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 $H^+_{extracellular}/H^+_{intracellular}$ 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 Mavrovouniotis (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.

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 $H^+_{extracellular}/H^+_{intracellular}$. 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

$$\boldsymbol{N} \cdot \boldsymbol{v} = \boldsymbol{0}, \quad (1)$$

where $\boldsymbol{N}$ is an $m \times r$ matrix of the stoichiometric coefficients for the r reactions and m metabolites in the model, and $\boldsymbol{v}$ 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 $\Delta_f G'^\circ$ 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 $\Delta_r G'^\circ$, $\Delta_r G'^\circ_{est}$, for these 52 reactions agree well with the measured $\Delta_r G'^\circ$ (Fig. 1). Literature values exist for $\Delta_f G'^\circ$ of 68 (11%) of the compounds in *iJR904*, and the $\Delta_f G'^\circ$ values from the literature are nearly identical to the estimated $\Delta_f G'^\circ$ (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 $\Delta_f G'^\circ$ 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),

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

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 $mol^{1/2}/L^{1/2}$ and 1.6 $mol^{1/2}/L^{1/2}$, respectively (45).