Genome-Scale Thermodynamic Analysis of *Escherichia coli* Metabolism

Christopher S. Henry, Matthew D. Jankowski, Linda J. Broadbelt, and Vassily Hatzimanikatis
Department of Chemical and Biological Engineering, McCormick School of Engineering and Applied Sciences, Northwestern University, Evanston, Illinois

ABSTRACT Genome-scale metabolic models are an invaluable tool for analyzing metabolic systems as they provide a more complete picture of the processes of metabolism. We have constructed a genome-scale metabolic model of *Escherichia coli* based on the *iJR904* model developed by the Palsson Laboratory at the University of California at San Diego. Group contribution methods were utilized to estimate the standard Gibbs free energy change of every reaction in the constructed model. Reactions in the model were classified based on the activity of the reactions during optimal growth on glucose in aerobic media. The most thermodynamically unfavorable reactions involved in the production of biomass in *E. coli* were identified as ATP phosphoribosyltransferase, ATP synthase, methylene-tetra-hydrofolate dehydrogenase, and tryptophanase. The effect of a knockout of these reactions on the production of biomass and the production of individual biomass precursors was analyzed. Changes in the distribution of fluxes in the cell after knockout of these unfavorable reactions were also studied. The methodologies and results discussed can be used to facilitate the refinement of the feasible ranges for cellular parameters such as species concentrations and reaction rate constants.

INTRODUCTION

Thermodynamic analysis of reaction systems provides a means of characterizing and describing the equilibrium state of the reactions in the system. Metabolic pathways are open systems, and they cannot exist in a state of thermodynamic equilibrium. However, thermodynamic analysis is invaluable in establishing the limits of activity of metabolic systems, and these limits are important for constraints-based modeling (1,2) and for understanding the design and evolution of metabolism.

The most prevalent constraints-based modeling technique, flux balance analysis (FBA), is based on the quasi-steady-state assumption that the net accumulation of every metabolite in a cell is zero (3), and the mass balance equations of each metabolite are used to formulate a set of linear constraints. Metabolic systems typically involve far more reactions than metabolites making these systems underdetermined, and as a result, these mass balance constraints are insufficient to uniquely determine the flux through all of the reactions in the metabolic network. Despite this limitation, FBA can be used to test the feasibility of possible flux distributions, and it has been utilized extensively to interpret NMR data for estimating intracellular fluxes (4), to provide a reference state for metabolic control analysis (5), to analyze metabolite production and growth rates in cell cultures (6,7), and to predict the effect of gene knockouts (8–10).

One method for improving the quality and accuracy of flux quantification through FBA is to provide tighter constraints on the flux solution space (1,11). Thermodynamic analysis provides a means of accomplishing this goal. Currently thermodynamic analysis has found only limited application in the study of metabolic networks. Beard and Qian have conducted studies on the topic of eliminating internal flux cycles (2,12,13). These are sets of reactions for which the overall reaction is zero, such as $A \rightarrow B \rightarrow C \rightarrow A$. According to the first law of thermodynamics, the overall thermodynamic driving force through this cycle must be zero, meaning no net-flux is possible through this cycle. These cycles are often referred to as type-3 extreme pathways (14). Through the introduction of the appropriate constraints, flux distributions from FBA will no longer involve any flux through type-3 extreme pathways. This analysis only requires that the stoichiometry of the system be known, but no quantitative information on the relative thermodynamic feasibility of the individual reactions and pathways in the metabolic chemistry is provided. Using the limited amount of experimental thermodynamic data currently available, Beard and colleagues also performed a study on the central carbon pathways of the hepatocyte cell, and they quantified the levels of metabolite concentrations and reaction fluxes using thermodynamic constraints (15).

Here we have applied thermodynamic analysis to study *Escherichia coli* metabolism described by the *iJR904* genome-scale metabolic model of *E. coli* (16). We employed the group contribution method of Mavrovouniotis (17,18) to estimate the thermodynamic feasibility of the reactions in *E. coli* metabolism. We utilized FBA to determine the thermodynamically unfavorable reactions that are essential for optimal growth yield, and we performed knockout studies of these reactions to determine the role these reactions play in cell growth and in the production of individual biomass precursors. We also studied the shift in the flux distribution when the activity of a thermodynamically unfavorable reaction was removed. The Methods and Results presented in this article are directly applicable to improving predictions of the effects of gene knockouts, refining the estimation of cellular parameters such as species concentrations or reaction rate constants,

Submitted August 2, 2005, and accepted for publication October 28, 2005.

C. S. Henry and M. D. Jankowski contributed equally to this work.

Address reprint requests to V. Hatzimanikatis, Tel.: 847-491-5357; E-mail: vassily@northwestern.edu.

0006-3495/06/02/1453/09 $2.00

doi: 10.1529/biophysj.105.071720

and analyzing a proposed pathway for thermodynamic infeasibilities.

METHODS

Definition of $\Delta_r G'^m$ for the assessment of thermodynamic feasibility

The most common measure used for assessing the thermodynamic feasibility of reactions is the Gibbs free energy change of reaction, $\Delta_r G'$, which can be calculated using Eq. 1,

$$\Delta_r G' = \sum_{i=1}^{m} n_i \Delta_f G_i'^{\circ} + RT \ln\left(\prod_{i=1}^{m} x_i^{n_i}\right), \quad (1)$$

where $\Delta_f G_i'^{\circ}$ is the standard Gibbs free energy of formation of compound i, R is the universal gas constant, T is the temperature assumed to be 298 K, m is the number of compounds involved in the reaction, x_i is the activity of compound i, and n_i is the stoichiometric coefficient of compound i in the reaction (n_i is negative for reactants and positive for products). Although the activities of most compounds in biological systems are unknown, the mean activity in the cell is on the order of 1 mM (19). Therefore, using $\Delta_r G'^{\circ}$ for the assessment of the thermodynamic feasibility of metabolic reactions is not ideal, since this assumes the activity of every metabolite is 1 M. We propose that a better measure of the thermodynamic feasibility of reactions in biological systems is the standard Gibbs free energy change of reaction based on a 1 mM reference state, $\Delta_r G'^m$, calculated by setting every x_i value in Eq. 1 equal to 1 mM. For a reaction with the same number of reactants and products ($\sum_i n_i = 0$), not including hydrogen or water, $\Delta_r G'^{\circ}$ is equal to $\Delta_r G'^m$. If $\sum_i n_i$ is not equal to zero, $\Delta_r G'^{\circ}$ and $\Delta_r G'^m$ can be substantially different. For example, for a reaction with one product molecule and two reactant molecules, such as threonine aldolase,

$$\text{glycine} + \text{acetaldehyde} \leftrightarrow \text{threonine}, \quad (2)$$

with an estimated $\Delta_r G'^{\circ}$, $\Delta_r G'^{\circ}_{est}$, of -1.9 kcal/mol, the $\Delta_r G'^m_{est}$ value is 2.2 kcal/mol or 4.1 kcal/mol greater than $\Delta_r G'^{\circ}_{est}$. Based on $\Delta_r G'^{\circ}_{est}$, this reaction is thermodynamically favorable, but based on $\Delta_r G'^m_{est}$, the reaction is mildly unfavorable. A second example is methylthioadenosine nucleosidase,

$$\text{5-methylthioadenosine} + H_2O \rightarrow \text{5-methylthio-D-ribose} + \text{adenine}, \quad (3)$$

which has one reactant and two products. The reaction is unfavorable at 1 M activities, with a $\Delta_r G'^{\circ}_{est}$ of 2.3 kcal/mol, although it is favorable at 1 mM activities with a $\Delta_r G'^m_{est}$ of -1.7 kcal/mol. Depending on the value of $\sum_i n_i$, the difference between $\Delta_r G'^{\circ}$ and $\Delta_r G'^m$ can be generalized as shown in Fig. 1.

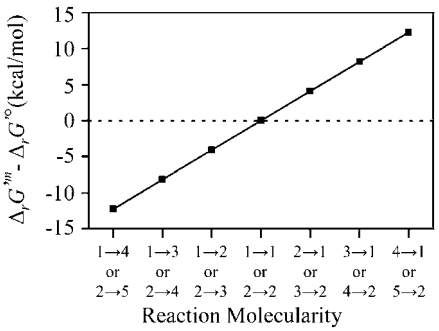

FIGURE 1 Effect of transformation of $\Delta_r G'^{\circ}$ into $\Delta_r G'^m$. The difference between $\Delta_r G'^{\circ}$ and $\Delta_r G'^m$ is shown for different reaction molecularities. The difference between $\Delta_r G'^{\circ}$ and $\Delta_r G'^m$ depends only on the difference between the number of reactant molecules and the number of product molecules.

Group contribution theory and estimation of $\Delta_r G'^{\circ}$

Although experimental measurements of $\Delta_f G'^{\circ}$ are unavailable for most compounds in *E. coli* metabolism, the group contribution methodology of Mavrovouniotis (17,18) provides a means by which the $\Delta_f G'^{\circ}$ of most metabolites can be estimated providing the estimated $\Delta_f G'^{\circ}$, or $\Delta_f G'^{\circ}_{est}$. Group contribution methods consider a single compound as being made up of smaller structural subgroups. The Gibbs free energy changes associated with the set of structural subgroups, $\Delta_{gr} G'^{\circ}$, commonly found in metabolites, are available in the literature along with special corrections for complex biochemical cofactors such as coA and NAD^+/NADH (17,18). To estimate $\Delta_f G'^{\circ}$ of the entire compound, the contributions of each of the subgroups to this property are summed along with an origin term

$$\Delta_f G'^{\circ}_{est} = \Delta_{gr} G'^{\circ}_0 + \sum_{i=1}^{N_{gr}} n_i \Delta_{gr} G'^{\circ}_i, \quad (4)$$

where $\Delta_{gr} G'^{\circ}_0$ is an origin term common to all compounds, N_{gr} is the number of subgroups, n_i is the number of instances of subgroup i in the compound, and $\Delta_{gr} G'^{\circ}_i$ is the contribution of subgroup i to $\Delta_f G'^{\circ}_{est}$ (17). All $\Delta_f G'^{\circ}_{est}$ values calculated using the group contribution methodology of Mavrovouniotis are based upon the standard condition of a solution with pH equal to 7 and with zero ionic strength.

For any reaction taking place in aqueous media, reactants will dissociate into several ionic forms (15,20). For example, ATP will dissociate and interconvert between the ionic forms: ATP^{4-}, $HATP^{3-}$, and H_2ATP^{2-}. In the cellular environment, the total amount of ATP present is the sum of all of these dissociated forms. In the fitting of thermodynamic energies of formation in the group contribution method of Mavrovouniotis, the total amount of ATP is represented by the single most common charged form found in a pH 7 solution, ATP^{4-} (17,18). Thus, the reaction for the hydrolysis of ATP into ADP and phosphate will be written as

$$ATP^{4-} + H_2O \leftrightarrow ADP^{3-} + Pi^{2-} + H^+. \quad (5)$$

The form of the reactants used in the group contribution method of Mavrovouniotis and in this work is the most common ionic form for a species in a solution at a pH of 7 such as the cytosol of an *E. coli* cell (21).

Using the group contribution methodology, we were able to determine $\Delta_f G'^{\circ}_{est}$ for 531 (85.9%) of the 618 compounds in the genome-scale *iJR904* metabolic model of *E. coli*, which allowed the calculation of $\Delta_r G'^m_{est}$ for 770 (82.6%) of the 932 reactions in the model. Values of $\Delta_f G'^{\circ}_{est}$ could not be determined for 87 compounds because these molecules contain substructures for which no energy value has been provided by Mavrovouniotis.

Construction of the *iHJ873 E. coli* model

To obtain a metabolic model of *E. coli* for which $\Delta_r G'^m_{est}$ of every reaction could be calculated, the reactions in the *iJR904* model containing compounds for which $\Delta_f G'^{\circ}_{est}$ could not be calculated were lumped into single reactions and these compounds were eliminated. For example, in the following series of reactions,

$$A \rightarrow B + C,$$
$$B + D \rightarrow E, \quad (6)$$

if $\Delta_f G'^{\circ}_{est}$ of compound B is unknown, we add the reactions involving B such that B is eliminated creating the lumped reaction of

$$A + D \rightarrow C + E. \quad (7)$$

After this lumping, the two reactions shown in Eq. 6 are removed from the model and replaced by the reaction shown in Eq. 7 and metabolite B is not explicitly accounted for in the network. Based on this lumping, we formulated the modified model, *iHJ873*, which contains 518 metabolites and 873 reactions that were fully characterized thermodynamically. Details of the

reactions and compounds removed from *iJR904* and the lumped reactions added to create *iHJ873* are provided in the Supplementary Material.

Classification of *iHJ873* model reactions

We performed flux variability analysis (FVA) (22) to determine the reactions involved in the maximum production of biomass from glucose in *E. coli* under aerobic conditions. Details of all flux analysis performed are listed in the Appendix. Under optimal growth conditions, reactions in *E. coli* may be classified as essential (requiring a nonzero flux for optimal growth to occur), substitutable (capable of carrying zero or nonzero flux at optimal growth), or blocked (do not carry any flux at optimal growth). In the *iHJ873* model, 250 (28.6%) reactions are essential, 51 (5.8%) reactions are substitutable, and 572 (65.5%) reactions are blocked. The total number of essential and substitutable reactions (301), which represents the total set of all reactions that participate in every alternative solution that produces optimal growth, agrees well with the average number of essential and substitutable reactions (294) reported for optimal growth phenotypes of *E. coli* utilizing a variety of nutrient sources (23). FVA also provides the direction of flux through the essential and substitutable reactions, allowing the reactants and products of all of these reactions to be redefined according to the direction of flux required for optimal growth (every flux will be positive). If a reaction can be active in both directions at optimal growth, the reactants and products and, consequently the reference directionality of the reactions, are defined according to their conventional nomenclature (16,24,25). Calculating $\Delta_r G'^{m}_{est}$ using this definition of reactants and products in the reaction means that a positive $\Delta_r G'^{m}_{est}$ value is indicative of a reaction that is thermodynamically unfavorable in the direction of flux required for optimal growth to occur at 1 mM activity conditions.

RESULTS

Distribution of $\Delta_r G'^{m}_{est}$ values for reactions in *iHJ873*

The distributions of $\Delta_r G'^{m}_{est}$ values for the essential and substitutable reactions in *iHJ873*, shown in histogram form in Fig. 2, *A* and *B*, indicate that 80.4% of the reactions have a $\Delta_r G'^{m}_{est}$ that is less than or equal to zero. However, there is uncertainty in $\Delta_f G'^{\circ}_{est}$, $U_{f,est}$, based on the group contribution methodology. The value $U_{f,est}$ is given as $\sim\pm 4$ kcal/mol (18), and the standard error is used for the uncertainty in $\Delta_r G'^{m}_{est}$, $U_{r,est}$, which is calculated as the Euclidean norm of the uncertainty for $\Delta_f G'^{\circ}_{est}$ of each compound involved in the reaction (*blue error bars* in Fig. 2, *C* and *D*) (26):

$$U_{r,est} = \sqrt{\sum_{i=1}^{m} n_i^2 U_{f,est}^2} = \sqrt{\sum_{i=1}^{m} 16 n_i^2}. \quad (8)$$

As indicated in Eqs. 1 and 8, $\Delta_r G'^{m}_{est}$ as well as the associated ranges of uncertainty depend on reaction molecularity (Fig. 2, *C* and *D*).

The reactions in *iHJ873* can be categorized thermodynamically based on their $\Delta_r G'^{m}_{est}$ value and the associated $U_{r,est}$. In category (*i*), $\Delta_r G'^{m}_{est} \pm U_{r,est} < 0$; 321 (36.8%) of all

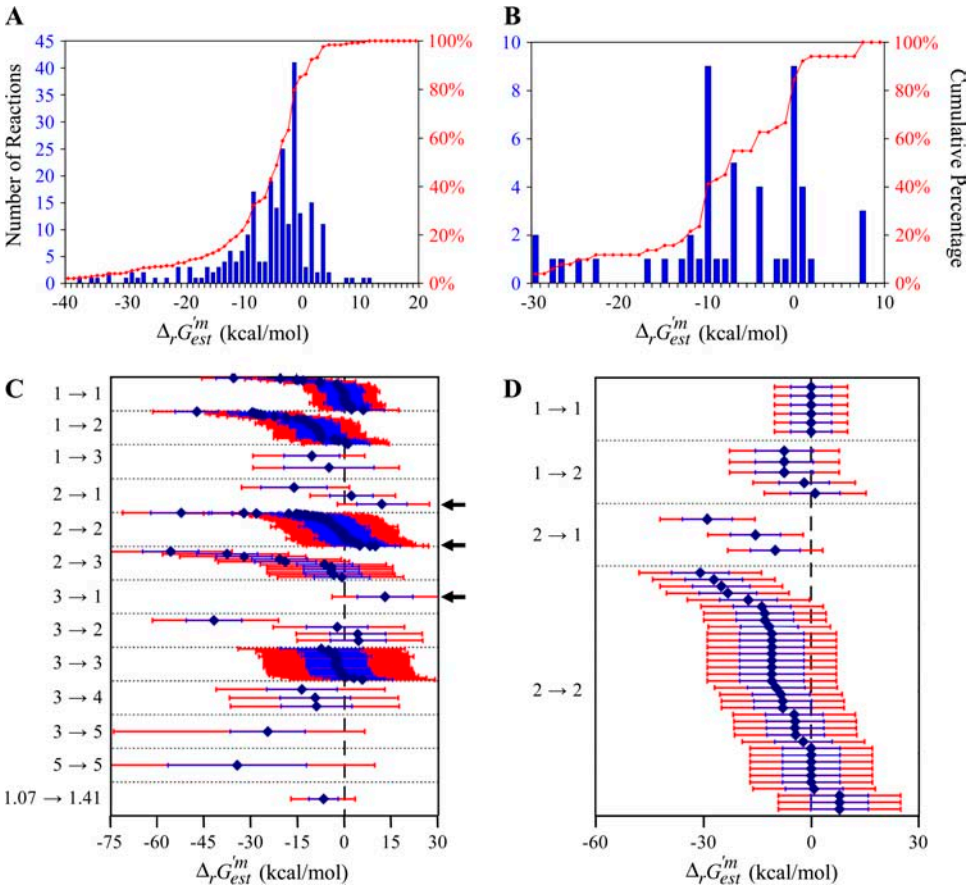

FIGURE 2 Thermodynamic feasibility of the reactions in *iHJ873*. Histograms of $\Delta_r G'^{m}_{est}$ values for the essential (*A*) and substitutable (*B*) reactions. The values of $\Delta_r G'^{m}_{est}$ for the essential (*C*) and substitutable (*D*) reactions. The blue error bars indicate the uncertainty of $\Delta_r G'^{m}_{est}$ calculated, given a 4 kcal/mol uncertainty in $\Delta_f G'^{\circ}_{est}$ provided in the literature (17). The red error bars indicate the range of values that $\Delta_r G'^{m}_{est}$ could take considering the uncertainty and concentrations in the cell ranging from 20 mM to 10^{-2} mM. The solid arrows in *C* mark the reactions for which $\Delta_r G'^{m}_{est} - U_{r,est} > 0$. These reactions must be unfavorable at the reference conditions.

of the reactions, and 90 (29.9%) of the essential and substitutable reactions in *iHJ873* are in this category. In category (*ii*), $\Delta_r G'^m_{est} \leq 0$ and $\Delta_r G'^m_{est} + U_{r,est} \geq 0$, and this category contains 429 (49.1%) of all of the reactions and 152 (50.5%) of the required and substitutable reactions. In category (*iii*), $\Delta_r G'^m_{est} > 0$ and $\Delta_r G'^m_{est} - U_{r,est} \leq 0$; this category consists of 114 (13.1%) of all reactions and 54 (17.9%) of the substitutable reactions. In category (*iv*), $\Delta_r G'^m_{est} \pm U_{r,est} > 0$ and five (0.6%) of all of the reactions and four (1.3%) of the essential and substitutable reactions are in this category. There are four different reactions that generate biomass in *iHJ873*, and these reactions are not part of any category, since the $\Delta_r G'^m_{est}$ of these reactions cannot be calculated.

The $\Delta_r G'^m_{est}$ values for the reactions in categories (*ii*) and (*iii*) are relatively close to zero, indicating that these reactions are close to equilibrium at reference conditions. Only the five reactions in category (*iv*) must be unfavorable at the standard conditions and millimolar metabolite activities. If we examine the distribution of the $\Delta_r G'^{\circ}_{est}$ values instead, we find that 232, 496, 138, and 3 of all of the reactions are in categories (*i*), (*ii*), (*iii*), and (*iv*), respectively. The smaller portion of reactions in the extreme categories (*i*) and (*iv*) indicates that the distribution of $\Delta_r G'^{\circ}_{est}$ values is narrower than the distribution of $\Delta_r G'^m_{est}$ values.

The values of $\Delta_r G'$ can deviate from $\Delta_r G'^m$ depending on how different the metabolite activities are from the reference value of 1 mM. Metabolite activities can range approximately between 10^{-2} mM and 20 mM (19). Based on these considerations, the maximum and minimum values for $\Delta_r G'_{est}$ were calculated using the equations

$$\Delta_r G'_{est,max} = \sum_{i=1}^{m} n_i \Delta_f G'^{\circ}_{est,i} + RT \sum_{i=1}^{Products} n_i \ln(x_{max}) + RT \sum_{i=1}^{Reactants} n_i \ln(x_{min}) + U_{r,est}, \quad (9)$$

$$\Delta_r G'_{est,min} = \sum_{i=1}^{m} n_i \Delta_f G'^{\circ}_{est,i} + RT \sum_{i=1}^{Products} n_i \ln(x_{min}) + RT \sum_{i=1}^{Reactants} n_i \ln(x_{max}) - U_{r,est}, \quad (10)$$

where $U_{r,est}$ is the uncertainty in $\Delta_r G'^m_{est}$ (Eq. 8), x_{min} is the minimal metabolite activity assumed to be 10^{-2} mM, and x_{max} is the maximum metabolite activity assumed to be 20 mM (19). Although metabolite activities can be lower than 10^{-2} mM, a decrease in the lower limit on metabolite activity will result in an increase in all $\Delta_r G'_{est,max}$ and a decrease in all $\Delta_r G'_{est,min}$ (Fig. 2). Of the five reactions in category (*iv*), which have the highest $\Delta_r G'^m_{est}$ values, metabolite activity profiles exist that can reduce $\Delta_r G'_{est}$ and thus make these reactions thermodynamically feasible. These cases are indicated in Fig. 2 *C* with arrows on the right side of the corresponding graphs.

The large fraction of essential and substitutable reactions in categories (*ii*) and (*iii*) with $\Delta_r G'^m_{est}$ values within the margin of error of the zero axis indicates that most reactions involved in growth are energetically balanced, and only small concentration gradients are required to make these reactions thermodynamically feasible. The large fraction of reactions with associated $\Delta_r G'^m_{est}$ values that are near zero is advantageous to the cell, because this prevents reactant and product concentrations from rising to toxic levels or falling to levels that would limit reaction rates.

ATP synthase and transport reactions

The standard conditions of pH 7 solution and zero ionic strength upon which all $\Delta_r G'^m_{est}$ values are based was applied to both the extracellular and intracellular environment when calculating $\Delta_r G'^m_{est}$ for reactions involving the transport of metabolites across the cellular membrane. As a result, $\Delta_r G'^m_{est}$ for these reactions is based on the assumption that the electrochemical potential, $\Delta\psi$, and pH gradient, ΔpH ($pH_{intracellular} - pH_{extracellular}$), across the cell membrane is zero. For example, the ATP synthase reaction in *E. coli* is typically written in the form of

$$4H^+_{extracellular} + Pi + ADP \rightarrow ATP + H_2O + 3H^+_{cytosolic}. \quad (11)$$

The $\Delta_r G'^m_{est}$ for the portion of this reaction that takes place inside the cell, $\Delta_r G'^m_{est,\,intracellular}$,

$$H^+_{cytosolic} + Pi + ADP \rightarrow ATP + H_2O, \quad (12)$$

can be found using Eq. 1. For ATP synthase, $\Delta_r G'^m_{est,\,intracellular}$ is 12 kcal/mol, which agrees well with the experimentally measured $\Delta_r G'^m_{intracellular}$ of 10.4 kcal/mol (27).

The energy contribution of the transmembrane transport portion of the ATP synthase reaction, $\Delta_r G'_{transport}$,

$$4H^+_{extracellular} \rightarrow 4H^+_{cytosolic}, \quad (13)$$

is the sum of the driving force of the ΔpH across the membrane for the transport of H^+ into the cell, $\Delta_{\Delta pH} G$, and the energy associated with the transport of an ion across the membrane, $\Delta_{\Delta\psi} G_{est}$,

$$\Delta_r G'_{est,\,transport} = \Delta_{\Delta\psi} G + \Delta_{\Delta pH} G. \quad (14)$$

At the standard conditions (pH 7, meaning $\Delta pH = 0$ and zero ionic-strength in intracellular and extracellular compartments meaning $\Delta\psi = 0$), $\Delta_r G'_{transport}$ for ATP synthase is 0.0 kcal/mol.

The overall $\Delta_r G'^m_{est}$ of a reaction energetically coupled to the transport of an ion across the cell membrane such as ATP synthase is

$$\Delta_r G'^m_{est} = \Delta_r G'^m_{est,\,transport} + \Delta_r G'^m_{est,\,intracellular}. \quad (15)$$

The value of $\Delta_r G'^m_{est}$ for the ATP synthase reaction (Eq. 11) at the standard conditions is 12 kcal/mol.

However, under physiological conditions ΔpH, $\Delta\psi$ and $\Delta_r G'_{transport}$ are not zero. The value $\Delta_{\Delta\psi} G_{est}$ depends upon $\Delta\psi$, which in turn depends on ΔpH according to the equations (28)

$$\Delta_{\Delta\psi}G(\text{kcal/mol}) = nF\Delta\psi, \tag{16}$$

$$\Delta\psi(\text{mV}) = 33.33\,\Delta pH - 143.33 \quad (\text{based on a fit of experimental data}), \tag{17}$$

where n is the net charge transported from outside the cell into the cell, and F is the Faraday constant in kcal/mV mol. The value $\Delta_{\Delta\text{pH}}G$ depends only on ΔpH according to the equation (28)

$$\Delta_{\Delta\text{pH}}G(\text{kcal/mol}) = -2.3\ \text{hRT}\,\Delta pH, \tag{18}$$

where h is the number of protons transported across the membrane. At an extracellular pH of 6, $\Delta_r G'_{\text{est, transport}}$ of ATP synthase is −15.6 kcal/mol, making the total $\Delta_r G'^{\text{m}}_{\text{est}}$ of ATP synthase −3.6 kcal/mol. The value of $\Delta_r G'^{\text{m}}_{\text{est}}$ for the ATP synthase reaction only becomes positive when the ΔpH is lower than −0.51, meaning the extracellular pH is higher than the intracellular pH and above the optimal pH for *E. coli* growth.

Identification and characterization of unfavorable reactions

Only five of the 873 reactions in the *iHJ873* model have a $\Delta_r G'^{\text{m}}_{\text{est}}$ that is greater than $U_{\text{r,est}}$, which indicates that every possible value of $\Delta_r G'^{\text{m}}_{\text{est}}$, given the uncertainty in the estimate, must be positive and these reactions are unfavorable at standard conditions and 1 mM activities. These five reactions are listed in Table 1. Four of these five unfavorable reactions are classified as essential for optimal growth to occur. These four reactions are

1. ATP phosphoribosyltransferase:
   ATP + 5-phospho-α-D-ribose-1-diphosphate → diphosphate + 1-(5-phosphoribosyl)-ATP. (19)
2. ATP synthase (without accounting for membrane potential, as discussed earlier; see Eq. 11).
3. Methylenetetrahydrofolate dehydrogenase:
   5, 10-methylenetetrahydrofolate + NADP + H^+ ↔ 5, 10-methenyltetrahydrofolate + NADPH. (20)
4. Tryptophanase:
   indole + ammonium + pyruvate ↔ H_2O + tryptophan. (21)

We simulated knockouts of each unfavorable reaction while maximizing the yield of each of the biomass precursors and the yield of biomass to study the effects of single-knockouts and simultaneous knockouts on the cell growth (see Methods). Although ATP synthase is typically an energetically favorable reaction due to the energy contribution of the pH gradient and electrochemical potential across the cell membrane, ATP synthase is also included in the knockout studies to investigate the response of the system in case ATP synthase becomes unfavorable due to the failure of the proton gradient coupling and transmembrane potential.

ATP phosphoribosyltransferase knockout

Only the precursor histidine is affected by the knockout of ATP phosphoribosyltransferase, and the production of histidine is not possible without the activity of this reaction, making histidine the limiting component preventing any growth without ATP phosphoribosyltransferase (Fig. 3 *A*). Experimental evidence confirms that ATP phosphoribosyltransferase is essential for the production of histidine, and mutant strains lacking this enzyme cannot grow without a histidine supplement (29). Experimental evidence also confirms that this reaction is thermodynamically unfavorable (29). The value $\Delta_r G'^{\text{m}}_{\text{est}}$ can range between 0.2 and 16.2 kcal/mol given the margin of uncertainty in the group contribution methodology. If the metabolite activities in the cell range between 20 mM and 10^{-2} mM, then $\Delta_r G'_{\text{est}}$ of this reaction can range between −0.81 kcal/mol and 17.2 kcal/mol. Therefore, the reactant to product activity gradients required to drive this unfavorable reaction are achievable within the range of the physiological intracellular activities.

As the first step in the histidine metabolism pathway, ATP phosphoribosyltransferase is an important point of control for the production of histidine. A mechanism even exists in the cell for feedback inhibition of ATP phosphoribosyltransferase by histidine (30). The unfavorable thermodynamics of this reaction provides another mechanism for product-inhibition of this enzyme as a means of limiting the flux that enters the histidine metabolism pathway.

ATP synthase knockout

The knockout of ATP synthase affects the optimal production of 49 of the 53 biomass precursors in the *iHJ873* model (Fig. 3 *B*). The production of the energy in the form of ATP during aerobic metabolism depends heavily upon the ATP synthase reaction, and without ATP synthase, the energy requirements for optimal growth are not satisfied. Although a lack of ATP synthase does not completely prevent cell growth, the cell can

TABLE 1 Unfavorable reactions

Name	Pathway	$\Delta_r G'^{\text{m}}_{\text{est}}$ kcal/mol	Classification
Tryptophanase	Tyrosine, tryptophan, and phenylalanine metabolism	13	Essential
ATP synthase	Oxidative phosphorylation	12	Essential
Methylene-tetra-hydrofolate dehydrogenase	Folate metabolism	9.9	Essential
ATP phosphoribosyltransferase	Histidine metabolism	8.2	Essential
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	Cofactor and prosthetic group biosynthesis	22	Blocked

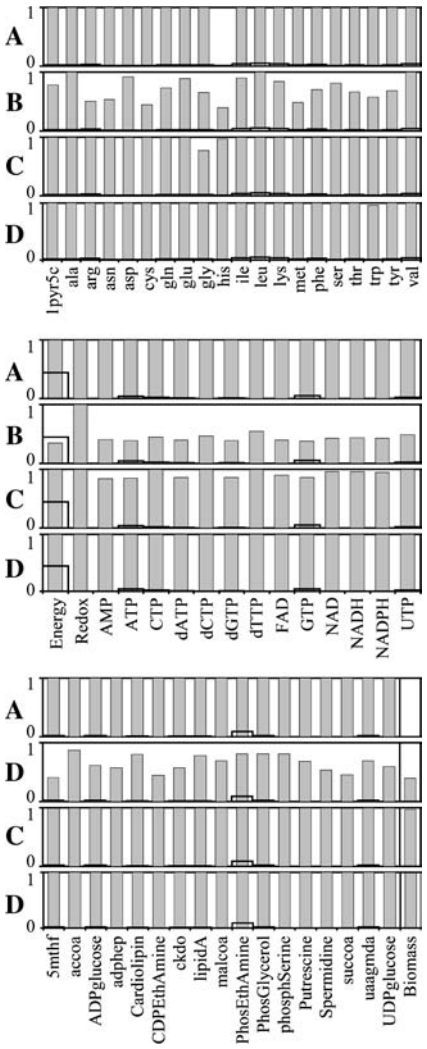

FIGURE 3 Impact of unfavorable reactions on the yield of biomass precursors. The maximum yields of the individual biomass precursors and of biomass relative to the corresponding maximum yields in the wild-type cells for different reaction knockouts: (*A*) ATP phosphoribosyltransferase, (*B*) ATP synthase, (*C*) methylene-tetra-hydrofolate dehydrogenase, and (*D*) tryptophanase. The bold lines in the plot show the relative precursor yield required for optimal growth to occur.

only grow at 39.2% of the optimal yield, and experimental evidence confirms this effect of ATP synthase on growth (31).

Methylene-tetra-hydrofolate dehydrogenase knockout

Although all biomass precursors can still be produced individually in sufficient quantity for optimal growth to occur with the knockout of methylenetetrahydrofolate dehydrogenase, this knockout does reduce the production of 14 biomass precursors by an average of 8.6% (Fig. 3 *C*). As a result, no precursors can be produced simultaneously in sufficient quantities for optimal growth to occur without this reaction, and growth yield is reduced to 97.3% of the optimum. Methylene-tetra-hydrofolate dehydrogenase is a key step in the folate-dependent one-carbon metabolism pathway. The one-carbon pool from folate cannot be synthesized without this reaction, and without this reaction, other sources of C_1 in metabolism must be utilized. According to the literature, this reaction is thermodynamically unfavorable with a $\Delta_r G'^{m}$ of 1.17 kcal/mol (32), which is within the margin of uncertainty of the group contribution $\Delta_r G'^{m}_{est}$ estimate of 9.94 kcal/mol. Given the range of physiological intracellular metabolite activities, $\Delta_r G'$ can deviate from $\Delta_r G'^{m}$ of 1.17 kcal/mol and range between −7.8 kcal/mol and 10.2 kcal/mol. The typical NADP/NADPH ratio found in *E. coli* is 6, and this ratio alone is already sufficient to reduce $\Delta_r G'^{m}$ of this reaction by 1.06 kcal/mol to a $\Delta_r G'$ of 0.11 kcal/mol.

Tryptophanase knockout

Only the maximum yield of the precursor tryptophan is affected by the knockout of tryptophanase (Fig. 3 *D*), and it is only reduced by 3.8%. Knockout out of tryptophanase has a nearly negligible effect on growth, reducing the yield by 0.03%. According to experimental evidence found in the literature, the $\Delta_r G'^{\circ}_{est}$ for this reaction is −4.98 kcal/mol (33), which transforms into a $\Delta_r G'^{m}_{est}$ value of 3.21 kcal/mol for this three-reactant, one-product reaction, confirming that this reaction is thermodynamically unfavorable under mM activity conditions. Although the estimate of $\Delta_r G'^{m}_{est}$ from group contribution theory, 13 kcal/mol, for this reaction is high relative to experimental values, the difference between the estimate and the experimental data, 9.8 kcal/mol, still falls near the standard uncertainty of this reaction, 8.9 kcal/mol.

Four-reaction knockout

To determine the cumulative effect on biomass production of knocking out multiple unfavorable reactions simultaneously, we performed knockout simulations in which the activities of all of the unfavorable reactions were removed in every possible combination (Fig. 4, *I*). The effect of the cumulative knockouts on energy production was also examined (Fig. 4, *II*). In the simultaneous knockout of ATP synthase and tryptophanase, the growth yield is the same as the lower growth yield from the single knockouts of the same reactions. In this case, the knockout is not additive and the reactions play independent roles in the production of biomass. However, in the simultaneous knockout of ATP synthase and methylene-tetra-hydrofolate dehydrogenase, the growth yield is lower than the yield achieved from either of the single knockouts of these reactions. The effect of the double knockout of these reactions is additive, demonstrating that the contribution of these reactions to growth is linked.

Effect of unfavorable reaction knockouts on reaction classification

Unlike ATP phosphoribosyltransferase, the activities of the unfavorable reactions ATP synthase, methylene-tetra-hydrofolate

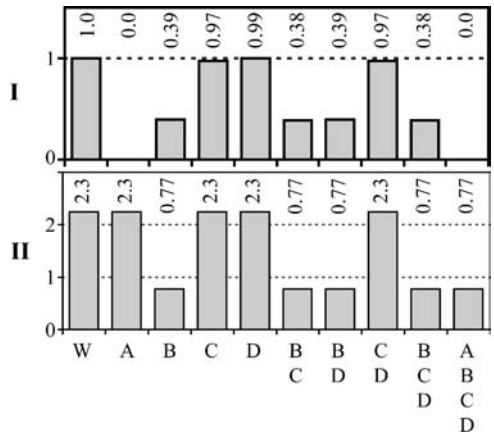

FIGURE 4 Impact of simultaneous knockouts of unfavorable reactions on growth and growth-limiting precursors. The maximum yields of biomass (*I*) and energy (*II*) for the combinations of simultaneous knockouts of the unfavorable reactions: *A*, ATP phosphoribosyltransferase, *B*, ATP synthase, *C*, methylene-tetra-hydrofolate dehydrogenase, and *D*, tryptophanase. The maximum yields of biomass (*I*) are scaled by the maximum yield of biomass in the wild-type. The maximum yields of energy (*II*) are scaled by the amount of energy required for optimal growth. Not all combinations including knockout of ATP phosphoribosyltransferase (*A*) are shown because zero growth is possible without ATP phosphoribosyltransferase and energy is not affected by knockout of ATP phosphoribosyltransferase.

dehydrogenase, and tryptophanase are not essential for cell growth to occur. These reactions have a wide range of effects on the metabolism of the cell, indicated by the widespread effect on the production of the biomass precursors. To study the effect a knockout of these reactions has on the distribution of fluxes in *E. coli*, FVA was utilized to determine how the reactions that are essential for optimal growth change when the activity of these reactions is knocked out. Table 2 summarizes the results of this study.

The wild-type and ATP synthase knockout share 231 essential reactions in common. There are 19 reactions that are essential in the wild-type and nonessential in the ATP synthase knockout. Fourteen of these reactions become substitutable in the ATP synthase knockout. These reactions are primarily clustered in the citrate cycle, glycolysis, and oxidative phosphorylation pathways. The remaining five of the 19 nonessential reactions in ATP synthase knockout, including ATP synthase, are blocked in the ATP synthase knockout. These reactions are found in the folate metabolism and pentose phosphate pathways.

TABLE 2 Minimal reaction sets for optimal growth in knockout and wild-type metabolism

Phenotype	Growth yield (gm biomass/ mmol glucose)	No. of essential reactions	No. of substitutable reactions
Wild-type	0.0923	250	51
ATP synthase KO	0.0362	234	118
Methylene-tetra-hydrofolate dehydrogenase KO	0.0897	249	37
Tryptophanase KO	0.0923	248	54

Three reactions that are essential in the ATP synthase knockout are blocked in the wild-type. Two of these reactions are involved in producing threonine while consuming one ATP, and the third reaction is an acetate transporter. Overall, the knockout of ATP synthase results in a deactivation of portions of the pentose phosphate pathways, citrate cycle, and glycolysis.

The wild-type and methylene-tetra-hydrofolate dehydrogenase knockout share 243 essential reactions in common. Six of the reactions that are essential in the methylene-tetra-hydrofolate dehydrogenase knockout are blocked in the wild-type. These reactions are involved in a variety of small-carbon metabolism pathways. Many of these reactions produce formate to compensate for the loss of the formate metabolism reactions with the knockout of methylene-tetra-hydrofolate dehydrogenase. Five reactions, in addition to methylene-tetra-hydrofolate dehydrogenase, that are essential in the wild-type are blocked in the methylene-tetra-hydrofolate dehydrogenase knockout. These reactions are involved in the alternate carbon metabolism, arginine and proline metabolism, and folate metabolism pathways. These reactions are associated with the decomposition of some small-carbon compounds and the production of formate and tetrahydrofolate. Overall, knockout of methylene-tetra-hydrofolate dehydrogenase results in the deactivation of the folate metabolism pathway and the activation of alternative pathways for the production of folate and other small carbon compounds.

Comparing the essential reactions in the wild-type at optimal growth to the tryptophanase knockout at optimal growth, every essential reaction in the tryptophanase knockout is also essential in the wild-type knockout. Other than tryptophanase, only the reaction tryptophan synthase from the aromatic amino acid metabolism pathways is essential in the wild-type and not essential in the tryptophanase knockout. This reaction becomes substitutable in the tryptophanase knockout.

DISCUSSION AND CONCLUSIONS

The group contribution methodology of Mavrovouniotis is demonstrated to be an effective means of estimating the free energy change of biochemical reactions. This methodology was utilized to calculate $\Delta_r G'^{m}_{est}$ for 82.6% of the reactions in the *iJR904* genome-scale metabolic model of *E. coli* developed by Palsson and co-workers. The *iJR904* model was modified to eliminate the 85 compounds for which no group contribution estimation of $\Delta_f G'^{\circ}_{est}$ was possible to create the *iHJ873* model, and $\Delta_r G'^{m}_{est}$ was determined for all of the reactions in *iHJ873*.

The $\Delta_r G'^{m}_{est}$ is an invaluable measure of the thermodynamic feasibility of the reactions in the metabolic pathways of the cell under physiological conditions. Four-hundred-and-twenty-nine (49.1%) of all of the reactions and 152 (50.5%) of the reactions that are essential or substitutable for optimal growth to occur have a negative $\Delta_r G'^{m}_{est}$ such that $\Delta_r G'^{m}_{est}$ +

$U_{r,est} > 0$. The majority of the reactions in the cell are thermodynamically favorable, with a $\Delta_r G'^m_{est}$ that is relatively close to zero under standard conditions and 1 mM metabolite activities. This result indicates that the cellular system is energetically buffered from large perturbations and a minimal thermodynamic driving force is utilized to drive reactions.

Only four reactions essential for optimal growth yield have a positive $\Delta_r G'^m_{est}$ such that $\Delta_r G'^m_{est} \pm U_{r,est} > 0$, indicating that these reactions must be unfavorable at standard conditions and 1 mM metabolite activity levels. These four reactions are ATP phosphoribosyltransferase in the histidine metabolism pathway, ATP synthase in the oxidative phosphorylation pathway, methylene-tetra-hydrofolate dehydrogenase in the folate metabolism pathway, and tryptophanase in the aromatic amino-acid metabolism pathway. Experimental data exists that confirms that ATP phosphoribosyltransferase, ATP synthase, methylene-tetra-hydrofolate dehydrogenase, and tryptophanase are unfavorable at standard conditions and mM activities. These unfavorable reactions represent crucial thermodynamic bottlenecks in the production of growth, constraining the activities of metabolites involved in these reactions so that sufficient activity gradient may be provided to drive the reactions.

The fact that out of the 250 reactions essential for growth, only four reactions have $\Delta_r G'^m_{est}$ values that are sufficiently large that they must be unfavorable at the reference conditions, indicates that the reactions involved in metabolism in *E. coli* are thermodynamically optimized to a great extent. It is important to note, however, that $\Delta_r G'^m_{est}$ values discussed in this article are estimates and not experimentally measured values, and in some cases, like tryptophanase, the estimates can differ from the experimental values. This emphasizes the importance of accounting for the uncertainty in the group contribution estimates before utilizing this data for any analysis. Although the group energy values upon which $\Delta_r G'^m_{est}$ are based were obtained from a fitting of experimentally measured $\Delta_r G'^{\circ}$ values, this experimental dataset consists of far fewer reactions than are involved in the genome-scale model of *E. coli*.

The thermodynamic data obtained from this methodology is essential for the determination of the thermodynamically feasible activity ranges for the metabolites involved in the active reactions in *E. coli* metabolism, as discussed in the literature (15,34). Such feasible ranges would be very useful for narrowing the constraints utilized in constraints-based models as well as the operating conditions explored in MCA and kinetic modeling (5,35,36). The $\Delta_r G'^m_{est}$ may also be used to formulate additional thermodynamic constraints for metabolic flux analysis (MFA) to ensure that flux distributions generated are thermodynamically feasible. Addition of thermodynamic constraints would aid in improving the predictions by metabolic models of the effect of gene knockout or other perturbations to the cellular metabolism. The error analysis discussed here will form an integral part of such thermodynamic constraints.

APPENDIX: DETAILS OF FLUX ANALYSIS

Metabolic flux analysis (MFA) defines the limits on the metabolic capabilities of a model organism under steady-state flux conditions (3). Steady-state flux conditions are described by constraining the net production of every metabolite in the system, given by the product of the stoichiometric matrix and flux vector, to 0, as shown in Eq. 22,

$$\boldsymbol{N} \cdot \boldsymbol{\nu} = \boldsymbol{0}, \tag{22}$$

where $\boldsymbol{N}$ is an $m \times r$ matrix of the stoichiometric coefficients for the r reactions and m metabolites in the model, and $\boldsymbol{\nu}$ is an $r \times 1$ vector of the steady-state fluxes through the r reactions in the model. A metabolic flux analysis was performed on the *iHJ873* model to determine the flux distributions that optimize various objective functions such as maximum yield on growth, maximum yield on biomass precursors, and maximum and minimum flux through every reaction in the system.

MFA studies were performed under a specific set of constraints on the metabolites the cell could uptake from or excrete to the cell surroundings. The ability of *E. coli* to grow optimally under aerobic conditions was studied using glucose as a primary carbon source. The uptake of glucose and oxygen from the environment into the cell was restricted to 10 and 20 mmol/g per dw per h, respectively (7). The uptake and excretion of sulfate, phosphate, and ammonium, CO_2, water, and hydrogen ion were left unrestricted and the ATP maintenance requirement was fixed at 7.6 mmol/g per dw per h (21,37,38). Under these conditions, the optimal growth on glucose was found to be 0.923 g biomass/g per dw per h, with a yield of 0.0923 gram biomass per mmol of glucose uptake (0.512 g biomass/g glucose). This optimal growth yield agrees well with the optimal growth yields for *E. coli* under similar conditions reported in the literature from MFA and experiments (38).

Flux variability analysis was used to classify the behavior of the reactions in the model using the methods described in the literature (22). There are 17 internal flux loops, or type-3 extreme pathways (14), in the *iJR904* model, and 13 of these internal flux loops also exist in the *iHJ873*. No flux should move through these internal flux loops in the FVA flux distributions, because no thermodynamic driving force can exist for such flux. To prevent any flux from moving through these internal flux loops, one reaction from each internal flux loop is blocked in the FVA (13 blocked reactions total). A list of the reactions that must be blocked is found in the *iJR904* literature (23) and in the Supplementary Material.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

We thank Professor Bernhard Palsson and colleagues at the University of California at San Diego for making the *iJR904* model readily available.

The work is supported by the United States Department of Energy, Genomes to Life Program.

REFERENCES

1. Covert, M. W., I. Famili, and B. O. Palsson. 2003. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology. *Biotechnol. Bioeng.* 84:763–772.
2. Beard, D. A., S. C. Liang, and H. Qian. 2002. Energy balance for analysis of complex metabolic networks. *Biophys. J.* 83:79–86.
3. Wiback, S. J., I. Famili, H. J. Greenberg, and B. O. Palsson. 2004. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.* 228:437–447.
4. Sauer, U., D. R. Lasko, J. Fiaux, M. Hochuli, R. Glaser, T. Szyperski, K. Wuthrich, and J. E. Bailey. 1999. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.* 181:6679–6688.

5. Wang, L. Q., I. Birol, and V. Hatzimanikatis. 2004. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys. J.* 87:3750–3763.
6. Papoutsakis, E. T., and C. L. Meyer. 1985. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. *Biotechnol. Bioeng.* 27:50–66.
7. Varma, A., and B. O. Palsson. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60:3724–3731.
8. Burgard, A. P., and C. D. Maranas. 2001. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* 74:364–375.
9. Edwards, J. S., and B. O. Palsson. 2000. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 16:927–939.
10. Alper, H., K. Miyaoku, and G. Stephanopoulos. 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.* 23:612–616.
11. 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.
12. Qian, H., D. A. Beard, and S. D. Liang. 2003. Stoichiometric network theory for nonequilibrium biochemical systems. *Eur. J. Biochem.* 270: 415–421.
13. Beard, D. A., E. Babson, E. Curtis, and H. Qian. 2004. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.* 228: 327–333.
14. Schilling, C. H., D. Letscher, and B. O. Palsson. 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203:229–248.
15. Beard, D. A., and H. Qian. 2005. Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *Am. J. Physiol. Endocrin. M.* 288:E633–E644.
16. Reed, J. L., T. D. Vo, C. H. Schilling, and B. O. Palsson. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4:54.51–54.12.
17. Mavrovouniotis, M. L. 1990. Group contributions for estimating standard Gibbs energies of formation of biochemical-compounds in aqueous-solution. *Biotechnol. Bioeng.* 36:1070–1082.
18. Mavrovouniotis, M. L. 1991. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* 266:14440–14445.
19. Albe, K. R., M. H. Butler, and B. E. Wright. 1990. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.* 143:163–195.
20. Alberty, R. A. 1998. Calculation of standard transformed Gibbs energies and standard transformed enthalpies of biochemical reactants. *Arch. Biochem. Biophys.* 353:116–130.
21. Gottschalk, G. 1988. Bacterial Metabolism. Springer-Verlag, New York.
22. Mahadevan, R., and C. H. Schilling. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5:264–276.
23. Reed, J. L., and B. O. Palsson. 2004. Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res.* 14: 1797–1805.
24. Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30.
25. Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27:29–34.
26. Box, G. E. P., J. S. Hunter, and W. G. Hunter. 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. Wiley, New York.
27. Fagan, M. H., and T. G. Dewey. 1985. Steady-state kinetics of ATP synthesis and hydrolysis coupled calcium-transport catalyzed by the reconstituted sarcoplasmic-reticulum ATPase. *J. Biol. Chem.* 260: 6147–6152.
28. Neidhardt, F. C., and R. Curtiss. 1996. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, DC.
29. Ames, B. N. 1961. First step of histidine biosynthesis. *J. Biol. Chem.* 236:2019–2026.
30. Martin, R. G. 1963. First enzyme in histidine biosynthesis—nature of feedback inhibition by histidine. *J. Biol. Chem.* 238:257–268.
31. Moser, T. L., D. J. Kenan, T. A. Ashley, J. A. Roy, M. D. Goodman, U. K. Misra, D. J. Cheek, and S. V. Pizzo. 2001. Endothelial cell surface F_1-F_o ATP synthase is active in ATP synthesis and is inhibited by angiostatin. *Proc. Natl. Acad. Sci. USA.* 98:6656–6661.
32. Uyeda, K., and J. Rabinowi. 1967. Enzymes of Clostridial purine fermentation—methylenetetrahydrofolate dehydrogenase. *J. Biol. Chem.* 242:4378–4385.
33. Tewari, Y. B., and R. N. Goldberg. 1994. An equilibrium and calorimetric investigation of the hydrolysis of L-tryptophan to indole plus pyruvate plus ammonia. *J. Sol. Chem.* 23:167–184.
34. Mavrovouniotis, M. L. 1996. Duality theory for thermodynamic bottlenecks in bioreaction pathways. *Chem. Eng. Sci.* 51:1495–1507.
35. Mavrovouniotis, M. L., and G. Stephanopoulos. 1990. Estimation of upper-bounds for the rates of enzymatic reactions. *Chem. Eng. Comm.* 93:211–236.
36. Bish, D. R., and M. L. Mavrovouniotis. 1998. Enzymatic reaction rate limits with constraints on equilibrium constants and experimental parameters. *Biosystems.* 47:37–60.
37. Varma, A., and B. O. Palsson. 1993. Metabolic capabilities of *Escherichia coli.* 1. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* 165:477–502.
38. Varma, A., and B. O. Palsson. 1993. Metabolic capabilities of *Escherichia coli*. 2. Optimal growth patterns. *J. Theor. Biol.* 165:503–522.