v_{growth}} Biomass \qquad (2)$$

where d_m is the biomass composition of metabolite X_m (defined from the literature[48]), and the growth flux is modeled as a single reaction that converts all the biosynthetic precursors into biomass.

Phenotype phase plane analysis. All the metabolic phenotypes (metabolic flux distributions) attainable from a defined metabolic genotype are mathematically confined to a flux cone (intersection of the null space[49] and the linear inequalities defined above[34,35]), where each solution in this space corresponds to a particular internal flux distribution or a particular phenotype[30]. The flux cone can be explored in two dimensions by using a phase plane analysis[8,31]. The uptake rates of two nutrients (such as the carbon substrate and oxygen) form two axes on an (x,y)-plane, and the optimal flux map can

RESEARCH ARTICLES

be calculated for all points in this plane. There is a finite number of fundamentally different optimal metabolic flux maps present in such a plane. The demarcations on the phase plane are identified by a shadow price (linear programming dual-variable) analysis[50]. Three-dimensional "surfaces" are constructed by defining the cellular growth rate as the third dimension, and for each point within the phase plane the optimal growth rate was calculated and displayed in the third dimension. The three-dimensional surface quantitatively demonstrates the sensitivity of the objective function to the uptake rate.

Note: Supplementary information can be found on the Nature Biotechnology *website in Web Extras (http://biotech.nature.com/web_extras).*

1. TIGR Microbial database: a listing of published microbial genomes and chromosomes and those in progress. (The Institute for Genomic Research, Rockville, MD, 2000). http://www.tigr.org/tdb/mdb/mdbcomplete.html
2. Karp, P.D. *et al.* The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* **28**, 56–59 (2000).
3. Selkov, E., Jr., Grechkin, Y., Mikhailova, N. & Selkov, E. MPW: the Metabolic Pathways Database. *Nucleic Acids Res.* **26**, 43–45 (1998).
4. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
5. Karp, P.D., Ouzounis, C. & Paley, S. HinCyc: a knowledge base of the complete genome and metabolic pathways of *H. influenzae*. *Proceedings of the ISMB-96 Conference* **4**, 116–124 (1996).
6. Overbeek, R. *et al.* WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* **28**, 123–125 (2000).
7. Edwards, J.S. & Palsson, B.O. Systems Properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416 (1999).
8. Edwards, J.S. & Palsson, B.O. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* **97**, 5528–5533 (2000).
9. Karp, P.D., Krummenacker, M., Paley, S. & Wagg, J. Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* **17**, 275–281 (1999).
10. Goryanin, I., Hodgman, T.C. & Selkov, E. Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics* **15**, 749–758 (1999).
11. Tomita, M. *et al.* E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**, 72–84 (1999).
12. Bailey, J.E. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnol. Prog.* **14**, 8–20 (1998).
13. Fell, D. *Understanding the control of metabolism.* (Portland Press, London; 1996).
14. Domach, M.M., Leung, S.K., Cohn, R.E. & Shuler, M.M. Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A. *Biotechnol. Bioeng.* **26**, 203–216 (1984).
15. Palsson, B.O. & Lee, I.D. Model complexity has a significant effect on the numerical value and interpretation of metabolic sensitivity coefficients. *J. Theor. Biol.* **161**, 299–315 (1993).
16. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
17. Liao, J.C. Modelling and analysis of metabolic pathways. *Curr. Opin. Biotechnol.* **4**, 211–216 (1993).
18. Novak, B. *et al.* Finishing the cell cycle. *J. Theor. Biol.* **199**, 223–233 (1999).
19. Kompala, D.S., Ramkrishna, D., Jansen, N.B. & Tsao, G.T. Investigation of bacterial growth on mixed substrates. Experimental evaluation of cybernetic models. *Biotechnol. Bioeng.* **28**, 1044–1056 (1986).
20. McAdams, H.H. & Arkin, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69 (1999).
21. Palsson, B.O., Joshi, A. & Ozturk, S.S. Reducing complexity in metabolic networks: making metabolic meshes manageable. *Fed. Proc.* **46**, 2485–2489 (1987).
22. Jamashidi, N., Edwards, J., Fahland, T., Church, G. & Palsson, B. A computer model of human red blood cell metabolism. *Bioinformatics,* in press (2001).
23. Lee, I.D. & Palsson, B.O. A Macintosh software package for simulation of human red blood cell metabolism. *Comput. Methods Programs Biomed.* **38**, 195–226 (1992).
24. Mulquiney, P.J., Bubb, W.A. & Kuchel, P.W. Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: in vivo kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using 13C and 31P NMR. *Biochem. J.* **342**, 567–580 (1999).
25. Fell, D.A. & Small, J.A. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *J. Biochem.* **238**, 781–786 (1986).
26. Edwards, J.S., Ramakrishna, R., Schilling, C.H. & Palsson, B.O. In *Metabolic engineering.* (eds Lee, S.Y. & Papoutsakis, E.T.) 13–57 (Marcel Dekker, New York, NY; 1999).
27. Pramanik, J. & Keasling, J.D. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **56**, 398–421 (1997).
28. Sauer, U., Cameron, D.C. & Bailey, J.E. Metabolic capacity of *Bacillus subtilis* for the production of purine nucleosides, riboflavin, and folic acid. *Biotechnol. Bioeng.* **59**, 227–238 (1998).
29. Bonarius, H.P.J., Schmid, G. & Tramper, J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* **15**, 308–314 (1997).
30. Varma, A. & Palsson, B.O. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology* **12**, 994–998 (1994).
31. Edwards, J.S. Functional genomics and the computational analysis of bacterial metabolism (PhD Thesis). (Department of Bioengineering, University of California San Diego, La Jolla, CA; 1999).
32. Chvatal, V. *Linear programming.* (W.H. Freeman, New York, NY; 1983).
33. Varma, A. & Palsson, B.O. Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. *J. Theor. Biol.* **165**, 503–522 (1993).
34. Schilling, C.H. & Palsson, B.O. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.* **203**, 249–83 (2000).
35. Schilling, C.H., Schuster, S., Palsson, B.O. & Heinrich, R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **15**, 296–303 (1999).
36. Schilling, C.H., Edwards, J.S., Letscher, D. & Palsson, B.O. Pathway analysis and flux balance analysis: a comprehensive study of metabolic systems. *Biotechnol. Bioeng.* in press (2001).
37. Sauer, U. *et al.* Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.* **181**, 6679–6688 (1999).
38. Klapa, M.I., Park, S.M., Sinskey, A.J. & Stephanopoulos, G. Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. Theory. *Biotechnol. Bioeng.* **62**, 375–391 (1999).
39. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
40. Richmond, C.S., Glasner, J.D., Mau, R., Jin, H. & Blattner, F.R. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27**, 3821–3835 (1999).
41. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P.A. & Collins, F.S. Detection of heterozygous mutations in BRCA1 using high-density oligonucleotide arrays and two-colour fluorescence analysis. *Nat. Genet.* **14**, 441–447 (1996).
42. Link, A.J., Robison, K. & Church, G.M. Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* **18**, 1259–1313 (1997).
43. Vanbogelen, R.A., Abshire, K.Z., Moldover, B., Olson, E.R. & Neidhardt, F.C. *Escherichia coli* proteome analysis using the gene-protein database. *Electrophoresis* **18**, 1243–1251 (1997).
44. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
45. Schilling, C.H., Edwards, J.S. & Palsson, B.O. Towards metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* **15**, 288–295 (1999).
46. Neidhardt, F.C. (ed.) *Escherichia coli and Salmonella: cellular and molecular biology.* (ASM Press, Washington, DC; 1996).
47. Maniatis, T., Fritsch, E.F. & Sambrook, J. *Molecular cloning: a laboratory manual.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 1982).
48. Neidhardt, F.C. & Umbarger, H.E. In *Escherichia coli and Salmonella: cellular and molecular biology.* (ed. Neidhardt, F.C.) 13–16 (ASM Press, Washington, DC; 1996).
49. Strang, G. *Linear algebra and its applications.* (Saunders, Fort Worth, TX; 1988).
50. Varma, A., Boesch, B.W. & Palsson, B.O. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* **59**, 2465–2473 (1993).
51. Covert, M.W. *et al.* Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.* in press (2001).