Quantitative Assessment of Thermodynamic Constraints on the Solution Space of Genome-Scale Metabolic Models

Joshua J. Hamilton, Vivek Dwivedi, and Jennifer L. Reed*
Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin

ABSTRACT Constraint-based methods provide powerful computational techniques to allow understanding and prediction of cellular behavior. These methods rely on physiochemical constraints to eliminate infeasible behaviors from the space of available behaviors. One such constraint is thermodynamic feasibility, the requirement that intracellular flux distributions obey the laws of thermodynamics. The past decade has seen several constraint-based methods that interpret this constraint in different ways, including those that are limited to small networks, rely on predefined reaction directions, and/or neglect the relationship between reaction free energies and metabolite concentrations. In this work, we utilize one such approach, thermodynamics-based metabolic flux analysis (TMFA), to make genome-scale, quantitative predictions about metabolite concentrations and reaction free energies in the absence of prior knowledge of reaction directions, while accounting for uncertainties in thermodynamic estimates. We applied TMFA to a genome-scale network reconstruction of *Escherichia coli* and examined the effect of thermodynamic constraints on the flux space. We also assessed the predictive performance of TMFA against gene essentiality and quantitative metabolomics data, under both aerobic and anaerobic, and optimal and suboptimal growth conditions. Based on these results, we propose that TMFA is a useful tool for validating phenotypes and generating hypotheses, and that additional types of data and constraints can improve predictions of metabolite concentrations.

INTRODUCTION

Genome-scale network reconstructions provide concise mathematical representations of an organism's biochemical capabilities, and serve as a platform for constraint-based techniques that can be used for understanding and predicting cellular behavior (1,2). The predictive accuracy of constraint-based methods depends on the degree to which constraints eliminate physiochemically and biologically infeasible behaviors.

Flux-balance analysis (FBA) (3) is commonly employed to predict the state of the network by identifying a steady-state flux distribution maximizing cellular growth, while also satisfying mass-balance and enzyme capacity constraints. Reaction directionality is typically assigned based on enzyme assays or biological considerations (e.g., no free ATP synthesis), with no consideration given to thermodynamic feasibility. The second law of thermodynamics states that a negative Gibbs energy of reaction ($\Delta_r G$) is needed to drive a forward reaction flux, v, leading to the thermodynamic feasibility constraint $v \cdot \Delta_r G > 0$ for nonzero v.

The first attempt to enforce thermodynamic feasibility on FBA was energy balance analysis (EBA) (4,5), which incorporated nonlinear constraints on chemical potentials into FBA to eliminate closed cycles. Closed cycles are sets of reactions for which the overall thermodynamic driving force is zero, and through which no net flux can occur (e.g., A → B → C → A). These closed cycles have been variously referred to as Type III pathways (6), internal flux cycles (7), and loops (8). Because its constraints are nonlinear, EBA is restricted to small models, though a scalable algorithm has recently been proposed (9). The same scientists also developed a method to compute and eliminate closed cycles in flux-balance models (10,11) based solely on stoichiometry, although the technique remains computationally demanding for genome-scale networks. Recently, a scalable, mixed-integer approach to eliminating closed cycles has also been developed (8).

Due to the limited availability of free energy data for reactions and metabolites (12,13), many of the above approaches do not directly account for the relationships among $\Delta_r G$, metabolite concentrations, and free energies of formation ($\Delta_f G$). Fortunately, group contribution methods (GCMs) (14–16) have been developed to estimate free energies of metabolites and reactions for which data are unavailable. A recent model of *Escherichia coli* used one such GCM to assign reaction directionalities based on thermodynamic feasibility (16,17). In other approaches, experimentally measured thermodynamic data have been combined with heuristics and/or group contribution data to define feasible reaction directions in *E. coli* (18,19). However, these approaches (17–19) neglect thermodynamic interactions between reactions in the network that arise due to shared metabolites. As a result, the directionality of a reaction is assigned independently of other reaction directions in the network. For example, two reactions may be feasible in both the forward and reverse directions, but due to a shared metabolite, the pair of reactions must proceed in the same direction. These approaches fail to capture this type of thermodynamic coupling between reactions.

Submitted October 5, 2012, and accepted for publication June 5, 2013.

*Correspondence: reed@engr.wisc.edu

Editor: Daniel Beard.

0006-3495/13/07/0512/11 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.06.011

GCMs have also been used in approaches that capture thermodynamic interactions by including metabolite concentrations as variables. EBA has been extended to predict intracellular metabolite concentrations in a small network (20), and two mixed-integer approaches have also been developed, in which thermodynamic constraints are imposed on top of predefined reaction directions. NET analysis (21) integrates quantitative metabolomics data with thermodynamic constraints to predict feasible free energy ranges for all reactions in the network. Another method, thermodynamics-based metabolic flux analysis (TMFA) (7), extends FBA with thermodynamic constraints, enabling the quantitative prediction of feasible ranges of metabolite concentrations and reaction free energies, without relying on metabolomic data. However, both of these methods have, to date, relied on prior knowledge of the reversibility or directionality of reactions (7,21–23), thereby restricting their predictive capabilities.

In this work, we examine the extent to which thermodynamics-based flux-balance methods can make genome-scale, quantitative predictions, in the absence of outside information on flux directions, considering both the presence and absence of uncertainty in thermodynamic estimates. To this end, we applied TMFA to the *i*JR904 model of *E. coli* (24). This model was used because group contribution estimates are available for a higher fraction of the metabolites in the *i*JR904 model than in newer models (16). We assessed the predictive performance of TMFA against a number of large-scale datasets (22,25–27), encompassing metabolite concentrations and gene deletion phenotypes, under both aerobic and anaerobic, and optimal and suboptimal growth conditions. Through this analysis, we highlight the importance of quantitative concentration measurements and thermodynamic coupling in achieving physiologically realistic predictions of growth rates and flux distributions. We were also able to generate hypotheses regarding metabolite concentrations and thermodynamic bottlenecks, and we discuss additional types of data and constraints that can improve predictions of metabolite concentrations.

MATERIALS AND METHODS

Overview and relationship to previous implementations

Given a stoichiometric matrix S and set of reactions J, flux-balance analysis (FBA) seeks a steady-state flux distribution (v) maximizing the flux through the biomass reaction (v_{BM}), while also satisfying mass-balance and enzyme capacity constraints for individual reactions, j:

$$\mathbf{max}\ v_{\mathrm{BM}}, \tag{1}$$

$$\mathbf{s.t.}\ S\cdot v = 0, \tag{2}$$

$$0 \le v_j \le v_{\max}\ \ \forall j \in J_{\mathrm{irrev}}, \tag{3}$$

$$v_{\min} \le v_j \le v_{\max}\ \ \forall j \notin J_{\mathrm{irrev}}. \tag{4}$$

Because the network is at steady state, net production of all metabolites is zero. Each reaction is further constrained to have flux within an appropriate range as given by enzyme capacities, with some reactions assumed to be irreversible ($j \in J_{\mathrm{irrev}}$). The limits through most reactions are set to $v_{\max} = 1000$ mmol/gDW/h and $v_{\min} = -1000$ mmol/gDW/h, except for measured fluxes (e.g., carbon uptake rates).

Previous implementations of TMFA (7,23) have imposed thermodynamic constraints on top of FBA, thereby allowing these constraints to further restrict reaction directions; however, a reaction deemed irreversible in FBA cannot become reversible even if indicated by the thermodynamic constraints. In contrast, our implementation makes no assumptions about reaction (ir)reversibility, allowing thermodynamic constraints to decide the directions of all reactions. We replaced Eqs. 3 and 4 with

$$v_{min} \le v_j \le v_{max}\ \ \forall j \in J \tag{5}$$

and rely on thermodynamic constraints alone to determine reaction directions.

Calculating free energies of reaction

Enforcing thermodynamic constraints requires knowledge of the standard transformed Gibbs energy of reaction ($\Delta_r G'^0$) for the reactions in the model. Due to a paucity of experimental data, group contribution methods (14,15) are used to provide estimates and uncertainties of the standard transformed Gibbs energy of formation ($\Delta_f G'^0$) for metabolites and of reaction ($\Delta_r G'^0$) for reactions.

GCMs assume that the $\Delta_f G'^0$ of a metabolite i is a linear combination of the formation energies of its constituent molecular substructures (or groups), k,

$$\Delta_f G_i'^0 = \sum_k n_{i,k}\Delta_{gr}G_k'^0, \tag{6}$$

where $\Delta_{gr} G_k'^0$ is the estimated contribution of group k to the overall $\Delta_f G'^0$, and $n_{i,k}$ is the number of groups k in the molecular structure of compound i. We used a software implementation of the GCM of Jankowski et al. (16), Henry et al. (28), and Finley et al. (29) to obtain estimates and uncertainties of $\Delta_f G'^0$ and $\Delta_r G'^0$ for metabolites and reactions in the *i*JR904 network (see the Supporting Methods and Table S1 and Table S2 in the Supporting Material).

The GCM method returns estimates of $\Delta_f G'^0$ for the predominant ionic species at biochemical standard state: pH 7, zero ionic strength, and temperature 298 K. Simulations were performed at conditions of 298 K and zero ionic strength, at a pH of 7 (7.4) for intra- (extra)cellular metabolites, using the ionic species represented in the *i*JR904 model stoichiometric matrix. These differences in pH and major ionic species required adjusting the GCM estimates of $\Delta_f G'^0$ (see the Supporting Material).

Using our new estimates of $\Delta_f G'^0$, we calculated $\Delta_r G'^0$from the stoichiometry of each reaction,

$$\Delta_r G_j'^0 = \sum_i S_{i,j}\Delta_f G_i'^0, \tag{7}$$

where $S_{i,j}$ is the stoichiometric coefficient of metabolite i in reaction j. We then calculated the transformed Gibbs energy of reaction ($\Delta_r G'$) as a function of metabolite concentration, x_i,

$$\Delta_r G_j' = \Delta_r G_j'^0 + RT\sum_i S_{i,j}\ln(x_i) + \Delta_t G_j'^0, \tag{8}$$

where R is the gas constant, T is the temperature (298 K), and $\Delta_t G'^0$ reflects the contribution to $\Delta_r G'$ from the transport of metabolites across the

membrane. Following an established derivation (7), we calculated $\Delta_t G'^0$ as a function of the electrochemical potential ($\Delta\Psi$) and pH gradient (ΔpH) across the membrane,

$$\Delta_t G'^0_j = c_j F \Delta\Psi - 2.3\, h_j RT \Delta\text{pH}, \quad (9)$$

where F is Faraday's constant, c_j is the net charge transported from outside to inside the cell, and h_j is the number of protons transported across the membrane (see Table S3 for details). Our constraints on intra- and extracellular proton concentrations resulted in values of −130 mV for $\Delta\Psi$, and 0.4 for ΔpH (7,30).

We performed simulations with and without uncertainties in the estimates for $\Delta_{gr} G'^0$ of each group. In TMFA, we fixed values of $\Delta_{gr} G'^0$ to their estimated values ($\Delta_{gr} G'^0_{\text{est}}$) as given by the group contribution method,

$$\Delta_{gr} G'^0_k = \Delta_{gr} G'^0_{k,\text{est}}. \quad (10)$$

Whereas in relaxed TMFA (RTMFA), we allowed $\Delta_{gr} G'^0$ of each group to vary within its 95% confidence interval, as determined by the standard error (SE) reported by the GCM software (i.e., $SE_{\Delta_{gr} G'^0_{k,\text{est}}}$):

$$\Delta_{gr} G'^0_{k,\text{est}} - 2SE_{\Delta_{gr} G'^0_{k,\text{est}}} \le \Delta_{gr} G'^0_k \le \Delta_{gr} G'^0_{k,\text{est}} + 2SE_{\Delta_{gr} G'^0_{k,\text{est}}}. \quad (11)$$

See the Supporting Methods in the Supporting Material for additional details.

Enforcing thermodynamic consistency

Thermodynamic consistency requires that reaction fluxes are consistent with predicted values of $\Delta_r G'$ (i.e., $v \cdot \Delta_r G < 0$). We employed a mixed-integer approach to enforce this requirement, in which a binary variable δ indicates if a reaction is operating in the forward ($\delta = 1$) or reverse ($\delta = 0$) direction. We then added the following equations to our model:

$$(1-\delta) v_{\min} \le v \le \delta v_{\max}, \quad (12)$$

$$-M\delta + \varepsilon \le \Delta_r G' \le M(1-\delta) - \varepsilon. \quad (13)$$

Here Eq. 13 is a big-M constraint (31) in which M is an upper limit on $\Delta_r G'$ (300 kcal/mol) and ε is a small nonzero number (10^{-6}). Equations 1, 2, 5–9, and 12–13 were used in TMFA and equations 1, 2, 5–8, and 10–13 were used in RTMFA.

Flux and thermodynamic variability analysis

We performed flux and thermodynamic variability analysis under a variety of conditions to identify thermodynamically feasible flux and metabolite concentration ranges. In flux variability analysis (FVA) (32), the flux v through each reaction is minimized and maximized subject to the constraints of TMFA. In thermodynamic variability analysis (TVA) (7), the concentration of each metabolite (or $\Delta_r G'$ of each reaction) is minimized and maximized subject to the constraints of TMFA. Additional details can be found in the Supporting Methods in the Supporting Material.

Using our FVA results, we defined sets of reactions that can operate only in the forward (J_{for}) or reverse (J_{rev}) directions, or are blocked (J_{bl}) entirely. By imposing the following constraints on our model, we were able to reduce simulation times by over an order of magnitude:

$$\delta_j = 1 \ \forall j \in J_{\text{for}}, \quad (14)$$

$$v_j \ge 0 \ \forall j \in J_{\text{for}}, \quad (15)$$

$$\delta_j = 0 \ \forall j \in J_{\text{rev}}, \quad (16)$$

$$v_j \le 0 \ \forall j \in J_{\text{rev}}, \quad (17)$$

$$v_j = 0 \ \forall j \in J_{\text{bl}}. \quad (18)$$

For some reactions, flux maximization using RTMFA was computationally intensive, so all FVA simulations were performed with a time limit of 5 min. As a result, there is a possibility that true flux ranges are larger than reported (see the Results).

Differences in phenotype: CONGA

We used an algorithm we previously developed, CONGA (33), to identify single, double, and triple gene deletions predicted to be lethal in FBA but not TMFA (or RTMFA). CONGA employs a bilevel optimization problem to identify gene deletion strategies maximizing the difference in maximum biomass flux between two different models, FBA and TMFA (or RTMFA). Because the maximum possible difference in biomass flux occurs when one model has zero biomass flux, CONGA identifies gene deletions that are lethal in only one model. Additional details can be found in the Supporting Methods in the Supporting Material and the original work (33).

Synthetic lethals and phenotype correction

Once we identified gene deletions lethal only in the FBA model, we employed the synthetic lethals (SL) Finder (34) to identify the reaction(s) responsible for the difference in predicted phenotype. SL Finder employs a bilevel optimization problem to identify reaction deletion strategies that minimize the maximum biomass flux through a network. Because the lowest maximum biomass is zero, SL Finder finds lethal reaction deletions. Additional details can be found in the Supporting Methods in the Supporting Material and the original work (34).

For some gene deletions predicted to be lethal in FBA, the TMFA prediction disagreed with experimental observation. This suggested that the SL reaction did not operate in vivo in the direction predicted by FBA. In these instances, we developed a constraint on metabolite concentrations that prevented the SL reaction from operating in the TMFA-predicted direction, thereby correcting the phenotype. We refer to such constraints as phenotype-correction constraints. Additional details can be found in the Supporting Methods in the Supporting Material.

Simulation conditions

All simulations were performed using CPLEX 12 (IBM, Armonk, NY) accessed via the General Algebraic Modeling System, Ver. 23.3.3 (GAMS; GAMS Development Corporation, Washington, DC). Simulations were performed on a Red Hat Enterprise Linux server with 2.66 GHz Intel Xeon processors and 8 GB of RAM. The TMFA formulation solves in a few seconds, whereas the RTMFA formulation takes ~1 h to prove global optimality.

Sources of experimental data and experimental methods

Details on the experimental datasets and methods used to validate our TMFA predictions are given in Supporting Methods in the Supporting Material.

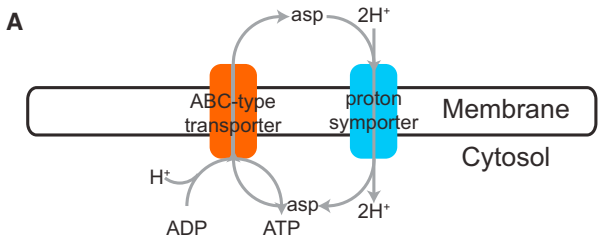
RESULTS

Optimization of aerobic growth on glucose using TMFA

We first used thermodynamics-based metabolic flux analysis (TMFA) to determine the maximum growth rate of *E. coli* under glucose aerobic conditions, using the *i*JR904 genome-scale metabolic model (24). Building on previous work (7), we constrained the concentration of intracellular metabolites to a range of 0.001–20 mM. We constrained the concentrations of extracellular nutrients to that of 0.4% glucose MOPS medium (35). The concentration of intracellular H^+ was fixed to a pH of 7, and extracellular H^+ to a pH of 7.4. The concentration of extracellular gases (O_2 and CO_2) was based on experimental measurement (36), with a requirement that the intracellular concentration be less than the extracellular concentration. Under these conditions, TMFA predicted a maximum growth rate of zero, suggesting that growth was not possible due to the thermodynamic infeasibility of one or more essential reactions. We expanded the concentration range of seven metabolites to enable thermodynamic feasibility of eight essential reactions (see Supporting Results in the Supporting Material). All simulations were performed with glucose as the limiting substrate, with a maximum uptake rate of 10 mmol/gDW/h. The full set of metabolite concentration and exchange flux constraints can be found in Table S6.

We first performed simulations neglecting the uncertainty in the standard Gibbs energy of formation of each group ($\Delta_{gr}\, G'^{0}$) (Eq. 10). The maximum growth rate predicted by TMFA exceeded that of FBA by only 2%, despite the increased network flexibility made possible by the lack of predefined reaction directions in TMFA.

We then introduced uncertainty in $\Delta_{gr}\, G'^{0}$ (Eq. 11) and observed that the maximum growth rate exceeded that of FBA by ~12%. This elevated growth rate highlights the additional network flexibility made possible by the uncertainty in free energy estimates arising from the GCM. In particular, the growth rate difference was due to mechanisms in the relaxed TMFA model (RTMFA) that enable ATP to be synthesized at lower energetic cost than occurs physiologically (and is reflected in FBA). For example, RTMFA should predict ATP synthesis via ATP synthase, using energy released from the transport of four protons across the plasma membrane. Instead, the RTMFA model identifies numerous cycles that synthesize ATP using energy released from the transport of fewer numbers of protons, by shuttling metabolites back and forth across the plasma membrane (Fig. 1 *A*). One such metabolite is aspartate (*asp*), which enables ATP to be synthesized using energy released from the transport of only two protons across the plasma membrane. However, this shuttling relies on the aspartate ABC-type transporter pumping out aspartate, which requires ADP to be present at a concentration higher than ATP (Fig. 1 *B*). By additionally constraining the concentration of ATP to its experimentally measured range (22), in which ATP is present at a higher concentration than ADP, we can force the RTMFA model to use ATP synthase to synthesize ATP. When a constraint on ATP concentration is added to the model, the predicted growth rate exceeds FBA by only 3%. This constraint on ATP concentration was used in all subsequent RTMFA simulations discussed in this work.



B

RTFMA	$[x]_{min}$ (mM)	$[x]_{max}$ (mM)	$[x]_{meas}$ (mM)
ADP	0.89	20	0.437 to 0.704
ATP	0.01	0.22	**8.13 to 11.4**

FIGURE 1 Examples of thermodynamically feasible but physiologically implausible behavior. (*A*) RTMFA predicts ATP can be synthesized by cycling small molecules across the membrane. One such cycle involves asp. (*B*) Metabolite concentration ranges predicted by RTMFA for which the cycle shown in panel *A* is thermodynamically feasible (columns *1* and *2*), and experimentally measured metabolite concentrations (column *3*). (*Highlighted type*) Constraining this concentration in the model renders the cycle thermodynamically infeasible.

Flux variability analysis: thermodynamically feasible reaction directions

We then used flux variability analysis (FVA) to determine thermodynamically feasible directions for all reactions in the network under glucose aerobic conditions (Fig. 2 *A*, and see Table S7 and Table S8). We classified reactions as fully bidirectional, or constrained in one of three ways: blocked entirely; capable of operating in the forward direction only; or capable of operating in the reverse direction only. We first performed FVA on a fully reversible version of the *i*JR904 model without any thermodynamic constraints (FBA_r), in which all reactions (except biomass) were allowed to be reversible. This allowed us to identify directionally constrained reactions based on stoichiometry and the external environment.

We observed that in FBA_r, the majority of reactions are bidirectional (60%), with the remainder being constrained in some way. On the other hand, the FBA model has a large set of irreversible reactions, causing the number of bidirectional reactions to decrease significantly, to a mere 5% of the network. The majority of reactions in the FBA model operate in the forward direction only, or not at all. When we neglect a priori reaction directionality assignments and

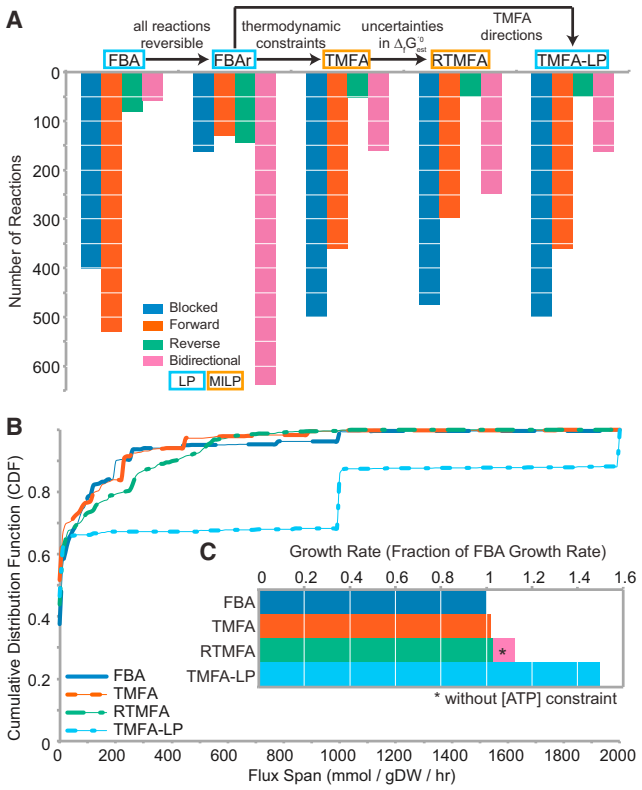


FIGURE 2 Comparison of thermodynamic formulations under glucose aerobic conditions. (*A*) Reactions are classified as blocked, forward, reverse, or bidirectional based on their thermodynamic feasibility. Differences between each formulation are given above the chart. (*B*) Cumulative distribution function (CDF) for the flux span for each formulation. (*C*) Plot of growth rate for each formulation, normalized to the FBA growth rate. (Formulations) FBA_r: FBA with all reaction directions fully reversible. TMFA: FBA_r with thermodynamic constraints, global metabolite bounds, and media constraints. RTMFA: cTMFA with uncertainties in $\Delta_f G'^{0}_{est}$ and a constraint on ATP concentration. TMFA-LP: LP FBA_r with reaction directions consistent with TMF.

instead allow reaction directions to be determined solely by thermodynamic constraints (TMFA), the fraction of directionally constrained reactions becomes 85%, compared to 95% in FBA, and 40% in FBA_r. This reveals that thermodynamic constraints play a major role in eliminating some of the network flexibility resulting from eliminating predefined reaction directions (FBA_r). We also see that, relative to FBA (with predefined reaction directions), TMFA enabled previously forward- or reverse-only reactions to become bidirectional, and previously blocked reactions to become feasible in the reverse direction.

Just as moving from FBA to TMFA led to a decrease in the number of constrained reactions, so too does moving from TMFA to RTMFA. The number of constrained reactions decreases from 85% in TMFA to 77% in RTMFA. The number of bidirectional reactions increases by 50%, as many previously forward-only reactions become feasible in the reverse direction.

When the reaction directionalities from FVA analysis of the TMFA model were used to further constrain fluxes in FBA_r (TMFA-LP), the predicted growth rate was 150% of that predicted by FBA (Fig. 2 *C*). This increase in growth rate indicates ATP- or other energy-generating cycles were present in the network that are not present when thermodynamic constraints are imposed directly. One such cycle involves the shuttling of sodium ions back and forth across the membrane, resulting in a proton gradient used to synthesize ATP (see Fig. S1 in the Supporting Material). This cycle is infeasible under TMFA, despite each reaction operating in a thermodynamically feasible direction. This emphasizes the need to actively impose thermodynamic constraints that account for thermodynamic interactions between reactions (7,8,21), as opposed to methods that impose thermodynamically feasible reaction directions without accounting for thermodynamic coupling between reactions (17–19).

Finally, we used our FVA results from each formulation to find a cumulative distribution function (CDF) for the flux span, the difference between the maximum and minimum flux through a given reaction (Fig. 2 *B* and see Table S9). The CDF for FBA is the sharpest, with >90% of the reactions having a span <300 mmol/gDW/h. The CDF for TMFA is similar to that of FBA, despite the increase in network flexibility. The CDF becomes more shallow for RTMFA (more reactions with larger spans), as a reaction's span can increase if it becomes directionally unconstrained. The CDF for TMFA-LP was the most shallow: jumps in the CDF value at flux spans of 1000 and 2000 mmol/gDW/h point to the existence of many thermodynamically infeasible closed cycles (e.g., A→ B → C → A) in which participating reactions operate at their maximum flux. These results highlight the important role thermodynamic interactions play in shaping a feasible flux space. We also observed that the sets of bidirectional and constrained reactions vary slightly across media conditions, though the overall CDFs remain qualitatively similar (data not shown).

Gene deletion studies: comparison of FBA to TMFA

Under glucose aerobic conditions, CONGA identified 22 single gene deletions for which TMFA and FBA made different growth phenotype predictions (Table 1). In 19 cases, TMFA predicted a nonlethal phenotype and FBA predicted a lethal one, while in the remaining three cases TMFA predicted a lethal phenotype and FBA predicted a nonlethal one. Relative to experimental data (25), this corresponded to a better prediction by TMFA in seven cases, and a worse prediction in the remaining 15 cases. Using RTMFA instead of TMFA introduces another two worse predictions and one better prediction (Table 1).

For TMFA to predict growth when FBA predicts no growth, TMFA must enable a reaction to proceed in a direction not allowed by FBA. We used a variant of SL Finder

TABLE 1 Single-gene deletions for which FBA and/or (R)TMFA predict different growth phenotypes under glucose aerobic conditions

	Gene locus	Gene name	Phenotype	FBA	TMFA	TMFA (corrected)	RTMFA	RTMFA (corrected)
Better in TMFA	b0907	*serC*	Nonlethal	Lethal	Nonlethal	Nonlethal	Nonlethal	Nonlethal
	b0928	*aspC*	Nonlethal	Lethal	Nonlethal	Nonlethal	Nonlethal	Nonlethal
	b2913	*serA*	Nonlethal	Lethal	Nonlethal	Nonlethal	Nonlethal	Nonlethal
	b3359	*argD*	Nonlethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b3429	*glgA*	Nonlethal	Lethal	Nonlethal	Nonlethal	Nonlethal	Nonlethal
	b3430	*glgC*	Nonlethal	Lethal	Nonlethal	Nonlethal	Nonlethal	Nonlethal
	b4388	*serB*	Nonlethal	Lethal	Nonlethal	Nonlethal	Nonlethal	Nonlethal
Worse in both	b0474	*adk*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b0720	*gltA*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b1207	*dnaR*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b2615	*nadK*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b2818	*argA*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b3607	*cysE*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b3608	*gpsA*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Nonlethal
	b3729	*glmS*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Nonlethal
	b3957	*argE*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b3958	*argC*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b3959	*argB*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b4226	*ppa*	Lethal	Lethal	Nonlethal	Lethal	Nonlethal	Lethal
	b3919	*tpiA*	Nonlethal	Nonlethal	Lethal	Lethal	Lethal	Lethal
Worse in TMFA	b1849	*purT*	Nonlethal	Nonlethal	Lethal	Lethal	Nonlethal	Nonlethal
	b2500	*purN*	Nonlethal	Nonlethal	Lethal	Lethal	Nonlethal	Nonlethal
Better in RTMFA	b0888	*trxB*	Lethal	Lethal	Lethal	Lethal	Nonlethal	Nonlethal
Worse in RTMFA	b1136	*icd*	Lethal	Lethal	Lethal	Lethal	Nonlethal	Lethal
	b2780	*pyrG*	Lethal	Lethal	Lethal	Lethal	Nonlethal	Lethal

Columns labeled "corrected" indicate RTMFA predictions after the additional phenotype-correction constraints (described in Table S10 in the Supporting Material) are included.

(34) to identify these reactions for each knockout mutant. In cases where TMFA makes a better prediction, we hypothesized that the SL reaction is active under the mutant phenotype. This hypothesis can be tested by knocking out the SL reaction from the single mutant: the resulting mutant should be incapable of growth. Conversely, when TMFA makes a worse prediction, the SL reaction may not operate in the predicted direction in vivo.

Of the seven deletions for which TMFA made a better prediction than FBA, we selected two for experimental validation: Δ*aspC* and Δ*argD*. Both mutants exhibit robust growth, and removing their SL reactions only requires a single gene deletion. For the Δ*aspC* mutant, SL Finder identified aspartase (encoded by *aspA*) as the SL reaction. After an *aspA::kan aspC* double mutant proved viable (see Fig. S2), we identified two studies reporting *tyrB* as an isozyme for *aspC* (37,38). A *tyrB::kan* Δ*aspC* double mutant proved nonviable (see Fig. S3), confirming that *tyrB* and not *aspA* rescues the Δ*aspC* mutant. We can thus correct the phenotype by adding *tyrB* as an isozyme for *aspC* in the model, and by imposing a constraint that prevents aspartase from operating in the reverse direction. SL Finder also predicted that Δ*argD* was rescued by ornithine transaminase, a reaction for which *argD* and *astC* are reported to have activity (39,40). However, allowing ornithine transaminase to be reversible results in TMFA making four worse predictions while correcting the single Δ*argD* prediction. This suggests that ornithine transaminase might not rescue Δ*argD*. Indeed, an *argD::kan* Δ*astC* double mutant proved viable (see Fig. S4), suggesting some other reaction or that isozyme rescues the Δ*argD* mutant.

In cases where TMFA made a worse prediction than FBA, we developed a phenotype-correction constraint that prevented the SL reaction from operating in the rescuing direction (see Table S10). In some instances, SL Finder predicted multiple reactions that acted together to rescue the phenotype, or that the same reaction rescued multiple phenotypes. For example, SL Finder predicted the reactions ornithine transaminase and *n*-acetylornithine deacetylase acted together to rescue the phenotypes of Δ*argA*, Δ*argB*, and Δ*argC*. All told, the 12 genes for which TMFA made an incorrect nonlethal prediction were associated with 10 SL reactions. We were able to identify metabolite concentration constraints for all 10 of these reactions (see Table S10), which, when implemented in TMFA, resulted in correct predictions for these 12 genes (Table 1). For RTMFA, there were 19 SL reactions associated with the 14 genes for which RTMFA made an incorrect nonlethal predictions. We were able to identify concentration constraints for 17 of these reactions (see Table S10), which, when implemented in RTMFA, resulted in correct predictions for 12 of these 14 genes (Table 1). With the exception of Δ*argD* noted above, correct TMFA and RTMFA predictions were unaffected by these additional phenotype-correction constraints.

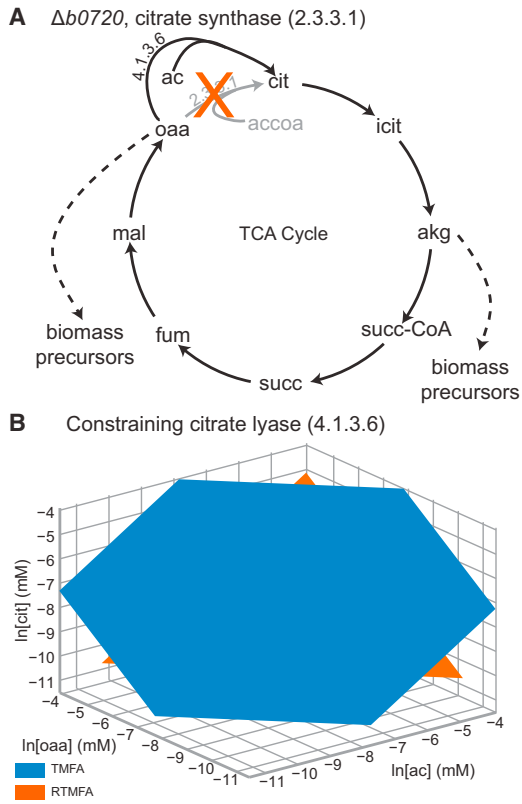


FIGURE 3 Example of reduced concentration spaces imposed by phenotype-correction constraints. (*A*) TMFA incorrectly predicts that the deletion of *gltA* (citrate synthase, 2.3.3.1) is nonlethal. When *gltA* is deleted (large *X*), TMFA predicts that citrate lyase (4.1.3.6) synthesizes citrate (*cit*) by operating in reverse. (*B*) To correct the phenotype, citrate lyase must be constrained to operate only in the forward direction. This requires the ac, oaa, and cit concentrations to lie beneath the shaded surface.

An example of phenotype-correction constraint is illustrated in Fig. 3. TMFA predicts that the deletion of citrate (cit) synthase (*gltA*, EC 2.3.3.1) is nonlethal, with citrate lyase (EC 4.1.3.6) rescuing growth. In this case citrate is synthesized from oxaloacetate (oaa) and acetate (ac), instead of oaa and acetyl-CoA (accoa) (Fig. 3 *A*). We found the phenotype-correction constraint for

$$\Delta gltA \text{ to be } \ln(ac) + \ln(oaa) - \ln(cit) < -2.76$$

for TMFA, and

$$\ln(ac) + \ln(oaa) - \ln(cit) < -7.77$$

for RTMFA. In Fig. 3 *B*, these constraints are indicated by the shaded planes, with concentrations in the half-space below the plane satisfying the constraint.

CONGA also identified a total of 20 double- and triple-deletions for which TMFA predicted a nonlethal phenotype and FBA a lethal one (see Table S11). We were able to find experimental phenotypes for 14 of these multi-gene deletions, and TMFA made a worse prediction in all cases. Taken together with the single-gene deletion data, this suggests that additional concentration measurements are needed to more accurately define the metabolic flux space and predict growth phenotypes.

We also identified three gene deletions for which TMFA falsely predicted a lethal phenotype (Δ*b1949*, Δ*b2500*, and Δ*b3919*) when FBA predicted a nonlethal one (Table 1). In these cases, we hypothesized that the SL reaction active in FBA was thermodynamically infeasible in TMFA. Indeed, in two of the three cases, when uncertainty in free energies was included, RTMFA made the correct prediction (the exception being Δ*b3919*). RTMFA also picked up an additional gene deletion (Δ*b0888*) for which TMFA and FBA both incorrectly predicted a lethal phenotype (Table 1). Thus, although there were cases in which thermodynamic assignment of some reaction directions led to inaccurate growth predictions, there were other cases where thermodynamic constraints were needed to explain observed growth phenotypes.

We also evaluated TMFA for aerobic growth on glycerol and anaerobic growth on glucose (see Table S5). We found that to enable aerobic growth in glycerol M9 medium TMFA required the same concentration constraints as the glucose case; however, for anaerobic growth on glucose, TMFA required expanded concentration ranges on a slightly different set of metabolites. We also performed CONGA on single-gene deletions under both conditions. Under glycerol aerobic conditions, we found 20 gene deletions for which FBA and TMFA made different predictions, with six better and 14 worse predictions (see Table S12). Under glucose anaerobic conditions, we found 26 gene deletions for which FBA and TMFA made different predictions. We then performed growth phenotype screens for the 20 deletion strains available in the Keio collection (see Fig. S5). Assuming the six deletion strains that are unavailable in the Keio collection involve essential genes (25), these screens reveal that TMFA makes a better prediction than FBA in eight cases and a worse prediction in 18 cases (see Table S13).

Thermodynamic variability analysis: ranges of metabolite concentration

Thermodynamic variability analysis (TVA) was used to study the ranges of metabolite concentrations that allow maximal growth on glucose minimal media in the absence of uncertainty in standard Gibbs energy of formation of each group ($\Delta_{gr} G'^{0}$). Using TMFA, we identified a total of 124 (out of 618) intracellular metabolites whose feasible concentration range was less than the default global bounds (0.001−20 mM, see Table 2, and see Table S14 and Fig. S5), indicating that the thermodynamic constraints impose restrictions on metabolite concentrations.

The study of Bennett et al. (22) examined exponential growth of *E. coli*, and reported measurements for 107

TABLE 2 Comparison of model-predicted and experimentally measured metabolite concentration ranges for glucose aerobic conditions at maximal growth

		Predicted by TMFA		
		Unconstrained by thermodynamics	Constrained by thermodynamics	Total
Experimentally Measured	Overlap	57	16	73
	No overlap	11	23	34
	No data	426	85	511
Total		494	124	618

"Constrained" indicates the concentration ranges are tighter than the global bounds (0.001–20 mM).

metabolites in the *i*JR904 model. Of these, predicted concentration ranges overlapped with their measured values in 73 instances, and failed to overlap in 34 (Table 2, Fig. 4, and see Fig. S6). Of these 34 metabolites (Fig. 4 *A*), 12 measurements did not overlap with their predicted values because the measurements fell outside the global bounds defined by our model. In addition, for 11 of these 12 metabolites, the concentration range predicted by TMFA spanned the full range allowed by our global bounds, indicating that changing the global bounds would likely resolve these conflicts. However, doing so is unlikely to result in tighter ranges on predicted metabolite concentrations. For the remaining 22 (out of 34) conflicting metabolites whose measurements did fall within the global bounds, thermodynamic consistency (i.e., predicted concentrations consistent with experimental measurement) could be achieved for all 22 metabolites by allowing for uncertainty using RTMFA. However, the predicted concentration range for all of these metabolites in RTMFA spanned the global range, indicating that RTMFA is unable to predict these metabolite concentrations with great precision. Of the 73 instances of overlap between measured and TMFA predicted values, 16 metabolites had a feasible concentration range that was restricted by thermodynamic constraints (Fig. 4 *B*), while 57 did not (i.e., the ranges were the same as the global bounds). These results suggest TMFA may be more suitable as a framework for incorporating measured concentration data into constraint-based models, rather than a tool to predict experimental concentration measurements.

A second study, from Ishii et al. (27), examined *E. coli* grown in continuous culture at 0.2 h^{-1}, and reported measurements for 88 metabolites in the *i*JR904 model. We performed TVA at this growth rate to examine the effect of suboptimal growth on predicted metabolite concentrations. In general, we find that fewer metabolites are predicted to have constrained ranges during suboptimal growth (97 compared to 124, see Table S15 and Fig. S7), most likely due to the increase in network flexibility associated with suboptimal growth. We also compared the model predictions to measurements taken from the dataset of Ishii et al. (27) and found that more predictions agree with experimental measurements (74 of 88 measurements, see Table S16, and Fig. S8, Fig. S9, and Fig. S10). Of the 68 metabolites measured in both studies, 42 predictions agreed with experimental measurements in both studies.

Examination of thermodynamic bottlenecks

Previous studies have utilized thermodynamic constraints to identify candidate reactions for regulation (those with transformed Gibbs energy of reaction ($\Delta_r G'$) far from zero) (21,23), and to identify thermodynamic bottlenecks in cellular metabolism (7,23). Thermodynamic bottlenecks were first defined as reactions that render metabolic pathways infeasible for a given system with known concentrations (41,42). The term was later used (7,23) to describe reactions for which $\Delta_r G'$ is close to equilibrium. Such reactions are feasible for only a narrow concentration

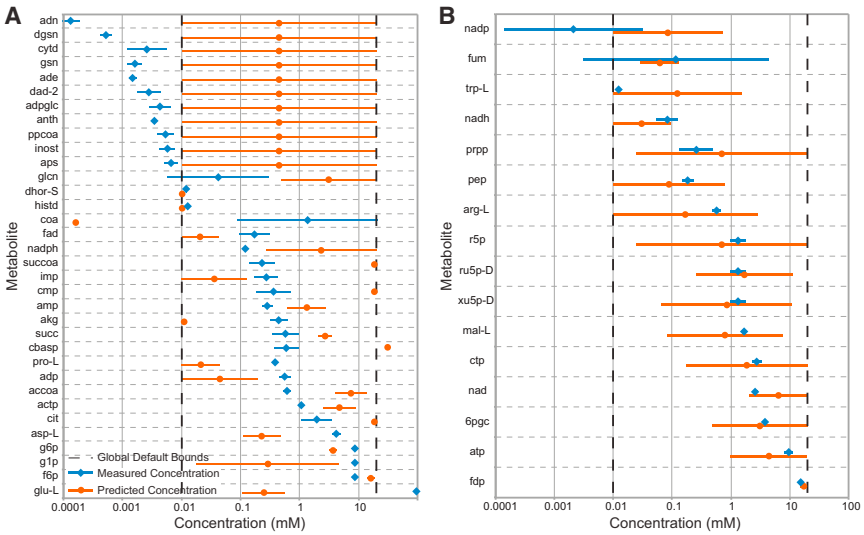


FIGURE 4 Comparison of model-predicted and experimentally observed metabolite concentrations for maximal growth. The dataset of Bennett et al. (22) contains metabolite concentrations for 107 metabolites in *i*JR904. (*A*) Plot of the 34 metabolites for which experimental and theoretical concentration measurements do not overlap. (*B*) Plot of the 16 metabolites for which experimental and theoretical concentration measurements overlap, and for which metabolite concentration ranges are constrained by thermodynamics. (*A* and *B*) Circles (*diamonds*) Mean predicted (measured) metabolite concentration . (*Horizontal bars*) Full concentration range. Metabolite abbreviations can be found in the Supporting Material.

range. Thus, the term "thermodynamic bottleneck" refers to a bottleneck in the space of potential metabolite concentrations. Thermodynamic models can provide quantitative values for the feasible concentration range of metabolites associated with a bottleneck. TVA was used to study the ranges of reaction Gibbs energies ($\Delta_r G'$) that allow maximal growth on glucose minimal media in the absence of uncertainty in $\Delta_{gr} G'^0$. Using TVA, we identified a total of 168 reactions that must operate in a single direction and whose free energy range includes equilibrium (see Table S17).

One such reaction operating very close to equilibrium is phosphoribosylaminoimidazole synthase (PRAIS), an intermediate step in purine biosynthesis. We used TVA to identify the range of $\Delta_r G'$ and metabolite concentrations consistent with cellular growth for each reaction and metabolite in the pathway (Fig. 5). TVA shows that the driving force for PRAIS is a large concentration gradient between 2-formamido-*n*(1)-(5-phospho-D-ribosyl)acetamidine (fpram) and 5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide (air), and PRAIS is only feasible for a narrow range of fpram and air concentrations. Conversely, reactions with a large positive or negative $\Delta_r G'$ should be relatively insensitive to metabolite concentration, and TMFA confirms this prediction. Reactions such as GLUPRT ($\Delta_r G' < 0$) remain feasible for concentration ranges spanning several orders of magnitude. Finally, PRPP synthetase, the primary regulatory control point for purine biosynthesis (30), also has a free energy range far from zero.

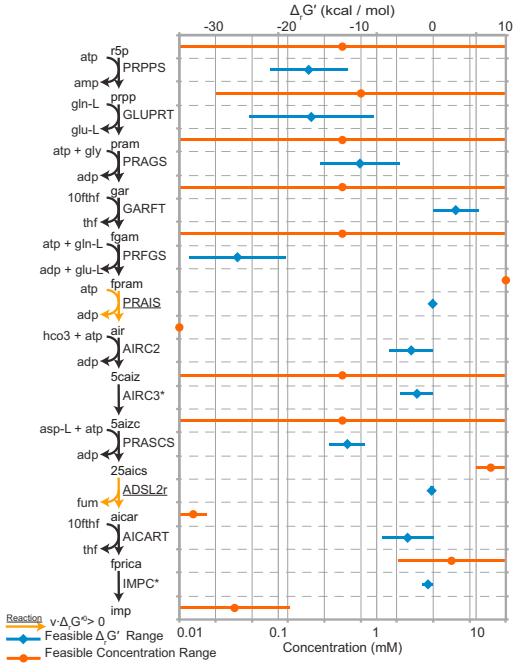


FIGURE 5 Model-predicted concentrations and $\Delta_r G'$ ranges for metabolites and reactions in the purine biosynthesis pathway. (*Circles* and *diamonds*) Mean predicted concentrations ($\Delta_r G'$). (*Horizontal bars*) Full concentration ($\Delta_r G'$) range. Underlined reaction abbreviations signify that $\Delta_r G'$ and $\Delta_r G'^0$ have opposite signs. Starred reaction abbreviations indicate that the reaction direction in the *i*JR904 model is opposite that shown in the figure, and $\Delta_r G'$ values have had their sign reversed to agree with the direction shown. Metabolite and reaction abbreviations can be found in the Supporting Material

DISCUSSION

TMFA modifies flux balance analysis with thermodynamic constraints, allowing for expanded predictive capability of constraint-based methods. TMFA ensures that all reactions operate in thermodynamically feasible directions, eliminates thermodynamically infeasible closed cycles, and accounts for thermodynamic coupling between reactions in the network. In this work, we demonstrate that thermodynamic constraints can provide a guide for predicting reaction directions in the absence of prior knowledge. We also systematically evaluated the impacts of these thermodynamic constraints on metabolic flux distributions and cellular growth rates, and we highlight the importance of explicitly accounting for thermodynamic coupling between reactions. We used TMFA to make both qualitative and quantitative predictions, and have validated these predictions against a variety of genome-scale datasets. We show how these predictions can generate hypotheses regarding reaction directions and thermodynamic bottlenecks. We found many instances in which predictions of metabolite concentrations lack precision and/or accuracy, in which case additional types of data or constraints can be incorporated into TMFA. Identifying what additional data are most useful is an important question that should be addressed in the years ahead.

In this work, we found that TMFA was able to achieve physiologically realistic predictions of growth rates and flux distributions in the absence of uncertainty in the estimated contributions of groups ($\Delta_{gr} G'^0_{k,\text{est}}$) to formation energies ($\Delta_f G'^0$), but concentration measurements for ATP were required in the presence of uncertainty. While it is encouraging that TMFA can reproduce wild-type growth rates with a minimum of experimental data, we found that additional concentration measurements may be necessary to refine growth rate predictions for other conditions (e.g., for knockout mutants). We also observed that slightly different global concentration bounds were necessary to support growth in aerobic versus anaerobic conditions, suggesting that concentration predictions are media-dependent.

TMFA can also be used to generate hypotheses regarding reaction directions and thermodynamic bottlenecks. For example, cofactor pairs such as ATP/ADP appear together in numerous reactions, resulting in constraints on the concentration ratio of the two metabolites (7). However, we found that constraining the ATP concentration was necessary to achieve physiologically realistic behavior in the presence of uncertainty in $\Delta_{gr} G'^0_{k,\text{est}}$ estimates. Thus, it may be that constraints on metabolite ratios serve to drive

reactions in their physiological directions. Physiological reaction directions are often assigned based on in vitro characterization of enzymes under conditions that may vary significantly from those found in vivo. Thus, we envision TMFA complementing other approaches (17–19,43) that are used to calculate thermodynamically feasible reaction directions for new genome-scale models. TMFA also promises to be a useful tool for metabolic engineering applications, by identifying thermodynamic bottlenecks in engineered pathways (44) or by pinpointing those reactions whose reversible operation enables new routes for chemical synthesis (45).

We also observed that the TMFA model is limited in its ability to predict metabolite concentrations. This may be because the majority of reactions in the *i*JR904 network are thermodynamically favorable, and thus relatively insensitive to metabolite concentrations. Obtaining tighter bounds on predicted metabolite concentrations may require the use of a penalty function (46), a thermodynamic objective (20), or the use of kinetic constraints. Recent studies have identified correlations between metabolite concentrations and physiochemical properties (47,48), including the K_M of metabolic enzymes (22). Alternatively, incorporation of known metabolite concentrations may enable TMFA to predict concentrations of metabolites for which data are unavailable. In light of these results, we suggest that, without additional constraints, TMFA is better suited for validating phenotypes and generating hypotheses than for quantitative prediction of metabolite concentrations.

Furthermore, our results suggest that additional types of data and constraints will be needed to improve TMFA's predictions of growth phenotypes and metabolite concentrations due to uncertainties in $\Delta_f G'^{0}_{\text{est}}$. A recently published GCM provides tighter estimates for $\Delta_f G'^{0}_{\text{est}}$ (49), whereas other approaches have successfully combined group contribution estimates with experimentally measured $\Delta_f G'^{0}$ values or equilibrium constants (18,21). We also observed that the inclusion of additional constraints on cofactor concentrations and formation energies (ATP, NAD, NADP, etc.) further constrains the flux space (data not shown). Thus, experimental measurements of $\Delta_f G'^{0}$ for cofactors may be a promising way to improve the accuracy of thermodynamic models. Finally, we note that our approach underestimates uncertainty in $\Delta_r G'^{0}$ by neglecting the error associated with structural groups unchanged by a reaction, an approach that is valid only if the contributions of $\Delta_{gr} G'^{0}$ to $\Delta_f G'^{0}$ are, in fact, linearly additive. Additional types of data will likely be necessary when considering the error associated with these unchanged groups.

Associated with a need to reduce uncertainty in $\Delta_f G'^{0}_{\text{est}}$ is a need to improve model run-time performance, as large mixed-integer programs such as TMFA can be quite cumbersome. A recent Master's thesis examines a number of thermodynamic approaches (EBA, TMFA, etc.) from a theoretical and practical perspective, and provides insights into how to improve solver performance of different formulations (50). We also observed that a priori thermodynamic constraints on reaction directions (Eqs. 14–18) reduced solution times by over an order of magnitude. Alternatively, rather than enforcing strict thermodynamic requirements, one could use thermodynamic and metabolomic data as a guide, and seek a solution that maximizes the consistency with the available data (e.g., by allowing thermodynamic and concentration constraints to be violated, and employing a penalty in the objective). One study suggests that metabolite concentrations remain stable in response to perturbations, implying a single set of metabolomics data (27) could be used to model a variety of conditions.

SUPPORTING MATERIAL

Methods, results, ten figures, and references (51–54) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)00685-1.

The authors acknowledge Mink Arunrattanamook and John de Friel for experimental contributions. The group contribution method software used in free energy calculations was kindly provided by Vassily Hatzimanikatis.

This work was funded by the National Science Foundation through a CAREER grant to J.L.R. (NSF No.1053712) and a Graduate Research Fellowship to J.J.H. (grant No. DGE-0718123). V.D. was supported by the Khorana Program at the University of Wisconsin-Madison.

REFERENCES

1. Oberhardt, M. A., B. Ø. Palsson, and J. A. Papin. 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5:320.
2. Lewis, N. E., H. Nagarajan, and B. Ø. Palsson. 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10:291–305.
3. Orth, J. D., I. Thiele, and B. Ø. Palsson. 2010. What is flux balance analysis? *Nat. Biotechnol.* 28:245–248.
4. Beard, D. A., S. D. Liang, and H. Qian. 2002. Energy balance for analysis of complex metabolic networks. *Biophys. J.* 83:79–86.
5. Qian, H., D. A. Beard, and S. D. Liang. 2003. Stoichiometric network theory for nonequilibrium biochemical systems. *Eur. J. Biochem.* 270:415–421.
6. Schilling, C. H., D. Letscher, and B. Ø. 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.
7. Henry, C. S., L. J. Broadbelt, and V. Hatzimanikatis. 2007. Thermodynamics-based metabolic flux analysis. *Biophys. J.* 92:1792–1805.
8. Schellenberger, J., N. E. Lewis, and B. Ø. Palsson. 2011. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* 100:544–553.
9. Heuett, W. J., and H. Qian. 2006. Combining flux and energy balance analysis to model large-scale biochemical networks. *J. Bioinform. Comput. Biol.* 4:1227–1243.
10. Beard, D. A., E. Babson, …, H. Qian. 2004. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.* 228:327–333.
11. Yang, F., H. Qian, and D. A. Beard. 2005. Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. *Metab. Eng.* 7:251–259.
12. Alberty, R. A. 2003. Thermodynamics of Biochemical Reactions. John Wiley & Sons, Hoboken, NJ.

13. 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.

14. Mavrovouniotis, M. L. 1990. Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.* 36:1070–1082.

15. Mavrovouniotis, M. L. 1991. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* 266:14440–14445.

16. Jankowski, M. D., C. S. Henry, …, V. Hatzimanikatis. 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* 95:1487–1499.

17. Feist, A. M., C. S. Henry, …, B. Ø. Palsson. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3:121.

18. Fleming, R. M. T., I. Thiele, and H. P. Nasheuer. 2009. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. *Biophys. Chem.* 145: 47–56.

19. Kümmel, A., S. Panke, and M. Heinemann. 2006. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics.* 7:512.

20. 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.

21. Kummel, A., S. Panke, and M. Heinemann. 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Sys. Biol.* 2: 2006.0034.

22. Bennett, B. D., E. H. Kimball, …, J. D. Rabinowitz. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5:593–599.

23. Garg, S., L. Yang, and R. Mahadevan. 2010. Thermodynamic analysis of regulation in metabolic networks using constraint-based modeling. *BMC Res. Notes.* 3:125.

24. Reed, J. L., T. D. Vo, …, B. Ø. Palsson. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4:R54.

25. Baba, T., T. Ara, …, H. Mori. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Sys. Biol.* 2: 2006.0008.

26. Joyce, A. R., J. L. Reed, …, S. Agarwalla. 2006. Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J. Bacteriol.* 188:8259–8271.

27. Ishii, N., K. Nakahigashi, …, M. Tomita. 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science.* 316:593–597.

28. Henry, C. S., M. D. Jankowski, …, V. Hatzimanikatis. 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* 90:1453–1461.

29. Finley, S. D., L. J. Broadbelt, and V. Hatzimanikatis. 2009. Thermodynamic analysis of biodegradation pathways. *Biotechnol. Bioeng.* 103:532–541.

30. Neidhardt, F. C. 1996. *Escherichia coli and Salmonella*: Cellular and Molecular Biology, 2nd Ed. ASM Press, Washington, DC.

31. Williams, H. P. 1999. Model Building in Mathematical Programming, 4th Ed. John Wiley & Sons, Hoboken, NJ.

32. 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.

33. Hamilton, J. J., and J. L. Reed. 2012. Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. *PLoS ONE.* 7:e34670.

34. Suthers, P. F., A. R. Zomorrodi, and C. D. Maranas. 2009. Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol. Syst. Biol.* 5:301.

35. Neidhardt, F. C., P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. *J. Bacteriol.* 119:736–747.

36. Yang, C., Q. Hua, …, K. Shimizu. 2003. Analysis of *Escherichia coli* anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout. *Biotechnol. Bioeng.* 84:129–144.

37. Gelfand, D. H., and R. A. Steinberg. 1977. *Escherichia coli* mutants deficient in the aspartate and aromatic amino acid aminotransferases. *J. Bacteriol.* 130:429–440.

38. Berg, C. M., M. D. Wang, …, L. Liu. 1988. Acquisition of new metabolic capabilities: multicopy suppression by cloned transaminase genes in *Escherichia coli* K-12. *Gene.* 65:195–202.

39. Billheimer, J. T., H. N. Carnevale, …, E. E. Jones. 1976. Ornithine δ-transaminase activity in *Escherichia coli*: its identity with acetylornithine δ-transaminase. *J. Bacteriol.* 127:1315–1323.

40. Billheimer, J. T., M. Y. Shen, …, E. E. Jones. 1979. Isolation and characterization of acetylornithine δ-transaminase of wild-type *Escherichia coli* W. Comparison with arginine-inducible acetylornithine δ-transaminase. *Arch. Biochem. Biophys.* 195:401–413.

41. Mavrovouniotis, M. L. 1993. Identification of localized and distributed bottlenecks in metabolic pathways. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1:275–283.

42. Mavrovouniotis, M. L. 1996. Duality theory for thermodynamic bottlenecks in bioreaction pathways. *Chem. Eng. Sci.* 51:1495–1507.

43. Haraldsdóttir, H. S., I. Thiele, and R. M. T. Fleming. 2012. Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophys. J.* 102:1703–1711.

44. Shen, C. R., E. I. Lan, …, J. C. Liao. 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* 77:2905–2915.

45. Dellomonaco, C., J. M. Clomburg, …, R. Gonzalez. 2011. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. *Nature.* 476:355–359.

46. Hoppe, A., S. Hoffmann, and H.-G. Holzhütter. 2007. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.* 1:23.

47. Bar-Even, A., E. Noor, …, R. Milo. 2011. Hydrophobicity and charge shape cellular metabolite concentrations. *PLOS Comput. Biol.* 7:e1002166.

48. Zhu, Q., T. Qin, …, H. Y. Zhang. 2011. Chemical basis of metabolic network organization. *PLOS Comput. Biol.* 7:e1002214.

49. Noor, E., A. Bar-Even, …, R. Milo. 2012. An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics.* 28:2037–2044.

50. Muller, A. C. 2012. Thermodynamic constraints for metabolic networks. Master's thesis, Department of Mathematics and Computer Science, Free University of Berlin, Berlin, Germany.

51. Kim, J., and J. L. Reed. 2010. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* 4:53.

52. Patrick, W. M., E. M. Quandt, …, I. Matsumura. 2007. Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* 24:2716–2722.

53. Yamamoto, N., K. Nakahigashi, …, H. Mori. 2009. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* 5:335.

54. Thomason, L. C., N. Costantino, and D. L. Court. 2007. *E. coli* genome manipulation by P1 transduction. *In* Current Protocols in Molecular Biology.. John Wiley & Sons, Hoboken, NJ. 1.17.1–1.17.8.