

Thermodynamics-Based Metabolic Flux Analysis

Christopher S. Henry,* Linda J. Broadbelt,* and Vassily Hatzimanikatis[†]

*Department of Chemical and Biological Engineering, McCormick School of Engineering and Applied Sciences, Northwestern University, Evanston, Illinois; and [†]Laboratory of Computational Systems Biotechnology, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

ABSTRACT A new form of metabolic flux analysis (MFA) called thermodynamics-based metabolic flux analysis (TMFA) is introduced with the capability of generating thermodynamically feasible flux and metabolite activity profiles on a genome scale. TMFA involves the use of a set of linear thermodynamic constraints in addition to the mass balance constraints typically used in MFA. TMFA produces flux distributions that do not contain any thermodynamically infeasible reactions or pathways, and it provides information about the free energy change of reactions and the range of metabolite activities in addition to reaction fluxes. TMFA is applied to study the thermodynamically feasible ranges for the fluxes and the Gibbs free energy change, $\Delta_r G'$, of the reactions and the activities of the metabolites in the genome-scale metabolic model of *Escherichia coli* developed by Palsson and co-workers. In the TMFA of the genome scale model, the metabolite activities and reaction $\Delta_r G'$ are able to achieve a wide range of values at optimal growth. The reaction dihydroorotase is identified as a possible thermodynamic bottleneck in *E. coli* metabolism with a $\Delta_r G'$ constrained close to zero while numerous reactions are identified throughout metabolism for which $\Delta_r G'$ is always highly negative regardless of metabolite concentrations. As it has been proposed previously, these reactions with exclusively negative $\Delta_r G'$ might be candidates for cell regulation, and we find that a significant number of these reactions appear to be the first steps in the linear portion of numerous biosynthesis pathways. The thermodynamically feasible ranges for the concentration ratios ATP/ADP, NAD(P)/NAD(P)H, and $H^+_{\text{extracellular}}/H^+_{\text{intracellular}}$ are also determined and found to encompass the values observed experimentally in every case. Further, we find that the NAD/NADH and NADP/NADPH ratios maintained in the cell are close to the minimum feasible ratio and maximum feasible ratio, respectively.

INTRODUCTION

Thermodynamics have been applied to many areas of analysis of biological systems (1–5), but thermodynamics have yet to be applied to a rigorous examination of entire metabolic networks. This has been primarily due to a scarcity of thermodynamic data on metabolic reactions, a lack of rigorous models of metabolic chemistry, and the absence of any extensive databases, which bring all of this information together. However, the availability of thermodynamic data has increased over time, and group contribution methodologies for estimating thermodynamic properties have also been introduced (6–9). Furthermore, several rigorous models of the metabolic chemistry of a variety of microorganisms have been developed including some genome-scale models (10–13). Recently, the application of thermodynamics to study the feasibility of metabolic pathways has been revisited. Beard, Qian, and co-workers have conducted studies on the topic of eliminating internal flux cycles from flux balance analysis solutions (14–16). These are sets of reactions such as $A \rightarrow B \rightarrow C \rightarrow A$. According to the first law of thermodynamics, the overall thermodynamic driving force through these cycles must be zero, meaning that no net flux is possible through these cycles. Beard and Qian have also used nonlinear thermodynamic and enzyme activity constraints to

determine the concentration profiles of metabolites in the central carbon chemistry of a hepatocyte cell (17). Maskow and Stockar used the pathway analysis method of Mavrouniotis (18,19) to study the thermodynamic feasibility of the lactic acid fermentation pathway, and they found that without careful consideration of ionic strength of solution, uncertainty in thermodynamic data, and cell pH, feasible pathways can be falsely labeled as infeasible or vice versa (20). However, these previous studies were performed on relatively small-scale pathways due to a lack of thermodynamic data for genome-scale models and utilized nonlinear optimization criteria to determine fixed values for the activities of the metabolites under an isolated set of conditions.

In a previous article, we utilized the group contribution method (7,8) to estimate the standard Gibbs free energy change, $\Delta_r G'^{\circ}$, of the reactions in a genome-scale model of *Escherichia coli*, and we used these estimates to assess the thermodynamic feasibility of the reactions in the model (21). We called this model *iHJ873*, which is based on the *iJR904* model developed by Palsson and co-workers. The *iHJ873* model was derived from the *iJR904* model by removing all of the reactions in the *iJR904* model that contain compounds for which the standard Gibbs free energy change of formation, $\Delta_f G'^{\circ}$, could not be estimated and replacing these reactions with lumped reactions. The *iHJ873* model contains fewer reactions than the *iJR904* model (873 vs. 931, respectively), but $\Delta_r G'^{\circ}$ of every reaction in the *iHJ873* can be estimated. The thermodynamic studies of the *iHJ873* model focused on the individual reactions in the model that

Submitted September 27, 2006, and accepted for publication November 3, 2006.

Address reprint requests to V. Hatzimanikatis, E-mail: vassily.hatzimanikatis@epfl.ch.

© 2007 by the Biophysical Society

0006-3495/07/03/1792/14 \$2.00

doi: 10.1529/biophysj.106.093138

were found to have a large positive $\Delta_r G'^\circ$ in the direction of flux. We simulated the impact of removing these unfavorable reactions on the growth of the cell, and we considered the biological implications that these particular reactions were thermodynamically unfavorable. In this article, we take this work a significant step forward by examining the metabolite concentrations required for every reaction essential for optimal growth to be simultaneously thermodynamically feasible. We propose a new methodology that we call thermodynamics-based metabolic flux analysis (TMFA) for integrating thermodynamic data and constraints into a constraints-based metabolic model to ensure that flux distributions produced by the model are thermodynamically feasible and to provide data on the thermodynamically feasible metabolite activity ranges for the metabolites in the cell. TMFA can also be used for the analysis of unmodified models that are lacking some thermodynamic data to allow for direct analysis of models such as *iJR904* without first creating lumped models like the *iHJ873*. We apply TMFA to analyze the *iJR904* model using new thermodynamic data estimated from an updated and expanded implementation of the group contribution method (M. D. Jankowski, C. S. Henry, L. J. Broadbelt and V. Hatzimanikatis, unpublished); we assess the sensitivity of TMFA to changes in $\Delta_r G'^\circ$ due to uncertainty and ionic strength; and we examine the thermodynamically feasible ranges for biologically important concentrations ratios such as ATP/ADP, NAD(P)/NAD(P)H, and $H_{\text{extracellular}}^+/H_{\text{intracellular}}^+$. Finally, we utilize the $\Delta_r G'$ ranges calculated for the *iJR904* reactions with TMFA to identify candidate reactions for cell regulation as it has been previously proposed (2).

METHODS

Metabolic flux analysis (MFA)

TMFA uses at its core the mass balance constraints of metabolic flux analysis (MFA) (13,23–25). MFA defines the limits on the metabolic capabilities of a model organism under steady-state flux conditions by constraining the net production rate of every metabolite in the system to zero as

$$N \cdot \nu = \mathbf{0}, \quad (1)$$

where N is an $m \times r$ matrix of the stoichiometric coefficients for the r reactions and m metabolites in the model, and ν is an $r \times 1$ vector of the steady-state fluxes through the r reactions in the model. MFA is combined with optimization to determine the limits on the ability of the cell to produce biochemicals such as ethanol (26,27), to predict the maximum possible growth yields of the cell (28–30), and to predict the responses to gene knockouts and additions (31,32).

The introduction of thermodynamics-based constraints in MFA will enforce the exclusion of thermodynamical infeasibilities from flux distribution solutions. One example of these infeasibilities would be flux distributions involving flux through the thermodynamically infeasible internal flux loops mentioned earlier. In addition, these constraints will allow the quantification of the ranges in the gradients of metabolite activities required to drive reactions in the direction of flux reported in all calculated flux distributions. Knowledge of the permissible ranges of metabolite activities is

essential for the development of kinetic models of metabolism and metabolic control analysis (33–38).

Estimation of $\Delta_r G'^\circ$ of reactions in the *iJR904* metabolic model

Formulation of the thermodynamic constraints in TMFA requires knowledge of $\Delta_r G'^\circ$ of the reactions in the model, and it must either be estimated or measured experimentally. Experimental data is available for only a small fraction of the reactions involved in a genome-scale metabolic model such as *iJR904*. Fortunately, the group contribution method provides a means of estimating $\Delta_r G'^\circ$ of nearly every reaction (7,8). In a previous article (21), the group contribution method was used to estimate $\Delta_r G'^\circ$ of 808 of the 931 reactions in the *iJR904* model. Recent improvement and expansion of the group contribution method (M. D. Jankowski, C. S. Henry, L. J. Broadbelt and V. Hatzimanikatis, unpublished) based on a refitting of the group contribution values using the thermodynamic data gathered in the NIST Standard Reference Database (39) and other literature (40–42) have allowed the estimation of $\Delta_r G'^\circ$ for 576 (92%) of the compounds and $\Delta_r G'^\circ$ for 891 (96%) of the reactions in the *iJR904* model. In addition, we have been able to quantify the ranges of uncertainty in the estimated energy values due to variances in experimental measurements and the fitting method. All new estimated thermodynamic data for the *iJR904* model are provided in the Supplementary Material.

In contrast, experimental $\Delta_r G'^\circ$ measurements taken within 10 K and one pH unit of the standard conditions of 298 K and a pH of 7 exist for only 52 (5.6%) of the 931 reactions in *iJR904*. The estimated $\Delta_r G'^\circ$, $\Delta_r G'_{\text{est}}$, for these 52 reactions agree well with the measured $\Delta_r G'^\circ$ (Fig. 1). Literature values exist for $\Delta_r G'^\circ$ of 68 (11%) of the compounds in *iJR904*, and the $\Delta_r G'^\circ$ values from the literature are nearly identical to the estimated $\Delta_r G'^\circ$ (Fig. 2). In all but three cases, the measured $\Delta_r G'^\circ$ values fall within the uncertainty of the estimated $\Delta_r G'^\circ$. All of the experimental $\Delta_r G'^\circ$ and $\Delta_r G'$ data shown in Figs. 1 and 2 were part of the dataset to which the group contribution energies were fit using multiple-linear regression (M. D. Jankowski, C. S. Henry, L. J. Broadbelt and V. Hatzimanikatis, unpublished).

In our previous article (21), all of the reactions in *iJR904* for which the value $\Delta_r G'^\circ$ could not be estimated were lumped to produce a set of net reactions for which the value $\Delta_r G'^\circ$ could be estimated. We then removed the reactions with unknown $\Delta_r G'^\circ$ and replaced them with the net reactions to produce the *iHJ873* model. In the work presented here, although we still lump together the 40 reactions for which $\Delta_r G'^\circ$ cannot be estimated, we do not remove these 40 reactions from the model stoichiometry. Instead, these reactions are treated in the manner discussed in the formulation of the thermodynamic constraints in TMFA described in a following section.

Adjustment of $\Delta_r G'^\circ$ for ionic strength

All TMFA studies performed in this article are in terms of metabolite activities instead of concentrations, making the results of these studies independent of ionic strength. However, ionic strength will have an effect on the $\Delta_r G'^\circ$ value of the reactions, and the zero ionic strength reference state upon which the estimated $\Delta_r G'^\circ$ values are based differs significantly from the ionic strength of the cytosol in which these reactions take place, between 0.15 and 0.20 M (20). We explore the sensitivity of $\Delta_r G'^\circ$ of the reactions in the genome-scale model to ionic strength using the extended Debye-Hückel equation (43,44),

$$\Delta_r G'_j(I) = \Delta_r G'_j(I=0) - 2.303 RTA \sum_i^m \frac{n_{ij} c_i^2 I^{1/2}}{1 + BI^{1/2}}, \quad (2)$$

where I is the ionic strength of the solution, c_i is the charge of species i , and A and B are parameters of the extended Debye-Hückel equation with universally applicable values of $0.5093 \text{ mol}^{1/2}/\text{L}^{1/2}$ and $1.6 \text{ mol}^{1/2}/\text{L}^{1/2}$, respectively (45).

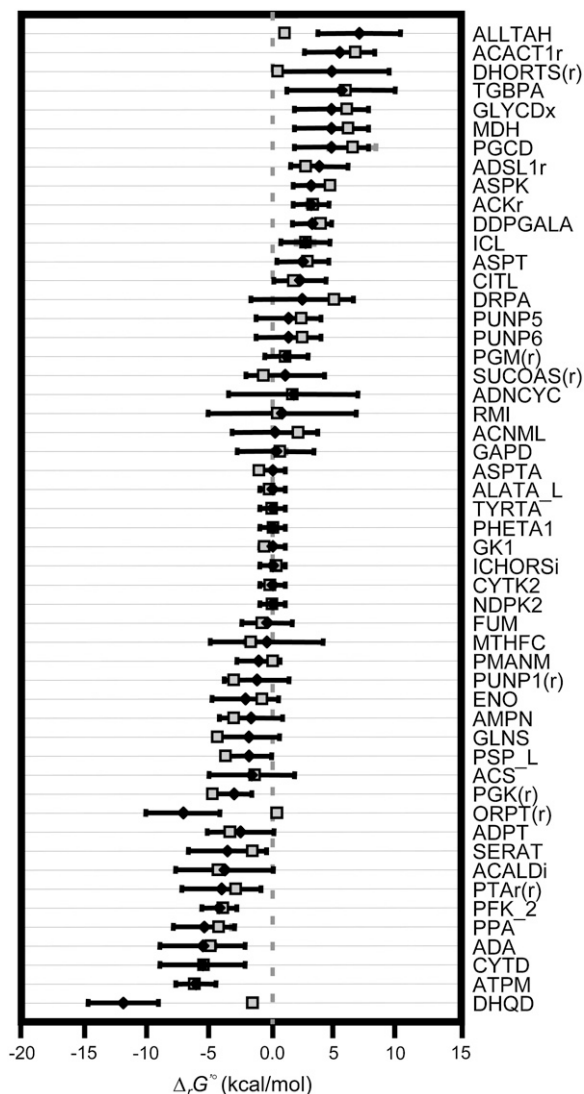


FIGURE 1 Comparison of estimated $\Delta_r G'^{\circ}$ to experimentally measured $\Delta_r G'^{\circ}$. Estimated $\Delta_r G'^{\circ}$ values (solid diamonds) are shown and compared to experimentally measured $\Delta_r G'^{\circ}$ values (shaded squares) for every reaction in the *iJR904* model for which data exists in the NIST database (39) within the temperature and pH limitations (288–308 K and pH of 6–8). In some cases, multiple data points existed in the NIST database at the suitable conditions. In these cases, the average $\Delta_r G'^{\circ}$ is shown along with the standard deviation in the data (shaded error bars). The uncertainty in the estimated $\Delta_r G'^{\circ}$ (solid error bars) calculated from the group contribution method (22) is also shown verifying that all but three of the measured $\Delta_r G'^{\circ}$ fall within the range of uncertainty of the estimated $\Delta_r G'^{\circ}$.

The charge used for each metabolite in the ionic strength studies performed is the charge of the predominant ionic form of each metabolite at pH 7 determined using pK_a estimation software (MarvinBeans pK_a estimation plug-in, Ver. 4.0.3, ChemAxon, Budapest, Hungary). This dramatically simplifies the calculation, although it might result in less accuracy.

Calculation of $\Delta_r G'_i$ for reactions involving transmembrane ion transport

As discussed in a previous article (21), the standard conditions of pH 7 solution and zero ionic strength upon which all $\Delta_r G'_{\text{est}}^{\circ}$ values are based is

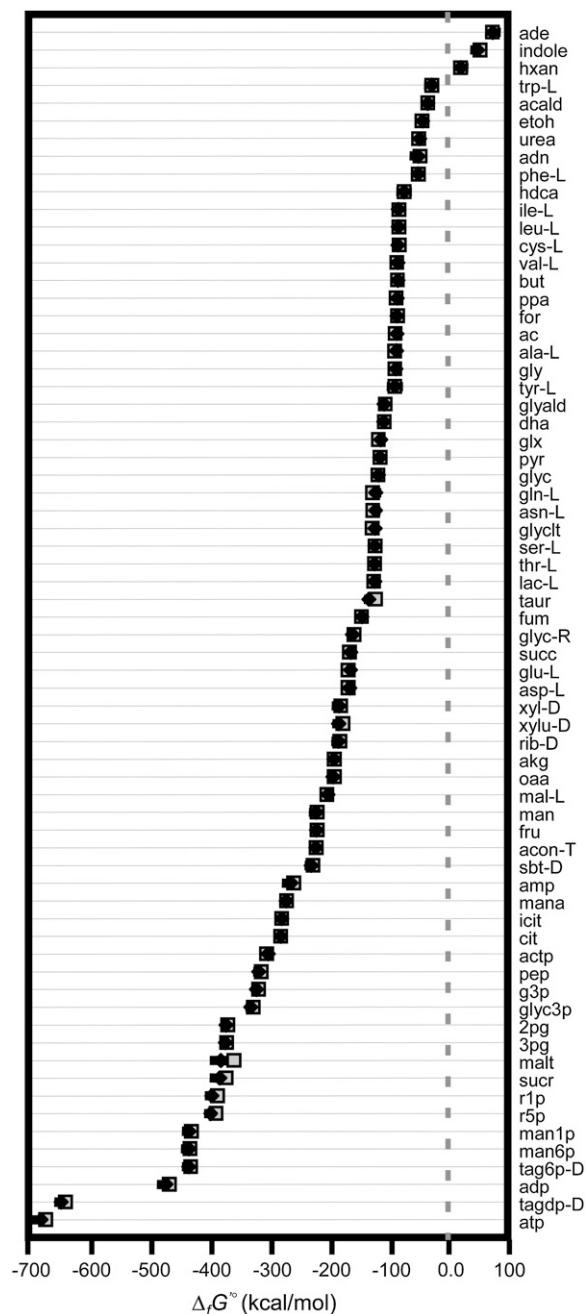


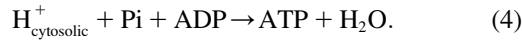
FIGURE 2 Comparison of estimated $\Delta_r G'^{\circ}$ to $\Delta_r G'^{\circ}$ available in the literature. Estimated $\Delta_r G'^{\circ}$ values (solid diamonds) are shown and compared to literature values of $\Delta_r G'^{\circ}$ (shaded squares) for every compound in the *iJR904* model for which literature data was found (40–42). The uncertainty in the estimated $\Delta_r G'^{\circ}$ (solid error bars) calculated from the group contribution method (22) are also shown verifying that all but one of the literature $\Delta_r G'^{\circ}$ values fall within the range of uncertainty of the estimated $\Delta_r G'^{\circ}$.

applied to both the extracellular and intracellular environment when calculating $\Delta_r G'_{\text{est}}^{\circ}$ for reactions involving the transport of metabolites across the cellular membrane. As a result, $\Delta_r G'_{\text{est}}^{\circ}$ for these reactions is based on the assumption that the electrochemical potential, $\Delta\psi$, and the pH gradient, ΔpH

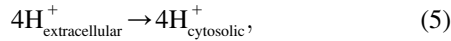
($pH_{\text{intracellular}} - pH_{\text{extracellular}}$), across the cell membrane is zero. For example, the ATP synthase reaction in *E. coli* is typically written in as



The $\Delta_r G'_{\text{est}}$ calculated from the group contribution method only applies to the portion of this reaction that takes place inside the cell, $\Delta_r G'_{\text{est, intracellular}}$,



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



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

$$\Delta_r G'_{\text{est, transport}} = \Delta_{\Delta\psi}G + \Delta_{\text{pH}}G \quad (6)$$

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

$$\Delta_r G'_{\text{est}} = \Delta_r G'_{\text{est, transport}} + \Delta_r G'_{\text{est, intracellular}} \quad (7)$$

Under physiological conditions $\Delta\psi$, and $\Delta_r G'_{\text{transport}}$ depend on ΔpH . $\Delta_{\Delta\psi}G_{\text{est}}$ depends upon $\Delta\psi$, which in turn depends on ΔpH according to the equations (47)

$$\Delta_{\Delta\psi}G(\text{kcal/mol}) = cF\Delta\psi, \quad (8)$$

$$\Delta\psi(\text{mV}) = 33.33\Delta pH - 143.33 \quad (9)$$

(based on a fit of experimental data),

where c is the net charge transported from outside the cell into the cell, and F is the Faraday constant in kcal/mV mol. $\Delta_{\text{pH}}G$ depends only on ΔpH according to the equation (47)

$$\Delta_{\text{pH}}G(\text{kcal/mol}) = -2.3 hRT \Delta pH, \quad (10)$$

where h is the number of protons transported across the membrane. While $\Delta_r G'_{\text{est}}$ used in the TMFA constraints for every reaction involving the transport of ions across the cell membrane will involve a $\Delta_r G'_{\text{transport}}$ term according to Eq. 7, the nature of the $\Delta_r G'_{\text{transport}}$ term depends on the ion being transported. If protons are transported across the cell membrane as in ATP synthase, $\Delta_r G'_{\text{transport}}$ will involve a contribution from both $\Delta_{\text{pH}}G$ and $\Delta_{\Delta\psi}G_{\text{est}}$ according to Eq. 6. If any other ions are being transported across the membrane, $\Delta_r G'_{\text{transport}}$ is equal to $\Delta_{\Delta\psi}G_{\text{est}}$.

Thermodynamics-based metabolic flux analysis (TMFA)

To produce flux distributions that are free of thermodynamic infeasibilities and to allow the exploration of feasible metabolite activities and reaction driving force, TMFA augments the mass balance constraints of MFA with additional thermodynamic constraints. Linear thermodynamics-based constraints have already been proposed and utilized with MFA to eliminate thermodynamically infeasible flux loops (14–16), and nonlinear constraints have been proposed to eliminate flux distributions that utilize reactions that cannot be thermodynamically feasible under physiological conditions (17). However, these constraints and methods have never been applied on a genome scale, and the nonlinear constraints make application to large-scale systems computationally challenging. In the proposed TMFA, the following mixed integer linear constraints are used to produce flux distributions free of any thermodynamic infeasibilities and provide data on feasible metabolite activity profiles and $\Delta_r G'$,

$$N \cdot \mathbf{v} = \mathbf{0}, \quad (11)$$

$$0 \leq v_i \leq z_i v_{\text{Max}}, \{i = 1, \dots, r\}, \quad (12)$$

$$\Delta_r G'_i - K + Kz_i < 0, \{i = 1, \dots, r | \Delta_r G'_i \text{ is known}\}, \quad (13)$$

$$\Delta_r G'_i + RT \sum_{j=1}^m n_{i,j} \ln(x_j) = \Delta_r G'_i, \{i = 1, \dots, r + L \quad (14)$$

| $\Delta_r G'_i$ is known\},

$$\Delta_r G'_i - Ky_i < 0, \{i = r, \dots, r + L\}, \quad (15)$$

$$y_i + \sum_{j=1}^r \alpha_{i,j} z_j \leq \sum_{j=1}^r \alpha_{i,j} \{i = r, \dots, r + L\}, \quad (16)$$

where N is the $m \times r$ stoichiometric matrix, \mathbf{v} is the $r \times 1$ flux vector, r is equal to the total number of reactions in N , and m is equal to the total number of metabolites. The value r is larger than the number of reactions in any model being used because each of the reversible reactions in the model must be split into separate irreversible forward and backward component reactions during the formation of N . This separation is performed so that the flux through each reaction i can be constrained to be greater than or equal to zero. A binary use variable, z_i , is also associated with each reaction i specified in the stoichiometric matrix N . The value of z_i is equal to one if the flux through reaction i , v_i , is positive, and z_i is equal to zero if v_i is zero. This condition is enforced by the constraint described in Eq. 12. The v_{Max} in Eq. 12 is the upper limit on the flux through any reaction typically set to a physiologically reasonable value such as 100 mmol/gm DW/h. Equations 11 and 12 represent the mass balance constraints carried over from MFA with the only difference being the separation of reversible reactions into forward and backward component reactions.

The new thermodynamic constraints of TMFA begin with Eq. 13, which ensures that the activity profiles and flux distributions generated by TMFA adhere to the second law of thermodynamics; a reaction flux cannot be positive unless $\Delta_r G'_i$ is negative. The K in Eq. 13 is a constant selected to be large enough that Eq. 13 is always satisfied if v_i and z_i are zero. The terms involving K in Eq. 13 ensure that the constraint is only applied to reactions with a non-zero flux. Equation 14 is the Gibbs free energy equation used to set the value of $\Delta_r G'_i$ for reaction i given the metabolite activities. In Eq. 14, $\ln(x_j)$ is the natural logarithm of the activity of compound j , and $n_{i,j}$ is the stoichiometric coefficient of compound j in reaction i . In Eq. 14, the energy contribution for any transport of ions across the cell membrane is accounted for in the $\Delta_r G'_{\text{est}}$ term as described in Eq. 7.

As $\Delta_r G'_{\text{est}}$ of a reaction must be known to formulate the thermodynamic constraints for a reaction, the constraints described by Eqs. 13 and 14 are not applied to the reactions in the model with unknown $\Delta_r G'_{\text{est}}$. Instead, these reactions are lumped into overall reactions with known $\Delta_r G'_{\text{est}}$, and special thermodynamic constraints are applied to the reactions for which $\Delta_r G'_{\text{est}}$ is unknown; these constraints are described by Eqs. 14–16. In these equations, L is the number of lumped reactions, and $\alpha_{i,j}$ is a coefficient equaling one if reaction j is one of the original reactions with unknown $\Delta_r G'_{\text{est}}$ that makes up the lumped reaction i . When Eq. 14 is applied to a lumped reaction, it sets the value for $\Delta_r G'_i$ of the lumped reaction. Equation 15 is the thermodynamic feasibility constraint for the lumped reactions similar to Eq. 13 except that the binary use variable for each lumped reaction, y_i , is set to zero when the reaction is thermodynamically feasible and one when the reaction is infeasible. Equation 16 excludes flux distributions that involve flux through the set of reactions that comprise an infeasible lumped reaction. The continuous independent variables of this optimization problem are $\ln(x_j)$, v_i , and $\Delta_r G'_i$, and the binary independent variables are z_i , and y_i .

In the TMFA of the *iJR904* model, the 40 reactions with unknown $\Delta_r G'_{\text{est}}$ are lumped into 20 reactions with known $\Delta_r G'_{\text{est}}$. An additional two internal flux loops were found within the stoichiometry of the reactions with unknown $\Delta_r G'_{\text{est}}$. Being internal flux loops, the $\Delta_r G'$ of these reactions is constrained to zero, i.e., Eqs. 15 and 16 constrain the flux through these

loops to zero. All 20 lumped net reactions and the two internal flux loop reactions are listed in the Supplementary Material.

TMFA with uncertainty in thermodynamic data

The values of $\Delta_r G'^{\circ}$ used in the thermodynamic constraints of TMFA can be based on either group contribution estimates or experimental observations, and they are subject to some uncertainty due to the error allowed when fitting group contribution energy values or in measuring $\Delta_r G'^{\circ}$ experimentally. The TMFA must account for this uncertainty to avoid overconstraining the ranges of possible values for metabolite activities and reaction $\Delta_r G'$ due to inaccurate $\Delta_r G'^{\circ}$ values. This uncertainty can be accounted for by allowing the $\Delta_r G'^{\circ}$ values used in the TMFA constraints to vary within the range of the uncertainty.

For example, when the $\Delta_r G'^{\circ}$ values used in TMFA are estimated using the group contribution method, the potential variance in $\Delta_r G'_{\text{est}}^{\circ}$ due to uncertainty is captured by allowing the energy contributions of the molecular substructures used in the group contribution method to vary. The error in the energy contribution of molecular substructure i , $E_{g,i}$, is introduced as a new variable in the TMFA formulation, and $E_{g,i}$ is allowed to vary within two standard errors, σ , reported for the group contribution energies in the new group contribution scheme (M. D. Jankowski, C. S. Henry, L. J. Broadbelt and V. Hatzimanikatis, unpublished),

$$-2\sigma_i \leq E_{g,i} \leq 2\sigma_i (i = 1, \dots, q), \quad (17)$$

where q is the number of molecular substructures used in the group contribution scheme, and σ_i is the standard error reported for the group contribution value of molecular substructure i . The thermodynamic constraint specified in Eq. 14 is now altered to include terms accounting for the uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$,

$$\Delta_r G'_{\text{est},i}^{\circ} + RT \sum_{j=1}^m n_{i,j} \ln(x_j) + \sum_{j=1}^q g_{i,j} E_{g,j} = \Delta_r G'_i, \quad (18)$$

where $g_{i,j}$ is the number of occurrences of molecular substructure j being created or destroyed during reaction i .

RESULTS

Optimization of aerobic growth on glucose using TMFA

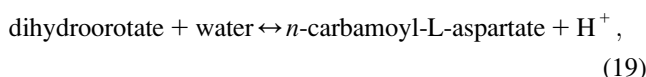
We applied TMFA to determine the maximum growth yield achievable in the *iJR904* genome-scale model. In our TMFA analysis, we chose to study aerobic growth on glucose as a carbon source. The concentrations of the extracellular compounds used as nutrient sources in the model were fixed to the typical concentrations found in growth media (48). These concentrations are listed in Table 1. The concentrations of all intracellular species were restricted within the ranges observed in the cell (between 10^{-5} M and 0.02 M) (49) with the following exceptions. The concentration of H^+ in the cell was held constant at 10^{-7} M, or a pH of 7. Because the oxygen concentration selected for the media is 8.2×10^{-6} M (Table 1) and the oxygen concentration in the cell cannot exceed the concentration in the media, the bounds on the oxygen concentration in the cell were set from 10^{-7} M to 8.2×10^{-6} M. The extracellular pH was allowed to vary between a pH of 4 and 11 to allow for exploration of the thermodynamically feasible extracellular pH ranges for

TABLE 1 Concentrations of nutrients in media

| Molecules | Concentration |
|-------------------------------|------------------------|
| Phosphate | 0.056 M |
| Sulfate | 0.0030 M |
| Ammonium | 0.019 M |
| Sodium | 0.16 M |
| Potassium | 0.022 M |
| Fe^{2+} | 0.062 M |
| CO_2 | 0.00010 M |
| Oxygen | 8.2×10^{-6} M |
| Glucose | 0.020 M |
| H^+ extracellular | pH 4–11 |

E. coli. Specifics on metabolite uptake and secretion constraints used in the analysis are listed in the Appendix.

Initially, uncertainty was not accounted for and σ_i in Eq. 17 was set to zero. With zero uncertainty, the maximum growth yield determined from TMFA was zero, suggesting that growth was not feasible due to the thermodynamic infeasibility of one or more essential reactions. Under these conditions, the reaction dihydroorotase,



which must operate in the reverse direction for any biomass production to occur, is thermodynamically infeasible with a $\Delta_r G'_{\text{est}}^{\circ}$ of 4.7 kcal/mol in the reverse direction. Because *E. coli* is capable of growing on glucose, this reaction must be thermodynamically feasible under physiological conditions. One possible way to overcome the infeasibility of this reaction is to allow for a wider range of values for the activities of the reactants and the products in this reaction. While the activities of water and H^+ are both lumped into $\Delta_r G'^{\circ}$ and thereby fixed, expanding the minimum and maximum bounds on the activities of dihydroorotate and *n*-carbamoyl-L-aspartate to 0.8×10^{-5} M and 0.025 M, respectively, allows sufficient driving force to exist for dihydroorotase to be thermodynamically feasible. Performing TMFA on the model with the bounds on the activities of dihydroorotate and *n*-carbamoyl-L-aspartate set to these expanded values results in an optimal growth yield of 0.0923 g biomass/mmol glucose. This is the same optimal growth yield observed for the *iJR904* model using mass balance constraints alone (50). Under these conditions, only dihydroorotase is sensitive to the thermodynamic constraints; there are no other active thermodynamic bounds.

However, another potential explanation for the infeasibility of dihydroorotase is inaccuracy in $\Delta_r G'_{\text{est}}^{\circ}$, the estimated values of $\Delta_r G'^{\circ}$. Allowing the error in the group contribution values to vary within the range of two standard deviations for the group contribution energies, the optimal growth yield for the *iJR904* model is achievable within the original selected activity range of 1×10^{-5} M and 0.02 M. While experimental evidence does verify that dihydroorotase is a

thermodynamically unfavorable reaction, the experimentally observed $\Delta_r G'^{\circ}$ is 1.4 kcal/mol, or 3.3 kcal/mol $< \Delta_r G'_{\text{est}}^{\circ}$. This deviation is within the range of two standard deviations for $\Delta_r G'_{\text{est}}^{\circ}$ of dihydroorotase, which is 4.7 ± 4.6 kcal/mol. These results suggest the importance of accounting for uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$, since ignoring uncertainty can lead to feasible reactions being incorrectly labeled as infeasible. Although dihydroorotase can be thermodynamically feasible in the direction required for growth at the experimentally observed $\Delta_r G'^{\circ}$ of 1.4 kcal/mol, this reaction is still less favorable than 91% of the reactions in the *iJR904* model. Interestingly, while dihydroorotase is not part of a multienzyme complex in *E. coli* (51), dihydroorotase is a part of a multienzyme complex in mammalian cells (52) indicating that mammalian cells could be utilizing substrate channeling as a means of overcoming the unfavorable $\Delta_r G'^{\circ}$ of this reaction without the use of large concentration gradients.

Essential, substitutable, and blocked reactions

Under optimal growth conditions, reactions in *E. coli* may be classified as essential (requiring a non-zero flux for optimal growth to occur), substitutable (capable of carrying zero or non-zero flux at optimal growth), or blocked (do not carry any flux at optimal growth). Flux variability analysis (FVA) (53) was used to classify the reactions involved in the maximum production of biomass from glucose in *E. coli*. In FVA, the flux of each reaction in the stoichiometric matrix is individually minimized and maximized subject to the mass balance constraints on the metabolites (Eq. 11) and the growth conditions being analyzed (optimal aerobic growth on glucose). When the reversible reactions are expressed as separate forward and backward component reactions, reactions with a positive minimum flux are essential; reactions with a minimum and maximum flux of zero are blocked; and reactions with a zero minimum flux and positive maximum flux are substitutable. When maximizing growth in the *iJR904* model using only mass balance constraints, 272 (29%) reactions are essential, 83 (9%) reactions are substitutable, and 576 (62%) reactions are blocked. When maximizing growth in the *iJR904* model using mass balance and thermodynamic constraints, 272 (29%) reactions are essential, 53 (6%) reactions are substitutable, and 606 (65%) reactions are blocked. The addition of thermodynamic constraints causes 30 of the reactions that were classified as substitutable using FVA alone to become blocked. These reactions are all part of thermodynamically infeasible flux loops in the *iJR904* model, demonstrating the effectiveness of the thermodynamic constraints in restricting the net flux through these loops to zero while still allowing flux through the individual reactions involved in the loops. Overall, every reaction involved in aerobic growth of *E. coli* on glucose that is not participating in an internal flux loop is indeed thermodynamically feasible under the activity conditions studied. These results hold true whether or not uncertainty in

$\Delta_r G'_{\text{est}}^{\circ}$ is accounted for as long as the expanded bounds on the activities of dihydroorotase and *n*-carbamoyl-L-aspartate are used.

Sensitivity of TMFA to the ionic strength reference state

A reference state of zero ionic strength is not an accurate representation of the intracellular environment in *E. coli*, as the ionic strength of the cytosol in *E. coli* is known to range between 0.15 and 0.20 M (20). To determine the effect of the high intracellular ionic strength on $\Delta_r G'_{\text{est}}^{\circ}$, we utilized Eq. 2 to adjust the reference ionic strength of $\Delta_r G'_{\text{est}}^{\circ}$ for all of the reactions in the *iJR904* model to a new reference state of 0.2 M. The effect of ionic strength on $\Delta_r G'_{\text{est}}^{\circ}$ was found to be very small for most reactions. For 95% of the reactions in *iJR904*, the difference between $\Delta_r G'_{\text{est}}^{\circ}$ ($I = 0$) and $\Delta_r G'_{\text{est}}^{\circ}$ ($I = 0.2$ M) is < 1 kcal/mol (Fig. 3). There are six exceptional reactions in the *iJR904* model for which $\Delta_r G'_{\text{est}}^{\circ}$ ($I = 0.2$ M) is > 10 kcal/mol $< \Delta_r G'_{\text{est}}^{\circ}$ ($I = 0$). However, these six reactions also have $\Delta_r G'_{\text{est}}^{\circ}$ ($I = 0$) values that are all < -80 kcal/mol, and these reactions are all large lumped reactions involved in membrane lipid metabolism.

In the case of thermodynamic bottleneck reactions such as dihydroorotase, even the small change to $\Delta_r G'_{\text{est}}^{\circ}$ that results from the adjustment for ionic strength is sufficient to make a reaction become infeasible. For dihydroorotase, $\Delta_r G'_{\text{est}}^{\circ}$ ($I = 0.2$ M) is 0.73 kcal/mol higher than $\Delta_r G'_{\text{est}}^{\circ}$ ($I = 0$). While this is sufficient to make this reaction infeasible under the concentration ranges studied when the uncertainty is set to zero, this is not sufficient to make this reaction infeasible

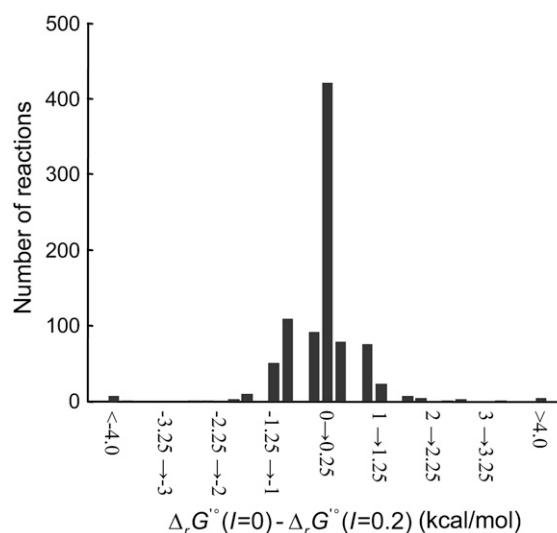


FIGURE 3 Effect of ionic strength on $\Delta_r G'^{\circ}$. A histogram of the differences between $\Delta_r G'^{\circ}$ at an ionic strength of zero ($\Delta_r G'^{\circ}$ ($I = 0$)) and $\Delta_r G'^{\circ}$ at an ionic strength of 0.2 M ($\Delta_r G'^{\circ}$ ($I = 0.2$ M)) for all of the reactions in the *iJR904* model. The bin sizes in the histogram are all 0.25 kcal/mol. Greater than 95% of the differences are within 1 kcal/mol of zero.

when the uncertainty is accounted for. Overall, the variation in $\Delta_r G'_{\text{est}}^{\circ}$ due to the ionic strength of solution falls well within the uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$.

Thermodynamic variability analysis (TVA): ranges of the Gibbs free energy

The permissible ranges of the new system variables of metabolite activity and $\Delta_r G'$ introduced in TMFA can be explored using linear optimization in a methodology we call thermodynamic variability analysis (TVA), which is analogous to FVA. In TVA, the activity of each metabolite and $\Delta_r G'$ of each reaction are minimized and maximized subject to the thermodynamic and mass balance constraints. TVA allows for the identification of the thermodynamic bottlenecks in the metabolic network. Thermodynamic bottlenecks are reactions for which $\Delta_r G'$ is constrained to be approximately zero, meaning these reactions must operate very close to equilibrium. Any small decrease in the concentrations of the reactants or increase in the concentrations of the products in these reactions is sufficient to force the flux to zero (18,19).

We first studied the system in the absence of uncertainty, i.e., TVA was performed with the uncertainty set to zero while using the expanded bounds on the dihydroorotase reactants and the default bounds on the remaining metabolite activities (*black error bars* in Fig. 4). Under these conditions, the $\Delta_r G'$ values of the reactions are constrained by the ranges of the values of the metabolite activities that allow every reaction required for optimal growth to be simultaneously thermodynamically feasible (Fig. 5 and Table 2). Only one reaction, dihydroorotase, has a $\Delta_r G'$ that is constrained to near zero, indicating that this reaction is the only thermodynamic bottleneck under these conditions.

We next performed TVA with the uncertainty in $\Delta_r G'^{\circ}$ accounted for by allowing the error in the group contribution energies to vary according to Eq. 17. The expanded bounds on activities of the metabolites in the dihydroorotase reaction were also used to keep the conditions consistent with the TVA study conducted with uncertainty set to zero and allow for comparison of the results. Due to the relatively large uncertainty ranges in $\Delta_r G'^{\circ}$, all of the reactions can achieve a wide range of $\Delta_r G'$ values, and none of the $\Delta_r G'$ values are constrained near zero, indicating that no reactions behave as thermodynamic bottlenecks under these conditions (*red error bars* in Fig. 4).

Thermodynamic variability analysis (TVA): ranges of metabolite activity

TVA was also used to study the ranges of the activities of the metabolites. When TVA was performed with the uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$ set to zero, the activities of the metabolites involved in the bottleneck reaction dihydroorotase were highly constrained (*black error bars* in Fig. 5). The activity of the primary product of the reaction, dihydroorotate, is fixed at

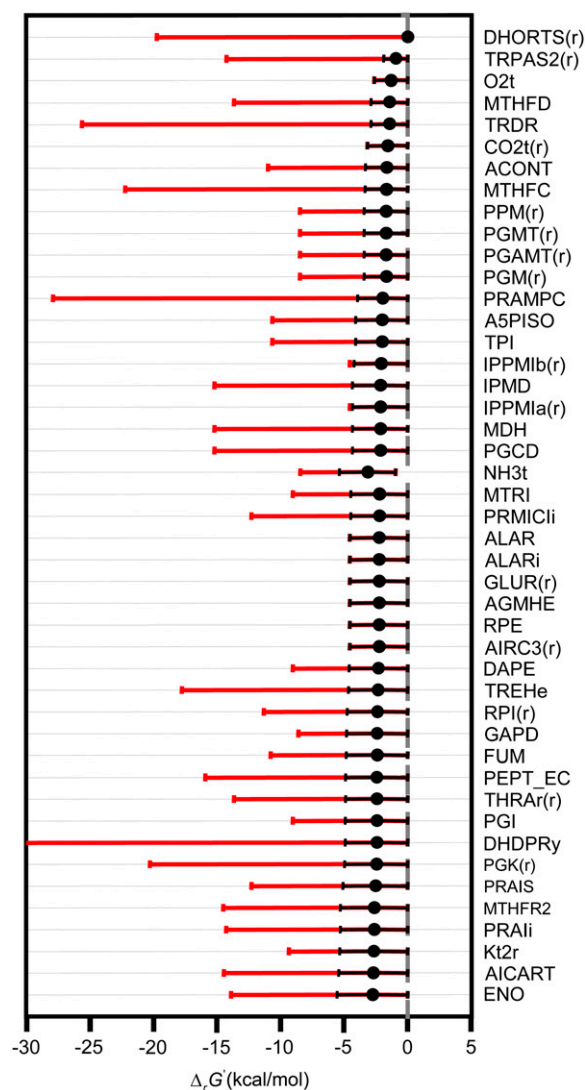


FIGURE 4 Ranges for $\Delta_r G'$ of required and substitutable reactions. The thermodynamically feasible ranges for $\Delta_r G'$ of the 45 essential and substitutable reactions with the narrowest feasible $\Delta_r G'$ range. The ranges are the widest when the uncertainty in $\Delta_r G'^{\circ}$ is accounted for by allowing the group contribution energy values to vary within their standard errors (*red error bars*). If the uncertainty is assumed to be zero, the feasible ranges for $\Delta_r G'$ decrease significantly (*black error bars*), and $\Delta_r G'$ of the reaction dihydroorotase becomes constrained to near zero. When the two error bars do not overlap completely, the larger error bar always corresponds to the range calculated, accounting for uncertainty.

the minimum concentration while the primary reactant, *n*-carbamoyl-L-aspartate, is fixed at the maximum concentration. When the uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$ is accounted for, concentrations of most compounds are nearly unconstrained (*red error bars* in Fig. 5). One important exception is oxygen. As mentioned earlier, the intracellular oxygen concentration is always fixed to the very low concentration range of $<10^{-5}$ M because the intracellular oxygen concentration must be lower than the extracellular concentration for diffusion into the cell to occur. Interestingly, this limita-

TABLE 2 Most thermodynamically constrained reactions

| Abbreviation | Reaction name |
|--------------|---|
| DHORTS(r) | Dihydroorotase (reverse) |
| TRPAS2(r) | Tryptophanase (reverse) |
| O2t | O ₂ transport via diffusion |
| MTHFD | Methylenetetrahydrofolate dehydrogenase |
| TRDR | Thioredoxin reductase |
| CO2t(r) | CO ₂ transport via diffusion (reverse) |
| ACONT | Aconitase |
| MTHFC | Methylenetetrahydrofolate cyclohydrolase |
| PPM(r) | Phosphopentomutase (reverse) |
| PGMT(r) | Phosphoglucomutase (reverse) |
| PGAMT(r) | Phosphoglucosamine mutase (reverse) |
| PGM(r) | Phosphoglycerate mutase (reverse) |
| PRAMPC | Phosphoribosyl-AMP cyclohydrolase |
| A5PISO | Arabinose-5-phosphate isomerase |
| TPI | Triose-phosphate isomerase |
| IPPMb(r) | 2-isopropylmalate hydratase (reverse) |
| IPMD | 3-isopropylmalate dehydrogenase |
| IPPMa(r) | 3-isopropylmalate dehydratase (reverse) |
| MDH | Malate dehydrogenase |
| PGCD | Phosphoglycerate dehydrogenase |
| NH3t | Ammonia reversible transport |
| MTRI | 5-methylthioribose-1-phosphate isomerase |
| PRMICi | Phosphoribosyl imidazolecarboxamide isomerase |
| ALARi | Alanine racemase |
| GLUR(r) | Glutamate racemase (reverse) |
| AGMHE | ADP-glycero-manno-heptose epimerase |
| RPE | Ribulose 5-phosphate 3-epimerase |
| AIRC3(r) | Phosphoribosylaminoimidazole carboxylase (reverse) |
| DAPE | Diaminopimelate epimerase |
| RPI(r) | Ribose-5-phosphate isomerase (reverse) |
| GAPD | Glyceraldehyde-3-phosphate dehydrogenase |
| FUM | Fumarase |
| PEPT_EC | Ethanolamine phosphotransferase |
| THRAr(r) | Threonine aldolase (reverse) |
| PGI | Glucose-6-phosphate isomerase |
| DHDPRy | Dihydrodipicolinate reductase |
| PGK(r) | Phosphoglycerate kinase (reverse) |
| PRAIS | Phosphoribosylaminoimidazole synthase |
| MTHFR2 | 5,10-methylenetetrahydrofolate reductase |
| PRAIi | Phosphoribosylanthranilate isomerase |
| AICART | Phosphoribosylaminoimidazolecarboxamide formyltransferase |
| S7PI | Sedoheptulose 7-phosphate isomerase |
| HCO3E | HCO ₃ equilibration reaction |
| FBA | Fructose-bisphosphate aldolase |
| ENO | Enolase |

tion does not have a large effect on the cell as only two essential reactions require oxygen as a reactant, cytochrome oxidase bo3 and 2,3-diketo-5-methylthio-1-phosphopentane degradation, and these reactions have very negative $\Delta_r G'_{est}$ values of -36.8 and -124 kcal/mol, respectively, ensuring thermodynamic favorability despite small reactant concentration.

Experimentally measured metabolite concentrations are available from a variety of literature sources (49,54,55) for 24 (3.8%) of the metabolites in *iJR904*. These concentrations are compared with the thermodynamically feasible activity

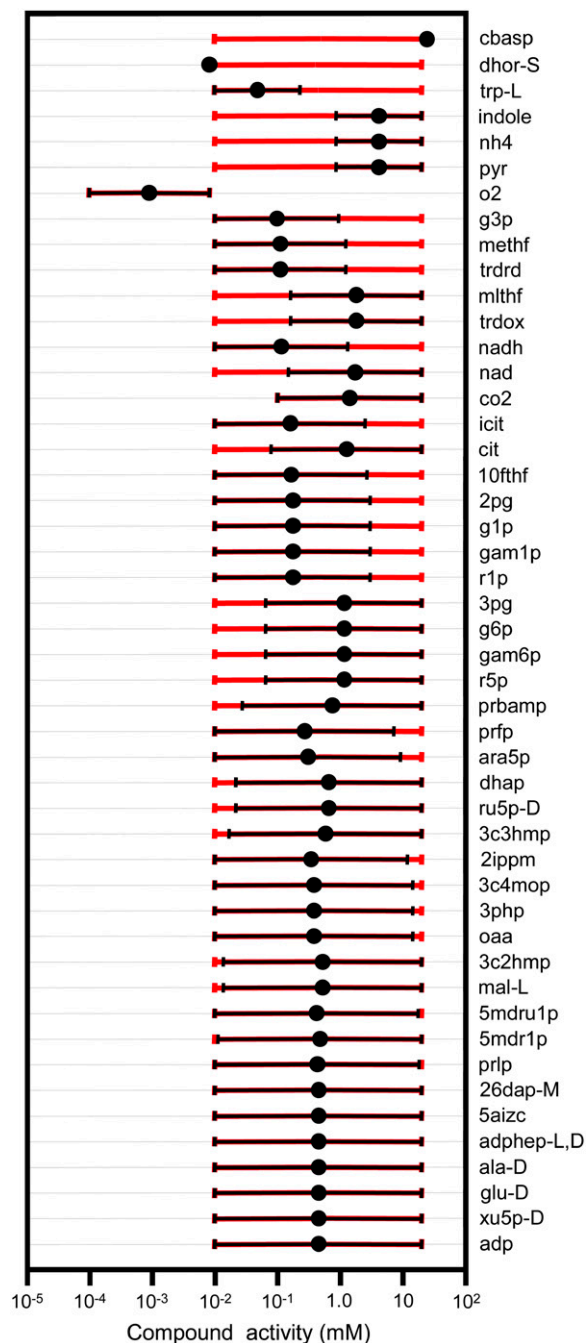


FIGURE 5 Thermodynamically feasible activity ranges. The thermodynamically feasible metabolite activity ranges are shown for all of the compounds with a feasible range that is less than the bounds placed on the metabolite activities (0.01–20 mM with the exception of the compounds listed in Table 1, *n*-carbamoyl-L-aspartate and dihydroorotate). When the uncertainty in the estimated $\Delta_r G'_{est}$ values is assumed to be zero, the metabolite activities are the most tightly constrained (*black error bars*). Tightly constrained metabolites include the main reactant and product of the bottleneck reaction dihydroorotase, *n*-carbamoyl-L-aspartate (cbasp) and dihydroorotate (dhor), respectively. When the uncertainty in the estimated $\Delta_r G'_{est}$ values is accounted for, the feasible activity ranges increase significantly (*red error bars*). When the two error bars do not overlap completely, the larger error bar always corresponds to the range calculated, accounting for uncertainty.

ranges determined from TMFA (Fig. 6), and for all but one metabolite (pyruvate) the experimentally measured concentration data falls within the thermodynamically feasible activity range when uncertainty in $\Delta_r G'_{\text{est}}$ is set to zero. Pyruvate falls within the thermodynamically feasible activity range when $\Delta_r G'_{\text{est}}$ is allowed to vary within the uncertainty. Surprisingly, the measured metabolite concentrations fall near the logarithmic mean ($\sqrt{X_{\text{Max}}X_{\text{Min}}}$) of the minimum and maximum feasible activities determined from TMFA with zero uncertainty for a significant number of important metabolites. In particular, 6pg, f6p, NADH, NAD⁺, akg, g3p, g6p, GTP, asp, and ADP all fall on or near the logarithmic mean of the minimum and maximum feasible activities. This is of particular significance because the concentrations of the metabolites NADH, NAD, and ADP are all under strict regulatory control. One purpose for maintaining the concentrations of these metabolites near the center of the thermodynamically feasible range could be to maximize flexibility. The activity of these metabolites can deviate significantly from the intrinsic value before any of the reactions in which the metabolites are involved become infeasible.

Predicting candidate reactions for regulation based on thermodynamics

As it has been recently proposed (2), the magnitude of $\Delta_r G'$ of a reaction has some implications for the possibility that this reaction is subject to regulation. It has been proposed that reactions with a $\Delta_r G'$ that is close to zero (the reactions we refer to as thermodynamic bottlenecks) have only limited potential for regulation as these reactions are very sensitive to minor perturbations in the concentrations of their reactants. In contrast, reactions with a highly-negative $\Delta_r G'$ have the thermodynamic potential to serve as regulatory control points for the pathways in which they participate because enzyme regulation will be the dominant mechanism for control of the flux through these reactions. The $\Delta_r G'$ range data generated by TVA can be used with this proposed criterion to determine all of the reactions in the genome-scale *iJR904* model that have the potential to be regulatory control points. Specifically, we identified the reactions in the *iJR904* model for which the maximum possible $\Delta_r G'$ as calculated with TVA is < -1.0 kcal/mol, meaning none of these reactions can reach equilibrium under the concentration ranges studied (Table 3). In our analysis of candidate reactions for regulation, we utilized the $\Delta_r G'$ range data produced assuming the uncertainty in $\Delta_r G'_{\text{est}}$ is zero, because we want to focus the analysis on the range of $\Delta_r G'$ values possible given variations in metabolite concentrations alone. While the analysis presented previously (2) focuses solely on the central carbon pathways, we find candidates for regulation in a variety of pathways in the cell. Because the metabolite concentration ranges that we explore are much larger than those used previously (2), we do not identify as many candidate reactions for regulation in central carbon

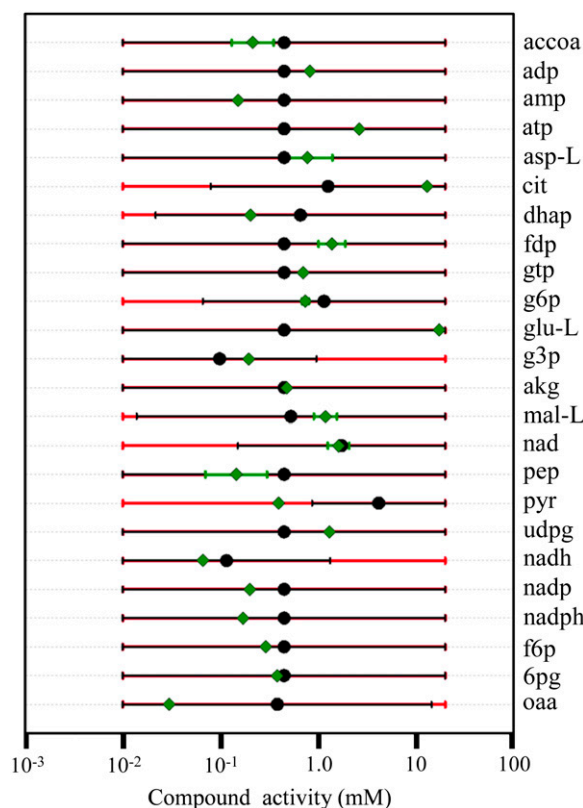


FIGURE 6 Comparison to metabolite concentrations reported in the literature. Experimentally measured concentration ranges (green diamonds and green error bars) reported in the literature (49,54,55) are shown along with the thermodynamically feasible activity ranges for the same metabolites found using TMFA with no uncertainty (black circles and error bars). The black circles show the logarithmic mean for the feasible activity range ($\sqrt{X_{\text{Max}}X_{\text{Min}}}$). The feasible activity ranges for the metabolites found using TMFA accounting for uncertainty in estimated $\Delta_r G'^{\circ}$ values are also shown (red error bars). When the two error bars do not overlap completely, the larger error bar always corresponds to the range calculated, accounting for uncertainty.

(we find six of the 14 reactions reported in Kummel et al. (2), only three of which carry flux; the rest do not carry flux in our analysis because the operating conditions used to generate our flux distributions are different from those used in (2)).

Typically, a key step for regulation in a metabolic pathway in the cell is the first step of a linear pathway (56). Therefore, we studied the locations of the candidate reactions for regulation identified based on thermodynamics, and 30 of the 86 reactions serve as the first step in the pathways in which they participate while five reactions serve as the final step. For example, in aromatic amino-acid biosynthesis, DDPA is the first step in the pathway that produces chorismate; CHORM is the first step in the pathways that produce phenylalanine and tyrosine; PPNDH is the first step of the pathway that produces only phenylalanine; and ANS is the first step of the pathway that produces tryptophan. All of these reactions belong to branching nodes in the aromatic amino-acid

TABLE 3 Candidate reactions for regulation

| Reaction | Pathway | Reaction | Pathway |
|----------------------|--------------------------------------|--------------------|------------------------------------|
| GLYCTO2 | Alternate carbon metabolism | C160SN* | Membrane lipid metabolism |
| GLYCTO3 | Alternate carbon metabolism | C181SN* | Membrane lipid metabolism |
| GLYCTO4 | Alternate carbon metabolism | C141SN* | Membrane lipid metabolism |
| PPA | Anaplerotic reactions | C161SN* | Membrane lipid metabolism |
| MDRPD | Arginine and proline metabolism | PSD_EC | Membrane lipid metabolism |
| DKMPPD | Arginine and proline metabolism | CYSTL* | Methionine metabolism |
| G5SADs | Arginine and proline metabolism | RNDR1 | Nucleotide salvage pathways |
| ORNDC* | Arginine and proline metabolism | RNDR2 | Nucleotide salvage pathways |
| ADMDCr* | Arginine and proline metabolism | RNDR4 | Nucleotide salvage pathways |
| PPTGS [†] | Cell envelope biosynthesis | RNDR3 | Nucleotide salvage pathways |
| UDCPDP | Cell envelope biosynthesis | RNTR1 | Nucleotide salvage pathways |
| KDOPP | Cell envelope biosynthesis | RNTR2 | Nucleotide salvage pathways |
| KDOPS* | Cell envelope biosynthesis | RNTR3 | Nucleotide salvage pathways |
| UAPGR | Cell envelope biosynthesis | RNTR4 | Nucleotide salvage pathways |
| USHD | Cell envelope biosynthesis | TMDS* | Nucleotide salvage pathways |
| PPCDC | Cofactor and prosthetic biosynthesis | CYTBO3 | Oxidative phosphorylation |
| GTPCI* | Cofactor and prosthetic biosynthesis | FDH2 | Oxidative phosphorylation |
| QULNS | Cofactor and prosthetic biosynthesis | FDH3 | Oxidative phosphorylation |
| ASPO3* | Cofactor and prosthetic biosynthesis | HYD1 | Oxidative phosphorylation |
| ASPO4* | Cofactor and prosthetic biosynthesis | HYD2 | Oxidative phosphorylation |
| ASPO5* | Cofactor and prosthetic biosynthesis | PGL [‡] | Pentose phosphate cycle |
| NNDPR | Cofactor and prosthetic biosynthesis | GLUPRT* | Purine and pyrimidine biosynthesis |
| DNTPPA* | Cofactor and prosthetic biosynthesis | IMPC(r) | Purine and pyrimidine biosynthesis |
| ADCL | Cofactor and prosthetic biosynthesis | PRFGS | Purine and pyrimidine biosynthesis |
| ASP1DC* | Cofactor and prosthetic biosynthesis | DHORD2 | Purine and pyrimidine biosynthesis |
| GTPCII2* | Cofactor and prosthetic biosynthesis | DHORD5 | Purine and pyrimidine biosynthesis |
| DB4PS* | Cofactor and prosthetic biosynthesis | OMPDC [†] | Purine and pyrimidine biosynthesis |
| DHPPDA2* | Cofactor and prosthetic biosynthesis | DHDPS* | Threonine and lysine metabolism |
| RBFSa | Cofactor and prosthetic biosynthesis | DAPDC [†] | Threonine and lysine metabolism |
| RBFSb | Cofactor and prosthetic biosynthesis | GLCpts* | Extracellular transport |
| SULR(r) [‡] | Cysteine metabolism | DHQS | Aromatic amino acid metabolism |
| CYSS [†] | Cysteine metabolism | CHORS | Aromatic amino acid metabolism |
| SULabc* | Cysteine metabolism (transport) | DHQD | Aromatic amino acid metabolism |
| GCALDD [†] | Folate metabolism | DDPA* | Aromatic amino acid metabolism |
| FTHFD | Folate metabolism | PPNDH* | Aromatic amino acid metabolism |
| PDH [‡] | Glycolysis/gluconeogenesis | CHORM* | Aromatic amino acid metabolism |
| GLCS1 | Glycolysis/gluconeogenesis | IGPS | Aromatic amino acid metabolism |
| IGPDH* | Histidine metabolism | ANS* | Aromatic amino acid metabolism |
| HISTD [†] | Histidine metabolism | THRD_L* | Val, Leu, and isoleu metabolism |
| IG3PS | Histidine metabolism | ACLS* | Val, Leu, and isoleu metabolism |
| PRATPP | Histidine metabolism | DHAD2 | Val, Leu, and isoleu metabolism |
| C140SN* | Membrane lipid metabolism | DHAD1 | Val, Leu, and isoleu metabolism |
| C120SN* | Membrane lipid metabolism | OMCDC | Val, Leu, and isoleu metabolism |

*Reaction is the first step in the pathway leading to the pathway end-product.

[†]Reaction is the final step in the pathway producing pathway end-products.

[‡]Reaction also identified in the literature (2) as a candidate for regulation.

biosynthesis pathways (56), and all four of these reactions were identified as candidates for regulation using TVA.

The analysis presented here has three advantages over the analysis performed in Kummel et al. (2). First, the thermodynamic data used in our analysis is far more complete, covering 891 of the reactions in the *iJR904* model as opposed to 132 in Kummel et al. (2). This allows for the identification of candidates for regulation in every known metabolic pathway in addition to central carbon. Second, TMFA combines the quantification of the fluxes in the cell metabolism with the assessment of thermodynamic feasibility in a single step. Thus no flux distributions are produced

that must later be corrected due to thermodynamic infeasibilities found. Finally, the metabolite concentration ranges used in our analysis are far less constrained, allowing a more complete exploration of the potential ranges for $\Delta_r G'$ of the reactions in the model.

Exploring the limits on physiologically important concentration ratios

Thermodynamically unfavorable reactions required to take place in cell metabolism are typically driven by hydrolysis of the triphosphate bond in ATP to generate ADP. As such, the

ratio of the ATP activity to the ADP activity is an important and carefully regulated quantity in the cell (54), and $\Delta_r G'$ of many reactions in the cell metabolism depend on this ratio. Similarly, oxidation and reduction reactions taking place in cell metabolism typically utilize NAD(P)^+ as electron sinks and NAD(P)H as electron sources. The $\Delta_r G'$ of these oxidation and reduction reactions depend on the ratio of NAD^+ to NADH and NADP^+ to NADPH , and these ratios have also been found to be under strict regulatory control (54).

We utilized TVA to determine if the physiological levels of the concentration ratios ATP/ADP, NAD/NADH, and NADP/NADPH maintained inside the cell are due to thermodynamic constraints by exploring the range of thermodynamically feasible values for these ratios. We also used TVA to study the limits on the ratio of the intracellular pH to the extracellular pH as all transport of ions across the cell membrane depends on this ratio. The minimum and maximum ratios found are shown in Table 4 along with the values for these ratios found in the literature (47,49,54). We found that without accounting for uncertainty in the group contribution method, all of the ratios found in the literature nearly fit within the thermodynamically feasible ranges for the ratios. Interestingly, the physiologically observed ratio for NAD/NADH is very close to the minimum ratio predicted by TVA, whereas the physiologically observed ratio of NADP/NADPH is very close to the maximum ratio predicted by TVA. This suggests that the values of these physiologically important quantities at optimal growth conditions are determined by the constraints imposed by the thermodynamic properties of the entire metabolic network. When accounting for uncertainty, the estimated ratios NADP/NADPH and NAD/NADH span the entire range of values possible given the concentration limits of 10^{-5} M and 0.02 M, whereas the ratios ATP/ADP and $\text{H}_{\text{int}}^+/\text{H}_{\text{ext}}^+$ remain close to the ranges estimated by the analysis without uncertainty.

The minimum value for the ATP/ADP ratio is constrained by thermodynamics because $\Delta_r G'$ of every kinase reaction in the cell depends on this ratio. As the ATP/ADP ratio decreases, $\Delta_r G'$ of every reaction driven by the dephosphorylation of ATP increases. This includes the many essential phosphorylase and kinase reactions in the nucleotide salvage

pathway responsible for the production of all the other triphosphate nucleotides in the cell. All of these reactions transfer one phosphate from ATP to another nucleotide such as GDP, UDP, or CDP to create ADP and either GTP, UTP, or CTP, respectively. Because the energies of the triphosphate bonds being created and destroyed in these reactions are essentially identical, $\Delta_r G'^{\circ}$ of these reactions are all approximately zero. Therefore, these reactions must be driven entirely by the concentration gradient between the reactants and products. Driving the ATP/ADP ratio to very low levels also drives the ratios GTP/GDP, UTP/UDP, CTP/CDP, and nearly every other triphosphate/diphosphate pair in the cell, to even lower levels to maintain a favorable concentration gradient. A low ATP/ADP ratio results in an unfavorable concentration gradient for every kinase reaction in the cell, reducing the energy generated by the kinase reactions.

The maximum $\text{H}_{\text{intracellular}}^+/\text{H}_{\text{extracellular}}^+$ ratio is constrained by thermodynamics because $\Delta_r G'$ of every reaction involving the transport of ions across the cell membrane depends on this ratio. Many transport reactions in *E. coli*, most notably ATP synthase, are powered by the transport of H^+ across the cell membrane. As the extracellular H^+ concentration decreases, the transmembrane proton concentration gradient becomes more unfavorable reducing the power generated by the transport of the proton into the cell until insufficient power is provided to drive these reactions in the direction required for growth.

DISCUSSION

The addition of thermodynamic constraints to MFA results in both improved accuracy and expanded applicability of flux balance analysis methods. The flux distributions generated using TMFA do not involve flux through any thermodynamically unfavorable flux loops, and no thermodynamically unfavorable reactions given the concentration ranges found in the cell are utilized. Furthermore, TVA allows the exploration of the thermodynamically feasible values for $\Delta_r G'$ of the reactions and the metabolite activities. Some differences do exist between the results presented here and the results presented in the previously published article on this topic (21). These differences are a result of the

TABLE 4 Thermodynamically feasible physiologically important concentration ratios

| | Ratio reported in literature | Min ratio with no uncertainty | Max ratio with no uncertainty | Minimum ratio | Maximum ratio |
|---|------------------------------|-------------------------------|-------------------------------|----------------------|----------------------|
| ATP/ADP | 3.2* | 0.17 | 2.0×10^{38} | 0.022 | 2.0×10^{38} |
| NAD/NADH | 19 [†] | 15 | 2.0×10^{38} | 0.00050 [¶] | 2.0×10^{38} |
| NADP/NADPH | 1.2 [†] | 0.0086 | 1.1 | 0.00050 [¶] | 2.0×10^{38} |
| $\text{H}_{\text{int}}^+/\text{H}_{\text{ext}}^+$ | 0.025–6.3 [‡] | 0.0010 [¶] | 5.0 | 0.0010 [¶] | 7.1 |

*From Albe et al. (49).

[†]From Andersen and Meyenburg (54).

[‡]From Neidhardt and Curtiss (47).

[¶]Equal to the minimum ratio possible given the bounds on the concentrations of these species.

[§]Equal to the maximum ratio possible given the bounds on the concentrations of these species.

combination of improvements in the thermodynamic estimates and an improved ability of TMFA to assess the feasibility of reactions by accounting for all possible metabolite concentrations rather than just mM concentrations used in the previous article.

Uncertainty remains an issue when utilizing the group contribution method to analyze the effect of thermodynamic constraints on metabolite activities and reaction $\Delta_r G'$. All of the feasible ranges for the metabolite activity ratios studied in this article with no uncertainty allowable in $\Delta_r G'^{\circ}$ encompassed the values observed in literature. When uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$ is accounted for, nearly all of the metabolite activities are completely unconstrained, and all of the experimentally measured metabolite activities fall within the range of possible values obtained from TVA. The feasible ranges for $\Delta_r G'$ and metabolite activities determined from TVA accounting for uncertainty in $\Delta_r G'^{\circ}$ are far larger than the true feasible ranges for these quantities. The ranges determined from TVA with no uncertainty allowable are much closer to the true feasible ranges.

Regardless of uncertainty in $\Delta_r G'_{\text{est}}^{\circ}$, very few of the metabolites in *E. coli* are affected by the thermodynamic constraints at all. The majority of the reactions in *E. coli* are thermodynamically favorable, allowing the reactions to remain active under a wide variety of metabolite concentrations and granting the cell a large degree of versatility. While this thermodynamic flexibility prevents the determination of exact values for $\Delta_r G'$ of the reactions, the estimated magnitudes of $\Delta_r G'$ are determined, which can provide valuable insight for the potential of a reaction to undergo regulation. To tightly constrain metabolite activities and reaction $\Delta_r G'$ based on thermodynamic constraints, either a thermodynamic optimization objective must be utilized as done in the work by Beard and Qian (17), or additional nonlinear constraints involving kinetics must be added to the TMFA formalism.

The activity ratio study performed does demonstrate that while the values for the activities of single metabolites have a wide range of feasible values, constraints do exist on the activity ratios of some metabolites of physiological significance. Important metabolite pairs that appear together in many different reactions such as ATP and ADP, NAD and NADH, NADP and NADPH, and intracellular and extracellular protons become thermodynamically coupled to the activity ratios of many other metabolites in the cell. This coupling results in a thermodynamic constraint on these ratios. Interestingly, while the physiological concentrations of key metabolites in the cell are maintained near the middle of the feasible activity range, the NAD/NADH ratio is maintained near the minimum feasible value while the NADP/NADPH ratio is maintained near the maximum feasible value.

While ionic strength can have a large effect on metabolite activities causing the activity of a metabolite to differ from the concentration of the metabolite by over 20%, ionic

strength has a much smaller impact on $\Delta_r G'^{\circ}$ of the reactions. Therefore, the thermodynamic feasibility of reactions can often be accurately assessed from $\Delta_r G'^{\circ}$ of the reactions based on a reference ionic strength of zero. The difference between $\Delta_r G'^{\circ}$ ($I = 0$) and $\Delta_r G'^{\circ}$ ($I = 0.2 \text{ M}$) is also typically much smaller than the uncertainty involved in $\Delta_r G'^{\circ}$ ($I = 0$) making uncertainty the more important factor to take into account. However, these adjustments are based on the extended Debye-Hückel equation, which is applicable for solutions with an ionic strength of $<0.1 \text{ M}$. Applying this relationship to adjust $\Delta_r G'^{\circ}$ to an ionic strength of 0.2 M may affect some of the reported results. The extent of these differences is the subject of ongoing investigations.

Finally, it is important to note that neither the key thermodynamic bottleneck dihydroorotase nor most of the candidates for regulation identified were a part of the central carbon chemistry of the cell. This emphasizes the importance of applying thermodynamic analysis to large-scale genome-based models to account for the highly coupled nature of thermodynamic constraints.

APPENDIX: MFA CONDITIONS

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 0.01 and 0.02 Mol/g dw h, respectively (29). 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 dw h (24,30,57). Under these conditions, the optimal growth on glucose was found to be 0.923 g biomass/g dw 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 (30).

SUPPLEMENTARY MATERIAL

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

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

REFERENCES

1. Alberty, R. A. 1998. Calculation of standard transformed formation properties of biochemical reactants and standard apparent reduction potentials of half reactions. *Arch. Biochem. Biophys.* 358:25–39.
2. Kummel, A., S. Panke, and M. Heinemann. 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* 2:2006–2034.
3. Roels, J. A. 1983. *Energetics and Kinetics in Biotechnology*. Elsevier Biomedical Press, Amsterdam and New York.
4. Westerhoff, H. V., and International Study Group of Biological Thermodynamics Control and Kinetics. 1994. *Biothermokinetics*. Intercept, Andover, Hampshire, England.

5. Nielsen, J. 1997. Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates. *Biochem. J.* 321:133–138.
6. Benson, S. W. 1968. Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters. Wiley, New York.
7. Mavrouniotis, M. L. 1990. Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.* 36:1070–1082.
8. Mavrouniotis, M. L. 1991. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* 266:14440–14445.
9. Alberty, R. A. 1998. Calculation of standard transformed Gibbs energies and standard transformed enthalpies of biochemical reactants. *Arch. Biochem. Biophys.* 353:116–130.
10. 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.* 4:54.1–54.12.
11. Edwards, J. S., and B. O. Palsson. 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA.* 97:5528–5533.
12. Schilling, C. H., M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184:4582–4593.
13. Price, N. D., J. L. Reed, and B. O. Palsson. 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2:886–897.
14. Beard, D. A., S. C. Liang, and H. Qian. 2002. Energy balance for analysis of complex metabolic networks. *Biophys. J.* 83:79–86.
15. Qian, H., D. A. Beard, and S. D. Liang. 2003. Stoichiometric network theory for nonequilibrium biochemical systems. *Eur. J. Biochem.* 270:415–421.
16. Beard, D. A., E. Babson, E. Curtis, and H. Qian. 2004. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.* 228:327–333.
17. Beard, D. A., and H. Qian. 2005. Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *Am. J. Physiol. Endocrinol. Metab.* 288:E633–E644.
18. Mavrouniotis, M. L. 1993. Identification of qualitatively feasible metabolic pathways. In *Artificial Intelligence and Molecular Biology*. L. Hunter, editor. AAAI Press/MIT Press, Menlo Park, California.
19. Mavrouniotis, M. L. 1993. Identification of localized and distributed bottlenecks in metabolic pathways. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 93:273–283.
20. Maskow, T., and U. V. Stockar. 2005. How reliable are thermodynamic feasibility statements in biochemical pathways? *Biotechnol. Bioeng.* 92:223–230.
21. Henry, C. S., M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis. 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* 90:1453–1461.
22. Reference deleted in proof.
23. Papoutsakis, E. T. 1984. Equations and calculations for fermentations of butyric-acid bacteria. *Biotechnol. Bioeng.* 26:174–187.
24. 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.
25. Edwards, J. S., M. Covert, and B. Palsson. 2002. Metabolic modeling of microbes: the flux-balance approach. *Environ. Microbiol.* 4:133–140.
26. 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.
27. Jin, Y. S., and T. W. Jeffries. 2004. Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 6:229–238.
28. 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.
29. 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.
30. Varma, A., and B. O. Palsson. 1993. Metabolic capabilities of *Escherichia coli*. 2. Optimal-growth patterns. *J. Theor. Biol.* 165:503–522.
31. 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.
32. Edwards, J. S., and B. O. Palsson. 2000. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 16:927–939.
33. Famili, I., R. Mahadevan, and B. O. Palsson. 2005. K-cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophys. J.* 88:1616–1625.
34. Wang, L. Q., I. Birol, and V. Hatzimanikatis. 2004. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys. J.* 87:3750–3763.
35. Fell, D. A., and H. M. Sauro. 1985. Metabolic control and its analysis. Additional relationships between elasticities and control coefficients. *Eur. J. Biochem.* 148:555–561.
36. Hatzimanikatis, V., and J. E. Bailey. 1996. MCA has more to say. *J. Theor. Biol.* 182:233–242.
37. Heinrich, R., and T. A. Rapoport. 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.* 42:89–95.
38. Kacser, H., and J. A. Burns. 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27:65–104.
39. Goldberg, R. N., Y. B. Tewari, and T. N. Bhat. 2004. Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics.* 20:2874–2877.
40. Alberty, R. A. 2003. Thermodynamics of Biochemical Reactions. Massachusetts Institute of Technology, Cambridge, MA.
41. Thauer, R. K., K. Jungermann, and K. Decker. 1977. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* 41:100–180.
42. Wagman, D. D. United States National Bureau of Standards, American Chemical Society and American Institute of Physics. 1982. The NBS Tables of Chemical Thermodynamic Properties: Selected Values for Inorganic and C1 and C2 Organic Substances in SI Units. American Chemical Society and the American Institute of Physics for the National Bureau of Standards, New York; Washington, DC.
43. Goldberg, R. N., and Y. B. Tewari. 1989. Thermodynamic and transport-properties of carbohydrates and their monophosphates—the pentoses and hexoses. *J. Phys. Chem. Ref. Data.* 18:809–880.
44. Larson, J. W., Y. B. Tewari, and R. N. Goldberg. 1993. Thermochemistry of the reactions between adenosine, adenosine 5'-monophosphate, inosine, and inosine 5'-monophosphate—the conversion of L-histidine to (urocanic acid plus ammonia). *J. Chem. Thermodyn.* 25:73–90.
45. Goldberg, R. N., and Y. B. Tewari. 1991. Thermodynamics of the disproportionation of adenosine 5'diphosphate to adenosine 5'-triphosphate and adenosine 5'-monophosphate. 1. Equilibrium-model. *Biophys. Chem.* 40:241–261.
46. Reference deleted in proof.
47. Neidhardt, F. C., and R. Curtiss. 1996. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, DC.
48. Yang, C., Q. Hua, T. Baba, H. Mori, and K. Shimizu. 2003. Analysis of *Escherichia coli* anaerobic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout. *Biotechnol. Bioeng.* 84:129–144.

49. Albe, K. R., M. H. Butler, and B. E. Wright. 1990. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.* 143: 163–195.
50. 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.
51. Washabaugh, M. W., and K. D. Collins. 1984. Dihydroorotase from *Escherichia coli*—purification and characterization. *J. Biol. Chem.* 259:3293–3298.
52. Mori, M., and M. Tatibana. 1978. Multi-enzyme complex of glutamine-dependent carbamoyl-phosphate synthetase with aspartate carbamoyltransferase and dihydroorotase from rat ascites-hepatoma cells. Purification, molecular properties and limited proteolysis. *Eur. J. Biochem.* 86:381–388.
53. 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.
54. Andersen, K. B., and K. V. Meyenburg. 1977. Charges of nicotinamide adenine-nucleotides and adenylate energy-charge as regulatory parameters of metabolism in *Escherichia coli*. *J. Biol. Chem.* 252:4151–4156.
55. Peng, L., M. J. Arauzo-Bravo, and K. Shimizu. 2004. Metabolic flux analysis for a ppc mutant *Escherichia coli* based on ¹³C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements. *FEMS Microbiol. Lett.* 235:17–23.
56. Umbarger, H. E. 1978. Amino acid biosynthesis and its regulation. *Annu. Rev. Biochem.* 47:532–606.
57. Gottschalk, G. 1988. *Bacterial Metabolism*. Springer-Verlag, New York.