Characterizing the metabolic phenotype: a phenotype phase plane analysis. Biotechnnol Bioeng 77:27–36

ARTICLE *in* BIOTECHNOLOGY AND BIOENGINEERING · JANUARY 2002

Impact Factor: 4.13 · DOI: 10.1002/bit.10047 · Source: PubMed

3 AUTHORS, INCLUDING:

Jeremy S Edwards
University of New Mexico
86 PUBLICATIONS **5,274** CITATIONS

Characterizing the Metabolic Phenotype: A Phenotype Phase Plane Analysis

Jeremy S. Edwards,[1] Ramprasad Ramakrishna,[2] Bernhard O. Palsson[3]

[1]Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716; telephone: (302) 831-8072; fax: (302) 831-1048; e-mail: edwards@che.udel.edu
[2]Physiome Sciences, Inc., Princeton, New Jersey
[3]Department of Bioengineering, University of California, San Diego, La Jolla, California

Received 14 June 2001; accepted 4 July 2001

Abstract: Genome-scale metabolic maps can be reconstructed from annotated genome sequence data, biochemical literature, bioinformatic analysis, and strain-specific information. Flux-balance analysis has been useful for qualitative and quantitative analysis of metabolic reconstructions. In the past, FBA has typically been performed in one growth condition at a time, thus giving a limited view of the metabolic capabilities of a metabolic network. We have broadened the use of FBA to map the optimal metabolic flux distribution onto a single plane, which is defined by the availability of two key substrates. A finite number of qualitatively distinct patterns of metabolic pathway utilization were identified in this plane, dividing it into discrete phases. The characteristics of these distinct phases are interpreted using ratios of shadow prices in the form of isoclines. The isoclines can be used to classify the state of the metabolic network. This methodology gives rise to a ''phase plane'' analysis of the metabolic genotype–phenotype relation relevant for a range of growth conditions. Phenotype phase planes (PhPPs) were generated for *Escherichia coli* growth on two carbon sources (acetate and glucose) at all levels of oxygenation, and the resulting optimal metabolic phenotypes were studied. Supplementary information can be downloaded from our website (http://epicurus.che.udel.edu). © 2002 John Wiley & Sons, Inc. *Biotechnol Bioeng* **77**: 27–36, 2002.
Keywords: bioinformatics; metabolism; flux-balance analysis; *Escherichia coli*

INTRODUCTION

Whole genome sequencing and bioinformatics provide the ability to generate a molecular ''parts catalog'' for a particular cell, and the availability of organism-scale molecular parts catalogs necessitates the development of integrative methods to analyze and interpret the systemic properties of cellular processes. To address this need, massively parallel experimental techniques are being developed to probe cellular functions. For example, DNA microarrays can be used to analyze mRNA expression at the whole-genome level (Brown and Botstein, 1999), and whole-genome methods for analyzing the proteome are also being developed (Gygi et al., 1999). These whole-genome experimental techniques are providing valuable information, but fall short of describing (and predicting) the integrated function of cellular processes. Therefore, the abundance of information calls for genome-scale computer models of cellular functions, and these models will find many applications in biotechnology. Herein, we focus on integrated models of one particular cellular function, namely metabolism.

Genome-scale metabolic maps have been reconstructed for several microorganisms (Edwards et al., 1999; Karp et al., 1996, 2000; Overbeek et al., 2000; Tatusov et al., 1996). These metabolic reconstructions are arguably the best-developed reconstructions of cellular processes because functional annotations based on sequence homology are likely the most reliable for metabolic processes given the long history of metabolic research. However, metabolic reconstructions for quantitative and qualitative simulations not only require genomic data, but they also require biochemical information and strain-specific data. Although the metabolic reconstructions presently available may be incomplete and in need of additional development, the need has arisen to analyze the integrated function of metabolic networks. Such analysis will rely on mathematical models and computer simulation to calculate the integrated function of biochemical networks.

Mathematical modeling of metabolic functions has been developed over the past several decades. A number of methods for analyzing metabolic networks have been described in the literature. These include metabolic control analysis (Fell, 1996; Kacser and Burns, 1973), flux-balance analysis (Bonarius et al., 1997; Edwards et al., 1999; Varma et al., 1994), metabolic pathway analysis (Karp et al., 1999; Liao et al., 1996; Mavrovouniotis and Stephanopoulos, 1992; Schilling et al., 1999; Schuster et al., 1999), cybernetic modeling (Kompala et al., 1986; Varner et al., 1998), biochemical systems theory (Savageau, 1969), temporal decomposition (Palsson

Correspondence to: J. S. Edwards
Contract grant sponsors: NIH; NSF
Contract grant numbers: GM-57089; MCB 98-73384; BES 98-14092

DOI 10.1002/bit.10047

et al., 1987), and so on. Dynamic modeling methods require kinetic parameters that describe the catalytic and regulatory functions of metabolic enzymes. Although collections of metabolic kinetic parameters are becoming available (www.empproject.com), it is not possible to obtain all the kinetic information needed to construct genome-scale models of metabolism. In an exceptional case, the human red cell (Joshi et al., 1989; Mulquiney et al., 1999), it is possible to acquire numerical values for all the kinetic constants. In many systems, it has been possible to simplify the network under consideration and to assume numerical values for the kinetic constants to construct large-scale kinetic models (Shu et al., 1989). However, in general, information on such kinetic constants is not available.

To alleviate this problem, an entirely different modeling approach can be taken to determine what a cell can and cannot do (rather than trying to calculate and predict the behavior) based on its metabolic reconstruction. Such analysis relies on the simultaneous mass balances constraints that are defined by the systems stoichiometry, and known capacity constraints on individual steps. Flux-balance analysis (FBA) takes into consideration a set of metabolic constraints to define a solution space in which all feasible metabolic flux distributions reside (Bonarius et al., 1997; Varma 1994). One can then study the properties of this space and, in particular, use linear optimization to calculate the optimal flux distribution that supports growth (Varma and Palsson, 1993).

The implementation of FBA and linear programming has thus proven useful for the initial analysis of genome-scale reconstructed metabolic maps, and is an approach with considerable potential for the analysis of integrated metabolic functions. At present, metabolic flux maps are typically calculated for a single growth condition, giving a limited view of the metabolic genotype–phenotype relation. An approach that provides a global perspective on the relation is required (see Fig. 1 for illustration of the phenotypic phase plane [PhPP] analysis). In this study we investigate the development of a PhPP analysis that maps all the growth conditions represented by two environmental variables into a single plane. We interpret and use the PhPP to get a broader view of the optimal flux distributions under varying constraints. Finally, PhPP analysis has proven useful for in silico studies and experimental design (Edwards et al., 2001).

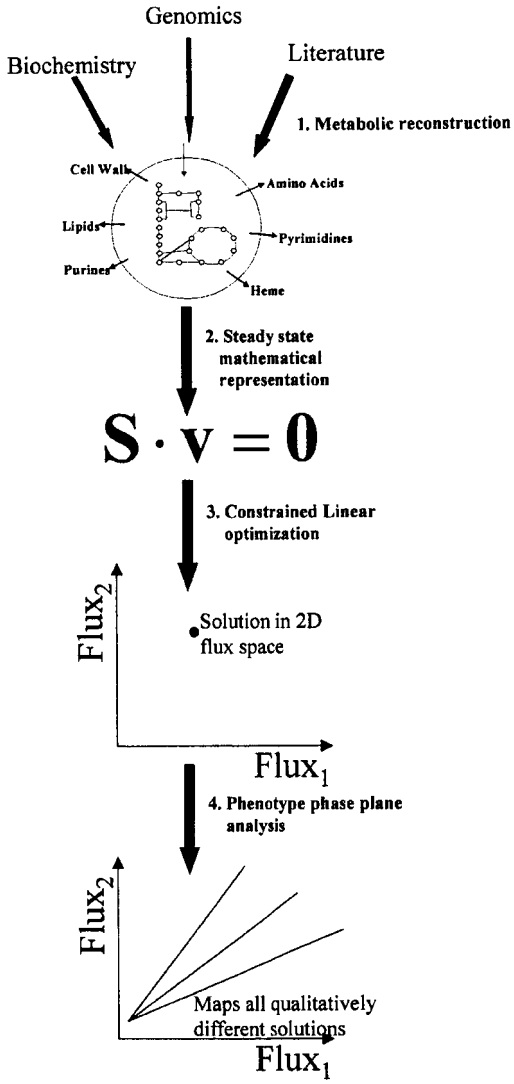

Figure 1. Phenotype phase plane analysis—extension of FBA. (1) Whole cell metabolic networks can be reconstructed based on biochemical and genomic information. (2) The metabolic network flux can be studied computationally based on the matrix equation, $\mathbf{S}\cdot\mathbf{v} = \mathbf{0}$ and constraints $\alpha_i \leq v_i \leq \beta_i$. (3) Constrained linear optimization can be applied and the optimal utilization of the metabolic network can be determined, for a given objective function. (4) The optimization provides information on the optimal metabolic fluxes for one set of constraints. Phenotype phase plane analysis can study the optimal utilization of the network as a function of the constraints. Phenotype phase plane analysis has proven extremely useful in the design and analysis of experimental data (see Edwards et al., 2001).

METHOD DEVELOPMENT

The capabilities of a metabolic network can be assessed using flux-balance analysis (FBA) (Bonarius et al., 1997; Edwards and Palsson, 1999; Varma and Palsson, 1994a, 1994b). FBA is based primarily on the conservation of mass in the metabolic network. Because the conservation requirement is implemented by stoichiometric balance equations, FBA relies on the stoichiometric characteristics of the metabolic network. Herein, we focus on the *Escherichia coli* MG1655 metabolic network reconstruction (Edwards et al., 2000), which was constructed from the biochemical literature (Neidhardt, 1996a, 1996b), genome sequence (Blattner et al., 1997), metabolic pathway databases (Karp et al., 2000; Ogata et al., 1998; Selkov et al., 1998) and was used to define the stoichiometric matrix. The entire list of reactions

included in the analysis is available (http://epicurus.che.udel.edu). In addition, experimentally determined strain-specific parameters, such as the biomass composition (Neidhardt et al., 1996a, 1996b) and the maintenance requirements (Varma and Palsson, 1994a) are also required.

The steady-state flux balance equation is $\mathbf{S} \cdot \mathbf{v} = \mathbf{b}^{\mathbf{v}}$, where $\mathbf{S}$ is the stoichiometric matrix, the vector $\mathbf{v}$ defines the metabolic fluxes, and $\mathbf{b}^{\mathbf{v}}$ is nominally zero (when the exchange fluxes are defined in $\mathbf{S}$; see the web extras to Edwards et al., 2001), thus enforcing simultaneous mass, energy, and redox balance constraints through a set of mass balances in the matrix equation. Variations of the $\mathbf{b}^{\mathbf{v}}$ vector from zero were used in the shadow price analysis (discussed in what follows). In the *E. coli* metabolic network, the number of metabolic fluxes was greater than the number of mass balances, thus leading to a plurality of feasible flux distributions that lie in the null space of the matrix $\mathbf{S}$. Additional constraints were also placed on the feasible value of each flux in the metabolic network ($\alpha_i \leq v_i \leq \beta_i$). These constraints were utilized to define the reversibility of the internal reactions and to set the uptake rate for the transport reactions. The transport of inorganic phosphate, ammonia, carbon dioxide, sulfate, potassium, and sodium were unrestrained, whereas the uptake of the carbon source and oxygen were restrained as specified in the respective PhPP. All metabolic byproducts (i.e., acetate, ethanol, formate, pyruvate, succinate, lactate, etc.), which are known to be transported out of the cell, were always allowed to leave the metabolic system. In this analysis, α_i for the internal fluxes was set to zero for all irreversible fluxes, and all reversible fluxes were unbounded in the forward and reverse direction (the reversibility of each reaction is defined on the website containing supplementary information, Edwards et al., 2001). The intersection of the null space and the region defined by the linear inequalities (flux constraints) defined the capabilities of the metabolic network and has been referred to as the flux cone (Schilling et al., 1999).

The determination of the optimal metabolic flux distribution that lies within the flux cone was formulated as a linear programming (LP) problem, in which the solution that minimizes a particular metabolic objective was identified (Bonarius et al., 1996; Edwards et al., 1999; Pramanik et al., 1997; Varma and Palsson, 1994a, 1994b). The growth flux was used as the objective function, and defined as a metabolic flux utilizing the biosynthetic precursors, X_m, in the appropriate ratios to produce biomass:

$$\sum_{\text{all } m} d_m \cdot X_m \xrightarrow{\text{growth}} \text{biomass}$$

where d_m is the biomass fraction of metabolite X_m. The biomass composition was defined based on Neidhardt et al. (1996a).

Phenotype Phase Plane Analysis

All steady-state metabolic flux distributions are mathematically confined to the flux cone defined for the given metabolic map, where each solution in the flux cone corresponds to a particular internal flux distribution or a particular metabolic phenotype (Varma and Palsson, 1994a). Under specified growth conditions, the optimal phenotype in the cone can be determined using linear programming (LP). If the constraints vary, the shape of the cone changes, and the optimal flux vector may qualitatively change; for example, inactive fluxes may be activated and vice versa. The phase plane analysis is developed to consider all possible variations in two constraining environmental variables.

Defining Phenotype Phase Planes

Uptake rates of two nutrients (such as the carbon substrate and oxygen) were defined as two axes on an (x,y)-plane, and the optimal flux distribution was calculated for all points in this plane. There are a finite number of qualitatively different optimal metabolic flux maps, or metabolic phenotypes, present in such a plane. The demarcations on the phase plane were defined by a shadow price analysis (Varma et al., 1993). This procedure led to the definition of distinct regions, or "phases," in the plane, for which the optimal use of the metabolic pathways was qualitatively different. Each phase was written as $\mathrm{P}n_{x,y}$, where P represents the phenotype, n is the number of the demarcated region for this phenotype (as shown in the corresponding figure), and x,y the two uptake rates on the axis of the plane.

Calculating the Phase Plane

The phase planes were constructed by calculating the shadow prices throughout the two-parameter space, and lines were drawn to demarcate regions of constant shadow prices. The shadow prices defined the intrinsic value of each metabolite toward the objective function. Changes in shadow prices were used to interpret the metabolic behavior.

Mathematically, the shadow prices are defined as:

$$\gamma_i = \frac{-dZ}{db_i^v} \tag{1}$$

and are associated with each metabolite in the network. The shadow prices defined the sensitivity of the objective function (Z) to changes in the availability of each metabolite (b_i^v defines the violation of a mass balance constraint and is equivalent to an uptake reaction). The shadow prices were either negative, zero, or positive, depending on the value of the metabolite. The direction and magnitude of the shadow price vector in each region of the phase plane was different (by definition of the

phase plane); thus, the state of the metabolic system was different in each region.

Isoclines

Isoclines were also defined to interpret the metabolic phenotype and to represent the locus of points within the two-dimensional space that provide for the same value of the objective function. The slope of the isoclines within each region was calculated from the shadow prices; thus, by definition, the slope of the isoclines was different in each region of the PhPP. A ratio of shadow prices was used to define the slope of the isoclines (ρ):

$$\rho = -\frac{\gamma_x}{\gamma_y}\bigg|_Z = -\left(\frac{-dZ/db_x^v}{-dZ/db_y^v}\right)\bigg|_z = -\frac{db_y^v}{db_x^v}\bigg|_Z \qquad (2)$$

The negative sign in Eq. (2) was introduced in anticipation of its interpretation. The condition dependent relative value of the substrates, defined as ρ, was used to interpret the constraining factors on the metabolic network. In regions where ρ was negative, there was dual limitation of the substrates. Under different conditions, the isoclines were also either horizontal or vertical in certain phase plane regions, representing regions of single substrate limitation, and the ρ value in these regions was zero or infinite, respectively. Certain regions in the PhPP also had a positive ρ. These regions were termed "futile," and increased uptake of one of the substrates had a negative effect on the objective function in these regions. Finally, due to stoichiometric limitations, there were infeasible steady-state regions in the PhPP.

Line of Optimality

The line of optimality (LO) was defined as the line representing the optimal relation between the two metabolic fluxes corresponding to the axis of the PhPP. For results presented herein, this line can be interpreted as the optimal oxygen uptake for the complete oxidation of the substrate in order to support the maximal biomass yield.

Example

To illustrate these concepts, we now present an example of a hypothetical metabolic system (Fig. 2; see Table I for a list of the reactions and the stoichiometry). This hypothetical network utilizes a single carbon source, which it metabolizes to a single biosynthetic precursor, **C**. This precursor is converted into biomass, via $\mathbf{R_z}$ (the objective function) using redox potential and high-energy phosphate bonds, and can be converted into two metabolic byproducts, **D** and **E**. An electron acceptor (oxygen) is used to convert redox potential into high-energy phosphate bonds, $\mathbf{R_{res}}$ (i.e., an electron transport system). In addition, there is a reaction, $\mathbf{R_3}$, which consumes 0.2**C** to generate NADH. Finally, one reaction, $\mathbf{R_{ft}}$, represents a futile cycle that dissipates ATP.

The PhPP for this hypothetical metabolic system was calculated (Fig. 2). The nonhatched region defined conditions for balanced flux maps (i.e., the flux vector must be within the flux cone) and defined the capabilities of the hypothetical metabolic network. A finite number of different metabolic phenotypes were found in the phase plane. P1 was a futile region where the electron acceptor was provided in excess. The metabolic network dissipated the excess oxygen by increasing the $\mathbf{R_3}$ flux, which generated NADH but also oxidized the precursor, **C**. The futile cycle reaction, $\mathbf{R_{ft}}$, was utilized to eliminate the excess ATP produced. The P1 upper limit represents the locus of points for which the carbon was completely oxidized to eliminate the excess electron acceptor, and thus no biomass was generated.

The metabolic flux map of this example system is shown in Figure 2 for conditions on the LO. The LO was a special case of P1 where the electron acceptor was no longer in excess and the futile cycle flux was zero. The LO represented the optimal utilization of this example metabolic network to produce biomass, and the shadow price for all metabolic byproducts was greater than zero at this point (Table II). The qualitative flux map indicated that, under conditions defined by the LO, there was no metabolic byproduct production and the futile cycle flux equaled zero.

In P2 a reduced metabolic byproduct (**D**) was secreted by the network due to low O_2 availability. The shadow price for the metabolite **D** was zero in region P2 (Table II), and the utilization of the metabolic pathways in this region was fundamentally different than in P1, Plo, P3, and P4. The metabolic pathway for the production and secretion ($\mathbf{R_4}$) of **D** was turned on under the conditions defined in this region, and the excess redox potential was eliminated through the secretion of **D**.

As the O_2 availability was decreased, the secretion of **D** was not sufficient to eliminate all the redox equivalents, and a different region (P3) resulted. In this region the value of the carbon source was reduced and the desire for O_2 was increased, thus the slope of the isoclines was reduced. Furthermore, the shadow price for the metabolite **E** was zero in this region (Table II), and therefore, it was secreted (in addition to **D**) as a metabolic byproduct. In P3, the shadow price for the redox equivalents was positive; thus, additional redox potential would lead to a reduction in the objective function. To compensate for the positive shadow price of redox potential, the cyclic reaction, $\mathbf{R_3}$, was utilized, thereby reducing redox production.

Finally, in P4, the futile cycle reaction was utilized, and all metabolic by-product formation was directed toward the formation of the more reduced byproduct, **E**. The shadow prices indicate that the metabolite **D** had value to the cell in this region, and therefore, it was not

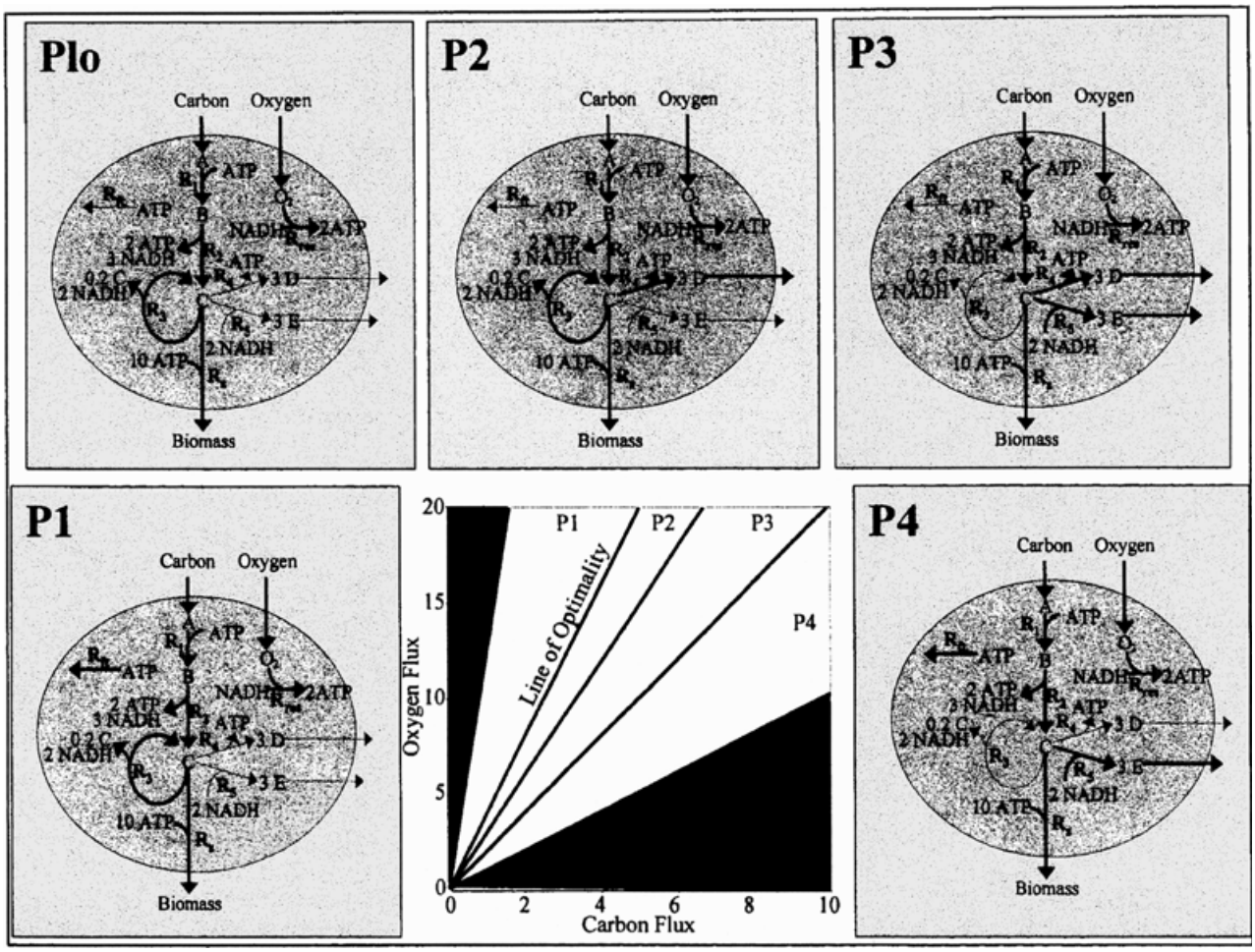

Figure 2. A hypothetical metabolic reaction network and its phenotype phase plane. The PhPP was calculated for the hypothetical reaction network shown. The PhPP was calculated as a function of the two metabolic fluxes representing the uptake of carbon and oxygen. Each of the regions are labeled and discussed in the text. The crosshatched regions represent infeasible steady states.

secreted. Region P4 defined a futile region where the carbon source was provided in excess. Thus, the carbon that was supplied beyond the line demarcating P3 from P4 was simply excreted from the cell, which costs cellular resources. The effect of the excess carbon on the growth flux was quantified with the shadow price analysis (Table II). When the oxygen uptake and the carbon uptake defined a point on the lower boundary of P4, the entire carbon source was directed toward the formation of metabolite **E**, and no biomass was generated. Thus, below this line (the crosshatched region) was an infeasible steady-state region of the metabolic network.

This example illustrates the utility of the PhPP in the interpretation of the metabolic physiology of the system.

Table I. Stoichiometry of the example reaction network.

$\xrightarrow{A_{uptake}} A$	$ATP \xrightarrow{R_n}$
$A + ATP \xrightarrow{R_1} B$	$C + 10ATP \xrightarrow{R_z}$ Biomass
$B \xrightarrow{R_2} 2ATP + 3NADH + C$	$\xrightarrow{O_{2uptake}} O_2$
$0.2C \xrightarrow{R_3} 2NADH$	$C \xleftrightarrow{C_{out}}$
$C \xleftrightarrow{R_4} ATP + 3D$	$D \xleftrightarrow{D_{out}}$
$C + 2NADH \xrightarrow{R_5} 3E$	
$NADH + O_2 \xrightarrow{R_{res}} 2ATP$	$E \xleftrightarrow{E_{out}}$

Table II. Shadow price of metabolites from the example shown in Figure 1.

	Carbon	A	B	C	D	E	O_2	NADH	ATP
P1	−1.30	−1.30	−1.30	−1.00	−0.33	−0.40	**0.10**[a]	**−0.10**	
Plo	−0.90	−0.90	−0.93	−0.67	−0.21	−0.27		−0.07	−0.03
P2	−0.21	−0.21	−0.30	−0.09	[b]	−0.04	−0.17	−0.01	−0.09
P3	−0.05	−0.05	−0.14	−0.09			−0.23	**0.05**[a]	−0.09
P4	**0.50**[a]	**0.50**[a]	**0.50**[a]	−1.00	−0.33		−0.50	**−0.50**[a]	

[a] Positive numbers are indicated in bold.
[b] Blank spaces are zero.

It shows that the optimal metabolic pathway utilization was condition-dependent, and that a finite number of qualitatively different optimal phenotypes can be derived from a single metabolic network.

RESULTS

The optimal utilization of the *E. coli* metabolic genotype was assessed for cellular growth in silico on two different substrates: acetate and glucose. For each substrate, a PhPP was computed to map the theoretical optimal metabolic characteristics for biomass production as a function of the environmental variables (carbon substrate and oxygen uptake rates).

Acetate. The acetate–oxygen PhPP was determined to consist of two phases, each corresponding to a futile region separated by the LO (Fig. 3). The shadow price of the key metabolites and byproducts is shown in Table III. By definition, the shadow prices are different in each region of the PhPP.

$P1_{ace,oxy}$ was defined as a futile region with excess oxygen. A defining characteristic of this futile region was the shadow price for the high-energy phosphate bonds ($\gamma_{ATP} = 0$). This indicated a condition where ATP production exceeded the biosynthetic/maintenance demands for cellular growth. The ρ value (2.0) in $P1_{ace,oxy}$ quantitatively indicated the effect of the excess oxygen, and the net stoichiometric coupling of oxygen and acetate to eliminate the excess oxygen was 2:1.

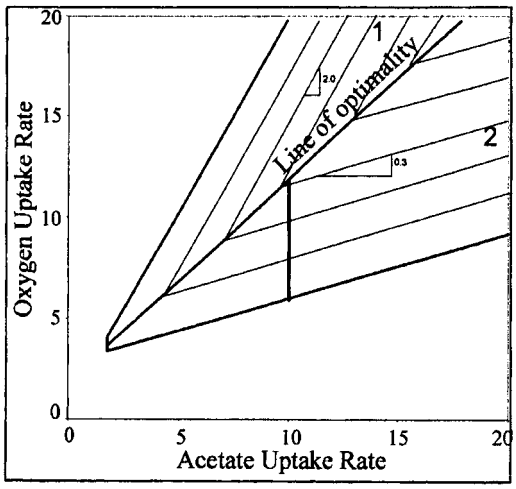

Figure 3. The acetate and oxygen phenotype phase plane for the *E. coli* metabolic genotype.

The line separating $P1_{ace,oxy}$ from $P2_{ace,oxy}$ was the LO. The in silico simulations for growth on acetate were compared with experimental data. The optimal biomass yield was determined in silico to be 0.322 g g^{-1} using an experimentally observed acetate uptake rate (Weikert et al., 1997) with no restrictions on the oxygen uptake rate. The experimentally observed biomass yield for this experiment was 0.319 g g^{-1} (Weikert et al., 1997).

Oxygen was in excess in $P1_{ace,oxy}$, whereas $P2_{ace,oxy}$ was a futile region with excess acetate (oxygen is the limiting substrate). In $P2_{ace,oxy}$, acetate was provided beyond the optimal amount required by the metabolic network to support growth (indicated by the p value), and FBA predicted that the excess acetate should be eliminated as ethanol, which required the consumption of redox potential. However, in $P2_{ace,oxy}$, additional redox potential had value to the cell (determined from the shadow price analysis), and thus the excess acetate led to reduced biomass production. Quantitatively, the positive ρ value ($\rho = 0.3$) defined the effect on biomass production of increasing the acetate uptake. Thus, for each additional millimole of acetate taken into the cell (beyond the line of optimality) 0.3 mmol of oxygen was required to eliminate this surplus acetate and maintain the same biomass production.

Glucose. The *E. coli* glucose–oxygen PhPP was calculated, and it exhibited six distinct optimal phenotypes (Fig. 4; shadow prices shown in Table III). The metabolic characteristics of each region were analyzed.

As with acetate, $P1_{glc,oxy}$ represented a futile region with excess oxygen ($\rho = 6.0$), and the LO separated $P1_{glc,oxy}$ from $P2_{glc,oxy}$ (Fig. 4). The optimal biomass yield of the in silico *E. coli* strain was shown to be on the LO. The comparison of experimentally observed *E. coli* growth rates with the LO demonstrated that the growth

Table III. Metabolite shadow prices within each region of the phenotype phases planes shown in Figures 2 and 3.

Region	P	Redox	GLC	O_2	Acetate	Ethanol	Formate	Succinate
Acetate PhPP	[b]							
1		−0.0115	−0.1383	**0.0230**[a]	−0.0461	−0.0691	−0.0115	−0.0807
2	−0.0106		−0.0607	−0.0317	**0.0106**[a]		−0.0013	−0.0026
Glucose PhPP								
1		−0.0115	−0.1383	**0.0230**[a]	−0.0461	−0.0691	−0.0115	−0.0807
2	−0.0093	−0.0021	−0.0504	−0.0238		−0.0135		−0.0168
3	−0.0096		−0.0434	−0.0289		−0.0096		−0.0108
4	−0.0097	**0.0014**[a]	−0.0406	−0.0319		−0.0068		−0.0080
5	−0.0102	**0.0051**[a]	−0.0356	−0.0407				−0.0013
6	−0.0102	**0.0051**[a]	−0.0356	−0.0433				

P, high-energy phosphate ($\gamma_{ATP} - \gamma_{ADP} - \gamma_{Pi}$); redox ($\gamma_{NADH} - \gamma_{NAD}$).

[a] Positive numbers are indicated in bold.

[b] Blank spaces are zero.

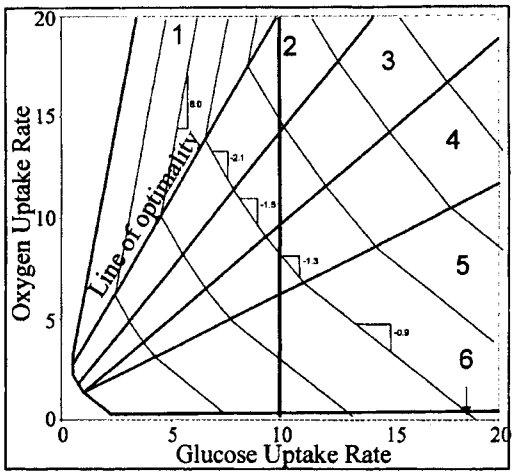

Figure 4. The glucose and oxygen phenotype phase plane for the *E. coli* metabolic genotype.

rate of *E. coli* was less than the LO (Edwards, 1999; Jensen et al., 1992; Weikert et al., 1997). Furthermore, the observed metabolic behavior led to acetate production, as was predicted to be the optimal metabolic byproduct for metabolic operation within $P2_{glc,oxy}$ by the PhPP analysis. In contrast to the in silico predictions for growth on acetate, the in silico analysis suggests that the cell can increase the growth rate by increasing the glucose uptake rate beyond the LO. However, this results in the production of acetate and a reduction in the biomass yield. Furthermore, this analysis assumes that all internal fluxes can be increased to account for the increased glucose uptake, which is not true for large increases in the uptake.

Below the LO were five optimal metabolic phenotypes. These phenotypes were characterized by oxygen limitations, and all biosynthetic requirements were not satisfied by oxidative metabolism. The metabolic phenotype was interpreted by examining the metabolic fluxes. The optimal metabolic fluxes were calculated for a fixed glucose uptake rate (10 mmol/g [DW]/h) and for all oxygen uptake rates in regions $P2_{glc,oxy}$ through $P6_{glc,oxy}$ (Fig. 5). It was shown that the optimal TCA cycle flux was sensitive to the oxygen uptake rate in $P2_{glc,oxy}$ (Fig. 5C), and acetate was formed as an optimal metabolic byproduct in $P2_{glc,oxy}$ ($\gamma_{ac} = 0$) (discussed earlier). We also found that the production of CO_2 was decreased as the oxygen uptake was decreased (Fig. 5D), and this was the result of shifting the optimal carbon

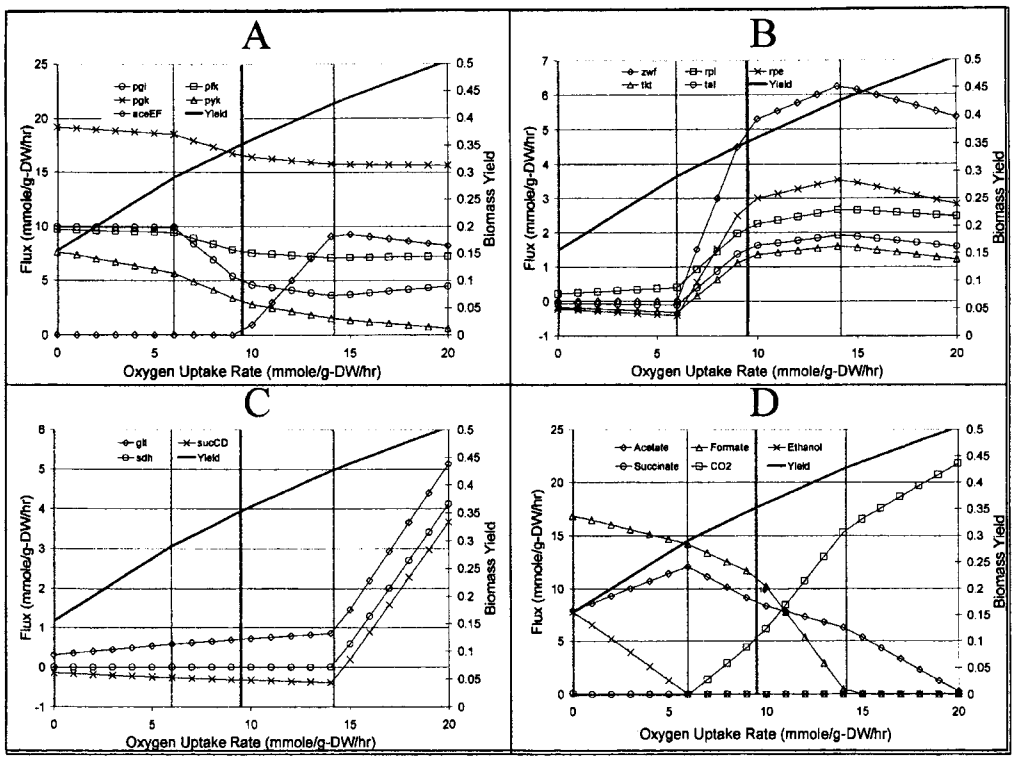

Figure 5. The metabolic fluxes with glucose as the carbon source and varying oxygen uptake rates. The glucose uptake rate was set to 10 mmol/g (DW)/h, as indicated by the line in Figure 4. (A) Glycolysis. (⊖) pgi, (⊗) pgk, (◈) aceEF, (⊟) pfk, (△) pyk, (–) yield. (B) Pentose phosphate pathway. (○) tal, (×) rpe, (◇) zwf, (□) rpi, (△) tkt, (–) yield. (C) TCA cycle, (○) sdh, (×) sucCD, (◇) glt, (–) yield. (D) Transport reactions, (○) Succinate, (×) ethanol, (◇) acetate, (□) CO_2, (△) formate, (–) yield. Abbreviations: aceA, isocitrate lyase; aceEF, pyruvate dehydrogenase; fbp, fructose-1,6-bisphosphatase; glt, citrate synthase; icd, isocitrate dehydrogenase; mae, malic enzyme; mdh, malate dehydrogenase; pfk, phosphofructokinase; pgi, phosphoglucose isomerase; pgk, phosphoglycerate kinase; pyk, pyruvate kinase; rpe, ribulose phosphate 3-epimerase; rpi, ribose-5-phosphate isomerase; sdh, succinate dehydrogenase; sucCD, succinyl-CoA synthetase; tal, transaldolase; tkt, transketolase; zwf, glucose 6-phosphate-1-dehydrogenase.

flux to the formation of more reduced byproducts. Conversely, the optimal fluxes in glycolysis (Fig. 5A) and the PPP (Fig. 5B) were not sensitive to the oxygen uptake rate in $P2_{glc,oxy}$, although the oxidative PPP flux was increased and the six-carbon glycolytic flux was decreased in $P2_{glc,oxy}$ as the oxygen uptake rate was decreased.

If oxygen was further limited, the ability to produce NADH exceeded the ability to utilize the redox potential for biomass synthesis, which led a different optimal metabolic phenotype $P3_{glc,oxy}$. Formate was the optimal sink for excess redox potential in $P3_{glc,oxy}$, and the value of oxygen to the cell (relative to glucose) was increased, but on a per-mole basis glucose had a higher value to the cell (ρ value $= -1.5$). In this region, the pyruvate kinase flux was increased as the oxygen uptake rate was decreased; furthermore, the pyruvate dehydrogenase flux was decreased (Fig. 5A) and the carbon was directed toward formate (Fig. 5D). Finally, the PPP fluxes were decreased, and the TCA cycle did not operate cyclically in this region—serving only to produce biosynthetic precursors (Fig. 5C).

In $P4_{glc,oxy}$, the NADH production exceeded the optimal respiratory requirements and, to balance redox metabolism, the transhydrogenase reaction was utilized to convert NADH to NADPH. In this region, the optimal flux in the oxidative branch of the PPP was sensitive to oxygen uptake (Fig. 5B), and the glycolytic fluxes increases with decreasing oxygen uptake rate (Fig. 5A). The absolute value of ρ in this region was further decreased relative to $P3_{glc,oxy}$, indicating the increased oxygen "value" relative to glucose.

Ethanol was secreted as a metabolic byproduct to optimally balance the redox potential of the cell in $P5_{glc,oxy}$. The glycolytic fluxes (Fig. 5A) were not sensitive to oxygen uptake rate in this region, whereas, the acetate:ethanol production (Fig. 5D) ratio was found to be sensitive to the oxygen uptake rate. More specifically, the acetate production decreased and the carbon was directed toward ethanol production as the oxygen uptake rate was decreased in $P5_{glc,oxy}$ (Fig. 5D). The oxidative branch of the PPP does not carry a flux in this region, and the nonoxidative branch of the PPP was used to generate the biosynthetic precursors (Fig. 5B). Finally, the absolute value of ρ (-0.87) dropped below unity in this region, and this indicated an oxygen value greater than glucose on per-mole basis.

In $P6_{glc,oxy}$ represents anaerobic growth, and succinate was also optimally produced as a metabolic byproduct. In the absence of oxygen as an electron acceptor, fumarate can accept electrons to balance the redox potential of the cell. During anaerobic growth, a major constraint facing the cellular metabolic network was the elimination of redox potential, not the production of high-energy phosphate bonds. Under completely anaerobic conditions, the ρ value (-0.8) was slightly decreased from $P5_{glc,oxy}$. However, under anaerobic conditions, glucose had value to the cell, which was in contrast to acetate (Fig. 3) for which an infeasible steady-state region characterized the anaerobic state.

DISCUSSION

The ability to generate genome-scale metabolic maps has created the need to develop mathematical methods for their analysis. FBA is an established method for the analysis of metabolic flux distributions. Here we have shown how this analysis can be extended from looking at single growth conditions, to the simultaneous consideration of all possible values of two constraining environmental factors. A phase plane representation is developed, interpreted, and applied to the analysis of the optimal uses of *E. coli* metabolism.

The phenotype phase plane analysis provides a deeper understanding of the metabolic flux distributions. Phenotype phase planes were calculated by considering all possible values of two variables that characterize the nutritional environment of the cell. The analysis shows that the potentially infinite metabolic flux maps can be classified into a finite number of metabolic phenotypes, and this classification is attained by using the shadow prices from the linear optimization. The properties of the discrete number of phenotypes that fall within defined regions can be interpreted using isoclines. The isoclines represent a new concept in FBA and they are lines of constant values of the objective function—the growth function in the present calculations. The isoclines were used to define futile regions, regions of single and dual substrate limitations, and a line of optimality. Thus, once the phenotype phase plane is calculated, one has a quick and clear overview of the optimal use of metabolism in the growth environment considered.

Although the phenotype phase planes are of clear utility and represent an advancement of the FBA methodology, there are two chief qualifiers that must be discussed. The first pertains to the quality of the reconstructed metabolic maps. For well-known organisms, such as *E. coli* and yeast, there exists a great wealth of biochemical data that can be used to confirm the functional assignment based on the genomic sequence. Furthermore, there are physiological data available. Reconstructed maps for well-known organisms are likely close to completion, although the discovery of additional metabolic capabilities of these organisms should not be ruled out. For less studied organisms that have been sequenced and annotated, the reconstructed maps are likely to be incomplete and contain errors. However, as more related organisms are sequenced, their annotations can be cross-referenced, and the quality of such maps will improve with time (Overbeek et al., 2000).

The second issue is the definition of the objective function for the linear programming analysis. Herein,

we have used the maximization of the growth flux, and data obtained under a restricted set of experimental conditions suggest that *E. coli* does express optimal growth phenotypes in nutritionally rich minimal media (Edwards et al., 2001; Pramanik and Keasling, 1997; Varma and Palsson, 1994). However, this objective function is not always applicable, and it is difficult to know what is optimal for *E. coli* under many other conditions (i.e., nongrowth conditions—stationary phase). For example, under poor growth conditions, organisms may choose to scavenge and store key elements, such as phosphate, as an optimal survival strategy. Furthermore, whether optimal growth in defined media is observed will be a function of the evolutionary history of the organism. Laboratory strains that have been grown on a defined media may have adopted for optimal growth performance under those conditions, and there has been good agreement between the FBA simulations and experimental data with laboratory strains (Edwards, 1999; Pramanik and Keasling, 1997; Varma and Palsson, 1994a, 1994b), and further studies of appropriate objective functions under different conditions are needed.

In summary, the ability to reconstruct genome-scale metabolic maps calls for the development of methods to analyze their integrative behavior. Flux-balance analysis has shown utility for such analysis. Phenotype phase planes have been developed to obtain a broad view of the metabolic genotype–phenotype relation based on FBA. The PhPP provides a concise and interpreted view of this relationship. The PhPP can be used to generate experimentally testable hypotheses to study the optimal responses to changing environmental conditions.

The authors thank Rafael Ibarra for proofreading the manuscript.

References

Blattner FR, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1474.

Bonarius HPJ, et al. 1996. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. Biotechnol Bioeng 50:299–318.

Bonarius HPJ, Schmid G, Tramper J. 1997. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. Trends Biotechnol 15:308–314.

Brown PO, Botstein D. 1999. Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37.

Edwards JS. 1990. PhD thesis, University of California, San Diego, San Diego CA, USA.

Edwards JS, Ibarra RU, Palsson BO. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130.

Edwards JS, Palsson BO. 1999. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. J Biol Chem 274:17410–17416.

Edwards JS, Palsson BO. 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 97:5528–5533.

Edwards JS, Ramakrishna R, Schilling CH, Palsson BO. 1999. In: Lee SY, Papoutsakis ET, editors. Metabolic engineering. New York: Marcel Dekker. p 13–57.

Fell D. 1996. *Understanding the control of metabolism*. London: Portland Press.

Gygsi SP, et al. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994–999.

Jensen PR, Michelson O. 1992. Carbon and energy metabolism of *atp* mutants of *Escherichia coli*. J Bacteriol 174:7635–7641.

Joshi A, Palsson BO. 1989. Metabolic dynamics in the human red cell. Part I–A comprehensive kinetic model. J Theor Biol 141:515–528.

Kacser H, Burns JA. 1973. The control of flux. Symp Soc Biol 27:65–104.

Karp PD, Krummenacker M, Paley S, Wagg J. 1999. Integrated pathway-genome databases and their role in drug discovery. Trends Biotechnol 17:275–281.

Karp PD, Ouzounis C, Paley S. 1996. HinCyc: A knowledge base of the complete genome and metabolic pathways of *H. influenzae*. Intelligent Systems for Molecular Biology 4:116–124.

Karp PD, et al. 2000. The EcoCyc and MetaCyc databases. Nucl Acids Res 28:56–59.

Kompala DS, Ramkrishna D, Jansen NB, Tsao GT. 1986. Investigation of bacterial growth on mixed substrates. Experimental evaluation of cybernetic models. Biotechnol Bioeng 28:1044–1056.

Liao JC, Hou SY, Chao YP. 1996. Pathway analysis, engineering and physiological considerations for redirecting central metabolism. Biotechnol Bioeng 52:129–140.

Mavrovouniotis M, Stephanopoulos G. 1992. Synthesis of biochemical production routes. Comput Chem Eng 16:605–619.

Mulquiney PJ, Kuchel PW. 1999. Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: computer simulation and metabolic control analysis. Biochem J 342:597–604.

Neidhardt FC, editor. 1996a. *Escherichia coli* and *Salmonella*: Cellular and molecular biology. Washington, DC: ASM Press.

Neidhardt FC, Umbarger HE, 1996b. *Escherichia coli* and *Salmonella*: Cellular and molecular biology. Vol. 1. Washington, DC: ASM Press. p 13–16.

Ogata H, Goto S, Fujibuchi W, Kanehisa M. 1998. Computation with the KEGG pathway database. Biosystems 47:119–128.

Overbeek R, et al. 2000. WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucl Acids Res 28:123–125.

Palsson BO, Joshi A, Ozturk SS. 1987. Reducing complexity in metabolic networks: Making metabolic meshes manageable. Fed Proc 46:2485–2489.

Pramanik J, Keasling JD. 1997. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng 56:398–421.

Savageau MA. 1969. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. J Theor Biol 25:370–379.

Schilling CH, Schuster S, Palsson BO, Heinrich R. 1999. Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. Biotechnol Prog 15:296–303.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17:53–60.

Selkov E, Jr., Grenchkin Y, Mikhailova N, Selkov E. 1998. MPW: The metabolic pathways database. Nucl Acids Res 26:43–45.

Shu J, Shuler ML. 1989. A mathematical model for the growth of a single cell of *Escherichia coli* on a glucose/glutamine/ammonium medium. Biotechnol Bioeng 33:1117–1126.

Tatusov RL, et al. 1996. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. Curr Biol 6:279–291.

Varma A, Boesch BW, Palsson BO. 1993. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59:2465–2473.

Varma A, Palsson BO. 1993. Metabolic capabilities of *Escherichia coli:* II. Optimal growth patterns. J Theor Biol 165:503–522.

Varma A, Palsson BO 1994a. Metabolic flux balancing: Basic concepts, scientific and practical use. Bio/Technology 12:994–998.

Varma A, Palsson BO. 1994b. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60:3724–3731.

Varner J, Ramkrishna D. 1998. Application of cybernetic models to metabolic engineering: Investigation of storage pathways. Biotechnol Bioeng 58:282–291.

Weikert C, Sauer U, Bailey JE. 1997. Use of a glycerol-limited, long-term chemostat for isolation of *Escherichia coli* mutants with improved physiological properties. Microbiology 143:1567–1574.