Energy Balance for Analysis of Complex Metabolic Networks

Daniel A. Beard,* Shou-dan Liang,† and Hong Qian‡

*Department of Bioengineering, University of Washington, Seattle, Washington 98915, †National Aeronautics and Space Administration Ames Research Center, Moffett Field, California 94035, and ‡Departments of Applied Mathematics and Bioengineering, University of Washington, Seattle, Washington 98195 USA

ABSTRACT Predicting behavior of large-scale biochemical networks represents one of the greatest challenges of bioinformatics and computational biology. Computational tools for predicting fluxes in biochemical networks are applied in the fields of integrated and systems biology, bioinformatics, and genomics, and to aid in drug discovery and identification of potential drug targets. Approaches, such as flux balance analysis (FBA), that account for the known stoichiometry of the reaction network while avoiding implementation of detailed reaction kinetics are promising tools for the analysis of large complex networks. Here we introduce energy balance analysis (EBA)—the theory and methodology for enforcing the laws of thermodynamics in such simulations—making the results more physically realistic and revealing greater insight into the regulatory and control mechanisms operating in complex large-scale systems. We show that EBA eliminates thermodynamically infeasible results associated with FBA.

INTRODUCTION AND BACKGROUND

Conservation principles impose constraints on the fluxes and chemical potentials associated with biochemical network reactions that are analogous to Kirchoff's current and voltage laws for electrical networks (Balabanian and Bickart, 1981). Flux balance analysis (FBA) Varma and Palsson, 1994; Bonarius et al., 1996, 1997; Pramanik and Keasling, 1997; Schilling et al., 1999, 2001; Schuster et al., 1999; Edwards and Palsson, 2000a, b, c; Edwards et al., 2001 invokes mass conservation, but does not consider the chemical potential associated with nonequilibrium steady-state chemical fluxes. The thermodynamic analysis presented here, energy balance analysis (EBA), provides additional constraints on the system that are analogous to the voltage loop law. In addition to predicting network fluxes that are thermodynamically feasible, EBA facilitates a detailed analysis of regulation of metabolic networks that is not available from FBA alone. These simple and fundamental theoretical results will have consequences in the analysis and manipulation of genomes, as we demonstrate for the genes involved in the *Escherichia coli* metabolism.

The central flux balance conservation statement is given by the equation

$$SJ = 0, \tag{1}$$

where $J \in \Re^n$ is the vector of n fluxes occurring in the network, $S \in \Re^{m\times n}$ is the stoichiometric matrix, and m is the number of reactants in the system. The matrix S stores the stoichiometric coefficients associated with each flux in the network. In the above formulation, both internal fluxes and boundary fluxes, which transport material into or out of the system, are included in S. Typically, a number of inequalities are introduced to constrain the boundary (also called injection) fluxes depending upon the external media (Edwards and Palsson, 2000a, b, c; Edwards et al., 2001; Ramakrishna et al., 2001),

$$\alpha_i \le J_i \le \beta_i, \tag{2}$$

where α_i and β_i represent the upper and lower bounds on the flux J_i.

As recently implemented with significant success (Bonarius et al., 1997; Edwards et al., 2001; Ramakrishna et al., 2001; Schilling et al., 2001), biological networks are assumed to optimize a certain biologically meaningful objective function, which is a linear combination of the fluxes,

$$Z = \sum_{i=1}^{n} c_i J_i = c^T J. \tag{3}$$

For example, in the analysis of *E. coli* intermediary metabolism, the objective is constructed from the production of biosynthetic precursors required to generate biomass (Edwards and Palsson, 2000a). Another example uses maximization of ATP production to simulate mitochondrial energy metabolism (Ramakrishna et al., 2001). Regardless of the application, optimization of a linear objective function (Eq. 3), together with the equality constraints (Eq. 1) and the inequality constraints (Eq. 2), represents a linear optimization problem, which can be solved with linear programming (Strang, 1986).

The power of FBA is illustrated by the tour de force assembly of the complete stoichiometry of the known reactions in *E. coli* metabolism provided by Edwards and Palsson (2000a, b, c). Edwards and Palsson show that the flux balance simulation of the organism-scale metabolic network predicts the metabolic capabilities of *E. coli*. Under various external constraints (e.g., aerobic versus anaerobic growth), FBA can distinguish between genes essential and not essential for growth in 68 of 79 cases studied (Edwards and Palsson, 2000a). This predictive capability of FBA is strik-

Submitted September 6, 2001 and accepted for publication February 4, 2002.

Address reprint requests to Daniel A. Beard, Box 352255, Univ. of Washington, Seattle, WA 98195. Tel: 206-685-9891; Fax: 206-685-3300; E-mail: dbeard@bioeng.washington.edu.

0006-3495/02/07/79/08 $2.00

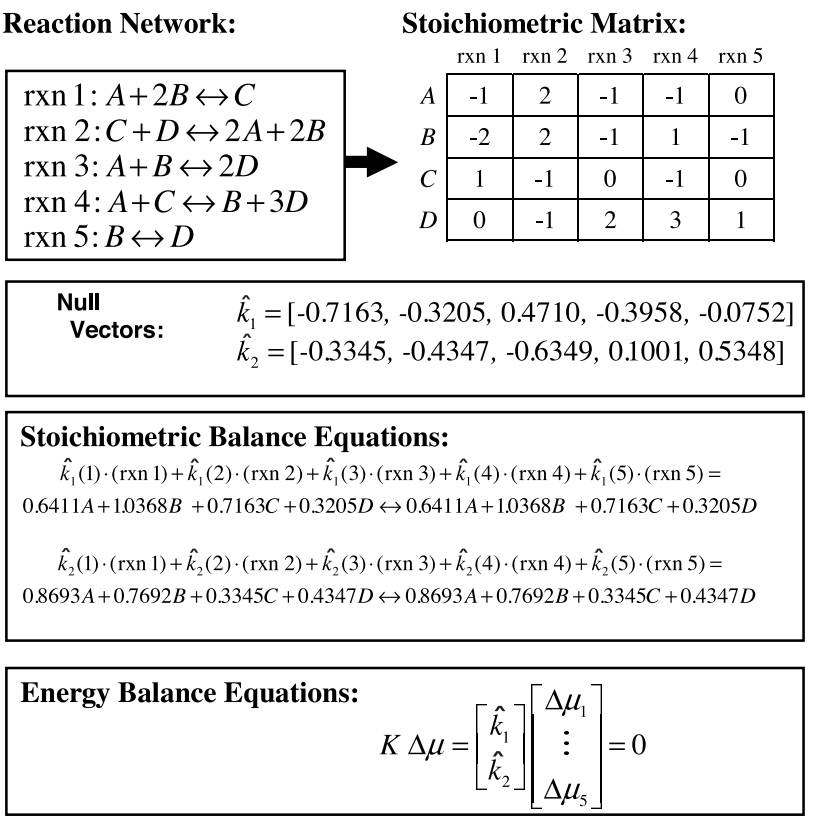

FIGURE 1 Illustration of energy-balance equations for a network of five reactions. The first step is determination of the stoichiometric matrix from the reactions in the network. For this example, the stoichiometric matrix has rank $r = 3$, resulting in two independent null-space vectors. The null-space vectors provide mutually orthogonal solutions to $S'J = 0$. In addition, the null-space vectors are the rows of the matrix K, the energy-balance matrix for which $K\Delta\mu = 0$.

ing considering the tremendously complex problems that can be modeled using little or no free parameters.

THEORY OF ENERGY BALANCE ANALYSIS

Lacking from FBA is the explicit consideration of the energy balance and thermodynamics of the network reactions. Because biochemical networks are composed of multiple-species reactions, energy-balance loop equations cannot be obtained from topological loops in the network, as is done in electrical circuit analysis (Balabanian and Bickart, 1981). We will show that the energy-balance equations are obtained from an analysis of the network stoichiometry.

If we combine redundant fluxes and remove the columns from S that correspond to boundary fluxes, the remaining matrix, denoted S', represents the complete set of possible internal chemical transformations. Using the singular value decomposition (Strang, 1986), S' can be decomposed as

$$S' = A\Lambda B^{\mathrm{T}}, \tag{4}$$

where Λ has the form,

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \lambda_{\mathrm{m}} & 0 & \cdots & 0 \end{bmatrix} \tag{5}$$

The first m columns of Λ form a diagonal matrix where the diagonal entries are the singular values λ_{i}. The entries of the remaining columns are all equal to zero. If n_{c} is the number of columns of S' and r is the rank of S', then columns $r + 1$ through n_{c} of the matrix B provide the $(n_{\mathrm{c}} - r)$-dimensional null space of S'. We introduce the matrix K, which stores the null space vectors of S',

$$K = \begin{bmatrix} B_{1,\mathrm{r}+1} & \cdots & B_{\mathrm{n_c},\mathrm{r}+1} \\ \vdots & & \vdots \\ B_{1,\mathrm{n_c}} & \cdots & B_{\mathrm{n_c},\mathrm{n_c}} \end{bmatrix}, \tag{6}$$

and we denote the ith row of K as $\hat{k}_{\mathrm{i}}$. Previous analyses based on the null space of stoichiometric matrices are presented by Alberty (1991) and Schilling and Palsson (1998).

Summing the n_{c} chemical reactions, each scaled by the corresponding entry of $\hat{k}_{\mathrm{i}}$, results in a perfectly balanced reaction equation (with same reactants in equal proportion on either side of the equation). An example of this analysis for a simple five-metabolite network with four reactants is illustrated in Fig. 1.

To study network thermodynamics, we consider the vector $\Delta\mu$ that lists the n_{c} chemical potential differences associated with the reaction fluxes. Because each $\hat{k}_{\mathrm{i}}$ provides weights for exactly balancing the chemical reaction equations (see Fig. 1), solutions to the equation

$$K\Delta\mu = 0 \tag{7}$$

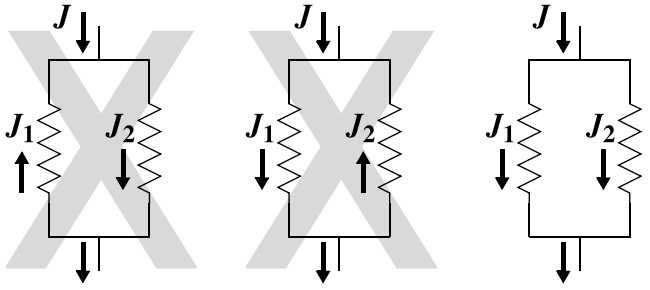

FIGURE 2 Electric circuit analogy for energy balance analysis. EBA tells us that the currents through the two parallel passive resistors must be in the same direction. Therefore, the first two possibilities, with current flow in opposite directions, are excluded.

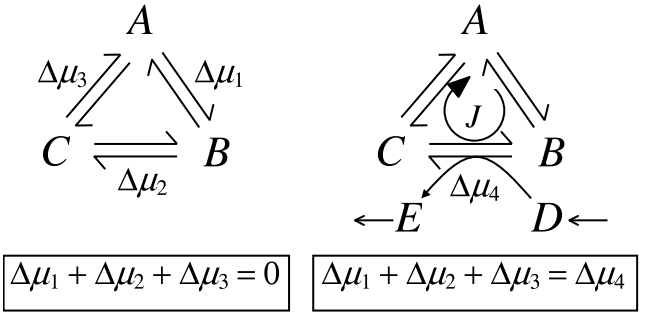

FIGURE 3 Energy balance for simple loops. The cycle $A\Delta\mu_1 B\Delta\mu_2 C\Delta\mu_3 A$ is illustrated in the left panel. There can be no steady-state flux for this cycle because the chemical potentials $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu_3$ must have the same sign for the flux to be nonzero, while energy balance maintains $\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 = 0$. When a driving force is added on the right panel, the energy-balance equation becomes $\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 = \Delta\mu_4$, and a nonzero cycle flux is possible.

balance the global potential energy of the network. The second law of thermodynamics takes the form of an inequality constraint for each flux,

$$J_i \cdot \Delta\mu_i < 0, \tag{8}$$

which ensures that entropy production is positive for each reaction (Qian, 2001a, b). Eqs. 7 and 8 represent thermodynamic constraints that should be considered in addition to the flux-balance constraints. Eq. 8 provides a link that couples mass balance and energy balance, and constraints the feasible flux space more strictly than Eqs. 1 and 2 alone.

The relationship between EBA and FBA is illustrated in Fig. 2 with an analogy to electric circuit analysis. Flux balance (or Kirchoff's current law) constrains the current through the two parallel passive resistors to sum to $J_1 + J_2 = J$, the total current into the circuit. Applying FBA alone, current can travel in either direction in either branch of the circuit. However, if we apply Kirchoff's loop law (energy balance), we find that the first two cases in Fig. 2 must be excluded, and the currents J_1 and J_2 must travel in the same direction.

NONEQUILIBRIUM STEADY-STATE THERMODYNAMICS

In studying network thermodynamics, it is essential to differentiate between free energy of equilibrium and the chemical potential of a nonequilibrium steady state (NESS). Although the basic thermodynamic theory for the equilibrium case can be found in a textbook, it is only recently that the theory for nonequilibrium steady states has been cogently established. See Qian (2001a, b).

Consider the following examples of how chemical potential $\Delta\mu$ is balanced in a NESS system. For a given NESS reaction, $A \rightarrow B$, there is a nonzero flux J and nonzero $\Delta\mu = \Delta\mu^o + k_B T \ln([B]/[A])$, where k_B represents Boltzmann's constant and T is absolute temperature. For each molecule of A that reacts to form B, $\Delta\mu$ equals the change in internal energy plus the change in entropy. The internal energy change is also exactly the amount of heat dissipated (or absorbed). Now consider a cycle, or loop,

$$A \xrightarrow{\Delta\mu_1} B \xrightarrow{\Delta\mu_2} C \xrightarrow{\Delta\mu_3} A$$

(see Fig. 3). The summation of chemical potential $\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 = 0$ is obvious. Alternatively, if the reaction $B \rightarrow C$ is coupled to the reaction

$$D \xrightarrow{\Delta\mu_4} E$$

(i.e., the full reaction is B + D→C + E) with clamped (externally fixed) concentrations for D and E, then the summation of chemical potential gives $\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 = \Delta\mu_4$.

When a molecule undergoes the reaction $A \rightarrow B$, with fixed concentrations [A] and [B], then the amount of heat dissipated per turnover is precisely equal to the amount of work required to sustain constant concentrations [A] and [B]. For a given flux J, total heat dissipation is $-J \cdot \Delta\mu$, which is always positive. When the energy balance equation for the loop illustrated in the right panel of Fig. 3 is multiplied by the flux, we obtain the equation $J\Delta\mu_1 + J\Delta\mu_2 + J\Delta\mu_3 = J\Delta\mu_4$, which represents the balance between energy dissipated by the circuit and energy supplied to the system. The energy supplied by pumping D (and E) molecules into (and out of) the system is given by $J\Delta\mu_4$, which is dissipated by the heat associated with the cycle flux, $J\Delta\mu_1 + J\Delta\mu_2 + J\Delta\mu_3$. For a reaction cycle (or resistance loop), the amount of heat dissipated by all of the components is equal to the energy supplied by concentration clamping (or the battery in an electrical circuit).

To determine effective loops in a more complex reaction network, we turn to the null space decomposition of the stoichiometric matrix. The null space decomposes the internal reactions in the system into generalized loops with no net change in stoichiometry from the left side to the right side of the summed reaction (see Fig. 1). For the example in Fig. 1, summing the five reactions, weighted by the components of the first null space vector for this system, yields the exactly balanced reaction,

$$0.6411A + 1.0368B + 0.7163C + 0.3205D \leftrightarrow 0.6411A + 1.0368B + 0.7163C + 0.3205D.$$

The $\Delta\mu$ for this reaction is obviously identically equal to zero, regardless of the concentrations of the reactants. Thus, Eq. 7 follows directly from the definition of chemical potential.

APPLICATION OF ENERGY BALANCE ANALYSIS

The EBA constraints are summarized as follows: For a flux vector J to be considered thermodynamically feasible, there must exist a vector $\Delta\mu$, for which $K\Delta\mu = 0$; and $\Delta\mu_i \cdot J_i < 0$, for all $\Delta\mu_i$ and J_i, where $\Delta\mu_i$ and J_i represent the ith entries of $\Delta\mu$ and J, respectively, and the rows of the matrix K correspond to a complete basis of null-space vectors of the stoichiometric matrix of internal reactions S'. The EBA constraints must be applied in addition to the FBA constraints to ensure that the predicted flux vector is thermodynamically feasible.

As an example, we return to the simple reaction network illustrated in Fig. 1, and consider the following problem: Determine the maximum steady-state production of reactant D, for a given set of maximal input fluxes of reactants A, B, and C. (Three related problems are considered below. The first problem assumes that reactant A is the only available input substrate. For this case, the maximal input fluxes of A, B, and C are set to be 1, 0, and 0 mmol sec^{-1}, respectively. The remaining two examples use B only and C only as input substrates.)

Optimal fluxes are obtained using the MATLAB (The Mathworks Inc., Natick, MA) software package. Production of reactant D is used as an objective function for optimization of flux vectors. For predictions based on FBA alone, linear programming is implemented as described by Edwards and Palsson (2000a). To implement both the linear FBA constraints and the nonlinear EBA constraints, we use the nonlinear constrained optimization package available in the MATLAB optimization package.

Predicted fluxes obtained using FBA, and FBA combined with EBA, are presented in Table 1. All fluxes are reported in units of mmol sec^{-1}. The predicted optimal fluxes obtained with FBA alone are thermodynamically infeasible, because no feasible potential vector exists for these fluxes. When EBA and FBA are applied simultaneously, thermodynamically feasible mass-balanced fluxes are predicted.

TABLE 1 Comparison of fluxes predicted by FBA and EBA for the example system illustrated in Fig. 2

	A input		*B* input		*C* input	
Identifier	FBA	FBA/EBA	FBA	FBA/EBA	FBA	FBA/EBA
Reaction 1	−0.060	0	0.094	0	−0.154	0
Reaction 2	−0.360	0	0.193	0	0.445	1
Reaction 3	0.040	1	0.392	0	0.643	2
Reaction 4	0.300	0	−0.100	0	0.401	0
Reaction 5	−0.334	−1	0.707	1	0.956	0
Input of *A*	1	1	0	0	0	0
Input of *B*	0	0	1	1	0	0
Input of *C*	0	0	0	0	1	1
Output of *D*	1	1	1	1	3	3

Columns labeled "FBA/EBA" refer to fluxes obtained with both the EBA and FBA fluxes implemented. The predicted internal fluxes for reactions 1–5, input fluxes of *A*, *B*, and *C*, and output flux of *D*, for three computational experiments are reported. For all cases, production of *D* is maximized under the following boundary constraints.

For "*A* input", the input of *A* is limited to be less than or equal to 1, and the inputs of *B* and *C* are limited to be less than or equal to 0. Similarly, for "*B* input" and "*C* input", the maximal input fluxes of A, B, and C, are (0, 1, 0), and (0, 0, 1), respectively. All fluxes are reported in units of mmol/sec^{-1}.

Because we consider the EBA constraints in addition to the FBA mass balance constraints, EBA reduces the size of the feasible flux space. However, we do find more than one optimal solution when we maximize the production of D. By varying the initial guess, the optimization routine arrives at different feasible solutions that maximize production. Therefore the fluxes reported in Table 1 do not represent unique solutions to the optimization problem, either for the FBA cases or for the FBA/EBA cases.

ANALYSIS OF *ESCHERICHIA COLI* METABOLISM

We have obtained the stoichiometric matrix provided by Edwards and Palsson (2000c), used to represent the flux balances in the *E. coli* metabolism. The web-posted supplementary information provides detailed descriptions of the reaction network, which contains 953 fluxes (735 internal; 218 boundary) acting on 536 metabolites. Again using the MATLAB (The Mathworks) optimization package, we reproduced the linear programming analysis presented in Edwards and Palsson (2000a), and optimized the production of biomass with glucose and oxygen uptake constrained to be less than or equal to 10 and 15 mmol g-$DW^{-1}h^{-1}$, respectively. The resulting flux produces a growth rate of 0.81 h^{-1} on glucose minimal media.

To compute the thermodynamic properties of this large-scale network, we first combine redundant fluxes (e.g., the phosphofructokinase A and B reactions are combined into a single column of S'), resulting in $n_c = 617$ internal reactions operating on 536 metabolites. The growth is then optimized under the flux balance constraints (Eqs. 1 and 2) and the constraint that the energy-balance equations (Eqs. 7 and 8)

are satisfied. We impose the additional constraint that $\Delta\mu$ must be finite for all nonzero fluxes: the free energy can go to zero only if the flux is zero, implying that a given reaction is in chemical equilibrium. This analysis predicts the same optimal growth rate (0.81 h^{-1} on glucose minimal media) as that reported above for FBA. Yet, as for the simple example illustrated above, the FBA prediction does not represent a unique solution to the optimization problem because redundancies in the metabolic network allow the optimal growth rate to occur for an infinite number of possible internal flux distributions. Because the EBA-constrained solution is (at least in this case) able to achieve the same optimal growth rate as that obtained by considering only the FBA constraints, the fluxes obtained by EBA represent one particular optimal solution to the FBA linear programming problem. However, optimal flux distributions obtained by considering flux balance alone are not necessarily thermodynamically feasible. The EBA constraints further restrict the set of feasible fluxes, and provide a more physically realistic flux distribution. Selected fluxes from glycolysis, TCA cycle, and respiration are tabulated in Fig. 4. Fluxes from the wild type (WT) on glucose media under aerobic and anaerobic conditions are labeled WT (oxygen) and WT (anaerobic), respectively.

ESTIMATION OF REACTION POTENTIALS

Beyond providing the constraints necessary to produce thermodynamically feasible fluxes, EBA may make it possible to make quantitative estimates of the chemical potentials in the system. As a first attempt to approximately identify the chemical potentials, we introduce the "flux resistances," defined as $r_i = -\Delta\mu_i/J_i$. To avoid the unphysical situation in which $\Delta\mu_i = 0$ for a finite J_i, which is equivalent to setting the flux resistance equal to zero, we assume that there exists a minimum flux resistance, r_{min} (which is equivalent to saying that there exists an upper limit to the effective reaction rate constants). Thus, the second law constraint is modified:

$$\Delta\mu_i \leq -r_{min}J_i \quad J_i > 0,$$
$$\Delta\mu_i \geq +r_{min}J_i \quad J_i < 0. \tag{9}$$

Realistically, each flux may have a different r_{min}. However, for our purposes, we find that a single value, $r_{min} = 400$ kcal mol^{-2}g-DW h, produces reasonable behavior and can be used to describe the entire network. The energy balance constraint can alternatively be written in terms of the chemical potentials (Eq. 7) or the flux resistances,

$$[K]\begin{bmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_{n_c} \end{bmatrix}\begin{bmatrix} r_1 \\ \vdots \\ r_{n_c} \end{bmatrix} = K \cdot \mathrm{diag}(J) \cdot r = 0. \tag{10}$$

To estimate the chemical potentials, we assume that the biological system tends to operate in a steady-state regime where the magnitudes of the chemical potentials are minimized. With the fluxes fixed using the values obtained from the above analysis, we use quadratic programming to find an optimal $\Delta\mu$ that minimizes the norm of the chemical potential vector $|\Delta\mu|^2$. In addition to the fluxes, Fig. 4 lists the predicted chemical potentials and the conductances, $c_i = r_i^{-1}$, of each reaction. The choice of $r_{min} = 400$ kcal mol^{-2}g-DW h results in reasonable predictions of chemical potential. For example, $\Delta\mu_i = -9.55$ kcal mol^{-1} for ATP hydrolysis. Small changes to the assumed value of r_{min} result in proportional changes in all of the predicted reaction potentials, while the relative values remain fixed.

The reaction conductance provides a measure of the activation level of the pathway. If $c_i = 0$, then the associated enzyme(s) is not present or is deactivated. Increases in flux or conductance, at a fixed chemical potential, indicate that a pathway is up regulated at either the expression level or the post-translational level. By changing the boundary constraints, we can study how the metabolic network responds to changes in substrate. In Fig. 4, we compare the predicted EBA fluxes and potentials for the WT cell grown under aerobic and anaerobic conditions. We identify a pathway as "down regulated" (colored blue) if the flux conductance decreases by a factor of 4 or more, and "up regulated" (colored green) when the conductance increases by more than a factor of 4. On the basis of this analysis, we identify three enzymes that require activation upon moving from aerobic to anaerobic conditions: fumarate reductace, pyruvate formate lysase, and pyridine nucleotide transhydrogenase. Other pathways show increases or decreases in flux compared to aerobic growth. However, these differences can be accounted for by changes in the potential and thus do not necessarily require regulation.

Following the work of Edwards and Palsson (2000a, b, c), we studied the effects of gene knockouts on the metabolic capabilities of *E. coli*. We found that *zwf*, *pgl*, and *gnd* knockouts result in the same predicted phenotype (Fig. 4), with only two up regulations compared to WT:succinate dehydrogenase and pyridine nucleotide transhydrogenase. Again, a number of predicted fluxes that differ from the WT do not require up regulation of the associated enzymes. The situation is similar for *pyk* and *pgi* knockouts. The *pyk* knockout requires an up regulation of phosphoenolpyruvate carboxylase to maintain growth of almost 99% that predicted for WT. No other significant regulations are predicted. The *pgi* knockout analysis predicts one significant up regulation and three down regulations and a similar growth rate. Thus, we find that much of the capacity for metabolic control is built in to the WT expression and activation levels. Few regulatory steps are required when nonessential genes are knocked out; the network is robust and tolerant to errors and deletions (Jeong et al., 2000).

However, the situation is different when essential genes are knocked out. The *eno* and *pfk* genes are examples of genes that FBA falsely identifies as nonessential to growth

Enzyme	Gene(s)	WT (oxygen) flux	$\Delta\mu_i$	c	WT (anaerobic) flux	$\Delta\mu_i$	c	*zwf* (or *pgl* or *gnd*) flux	$\Delta\mu_i$	c	*pyk* flux	$\Delta\mu_i$	c	*pgi* flux	$\Delta\mu_i$	c	*eno* flux	$\Delta\mu_i$	c	*pfk* flux	$\Delta\mu_i$	c
Glucose phosphotransferase	*ptsG*	10	-4	2.5	10	-4	2.5	10	-4	2.5	10	-4	2.5	10	-4	2.5	0	-5.644	0	1.341	-0.536	2.5
Glucokinase	*glk*	0	-7.008	0	0	-3.765	0	0	-8.369	0	0	-4.324	0	0	-4.136	0	10	-4	2.5	5.175	-2.851	1.815
PGI (G-6-P / F-6-P)	*pgi*	2.330	-1.208	1.929	3.091	-5.485	0.563	3.084	-5.417	0.569	1.205	-1.877	0.642	0	-7.398	0	-0.332	0.538	0.617	-3.791	6.152	0.616
PGI (bDG-6-P / G-6-P)	*pgi*	-1.510	0.604	2.5	-6.856	2.742	2.5	-6.772	2.70	2.5	-2.347	0.938	2.5	0	3.699	0	0.673	-0.269	2.5	7.690	-3.076	2.5
PGI (bDG-6-P / F-6-P)	*pgi*	1.510	-0.604	2.5	6.856	-2.742	2.5	6.772	-2.70	2.5	2.347	-0.938	2.5	0	-3.699	0	-0.673	0.269	2.5	-7.690	3.076	2.5
Glucose-1-phosphatase	*agp*	0	-1.270	0	0	-3.672	0	0	-1.92	0	0	-0.951	0	0	-0.955	0	0	-6.604	0	0.0006	-1.805	3e-4
Phosphofructokinase	*pfkB*	7.147	-2.859	2.5	9.701	-3.880	2.5	9.178	-3.671	2.5	7.063	-2.825	2.5	5.881	-4.872	1.206	5.759	-13.36	0.430	0	-6.462	0
Fructose-1,6-bisphosphatase	*fbp*	0	-6.690	0	0	-7.230	0	0	-8.552	0	0	-3.400	0	0	-1.173	0	0	-3.844	0	0	0	Eq.
Fructose-1,6-bp aldolase	*fba*	7.147	-2.859	2.5	9.701	-3.880	2.5	9.178	-3.671	2.5	7.063	-2.825	2.5	5.881	-3.699	1.589	5.759	-9.519	0.605	0	-19.26	0
Triosphosphate isomerase	*tpia*	-7.056	2.822	2.5	-9.668	3.867	2.5	-9.089	3.635	2.5	-6.973	2.789	2.5	-5.791	3.699	1.565	-5.693	9.519	0.598	0.059	-0.023	2.5
GAP-dehydrogenase	*gapAC1C2*	15.77	-6.309	2.5	19.21	-7.686	2.5	17.84	-7.139	2.5	15.71	-6.284	2.5	14.53	-5.812	2.5	14.77	-5.910	2.5	5.627	-19.26	0.292
Phosphoglycerate kinase	*pgk*	15.77	-6.309	2.5	19.21	-7.686	2.5	17.84	-7.139	2.5	15.71	-6.284	2.5	14.53	-5.812	2.5	14.77	-5.910	2.5	5.627	-19.26	0.292
Phosphoglycerate mutase	*gpmAB*	14.32	-5.728	2.5	18.73	-7.494	2.5	16.44	-6.576	2.5	14.27	-5.711	2.5	13.19	-5.276	2.5	0	0	Eq.	5.310	-9.334	0.568
Enolase	*eno*	14.32	-5.728	2.5	18.73	-7.494	2.5	16.44	-6.576	2.5	14.27	-5.711	2.5	13.19	-5.276	2.5	0	9.999	0	5.310	-9.334	0.568
Phoshphoenolpyr. synthetase	*ppsA*	0	-4.379	0	0	-7.760	0	0	-4.965	0	0	-2.828	0	0	-2.082	0	2.152	-0.861	2.5	0.0008	-0.000	2.5
Pyruvate kinase	*pykA,F*	1.314	-0.804	1.634	7.588	-3.035	2.5	3.530	-1.412	2.5	0	-2.528	0	0.236	-2.729	0.086	0	-14.85	0	0.0023	-1.757	0.001
Pyruvate dehydrogenase	*lpdA, aceEF*	9.236	-3.694	2.5	0	-2.403	0	11.51	-4.607	2.5	9.257	-3.703	2.5	5.046	-2.018	2.5	0	0	Eq.	0	-11.28	0
Glucose-1-P adenyltransferase	*glgC*	0.132	-2.827	0.046	0.046	-3.578	0.013	0.128	-3.467	0.036	0.130	-1.725	0.075	0.13	-1.666	0.078	0.0947	-2.512	0.037	0.086	-0.034	2.5
Glycogen synthase	*glgA*	0.132	-2.827	0.046	0.046	-3.578	0.013	0.128	-3.467	0.036	0.130	-1.725	0.075	0.13	-1.666	0.078	0.0947	-2.512	0.037	0.086	-0.034	2.5
Glycogen phosphorylase	*glgP*	0	-2.827	0	0	-3.578	0	0	-3.467	0	0	-1.725	0	0	-1.666	0	0	-2.512	0	0.0007	-0.000	2.5
Glucose-6-P dehydrogenase	*zwf*	6.010	-2.404	2.5	0	-6.682	0	0	-4.401	0	6.301	-2.520	2.5	9.854	-3.941	2.5	10.9	-4.360	2.5	17.90	-7.161	2.5
6-Phosphogluconolactonase	*pgl*	6.010	-2.404	2.5	0	-6.682	0	0	-4.401	0	6.301	-2.520	2.5	9.854	-3.941	2.5	10.9	-4.360	2.5	17.90	-7.161	2.5
6-Phosphogluconate dehydrog.	*gnd*	6.010	-2.404	2.5	0	-6.682	0	0	-4.401	0	6.301	-2.520	2.5	9.854	-3.941	2.5	10.9	-4.360	2.5	17.90	-7.161	2.5
Ribose-5-P isomerase	*rpiAB*	2.630	-1.052	2.5	0.220	-2.227	0.099	0.607	-1.467	0.413	2.718	-1.087	2.5	3.901	-1.560	2.5	4.082	-1.633	2.5	6.373	-2.549	2.5
Ribose-5-P 3-epimerase	*rpe*	3.361	-1.344	2.5	-0.227	0.091	2.5	-0.625	0.250	2.5	3.564	-1.425	2.5	5.934	-2.373	2.5	6.804	-2.721	2.5	11.51	-4.606	2.5
Transketolase rxn. 1	*tktAB*	1.853	-0.741	2.5	-0.052	0.021	2.5	-0.145	0.058	2.5	1.952	-0.781	2.5	3.137	-1.255	2.5	3.526	-1.410	2.5	5.870	-2.348	2.5
Transketolase rxn. 2	*tktAB*	1.507	-0.603	2.5	-0.174	0.069	2.5	-0.480	0.192	2.5	1.611	-0.644	2.5	2.797	-1.119	2.5	3.278	-1.311	2.5	5.646	-2.258	2.5
Transaldolase	*talB*	1.842	-0.736	2.5	-0.056	0.022	2.5	-0.156	0.062	2.5	1.941	-0.776	2.5	3.126	-1.250	2.5	3.517	-1.407	2.5	5.863	-2.345	2.5
Citrate synthase	*gltA*	1.420	-3.184	0.445	0.310	-0.124	2.5	3.235	-1.294	2.5	1.244	-2.528	0.492	0.866	-5.320	0.162	0.631	-12.09	0.052	0.570	-16.20	0.035
Aconitase	*acnAB*	1.420	-3.184	0.445	0.310	-0.124	2.5	3.235	-1.294	2.5	1.244	-2.528	0.492	0.866	-5.320	0.162	0.631	-12.09	0.052	0.570	-16.20	0.035
Isocitrate dehydrogenase	*icdA*	1.420	-0.568	2.5	0.310	-0.124	2.5	3.235	-1.294	2.5	1.244	-1.188	1.046	0.866	-1.926	0.449	0.631	-5.001	0.126	0.570	-9.867	0.057
2-Ketoglutarate dehydrogenase	*sucAB, lpdA*	0.538	-0.215	2.5	0	0	eq.	2.381	-0.952	2.5	0.375	-0.150	2.5	0	-0.983	0	0	0	Eq.	0	-7.729	0
Succinyl-CoA synthetase	*sucCD*	0.153	-0.061	2.5	-0.135	7.154	0.019	2.008	-0.803	2.5	-0.009	0.659	0.014	-0.378	0.151	2.5	-0.275	1.531	0.179	-0.249	0.0997	2.5
Succinate dehydrogenase rxn 1	*sdhABCD*	0.538	-3.184	0.169	0	0.035	N.P.	2.381	-0.952	2.5	0.375	-2.528	0.148	0	0	Eq.	0	-12.09	0	0	0	Eq.
Fumarate reductase	*frdABCD*	0	3.184	N.P.	0.089	-0.035	2.5	0	0.952	N.P.	0	2.528	N.P.	0	0	N.P.	0	12.09	N.P.	0	0	N.P.
Fumarase	*fumABC*	1.417	-0.566	2.5	0.220	-0.088	2.5	3.232	-1.293	2.5	1.241	-0.496	2.5	0.863	-0.345	2.5	0.629	-0.251	2.5	0.577	-0.871	0.662
Malate dehydrogenase	*mdh*	1.417	-0.566	2.5	0.220	-0.088	2.5	3.232	-1.293	2.5	-0.068	2.107	0.032	0.863	-0.345	2.5	0.629	-0.251	2.5	0.576	-5.774	0.099
D-lactate dehydrog. (pyr -> lac)	*Dld, ldhA*	0	-22.55	0	0	-5.062	0	0	-26.86	0	0	-16.77	0	0	-16.53	0	0	-36.07	0	0.001	-16.71	6e-5
D-lactate dehydrog. (cyto.)	*dld*	0	-22.55	0	0	-5.062	0	0	-26.86	0	0	-16.77	0	0	-16.53	0	0	-36.07	0	0.001	-16.71	6e-5
Acetaldehyde dehydrogenase	*adhE*	0	0	Eq.	7.854	-3.141	2.5	0	0	Eq.	0	0	Eq.	0	0	Eq.	0	0	Eq.	0	0	2.495
Pyruvate formate lysase	*pflABCD*	0	-8.704	0	16.85	-6.742	2.5	0	-10.27	0	0	-8.291	0	3.149	-5.864	0.537	10.09	-7.433	1.358	0	-8.729	0
Formate hydrogen lysase	*fdhF, hycBE*	0.027	-0.011	2.5	8.243	-3.297	2.5	0.013	-0.005	2.5	0.013	-0.005	2.5	1.059	-0.423	2.5	4.309	-1.723	2.5	0	0	Eq.
Phosphotransacetylase	*pta*	5.146	-2.058	2.5	7.750	-3.100	2.5	5.698	-2.279	2.5	5.382	-2.153	2.5	4.705	-1.882	2.5	7.552	-3.021	2.5	-0.280	3.2245	0.086
Acetate Kinase	*ackA, pntT*	5.146	-2.058	2.5	7.750	-7.320	1.058	5.698	-2.279	2.5	5.382	-2.153	2.5	4.705	-1.882	2.5	7.552	-3.021	2.5	-0.280	3.2245	0.086
Acetyl-CoA synthetase	*acs*	0	0	Eq.	0	0	Eq.	0	0	Eq.	0	0	Eq.	0	0	2.5	0	0	Eq.	0.0006	-1.813	0.000
Phosphoenolpyr. carboxykinase	*pckA*	0	-3.403	0	0	0	Eq.	0	-3.406	0	0	-3.972	0	0	-2.861	0	0	0	Eq.	0.0006	-4.421	1e-4
Phosphoenolpyr. carboxylase	*ppc*	2.300	-6.145	0.374	0.899	-11.11	0.081	2.227	-8.816	0.252	3.583	-2.253	1.589	2.260	-3.185	0.709	1.647	-17.20	0.095	3.510	-2.041	1.720
Malic enzyme (NADP)	*maeB*	0	0	eq.	0	-1.677	0	0	0	2.5	1.3105	-0.524	2.5	0	0	Eq.	0	-6.501	0	0.0005	-0.000	2.5
Malic enzyme (NAD)	*maeA*	0	-4.774	0	0	-3.123	0	0	-6.111	0	0	-4.393	0	0	-5.935	0	0	-15.10	0	0.0005	-11.95	4e-4
Isocitrate lysase	*aceA*	0	-3.349	0	0	0	Eq.	0	-5.333	0	0	-1.339	0	0	-3.393	0	0	-7.088	0	0	-6.335	0
Malate synthase A	*aceB, glcB*	0	-3.349	0	0	0	Eq.	0	-5.333	0	0	-1.339	0	0	-3.393	0	0	-7.088	0	0	-6.335	0
Inorganic pyrophosphatase	*ppa*	2.666	-1.066	2.5	0.939	-0.375	2.5	2.581	-1.819	1.418	2.626	-1.050	2.5	2.619	-1.047	2.5	1.929	-9.672	0.199	1.733	-6.392	0.271
NADH dehydrogenase II	*ndh*	0	-45.10	0	0	-10.12	0	0	-53.73	0	0	-33.54	0	0	-33.06	0	0	-72.15	0	0.0007	-33.43	2e-5
NADH dehydrogenase I	*nuoABEFG*	29.20	-11.68	2.5	0	0	Eq.	27.37	-10.94	2.5	29.37	-11.75	2.5	29.75	-11.9	2.5	29.81	-11.92	2.5	29.83	-11.93	2.5
Formate dehydrogenase-N,O	*fdnGHI*	0	-20.99	0	0	0	Eq.	0	-23.61	0	0	-16.50	0	0	-17.12	0	0	-30.30	0	0	-23.70	0
Glycerol-3-P dehydrogenase	*fdoIHG*	0	-24.94	0	0	-5.785	0	0	-29.92	0	0	-18.70	0	0	-19.5	0	0	-40.38	0	0.0007	-22.69	3e-5
Cytochrome oxidase bo3	*cyoABCD,c*	30	-12	2.5	0	0	Eq.	30	-12	2.5	30	-12	2.5	30	-12	2.5	30	-12	2.5	29.99	-11.99	2.5
Cytochrome oxidase bd	*cydABCE,a*	0	-16.77	0	0	-1.446	0	0	-18.11	0	0	-15.11	0	0	-15.02	0	0	-20.60	0	0.005	-15.06	3e-4
Succinate dehydrogenase rxn 2	*sdhABCD*	0.545	-25.06	0.021	-0.086	0.034	2.5	2.388	-30.24	0.079	0.383	-18.31	0.020	0.007	-20.57	0.000	0.005	-43.50	0.000	0.006	-29.39	2e-4
Thioredoxin reductase	*trxB*	0.257	-0.102	2.5	0.090	-0.036	2.5	0.249	-0.099	2.5	0.253	-0.101	2.5	0.252	-0.101	2.5	0.184	-0.073	2.5	0.166	-0.066	2.5
PNT 1	*pntAB*	0	0	Eq.	0	0	Eq.	0	0	Eq.	1.891	-0.756	2.5	7.281	-2.912	2.5	0	0	Eq.	22.20	-8.881	2.5
PNT 2	*pntAB*	0	-14.32	0	4.455	-4.339	1.026	9.779	-18.33	0.533	0	-8.582	0	0	-6.157	0	0.927	-25.81	0.035	0.001	-0.331	0.004
ATP hydrolysis (maintenance)		7.6	-9.549	0.795	7.6	-11.11	0.684	7.6	-12.22	0.621	7.6	-6.226	1.220	7.6	-6.046	1.256	7.6	-17.20	0.441	7.6	-6.462	1.176
Growth		**0.81**	**-0.32**	**2.5**	**0.285**	**-0.11**	**2.5**	**0.784**	**-0.31**	**2.5**	**0.798**	**-0.32**	**2.5**	**0.796**	**-0.32**	**2.5**	**0.58**	**-0.23**	**2.5**	**0.524**	**-0.21**	**2.5**

Up Regulated
Down Regulated
Reverse Direction
Knock Out

FIGURE 4 Regulation of reactions in glycolysis, TCA cycle, and respiration. Predicted biochemical fluxes (in μmol g-DW^{-1}h^{-1}kcal^{-1}), chemical potentials (in kcal mol^{-1}), and conductances (in 10^{-3} g-DW^{-1}h^{-1}kcal^{-1}) are reported for 64 central metabolic reactions, for WT aerobic and anaerobic

on glucose minimal media (Edwards and Palsson, 2000a). The fact that *eno* and *pfk* are essential for growth in vivo (Fraenkel, 1996) may be related to the demands that the knockouts of these genes place on metabolic regulatory mechanisms. These demands are much heavier than those imposed by nonessential knockouts. Our analysis predicts that maintaining growth with these knockouts requires a greater number of pathway regulations than for the nonessential knockouts. Of the 64 reactions listed in Fig. 4, 15 are predicted to be up regulated and 15 down regulated for the *pfk* knockout. Thus, the predicted phenotype is markedly different from the WT for nearly half of the reactions associated with glycolysis, TCA cycle, and respiration, whereas the *pyk* and *pgi* knockouts differ very little from WT.

DISCUSSION AND CONCLUSIONS

The predictions of reaction potentials and conductivities presented above are based on one major simplification—that the entire network can be characterized by a single r_{min}, or equivalently a maximum pathway conductance. More realistically, constraints could be placed on each pathway based on a priori knowledge of the biochemistry. One promising aspect of EBA is that it is possible to incorporate as much, or as little, as is known about the individual reactions. For example, if we know that the ratio [ATP]/[ADP] in the cell is greater than the equilibrium ratio (Stryer, 1995), then we know that the free energy of ATP hydrolysis satisfies the inequality

$$\Delta\mu_{ATP\rightarrow ADP} \leq \Delta\mu^{o}_{ATP\rightarrow ADP} = -7.3 \text{ kcal mol}^{-1}. \quad (11)$$

In general, consider the reaction $A \rightarrow B$. If the ratio $[A]/[B]$ is greater than 1, then the potential energy of that reaction is constrained $\Delta\mu_i < \Delta\mu_i^o$. Alternatively, if the concentration ratio is less than 1, then $\Delta\mu_i > \Delta\mu_i^o$. If the reactant concentrations are known, then the constraint becomes an equality: $\Delta\mu_i = -k_B T \ln(K_{eq}[A]/[B])$, where K_{eq} is the equilibrium constant of the reaction.

The arbitrary parameter r_{min} was introduced to estimate reaction potentials. However, the EBA constraints used to determine the thermodynamic feasibility of a network flux do not depend on the choice of r_{min}. The feasible flux space allowed by EBA and FBA together is more restricted than the space predicted by FBA alone. Therefore, predictions made without considering energy balance in general do not satisfy the EBA constraints and are not physically realistic.

Together, flux balance and energy balance provide basic laws for the analysis of biochemical networks. These laws, akin to the basic principles of circuit theory for electrical networks, make the analysis and design of large-scale biochemical systems practical. The engineering approach to analysis and design of such complex systems is the identification of modular components that are separable within the system (Hartwell et al., 1999). Flux balance and energy balance analysis provide a basis for understanding how these modules function and interact.

After the completion of this work, we were reminded by Professor E. Selkov and a reviewer of the classic work of Oster, Perelson, and Katchalsky on network thermodynamics. Although we have not found any mentioning of null space and its relation to energy balance in the in work, we did discover that a paper by Oster and Perelson (1974a,b) on an algebreic topologic theory of chemical reaction dynamics contains a theorem that states a related result.

We thank Dr. J. B. Bassingthwaighte for constant insight and support and Dr. J. S. Edwards and Dr. B. O. Palsson for providing the *E. coli* metabolism stoichiometric matrix and for helpful discussion.

The work is in part supported by National Institute of Health grants NCRR-1243 and NCRR-12609, and National Aeronautics and Space Administration grant NCC2-5463 to H.Q.

REFERENCES

Alberty, R. A. 1991. Equilibrium compositions of solutions of biochemical species and heats of biochemical reactions. *Proc. Natl. Acad. Sci. U.S.A.* 88:3268–3271.

Balabanian, N., and T. A. Bickart. 1981. Linear Network Theory: Analysis, Properties, Design, and Synthesis. Matrix, Beaverton, OR.

Bonarius, H. P. J., V. Hatzimanikatis, K. P. J. Meesters, C. D. De Gooijer, G. Schmid, and J. Tramper. 1996. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnol. Bioeng.* 50:299–318.

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

Edwards, J. S., and B. O. Palsson. 2000a. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U.S.A.* 97:5528–5533.

Edwards, J. S., and B. O. Palsson. 2000b. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 16:927–939.

Edwards, J. S., and B. O. Palsson. 2000c. Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* 1:1; Supplementary data available at http://gcrg.ucsd.edu/supplementary_data/BP2000/main.htm.

Edwards, J. S., R. U. Ibarra, and B. O. Palsson. 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19:125–130.

Fraenkel, D. G. 1996. Glycolysis. In *Escherichia Coli* and *Salmonella* Cellular and Molecular Biology. vol. 1. F. C. Neidhardt, editor. ASM Press, Washington, DC. 189–198.

growth, and for *zwf, pgl, gnd, pyk, pgi, eno*, and *pfk* knockouts. Predicted growth rate for each case is reported in units of hour one. Green and blue shading indicate up and down regulation relative to WT, respectively. Yellow indicates that a reversible flux has changed direction. Gray shading indicates that the flux cannot be identified as a site of regulation because, although the conductance has changed moderately (by less than a factor of 10), the magnitudes of the potential and the flux have both either increased or decreased. "N.P." denotes enzymes that are not present under given conditions. "Eq." refers to reactions that are at equilibrium, for which the conductance cannot be estimated. All simulations are performed for glucose-minimal media.

Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray. 1999. From molecular to modular cell biology. *Nature.* 402:C47–C52.

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. 2000. The large-scale organization of metabolic networks. *Nature.* 407: 651–654.

Oster, G. F., and A. S. Perelson. 1974a. Chemical reaction dynamics. Part I: Geometrical structure. *Arch. Rational Mech. Anal.* 55:230–274.

Oster, G. F., and A. S. Perelson. 1974b. Chemical reaction dynamics. Part II: Reaction networks. *Arch. Rational Mech. Anal.* 57:31–98.

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

Qian, H. 2001a. Mathematical formalism for isothermal linear irreversibility. *Proc. R. Soc. Lond. A.* 457:1645–1655.

Qian, H. 2001b. Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. *Phys. Rev. E.* 65:016102, 1–5.

Ramakrishna, R., J. S. Edwards, A. McCulloch, and B. O. Palsson. 2001. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol.* 280: R695–R704.

Schilling, C. H., and B. O. Palsson. 1998. The underlying pathway structure of biochemical reaction networks. *Proc. Natl. Acad. Sci. U.S.A.* 8:4193–4198.

Schilling, C. H., S. Schuster, B. O. Palsson, and R. Heinrich. 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15:296–303.

Schilling, C. H., J. S. Edwards, D. Letscher, and B. O. Palsson. 2001. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* 71:286–306.

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

Strang, G. 1986. Introduction to Applied Mathematics. Wellesley-Cambridge Press, Wellesley, MA. 665–672.

Stryer, L. 1995. Biochemistry. forth edition. W. H. Freeman and Co., New York. 443–462.

Varma, A., and B. O. Palsson. 1994. Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnology.* 12:994–998.