Biophysical Chemistry 145 (2009) 47–56

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Quantitative assignment of reaction directionality in constraint-based models of metabolism: Application to *Escherichia coli*

R.M.T. Fleming [a,b,*], I. Thiele [b,c], H.P. Nasheuer [d,e]

[a] *Science Institute, University of Iceland, Reykjavík, Iceland*
[b] *Center for Systems Biology, University of Iceland, Reykjavík, Iceland*
[c] *Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavík, Iceland*
[d] *Cell Cycle Control Laboratory, School of Natural Sciences, National University of Ireland, Galway, Ireland*
[e] *Systems Biology Ireland, Galway, Ireland*

ARTICLE INFO

Article history:
Received 9 July 2009
Received in revised form 21 August 2009
Accepted 23 August 2009
Available online 1 September 2009

Keywords:
Systems biology
Constraint-based modeling
Thermodynamics

ABSTRACT

Constraint-based modeling is an approach for quantitative prediction of net reaction flux in genome-scale biochemical networks. *In vivo*, the second law of thermodynamics requires that net macroscopic flux be forward, when the transformed reaction Gibbs energy is negative. We calculate the latter by using (i) group contribution estimates of metabolite species Gibbs energy, combined with (ii) experimentally measured equilibrium constants. In an application to a genome-scale stoichiometric model of *Escherichia coli* metabolism, iAF1260, we demonstrate that quantitative prediction of reaction directionality is increased in scope and accuracy by integration of both data sources, transformed appropriately to *in vivo* pH, temperature and ionic strength. Comparison of quantitative versus qualitative assignment of reaction directionality in iAF1260, assuming an accommodating reactant concentration range of 0.02–20 mM, revealed that quantitative assignment leads to a low false positive, but high false negative, prediction of effectively irreversible reactions. The latter is partly due to the uncertainty associated with group contribution estimates. We also uncovered evidence that the high intracellular concentration of glutamate in *E. coli* may be essential to direct otherwise thermodynamically unfavorable essential reactions, such as the leucine transaminase reaction, in an anabolic direction.

1. Introduction

Biological systems can be modeled at a large scale by taking an approach which balances computationally tractability with physically and biochemical realistic representation. Constraint-based modeling is a flexible and scalable approach for *in silico* phenotype prediction [1]. It relies on an accurate biochemical network reconstruction which is a biochemically, genetically and genomically structured representation of experimental biochemical and molecular biological literature [2]. In the case of metabolic networks, biochemical characterization of an enzyme establishes the substrate(s) and product(s) and genetic studies *establish* the gene–protein-reaction associations which tie a particular metabolic function in a model to a particular genomic location. A biochemical network reconstruction is then converted into a prototype computational model such that predictions may be compared with experimental data. In many cases, initial *in silico* tests suggest further refinements to the reconstruction underlying the prototype computational model. Iterative refinement of a constraint-based model, by comparison of prediction with experiment, supports its use for *a priori in silico* prediction of phenotypic capabilities for *a posteriori in vivo* experimental validation.

There have been many practical biological uses of constraint-based models, including study of bacterial evolution [3], analysis of network properties [4–7], study of phenotypic behavior [8,9], biological discovery [10–12], and metabolic engineering [13–15]. The growing scope of applications of genome-scale metabolic reconstructions in metabolic engineering and other fields has recently been reviewed [16]. The predictive fidelity of a constraint-based model is dependent on the accuracy of the constraints used to eliminate physicochemically and biochemically infeasible network states. Generally, the resulting constraint equations define an under-determined feasible set of network states. Therefore, in unicellular organisms, a biological objective, such as maximization of growth rate, can be used to predict a single network state within this feasible set depending on the objective. The sensitivity of *in silico* predictions to the choice of objective function was treated in detail by Savinell and Palsson [17], and more recently by comparison of predictions with fluxomics data [18].

In this work, we focus on the assignment of reaction directionality in stoichiometric, metabolic models since it has a significant effect on the feasible set of functional states [19–22]. There are two forms of thermodynamic constraints on reaction directionality. Local thermodynamic constraints apply on a reaction, by reaction basis. Essentially, a negative reaction Gibbs energy dictates a net forward reaction flux. This application of thermodynamics to the direction of biochemical reactions has a long history [23], with the first comprehensive treatment

* Corresponding author. Science Institute, University of Iceland, Reykjavík, Iceland. Tel.: +354 618 6245.
E-mail address: ronan.mt.fleming@gmail.com (R.M.T. Fleming).

0301-4622/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2009.08.007

by Burton et al. [24]. Non-local thermodynamic constraints apply to sets of reactions [25] and arise due to the necessity to satisfy energy conservation, in addition to the second law of thermodynamics. Non-local thermodynamic constraints have been applied to small systems of biochemical reactions [26], but that approach "cannot be efficiently applied directly to genome-scale problems" [27] due to limitations imposed by computational complexity. Here, our focus is on local thermodynamic constraints for a system of reactions at genome scale. We summarize the theory underlying quantitative assignment of local reaction directionality at physiologically relevant conditions. Then, we apply this theory to a stoichiometric model of *Escherichia coli* metabolism [28]. Our study relies on the extensive body of work on the thermodynamics of biochemical reactions by Alberty [29,30] which we apply to a genome-scale model for the first time. We complemented Alberty's approach with ongoing efforts by Henry et al. [19,20] and Jankowski et al. [31] which seek to estimate the standard Gibbs energy of metabolite species based on a group contribution methodology.

2. Methods

2.1. Standard Gibbs energy of formation of metabolite species

There exist two complimentary quantitative methods for assigning reaction directionality based on different ways of calculating standard Gibbs energy of formation for metabolite species. The first involves back-calculation of standard Gibbs energy of formation using experimentally measured equilibrium constants [29]. In the absence of apparent equilibrium constants, standard Gibbs energy of formation for certain metabolite species cannot be back-calculated. In this case, a second complementary method involves estimation of standard Gibbs energy of formation of a metabolite species by summation of terms representing contributions to standard Gibbs energy of formation from different structural subgroups within that metabolite species [31,32] (see Supplementary material).

In constraint-based modeling of metabolic networks, it is typical to assume that all metabolite species concentrations are uniform throughout a single compartment enclosed by a lipid bilayer [33,34]. This assumption is based on a local equilibrium hypothesis; that local variation in metabolite concentration is insignificant with respect to average compartment concentration [35]. In theory, a chemical reaction will perturb the spatial distribution of metabolite concentration, especially for certain metabolite intermediates at low concentration. However, unless the rate of such a reaction is diffusion limited, one can expect such perturbations to be insignificant on longer timescales. A significant number of publications on possible deviations of intracellular reaction thermodynamics from a local equilibrium hypothesis have been published and are reviewed in [36]. However, at present, due to the dearth of spatio-temporally resolved concentration data, one cannot be sure that any deviation from a local equilibrium hypothesis is actually significant *in vivo*. We assume that the metabolite concentration is spatially invariant, and that, temperature and pressure are constant. Therefore, one can define a single Gibbs energy of formation for each metabolite species which is valid throughout a single compartment.

2.1.1. Adjusting to in vivo ionic strength, temperature and pH

The standard Gibbs energy of formation of a metabolite species, $\Delta_f G_j^o$, is an experimental measurement of the change in chemical potential which accompanies the synthesis of 1 mol of a metabolite species from its constituent elements in standard conditions. Standard conditions, denoted with the superscript, ^{o}, specify the synthesis of one mole of a metabolite species (1 M, zero *pH*, zero ionic strength) from its constituent elements in their reference state (temperature 298.15 K, pressure 1 bar). Throughout this study, we assume that atmospheric pressure and all thermodynamic potentials are expressed in kJ mol^{-1}.

In our treatment of thermodynamic potentials, we assume that, with respect to metabolite species, the interior of a cell can be approximated by a buffer solution. This assumption neglects the fact that in general, there are effects on the equilibrium constants of *in vivo* reactions due to macromolecular crowding [37], confinement, adsorption and high fluid phase viscosity [38]. These effects are all a function of size and charge of the molecules involved. We confine ourselves to consideration of equilibrium constants for metabolic overall reactions between relatively small metabolites. Implicitly, with the exception of spontaneous reactions, all such reactions involve enzymes in elementary steps within the overall reactions. Enzymes are relatively large molecules but do not affect the equilibrium constant of an overall enzyme catalyzed reaction. For charged metabolites, there could be significant effects due to macromolecular adsorption and it is known that a higher intracellular fluid phase viscosity slows the rate of diffusion of metabolites [39]. However, in this study, we omit a treatment of these effects. The effect of macromolecular crowding and confinement is much more significant for macromolecules than metabolites and we do not consider these further. We do however consider the effects of *in vivo* pH, temperature and ionic strength.

In any quantitative system of relative measurements it is essential to have a standard reference point. However, any thermodynamic potential given for an *ex vivo* chemical standard reference condition must be adjusted to reflect *in vivo* conditions. The procedure for adjusting standard Gibbs energy of formation for each metabolite species, to *in vivo* ionic strength, temperature and *pH*, has been discussed theoretically and illustrated with computational implementation in textbooks by Alberty [29,30]. We follow closely the same theoretical procedure, which is detailed in the Supplementary material. There have been attempts to indirectly measure *in vivo* ionic strength, without actually measuring the concentrations of all charged metabolite species [40]. Such work would significantly refine our ability to model biochemical thermodynamics. However, at present, a pragmatic approach is to assume a physiological range of ionic strength between 0.05 M and 0.35 M [29], and then to adjust each metabolite species standard Gibbs energy of formation by assuming a particular ionic strength in this range.

We assume an ionic strength of 0.25 M then use the extended Debye–Hückel equation [41], to estimate the activity coefficient associated with each metabolite species, γ_j. The $RT \ln \gamma_i$ term is then absorbed into the standard Gibbs energy of formation at *in vivo* ionic strength. The advantage of this approach is that metabolomic data can be incorporated directly without adjusting it to represent ionic strength and charge dependent activities. The chemical standard reference temperature is typically room temperature, 298.15 K (25 °C). However, generally for biochemical thermodynamics, reactions occur at body temperature, 310.15 K (37 °C). It is possible to adjust metabolite species standard Gibbs energy of formation, within a narrow range of the chemical standard reference temperature, provided that the standard enthalpy of formation of a metabolite species is known, $\Delta_f H_j^o$ (see Supplementary material).

Through a variety of homeostatic mechanisms [42], *E. coli* can grow over a wide range of external *pH* values, (5.5–9.0), yet maintains cytoplasmic *pH* between 7.6 and 7.8 [43,44]. In contrast, the *pH* in the periplasmic space equals that of the external medium [44]. Metabolites, cofactors and ions make up approximately 3.5% of the dry weight of a typical *E. coli* cell [42]. While proteins account for approximately 55% of the dry weight. Cytoplasmic proteins represent a large buffering capacity, relative to the total pool of metabolites, cofactors and ions [45]. Effectively, the large buffering capacity of intracellular proteins maintains a constant concentration of cytoplasmic hydrogen ions. In a thermodynamic treatment of any system, knowledge of the values of a constant intensive variables dictates the appropriate thermodynamic potential to be used. This concept applies regardless of whether the intensive variable is temperature or pH. As stressed by Alberty [29], the appropriate thermodynamic potential may be

mathematically defined using a Legendre transform. However, knowledge of the latter is not essential to understand why it is appropriate to define a new thermodynamic potential, rather it is an elegant mathematical device for doing so. In the Supplementary material, we show how to define a new thermodynamic potential for constant *pH* by adapting the classical thought experiment of a 'bath' that maintains an intensive variable constant.

2.1.2. Pseudo-isomer groups at specified pH

A reactant represents the properties of a set of related metabolite species. When *pH* is a known constant, a reactant may be treated as a pseudo-isomer group of metabolite species each of which is in a different state of hydrogen ion dissociation. e.g. at physiological ranges of *pH*, the reactant ATP consists of a series of different hydrogen ion dissociated forms. The concentration of the reactant ATP is given by the sum of the concentrations of the metabolite species in its pseudo-isomer group

$$[ATP] = [ATP^{4-}] + [HATP^{3-}] + [H_2ATP^{2-}]$$

Thus far, established practice in reconstruction of metabolic networks has been to approximate each reactant with a single predominant ionic form [28,31]. Instead of this approximation, we follow the method of Alberty [29], and quantitatively amalgamate each property of a set of metabolite species into a property of the corresponding reactant. This retains the simplicity of qualitatively dealing with a single reactant but quantitatively representing the properties of a pseudo-isomer group.

Thermodynamically, at the timescale relevant for enzyme catalyzed reactions, the metabolite species which make up a single pseudo-isomer group may be considered to be in equilibrium with each other [46]. The hydrogen ion transfer rate constants range between 10^9 and $10^{11}\,s^{-1}\,M^{-1}$ as compared to the fastest enzyme turnover rates of ~$10^7\,s^{-1}\,M^{-1}$ [47]. Therefore, when modeling enzyme catalyzed reactions, it is reasonable to assume that the different metabolite species of a reactant are at equilibrium [29]. Thermodynamic state variables [41] for each metabolite species, such as standard Gibbs energy of formation, $\Delta_f G_j^o$, must be individually adjusted from chemical standard conditions, to *in vivo pH*, ionic strength, and temperature. Thereafter, assuming equilibrium between metabolite species in a pseudo-isomer group, the adjusted state variables for each metabolite species in a pseudo-isomer group are then combined, and represented by a reactant state variable. The reactant state variable we require is the standard transformed Gibbs energy of a reactant, $\Delta_f G_i^0$. Note that we distinguish a reactant, with index *i*, from a metabolite species with index *j*. Note also, the slight change in superscript, from ^{o} to ^{0}, which is used to remind the reader that the latter is now a function of ionic strength (see Supplementary material 3).

The standard transformed Gibbs energy of formation of a reactant can be calculated by assuming that each metabolite species in a group acts as an isomer and applying established techniques for isomer group thermodynamics [46]. The equation for standard transformed reactant Gibbs energy of formation, $\Delta_f G_i^0$, in terms of metabolite species standard transformed Gibbs energy of formation, $\Delta_f G_i^0$, is

$$\Delta_f G_i'^0 = -RT \ln\left(\sum_j \exp\left(\frac{-\Delta_f G_j'^0}{RT}\right)\right) \quad (1)$$

where the summation is over all the metabolite species, *j*, within the pseudo-isomer group, *i*. Each metabolite species standard transformed Gibbs energy of formation is individually a function of ionic strength, *pH* and temperature. It is important to realize that Eq. (1) does not represent a linear average of the standard transformed Gibbs energy of formation of each metabolite species. In other words, reactant standard transformed Gibbs energy of formation is not additive in the standard transformed Gibbs energy of its metabolite species.

At specified ionic strength, *pH* and temperature, the mole fraction of a metabolite species, r_j, with respect to its pseudo-isomer group, may be calculated using

$$r_j = \frac{\exp\left(\frac{-\Delta_f G_j'^0}{RT}\right)}{\sum_j \exp\left(\frac{-\Delta_f G_j'^0}{RT}\right)} \quad (2)$$

where, again, the summation is over all the metabolite species, *j*, within a particular pseudo-isomer group, *i* [46]. This mole fraction r_j, represents the fraction of a reactant which is present as a particular metabolite species *j*. It is evident from Eq. (2) that the mole fraction of a metabolite species is a non-linear function of its standard transformed Gibbs energy of formation. If each metabolite species in a pseudo-isomer group could be adjusted identically for temperature, ionic strength, and *pH* then it would be possible to adjust the properties of the reactant after it was approximated with the pseudo-isomer group approach. Mathematically, this follows since subtraction of any constant term from each $\Delta_f G_i^0$ in Eq. (2) has no effect on the mole fraction.

2.1.3. In vivo transformed reactant Gibbs energy

Transformed reactant Gibbs energy, $\Delta_f G_i$, is given by $\Delta_f G_i = \Delta_f G_i^0 + RT \ln(x_i)$, where $\Delta_f G_i^0$ is the standard transformed reactant Gibbs energy and x_i is the concentration of reactant *i* (Molar), which is the sum of the concentrations of its constituent metabolite species *j*. Quantitative metabolomic data [48], which can measure simultaneously a wide selection of reactant concentrations, is increasingly becoming available [49]. Quantitative metabolomic data can be used to estimate *in vivo* reactant transformed Gibbs energy [21]. Experimentally measured reactant concentrations indicate that the concentration of most amino acids [50], and reactants in central metabolism fall within the 0.02–20 mM range [21,51]. Therefore, in the absence of metabolomic data for a particular reactant, we define the minimum and maximum standard transformed reactant Gibbs energy of formation, $\Delta_f G'_{i,\min} \equiv \Delta_f G_i^0 + RT \ln(x_{i,\min})$ and $\Delta_f G'_{i,\max} \equiv \Delta_f G_i^0 + RT \ln(x_{i,\max})$, where we assume $x_{\min} = 0.02$ mM and $x_{\max} = 20$ mM. There are exceptions for water and dissolved gases (see Supplementary material 7).

2.1.4. Quantitative assignment of reaction directionality

Although, in principle, all microscopic processes in biochemistry are reversible, certain reactions are effectively irreversible at typical *in vivo* reactant concentrations. Let $\mathbf{S}_k \in \mathbb{Z}^{m,1}$ denote a column from a stoichiometric matrix [1], then the transformed Gibbs energy for a biochemical reaction is $\Delta_r G'_k = \mathbf{S}_k^T \cdot \boldsymbol{\Delta}_f \mathbf{G}'$. In most modeling situations, metabolomic data will not be available for all reactants in a model. In this situation, we assume a physiological concentration range for each reactant. We define the minimum and maximum transformed Gibbs energy for a biochemical reaction with $\Delta_r G'_{k,\min} = \inf\{\mathbf{S}_k^T \cdot \boldsymbol{\Delta}_f \mathbf{G}'\}$ and $\Delta_r G'_{k,\max} = \sup\{\mathbf{S}_k^T \cdot \boldsymbol{\Delta}_f \mathbf{G}'\}$, where inf denotes the infimum, sup denotes the supremum, and reactant standard transformed Gibbs energy of formation is bounded $\boldsymbol{\Delta}_f \mathbf{G}'_{i,\min} \leq \boldsymbol{\Delta}_f \mathbf{G}'_i \leq \boldsymbol{\Delta}_f \mathbf{G}'_{i,\max}$. A reaction may be assigned to be quantitatively forward, if $\Delta_r G'_{k,\max} < 0$, or quantitatively reverse if $\Delta_r G'_{k,\min} > 0$. We assigned a reaction to be quantitatively reversible if the physiological range of biochemical reaction transformed Gibbs energy spans the zero line, i.e. $\Delta_r G'_{k,\min} < 0$ and $\Delta_r G'_{k,\max}$.

2.2. Physiological standard transformed reaction Gibbs energy

Given a set of reactions assigned to be quantitatively reversible, we now propose a metric to rank them in a descending sequence according to probability that each reaction is irreversible in the forward direction. Standard reaction Gibbs energy is defined for 1 M concentration for each reactant, which far exceeds physiological concentration ranges in *E. coli*. We define a new physiological group contribution

estimate for each reactant standard transformed Gibbs energy of formation, $\Delta_f G'^{m}_{est,i} \equiv \Delta_f G'^{0}_{est,i} + RT\ln(x_{i,m})$, where $\Delta_f G'^{0}_{est,i}$ is the 1 M group contribution estimate of reactant standard Gibbs energy of formation, transformed to *in vivo* conditions, and $x_{i,m}$ is the geometric mean of the concentration range for each reactant, $x_{i,m} \equiv \exp((\ln(x_{i,\min}) + \ln(x_{i,\max}))/2)$. This refines the approach of Henry et al. [19], since $x_{i,m}$ represents a new standard concentration for each reactant, depending on its physiological range *in vivo*. We choose the arithmetic mean of minimum and maximum concentration since logarithmic concentration is linear in Gibbs energy.

In the group contribution method, the standard error in estimated reactant standard Gibbs energy of formation, $SE_{\Delta_f G^0_{est}}$, assumes that the actual value of standard Gibbs energy of formation is normally distributed. Likewise, we assume that the actual value of physiological standard transformed reaction Gibbs energy, $\Delta_r G'^{m}$, is also normally distributed. This is justified by comparing the relative magnitude of $SE_{\Delta_f G^0_{est,i}}$ with $RT\ln(x_{i,\max}/x_{i,m})$ for reactants in iAF1260. Assuming a concentration range of 0.02–20 mM, the former is greater than the latter for 85% of reactants, see Fig. 1. That is, for the majority of metabolites, the uncertainty in estimated reactant standard Gibbs energy of formation is more significant than the uncertainty associated with a lack of metabolomic data. The qualitatively reversible reactions, that are assigned to be quantitatively reversible, can be ranked based on the probability that the real physiological standard transformed reaction Gibbs energy is negative, $P(\Delta_r G'^{m}<0)$. This probability that a reaction's direction is forward is given by the cumulative of a normal distribution

$$P(\Delta_r G^{m}<0) = \frac{1}{2}\left(1 + erf\left(\frac{\Delta_r G'^{m} - \Delta_r G'^{m}_{est}}{\sqrt{2}(\mathbf{S}_k^T \cdot (SE_{\Delta_f G^0_{est}}))}\right)\right) \quad (3)$$

where *erf* denotes the error function [52]. Ranking the reactions by the probability that physiological standard transformed reaction Gibbs energy is negative, $P(\Delta_r G'^{m}<0)$, is superior to ranking reactions by mean estimated physiological standard transformed reaction Gibbs energy since the former accounts for the uncertainty in the group contribution method. (For the sake of clarity we omit a *k* subscript from $P(\Delta_r G'^{m}<0)$, but it is specific to each reaction). Note that here probability is used in the Bayesian sense to reflect the confidence with which a directionality assignment can be made. If $P(\Delta_\gamma G'^{m}<0) = 0.5$, then we have no confidence to assign direction either way.

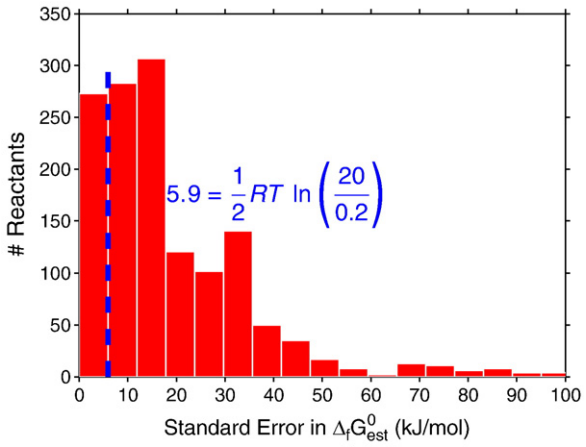

Fig. 1. Uncertainty in reactant Gibbs energy. The majority of reactants in the *E. coli* iAF1260 model have a standard error in estimated standard Gibbs energy of formation which exceeds half the maximum error associated with assuming that reactant concentrations lie between 0.02 and 20 mM. This means that, for the majority of reactants, the uncertainty in reactant Gibbs energy is mostly from uncertainty in estimation of standard Gibbs energy, and not from uncertainty in concentration.

3. Results

3.1. Pseudo-isomer groups

We applied quantitative assignment of reaction directionality to the genome-scale model of *E. coli* metabolism, iAF1260 (1668 reactants, 2076 reactions) [28]. This highlighted a few pertinent methodological lessons which apply regardless of the organism of interest. The assumption that a metabolite is present as a single predominant metabolite species does not always apply. Fig. 2 illustrates that certain reactants in *E. coli* do have significant mole fractions present as non-predominant metabolite species. For instance, acetyl-phosphate is almost equally present as $C_2H_3O_5P^{2-}$ and $C_2H_4O_5P^{-}$, while a small mole fraction is predicted to be present as uncharged $C_2H_5O_5P$. Each metabolite species has a different standard Gibbs energy of formation and responds individually to changes in *pH*, temperature and ionic strength. Where the necessary data was available, we calculated the standard transformed Gibbs energy of formation of each metabolite species as a function of *pH*, temperature and ionic strength. The standard transformed reactant Gibbs energy, and hence standard (Legendre) transformed reaction Gibbs energy is a function of the properties of each metabolite species in the respective pseudo-isomer groups, not just the predominant hydrogen ion dissociated form.

Another approach reported in the literature assumes that all reactants can be approximated by a single predominant metabolite species at a given *pH* [19,20]. There, in order to represent a known constant hydrogen ion concentration, $[H^+]$, a new standard hydrogen ion Gibbs energy of formation is defined, $\Delta_f G^o(H^+) \equiv RT\ln[H^+]$, and the Legendre transformation of all metabolite species at constant *pH* is omitted [19,20]. Thereafter, standard 'transformed' reaction Gibbs energy is calculated with the newly defined standard hydrogen ion Gibbs energy of formation and standard Gibbs energy of formation for the remaining reactants, each approximated by a single metabolite species. Fig. 3 illustrates that there is a significant difference between the standard (Legendre) transformed reaction Gibbs energy and the standard 'transformed' Gibbs energy of the same reaction, calculated when one assumes that each reactant can be approximated with one predominant metabolite species. As expected, in exceptional cases, where all reactants involved in a biochemical reaction exist as one metabolite species, the standard 'transformed' Gibbs energy of a reaction is identical to the standard transformed Gibbs energy of the same reaction.

3.2. Qualitative versus quantitative reaction directionality in E. coli

To quantitatively assign reaction directionality, where available, we used standard reactant Gibbs energy of formation back-calculated from experimentally measured equilibrium constants in preference to group contribution estimates. However, a minority of the reactants in *E. coli* (311/1668) have standard Gibbs energy of formation back-calculated from experimentally measured equilibrium constants [30]. Most of the remaining reactant standard Gibbs energy of formation (1127/1668) can be estimated using group contribution methodology [31]. Certain reactants have no estimated standard Gibbs energy of formation due to structural subgroups in metabolite species not covered by the current version of this method (230/1668 or 163/1039 unique iAF1260 reactants).

We compared quantitative assignment of reaction directionality with qualitatively assigned reaction reconstruction directionality in the genome-scale metabolic model of *E. coli*, iAF1260 [28]. The heuristics for the latter are given in the Supplementary material 8. Overall, quantitative assignment of reaction directionality results in a relaxation of qualitative directionality constraints, see Table 1. Directionality changes are in three categories: i) unchanged directionality, ii) relative relaxation of reaction directionality (qualitatively forward yet quantitatively reversible), and iii) tightened or reversed reaction directionality. Unless otherwise specified, we refer to quantitative

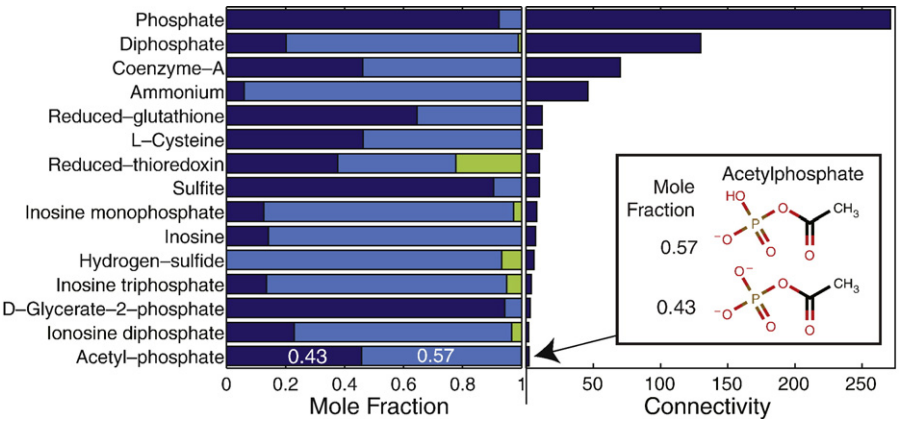

Fig. 2. Non-predominant metabolite species. At typical *in vivo* conditions, a minority of cytoplasmic reactants in *E. coli* (15/138 with available data), have significant mole fractions (>0.05) present as non-predominant metabolite species. The length within each bar (left), apportioned to each color equals the mole fraction of the reactant present as a particular metabolite species. A bar with three colors indicates a reactant simultaneously present as three different metabolite species differing only in their state of protonation. Accurate prediction of standard transformed Gibbs energy of formation is especially important for reactants, such as Co-enzyme-A, which is almost equally present in the two different metabolites species forms, and participates as a reactant in many reactions (connectivity bar to the right). Another example is acetyl-phosphate which is also almost equally distributed, as $C_2H_3O_5P^{2-}$ and $C_2H_4O_5P^{-}$ (inset). The mole fractions, corresponding to a *pH* 7.7 [43,44], ionic strength 0.25 M and temperature 310.15 K, were calculated as described in Section 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

assignment at a temperature of 310.15 K, ionic strength of 0.25 M, *pH* of 7.7 [43,44], and an accommodating physiological concentration range of 0.02–20 mM. Initially, we assumed thermodynamic reversibility for internal reactions (309/2077) where data is missing for at least one reactant.

3.3. Quantitative relaxation of qualitatively assigned reaction directionality

In iAF1260, most of the internal reactions (1524/2077) have been qualitatively assigned to be irreversible in the forward direction. Of these forward reactions, about a third are also quantitatively forward (638/1524). Another third of qualitatively forward reactions are assigned to be quantitatively reversible (603/1524), see Table 1. Taking into account the uncertainty in group contribution estimation of metabolite species standard Gibbs energy of formation, the fraction of qualitatively forward reactions changed to quantitatively reversible almost doubles (1011/1524). Relaxation of directionality constraints for so many reactions would significantly increase the flexibility of the resulting iAF1260 constraint-based model. Such relaxation significantly effects model predictions, therefore, we conducted a detailed analysis of these qualitatively forward, yet quantitatively reverse reactions.

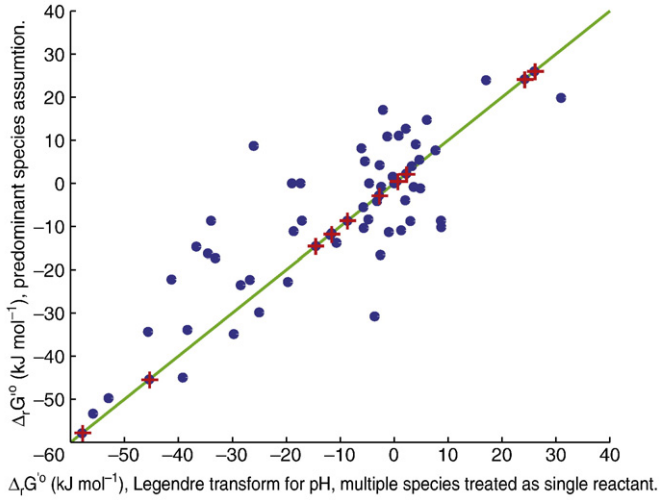

Fig. 3. The importance of a true Legendre transform. When cytoplasmic *pH* is held constant by buffering, a Legendre transform of metabolite species standard Gibbs energy of formation defines a standard transformed Gibbs energy of formation for each metabolite species. Reactant standard transformed Gibbs energy of formation, and therefore reaction standard transformed Gibbs energy, depends non-additively on the standard transformed metabolite species Gibbs energy of formation. Due to this non-additivity, when reactants are present as multiple species, differing in their state of protonation, it is erroneous to replace transformation of metabolite species standard Gibbs energy of formation, with adjustment to the hydrogen ion standard Gibbs energy of formation, when calculating reaction standard transformed Gibbs energy. The latter gives rise to an erroneous estimate of standard 'transformed' reaction Gibbs energy, as illustrated above for the reactions in the central metabolic *E. coli* model [61] at *pH* 7.7, zero ionic strength, atmospheric pressure and temperature 298.15 K. The exception is when each reactant involved in a reaction is present as only one metabolite species. In this case both approaches agree (stars on the diagonal).

Table 1
Comparison of qualitative versus quantitative assignment of reaction directionality for the genome-scale constraint-based model of *E. coli*, iAF1260 [28].

Qualitative directionality assignment	Quantitative directionality assignment	$\Delta_f G^0_{K_{eq}} \cup \Delta_f G^0_{est}$			$\Delta_f G^0_{est}$ only		
Forward	Reversible	603[a] (103, 376, 124)	1011 (103, 774, 134)	612 (89, 399, 89)	609 (0, 406, 203)	1133 (0, 853, 280)	642 (0, 440, 202)
Reversible	Forward	10	6	8	7	1	7
Reversible	Reverse	3	1	2	11	0	11
Forward	Reverse	11	5[b]	12	57	35	56
	Temp. (K)	310.15	310.15	298.15	310.15	310.15	298.15
	Ionic strength (M)	0.25	0.25	0	0.25	0.25	0
	pH	7.7	7.7	7	7.7	7.7	7
	$SE_{\Delta_f G^0_{est}}$	0	1	0	0	1	0

Quantitative assignments were calculated at specific temperature, ionic strength and *pH*. Each cell represents the number of reactions with a particular type of change in directionality assignment. Reaction directionality was calculated with metabolite group contribution estimates [31] alone (right) or with both group contribution estimates and reactant standard transformed Gibbs energy of formation back-calculated from experimentally measured equilibrium constants, when available (left) [30]. By incorporating the uncertainty associated with group contribution estimates ($SE_{\Delta_f G^0_{est}} = 1$), reaction directionality is considerably relaxed.

[a] The numbers of qualitatively forward, quantitatively reverse reactions are further decomposed depending on the provenance of the metabolite data ($\#\Delta_f G^0_{K_{eq}}$, $\#\Delta^0_{f K_{eq}} \cup \#\Delta_f G^0_{est}$, $\#\Delta_f G^0_{est}$).

[b] The lowest number of reactions with reversal of directionality, i.e. forward irreversible changed to reverse irreversible, is achieved when using the following conditions: (i) use of reactant standard transformed Gibbs energy of formation back-calculated from experimentally measured equilibrium constants, (ii) incorporation of uncertainty associated with group contribution estimates.

3.3.1. Quantitative relaxation of reaction directions with no uncertainty in $\Delta_\gamma G'^0$

A tenth of qualitatively forward yet quantitatively reversible reactions (103/1011) exclusively involved reactant standard transformed Gibbs energy of formation back-calculated from equilibrium constants. The majority of these reactions do indeed have a negative standard transformed reaction Gibbs energy, but within our accommodating physiological range of reactant concentration, they are quantitatively reversible (see Supplementary Fig. 10). Strictly, these reactions are quantitatively reversible, i.e., $\Delta_\gamma G'_{\min} < 0$ and $\Delta_\gamma G'_{\max} > 0$, but most are probably forward at *in vivo* concentrations.

Many of the quantitatively reversible reactions transport metabolites between compartments. Although the net electrical charge within a compartment is assumed to be neutral, an electrical potential difference can exist between compartments. In *E. coli*, the cytoplasmic side of the cytoplasmic membrane is more negative than the periplasmic side [53]. A thorough treatment of electrochemical potential difference is beyond the scope of this current work. However, even when cytoplasm and periplasm are assumed to have an identical *pH*, an electrical potential difference still exists across the cytoplasmic membrane. We assumed an electrical potential difference of 90 mV, which corresponds to a Gibbs energy change of -8.7 kJ mol^{-1} per negative charge transported from cytoplasm to periplasm [53]. Therefore, the passive transport of negatively charged reactants, e.g., formate, from cytoplasm to periplasm is thermodynamically favorable at equal concentrations of formate in both compartments. However, if the concentration of formate is high in the periplasm, and low in the cytoplasm, this can reverse the direction of formate diffusion. Therefore, many transport reactions are quantitatively assigned to be reversible within a physiological range of reactant concentration. This is the case regardless of the provenience of the reactant standard Gibbs energy of formation, see Supplementary Fig. 11.

3.3.2. Quantitative relaxation of reaction directions with uncertainty in metabolite $\Delta_f G'^0_{est}$

A small minority (134/1011) of the qualitatively forward yet quantitatively reversible reactions are assigned based solely on group contribution estimates of reactant standard transformed Gibbs energy of formation. The uncertainty in these estimates combines with variability in metabolite concentration to increase the feasible range of reaction Gibbs energy. The remaining majority (774/1011) of the qualitatively forward yet quantitatively reversible assignments are based on standard transformed Gibbs energy of formation for reactants estimated by the group contribution method combined with standard transformed Gibbs energy of formation for reactants back-calculated from equilibrium constants. The latter reactants include most of the widely used co-enzymes [54] and prosthetic groups. In order to understand the reasons for relaxation of so many reaction directions ((134 + 774)/1011), we categorized the reactions. Fig. 4 illustrates all reactions that are qualitatively forward yet quantitatively reversible (908/1011), using at least one group contribution estimate of reactant standard transformed Gibbs energy of formation. Apart from transport reactions, the reactions in Fig. 4 are ordered by descending probability of the reaction being forward, $P(\Delta_r G'^m < 0)$, at physiological standard concentrations of reactants.

3.3.3. Thermodynamically favorable, qualitatively forward transport reactions

For most transport reactions, there is no uncertainty in physiological standard transformed reaction Gibbs energy since the same reactant is present on both sides of the reaction. In Fig. 4, transport reactions are sorted by ascending physiological standard transformed reaction Gibbs energy since $P(\Delta_r G'^m < 0)$ is undefined if the denominator in Eq. (3) is zero. A non-zero standard transformed reaction Gibbs energy for a transport reaction is due to net transport of a charged metabolite between the cytoplasm and periplasm. Of the qualitatively forward yet quantitatively reversible reactions, 155/1011 result in net transport of a negative (or neutral) charge from the cytoplasm to the periplasm and therefore have a negative (or zero) physiological standard transformed reaction Gibbs energy, $\Delta_r G'^m \leq 0$. These transport reactions are qualitatively assigned to be forward, but within the physiological range of reactant concentration, they are quantitatively reversible.

3.3.4. Thermodynamically unfavorable, qualitatively forward transport reactions

The forward direction of a transport reaction, as for any reaction, is defined by convention. Assuming the concentration range of the

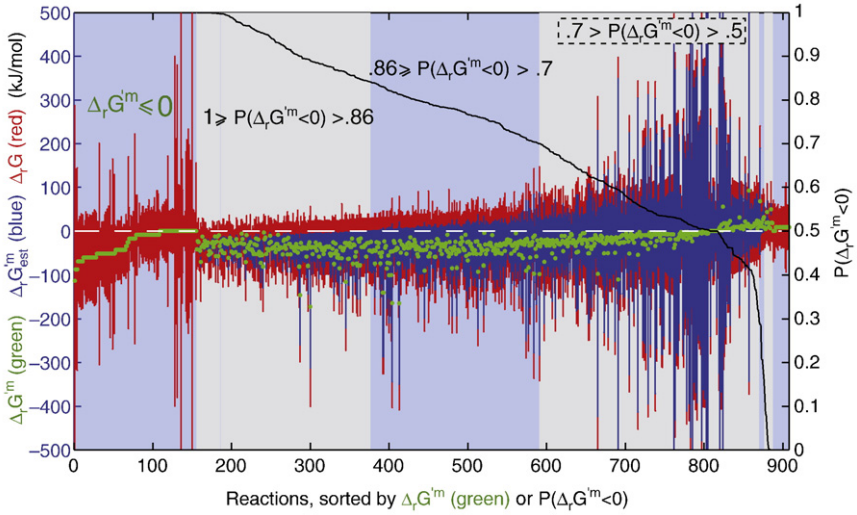

Fig. 4. Qualitatively forward, quantitatively reverse reactions. The reactions that are qualitatively assigned to be forward in iAF1260, yet quantitatively reversible, using at least one group contribution estimate of reactant standard transformed Gibbs energy of formation. The feasible range of $\Delta_r G'$ and $\Delta_r G'^m_{est}$ are given as red and blue bars respectively. Far left are transport reactions with negative or zero physiological standard transformed reaction Gibbs energy, $\Delta_r G'^m \leq 0$, but reversible depending on concentration of reactants. Reactions with uncertainty due to estimation of standard transformed reaction Gibbs energy are rank ordered by decreasing probability that physiological standard transformed reaction Gibbs energy is negative. In mathematical notation this probability is represented by the symbol $P(\Delta_r G'^m < 0)$, as defined in Eq. (3), and used *in situ* above to denote the intervals as follows: Reactions with $P(\Delta_r G'^m < 0) > 0.7$ were assumed to be irreversible in the forward direction, and reactions with $P(\Delta_r G'^m < 0) < 0.3$ or $\Delta_r G'^m > 0$ were assumed to be irreversible in the reverse direction. See Supplementary Fig. 12 for a detailed illustration of the latter reactions. Reactions with $0.7 \geq P(\Delta_r G'^m < 0) > 0.3$ were allowed to be quantitatively reversible in lieu of the large uncertainty in estimation of standard transformed reaction Gibbs energy.

reactant is the same in both compartments, reactions which transport a positively charged reactant from the cytoplasm to the periplasm have a positive physiological standard transformed reaction Gibbs energy, $\Delta_r G'^m > 0$. Likewise, reactions which transport a negatively charged reactant from periplasm to the cytoplasm have a positive standard transformed reaction Gibbs energy. At equal cytoplasmic and periplasmic reactant concentrations these reactions would operate in reverse, despite being qualitatively assigned to be forward reactions. These 21 transport reactions are detailed in Supplementary Fig. 12. Any facilitated diffusion reaction in this group would have to have a higher periplasmic than cytoplasmic concentration in order to proceed in the forward direction as qualitatively assigned, e.g., undecaprenyl phosphate transport cytoplasm to periplasm.

At an equal *pH* of cytoplasm and periplasm, any proton symport reaction, transporting a reactant with a charge of −2 into the cell, would require more than one proton to translocate across the periplasmic membrane in order to make the reaction thermodynamically favorable. It may be that the proton stoichiometry of these reactions needs to be revised in order to take this into account. For example the reaction 'GlcNAc anhMurNAc tetrapeptide transport in via proton symport', which symports GlcNAc anhMurNAc tetrapeptide (N-Acetyl-D-glucosamine (anhydrous) N-Acetylmuramyl-tetrapeptide), a biopolymer in the bacterial cell wall, with a charge of −2, from the periplasm to the cytoplasm, see Supplementary Fig. 12. This result indicates how thermodynamic consideration of reaction directionality can highlight reactions with thermodynamically infeasible stoichiometry.

3.3.5. Qualitatively forward reactions, quantitatively reversible within a physiological concentration range

The probability that physiological standard transformed reaction Gibbs energy less than zero, $P(\Delta_r G'^m < 0)$, Eq. (3), attempts to quantify the probability that a reaction is indeed irreversible in the forward direction at physiological standard concentrations of reactants. Reactions that are most probable to be irreversible in the forward direction are those where the physiological standard transformed reaction Gibbs energy is negative, even one standard deviation from the mean, i.e. $1 \geq P(\Delta_r G'^m < 0) \geq 0.86$. Of the qualitatively forward, yet quantitatively reversible reactions, a quarter (221/1011) are probably forward by this reasoning. By the same reasoning for the opposite direction, $0.14 \geq P(\Delta_r G'^m < 0) \geq 0$, there are 11 reactions that are probably reverse, but are qualitatively assigned to operate in the forward direction, see Supplementary Fig. 12. These quantitative reaction directionality assignments are interesting because they possibly change the direction of qualitative assignment.

The intermediate range of probability of forward physiological standard transformed reaction Gibbs energy, $0.86 \geq P(\Delta_r G'^m < 0) > 0.14$, still accounts for half (500/1011) of the qualitatively forward, yet quantitatively reverse reactions. Fig. 4 illustrates that a large uncertainty in physiological standard transformed reaction Gibbs energy precludes definitive quantitative forward assignment for many of these reactions. We arbitrarily set the 214 reactions with $0.86 \geq P(\Delta_r G'^m < 0) > 0.7$ to be quantitatively forward, the 279 reactions with $0.7 \geq P(\Delta_r G'^m < 0) > 0.3$ to be quantitatively reversible and the remaining 7 reactions with $0.3 \geq P(\Delta_r G'^m < 0) > 0.14$ to be quantitatively reverse.

Whatever upper cutoff is chosen for qualitatively forward reactions, then in order to be logically consistent, one minus the same cutoff must be used to quantitatively assign other reactions to be reverse. Our conservative upper and lower cutoff was chosen to compensate for the uncertainty in group contribution estimates of reactant standard Gibbs energy. Even with these conservative cutoffs, the resulting model cannot produce biomass in glucose minimal medium, because certain qualitatively forward reactions are essential in this direction and therefore cannot be reversed. Computational analysis can reliably predict, which reaction directions are essential *in vivo* for a given boundary condition [8,55]. We identified the 7 reactions that were essential for production of biomass in glucose minimal medium, yet seemed thermodynamically unfavorable according to $0.3 \geq P(\Delta_r G'^m < 0)$ (see Supplementary Fig. 12). This highlights the importance of incorporating metabolite concentration data to refine thermodynamic assignment of reaction directionality [21] since it is ultimately the transformed Gibbs energy of a biochemical reaction, $\Delta_r G'$, that determines directionality.

3.3.6. L-glutamate and reaction directionality

The amino acid L-glutamate is the major nitrogen donor in the cell, distributing ~88% of the total nitrogen that ends up in biomass, largely via transamination reactions. It has been measured at a relatively high ~100 mM concentration in a range of growth conditions [56]. L-glutamate is a reactant in three of the qualitatively forward reactions, that could possibly be assigned reverse based on consideration of physiological standard transformed reaction Gibbs energy alone (Supplementary Fig. 12). L-glutamate is a substrate for glutamate 5 kinase (E.C. 2.7.2.11) and leucine transaminase (E.C. 2.6.1.6), yet a product of the reaction catalyzed by 1-pyrroline 5-carboxylate dehydrogenase (EC 1.5.1.12). The reaction catalyzed by leucine transaminase is essential for production of biomass from glucose minimal medium *in silico* and *in vivo* [57]. A high concentration of L-glutamate favors the forward direction of this reaction. In the same conditions, the other two reactions involving L-glutamate are not essential *in silico*.

Leucine transaminase catalyzes the transfer of the amine group in L-glutamate to 4-methyl-2-oxopentanoate to form L-leucine. This reaction can also be catalyzed by branched-chain-amino acid transaminase (EC 2.6.1.42) which catalyzes the first step in the catabolism of branched-chain amino acids, being leucine, isoleucine and valine. At 298.15 K, *pH* 7.21, and ionic strength 0.31 mol kg^{-1}, the apparent equilibrium constant of

$$\mathrm{L-leucine} + \mathrm{2-oxolutarate} \rightleftharpoons \mathrm{4-methyl-2-oxopentanoate} + \mathrm{L-glutamate}$$

was reported to be 2.42 ± 0.25 with respect to the forward direction [58]. Without error this corresponds to a standard transformed Gibbs energy of -2.19 kJ mol^{-1}. This indicates that the reverse direction, biosynthesis of L-leucine, is thermodynamically unfavorable at equal concentrations of substrates and products. This indicates that the high concentration of glutamate observed in the cell may well be thermodynamically necessary to drive the biosynthesis of L-leucine. On the contrary, depending on the needs of the cell, all else being equal, the reaction will run in the opposite direction at low glutamate concentration. This illustrates how thermodynamic data can be used to interpret the meaning of metabolomic data [21].

3.4. Quantitative tightening of qualitatively assigned reaction directionality

Using quantitative assignment, a small minority of reaction directions are tightened or reversed, depending on the quantitative method used. Using group contribution estimates alone, there are at least 35 qualitatively reversible reactions that are deemed to be quantitatively irreversible, see Table 1. This number rises to 56 reactions when the uncertainty associated with the group contribution methodology is not taken into account. In contrast, the use of standard Gibbs energy of formation back-calculated from experimentally measured equilibrium constants, in place of group contribution estimates, reduces the number of reaction directions that are tightened or reversed. In particular, of 553 qualitatively reversible internal reactions in iAF1260, only 6 are deemed to be quantitatively irreversible using standard transformed Gibbs energy of formation back-calculated from experimentally measured equilibrium constants, in preference to group contribution estimates. Of course, group contribution estimates are still required for the majority of reactants. Of 1524 qualitatively forward reactions, only 5 are deemed to be quantitatively irreversible in the reverse direction.

The tightened or reversed directionality assignments, directly conflict with qualitative assignments and therefore we investigated the literature on biochemical and/or thermodynamic characterization experiments for each of these reactions. The detailed results of this study are given in Supplementary material 9. Quantitative tightening of reaction directionality was supported for a minority of reactions, e.g. D-Erythrose-4-phosphate dehydrogenase reaction. Other reactions had experimental evidence indicating that quantitative assignment was too tight, which may perhaps be improved by more accurate modeling of the effect of Magnesium on the standard transformed Gibbs energy of ATP, e.g. galactokinase reaction. As discussed by Alberty [29], the concept of a pseudo-isomer group may be extended to different metal ion dissociated forms, e.g., when $pMg = -\log_{10}[Mg^{2+}]$ is a known constant. The group contribution method seemed to have difficulty in estimating the standard Gibbs energy of metabolites with an imidazole ring structure, giving rise to erroneous, over tight assignments for certain reactions, e.g. IMP cyclohydrolase. Similarly, certain qualitatively forward reactions were also incorrectly predicted to be quantitatively reverse for reactions involving other metabolites with complex structure, e.g. glycogen. Ongoing refinements to the group contribution method can be expected to iron out these issues with structurally complex metabolite species [31].

3.5. Quantitative assignment of reaction directionality and biomass production in E. coli

Using the stoichiometry of the latest genome-scale metabolic model, iAF1260 [28], we quantitatively set reaction directionality according to local thermodynamic constraints and compared growth rate prediction against those experimentally measured. A first pass thermodynamic assignment of reaction directionality, based on $\Delta_r G'$ alone, resulted in a minority of reactions being assigned thermodynamically forward (242/2077), and a few reactions being thermodynamically reverse (6/2077). A few of these reversed reactions were incorrectly quantitatively assigned, essential for growth, and therefore qualitatively set to forward, ignoring the thermodynamic assignment. Using flux balance analysis [33,59] to maximize reaction representing the production of biomass, at first, the biomass production rate of the resulting model far exceeded the *in vivo* growth rate on glucose minimal medium. This indicates that significant reaction directionality constraints were missing since the published iAF1260 model does reproduce experimentally reported growth rates in various media.

At this point, we attempted to identify the minimum number of biochemically reasonable qualitative assignments necessary in order to match the growth rate observed *in vivo*. The qualitative assignments were based on the experimental biochemical literature used to assign reaction directionality in the *iAF1260* reconstruction [28]. In addition, care was taken to qualitatively assess prediction of internal fluxes, such as flux through the electron transport chain, as expected during aerobic growth. First, we assigned the qualitative reconstruction directions to reactions without any thermodynamic data (310/2011). Similarly for reactions involving a quinone, since quinones are structurally complex metabolite species with significant uncertainty in group contribution estimates. Then, using the probabilistic criteria discussed above, we made these further quantitative assignments: $P(\Delta_r G'^m<0)>0.7 \iff$ forward, $0.7 \geq P(\Delta_r G'^m<0)>0.3 \iff$ reversible, and $0.3 \geq P(\Delta_r G'^m<0) \iff$ reverse, except for *in silico* essential reactions.

In addition, it was necessary to eliminate artifactual excess ATP synthesis due to (i) reversal of ATP consuming ATP-binding cassette transport reactions, and (ii) net shuttling of protons from the cytoplasm into the periplasmic compartment via transport reactions, leading to excess ATP synthase flux. We assigned reconstruction directions to all transport reactions involving proton symport or antiport, and all ATP-binding cassette transport reactions. The majority of transport reactions qualitatively assigned a forward direction in this way do indeed have a negative physiological standard transformed reaction Gibbs energy, see Supplementary Fig. 11. For transport reactions, compartmentally resolved reactant concentration, if available, would significantly influence reaction directionality, since, for many such reactions, the standard Gibbs energy is close to zero.

We qualitatively assigned reaction directionality to cytoplasmic reactions involving the cofactors ATP, GTP, CTP and UTP. In addition, we incorporated experimentally reported concentration ranges for key cofactors (ATP, ADP, AMP, NAD, NADH, NADP, and NADPH) measured for *E. coli* growing aerobically on glucose minimal medium [51]. At glucose and oxygen uptake rates of 11 and 18.2 mmol g^{-1} h^{-1} respectively, the experimentally observed growth rate is 0.82 h^{-1} [60]. The aforementioned qualitative reaction directionality assignments, which override the quantitative assignments, reduce the growth rate to 1.1 h^{-1}, which compares favorably with that observed experimentally. At this margin of error, setting the effective stoichiometry of oxidative phosphorylation by constraints on the relative fluxes through various NADH dehydrogenases and terminal oxidases has a significant effect on growth rate [28]. It is clear that the latter modeling issue awaits a thorough investigation. In summary, quantitative assignment of reaction directionality must be accompanied by qualitative assignment, for certain classes of reactions, in order to match experimental data with constraint-based models of metabolism.

4. Discussion

In principle, the application of the second law of thermodynamics to metabolic reactions results in a constraint on the direction of all network reactions. However, this assumes that all reactant concentrations and standard Gibbs energies are known. Using experimentally measured apparent equilibrium constants, Alberty has published tables of standard transformed reactant Gibbs energies for 200 reactants [29,30]. In addition, group contribution estimates of standard reactant Gibbs energy have recently become available for two thirds of the 15,000 reactants in the KEGG compound database [31]. We have integrated these estimates, with Alberty's tables in order to quantitatively constrain the directions in the *E. coli* genome-scale metabolic model.

Standard Gibbs energy applies to chemical standard conditions, which are markedly different from *in vivo* conditions. Therefore, following the methods developed by Alberty [29], we have theoretically summarized and practically implemented the main steps in transformation of chemical standard thermodynamic potentials into biochemical standard thermodynamic potentials. The large buffering capacity of proteins, relative to the size of the soluble reactant pool means that *pH* is effectively held constant. Therefore, we use the transformed reaction Gibbs energy to provide the criterion for spontaneous change, and therefore the direction of net flux, at *in vivo* temperature, pressure, *pH* and ionic strength.

Even in the absence of quantitative metabolomic data, thermodynamic assignment of reaction directionality can be made by assuming an accommodating range of reactant concentration (0.02–20 mM). In *E. coli*, our thermodynamic assignment of reaction directionality rarely conflicts with qualitatively assigned reconstruction direction based on experimental literature. This indicates that the same algorithms may be reliably used to constrain reaction directionality for other organisms without the same breadth of experimental evidence as exists for *E. coli*. However, a large proportion of reaction directions, which were qualitatively assigned to be irreversible in *E. coli*, become quantitatively reversible, by transformed reaction Gibbs energy alone. Partially, this is due to the accommodating range of reactant concentration, especially for transport reactions. However, the main reason is that the uncertainty associated with computational estimation of standard reactant Gibbs energy can give rise to a significant uncertainty in standard transformed reaction Gibbs energy. This is especially so when structurally complex reactants are involved.

We partially overcame the uncertainty issue by formulating a probabilistic metric expressing our confidence that a given reaction is irreversible, given the uncertainty in standard reactant Gibbs energy. Using this Bayesian probabilistic metric, it is possible to quantitatively assign a large number of reactions to be irreversible, at a given confidence cutoff. Even still, some qualitative assignment of reaction directionality is necessary in order to sufficiently constrain the genome-scale *E. coli* model to agree with experimentally observed growth rate in glucose minimal medium. In particular, this is so for quinone coupled oxidative phosphorylation reactions transport reactions, and reactions involving nucleotide cofactors.

Quantitative metabolomics data, combined with standard transformed reactant Gibbs energy calculations, would increase the number of predicted effectively irreversible reactions. However, metabolite concentration data only becomes important for determining reaction directionality when the uncertainty in group contribution estimation of reactant Gibbs energy is low (6 kJ mol^{-1}). Cofactor pair concentration ratios are particularly useful since these reactants are involved in such a large proportion of reactions and since experimentally derived standard reactant Gibbs energy are available for many cofactors from the tables compiled by Alberty [29]. Beyond cofactors, the concentration of other key metabolites is also important for determining the direction of essential anabolic reactions. For instance, we provide evidence that the high concentration of cytoplasmic L-glutamate, observed in *E. coli* [56], may well be thermodynamically necessary to drive the biosynthesis of L-leucine.

We observed that reaction directionality is relatively insensitive to the specific adjustments for temperature, *pH* and ionic strength necessary to represent *in vivo* conditions. This is more due to the uncertainty in standard reactant Gibbs energy than the absence of metabolomic data. Integration of standard reactant Gibbs energy back-calculated from equilibrium constants significantly reduces the number of false positive irreversible reactions, as compared to the use of group contribution estimates alone. Nevertheless, group contribution estimates are essential to match the broad scope of current genome-scale models, therefore it is evident that these approaches to the thermodynamics of biochemical reactions are complementary.

Acknowledgements

This work was supported by NIH grant Grant 5R01GM057089-11, Science Foundation Ireland (Systems Biology Ireland), and a National University of Ireland, Galway, Postgraduate Fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bpc.2009.08.007.

References

[1] B.Ø. Palsson, Systems Biology: Properties of Reconstructed Networks, Cambridge University Press, Cambridge, 2006.
[2] I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc, (in press), 2009.
[3] C. Pál, B. Papp, M.J. Lercher, P. Csermely, S.G. Oliver, L.D. Hurst, Chance and necessity in the evolution of minimal metabolic networks, Nature 440 (7084) (2006) 667–670.
[4] R. Mahadevan, C.H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models, Metab. Eng. 5 (4) (2003) 264–276.
[5] A.P. Burgard, E.V. Nikolaev, C.H. Schilling, C.D. Maranas, Flux coupling analysis of genome-scale metabolic network reconstructions, Genome Res. 14 (2) (2004) 301–312.
[6] C.L. Barrett, C.D. Herring, J.L. Reed, B.Ø. Palsson, The global transcriptional regulatory network for metabolism in *Escherichia coli* exhibits few dominant functional states, Proc. Natl. Acad. Sci. U. S. A. 102 (52) (2005) 19103–19108.
[7] A. Samal, S. Jain, The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response, BMC Syst. Biol. 2 (2008) 21.
[8] J.S. Edwards, R.U. Ibarra, B.Ø. Palsson, *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data, Nat. Biotechnol. 19 (2) (2001) 125–130.
[9] R.U. Ibarra, J.S. Edwards, B.Ø. Palsson, *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth, Nature 420 (6912) (2002) 186–189.
[10] M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard, B.Ø. Palsson, Integrating high-throughput and computational data elucidates bacterial networks, Nature 429 (6987) (2004) 92–96.
[11] M.J. Herrgård, S.S. Fong, B.Ø. Palsson, Identification of genome-scale metabolic network models using experimentally measured flux profiles, PLoS Comp. Bio. 2 (7) (2006) e72.
[12] J.L. Reed, T.R. Patel, K.H. Chen, A.R. Joyce, M.K. Applebee, C.D. Herring, O.T. Bui, E.M. Knight, S.S. Fong, B.Ø. Palsson, Systems approach to refining genome annotation, Proc. Natl. Acad. Sci. U. S. A. 103 (46) (2006) 17480–17484.
[13] A.P. Burgard, P. Pharkya, C.D. Maranas, Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization, Biotechnol. Bioeng. 84 (6) (2003) 647–657.
[14] K.R. Patil, I. Rocha, J. Förster, J. Nielsen, Evolutionary programming as a platform for *in silico* metabolic engineering, BioMed. Central Bioinformatics 6 (2005) 308.
[15] J.H. Park, K.H. Lee, T.Y. Kim, S.Y. Lee, Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation, Proc. Natl. Acad. Sci. U. S. A. 104 (19) (2007) 7797–7802.
[16] A.M. Feist, B.Ø. Palsson, The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*, Nat. Biotechnol. 26 (6) (2008) 659–667.
[17] J.M. Savinell, B.O. Palsson, Network analysis of intermediary metabolism using linear optimization. i. development of mathematical formalism, J. Theor. Biol. 154 (4) (1992) 421–454.
[18] R. Schuetz, L. Kuepfer, U. Sauer, Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*, Mol. Syst. Biol. 3 (2007) e119.
[19] C.S. Henry, M.D. Jankowski, L.J. Broadbelt, V. Hatzimanikatis, Genome-scale thermodynamic analysis of *Escherichia coli* metabolism, Biophys. J. 90 (4) (2006) 1453–1461.
[20] C.S. Henry, L.J. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis, Biophys. J. 92 (5) (2007) 1792–1805.
[21] A. Kümmel, S. Panke, M. Heinemann, Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data, Mol. Syst. Biol. 2 (2006) e34.
[22] A. Kümmel, S. Panke, M. Heinemann, Systematic assignment of thermodynamic constraints in metabolic network models, BioMed. Central Bioinformatics 7 (1) (2006) 512.
[23] R.A. Alberty, A short history of the thermodynamics of enzyme-catalyzed reactions, J. Biol. Chem. 279 (27) (2004) 27831–27836.
[24] K. Burton, H.A. Krebs, H.L. Kornberg, Energy Transformations in Living Matter, Springer-Verlag, Berlin, 1957.
[25] F. Yang, H. Qian, D.A. Beard, Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry, Metab. Eng. 7 (4) (2005) 251–259.
[26] Yang Feng, Daniel A. Beard, Thermodynamically based profiling of drug metabolism and drug-drug metabolic interactions: a case study of acetaminophen and ethanol toxic interaction, Biophys. Chem. 120 (2) (2006) 121–134.
[27] W.J. Heuett, H. Qian, Combining flux and energy balance analysis to model large-scale biochemical networks, J. Bioinform. Comput. Biol. 4 (6) (2006) 1227–1243.
[28] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, B.Ø. Palsson, A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information, Mol. Syst. Biol. 3 (1) (2007) e121.
[29] R.A. Alberty, Thermodynamics of Biochemical Reactions, Wiley-Interscience, Hoboken, NJ, 2003.
[30] R.A. Alberty, Biochemical Thermodynamics: Applications of Mathematica, Wiley-Interscience, Hoboken, NJ, 2006.
[31] M.D. Jankowski, C.S. Henry, L.J. Broadbelt, V. Hatzimanikatis, Group contribution method for thermodynamic analysis of complex metabolic networks, Biophys. J. 95 (3) (2008) 1487–1499.
[32] M.L. Mavrovouniotis, Estimation of standard Gibbs energy changes of biotransformations, J. Biol. Chem. 266 (22) (1991) 14440–14445.
[33] A. Varma, B.Ø. Palsson, Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors, J. Theor. Biol. 165 (4) (1993) 477–502.
[34] A. Varma, B.Ø. Palsson, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110, Appl. Environ. Microbiol. 60 (10) (1994) 3724–3731.
[35] J. Ross, Thermodynamics and fluctuations far from equilibrium, volume 80 of Springer Series in Chemical Physics, Springer, New York, 2008.
[36] S. Schnell, T.E. Turner, Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws, Prog. Biophys. Mol. Biol. 85 (2–3) (2004) 235–260.
[37] A.P. Minton, How can biochemical reactions within cells differ from those in test tubes? J. Cell Sci. 119 (14) (2006) 2863.
[38] A.S. Verkman, Solute and macromolecule diffusion in cellular aqueous compartments, Trends Biochem. Sci. 27 (1) (2002) 27–33.
[39] H.P. Kao, J.R. Abney, A.S. Verkman, Determinants of the translational mobility of a small solute in cell cytoplasm, J. Cell Biol. 120 (1) (1993) 175–184.
[40] Biemans-Oldehinkel Esther, Nik A.B.N. Mahmood, Poolman Bert, A sensor for intracellular ionic strength, Proc. Natl. Acad. Sci. U. S. A. 103 (28) (2006) 10624–10629.

[41] K.A. Dill, S. Bromberg, Molecular driving forces: Statistical thermodynamics in Chemistry and Biology, Garland Science, London, 2003.
[42] F. C. Neidhardt (Ed. in Chief), R. Curtis, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (eds). *Escherichia coli* and Salmonella: Cellular and Molecular Biology. American Society for Microbiology Press, Washington, DC, 1996.
[43] E. Padan, S. Schuldiner, Intracellular pH and membrane potential as regulators in the prokaryotic cell, J. Membr. Biol. 95 (3) (1987) 189–198.
[44] J.C. Wilks, J.L. Slonczewski, pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry, J. Bacteriol. 189 (15) (2007) 5601–5607.
[45] J.L. Slonczewski, R.M. Macnab, J.R. Alger, A.M. Castle, Effects of pH and repellent tactic stimuli on protein methylation levels in Escherichia coli, J. Bacteriol. 152 (1) (1982) 384–399.
[46] W.R. Smith, R.W. Missen, Chemical Reaction Equilibrium Analysis: Theory and Algorithms, Wiley, New York, 1982.
[47] A. Fersht, Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding, W.H. Freeman, New York, 1999.
[48] O. Fiehn, Metabolomics: the link between genotypes and phenotypes, Plant. Mol. Biol. 48 (1–2) (2002) 155–171.
[49] M.J. Brauer, J. Yuan, B.D. Bennett, W. Lu, E. Kimball, D. Botstein, J.D. Rabinowitz, Conservation of the metabolomic response to starvation across two divergent microbes, Proc. Natl. Acad. Sci. U. S. A. 103 (51) (2006) 19302–19307.
[50] N. Ishii, K. Nakahigashi, T. Baba, M. Robert, T. Soga, A. Kanai, T. Hirasawa, M. Naba, K. Hirai, A. Hoque, P.Y. Ho, Y. Kakazu, K. Sugawara, S. Igarashi, S. Harada, T. Masuda, N. Sugiyama, T. Togashi, M. Hasegawa, Y. Takai, K. Yugi, K. Arakawa, N. Iwata, Y. Toya, Y. Nakayama, T. Nishioka, K. Shimizu, H. Mori, M. Tomita, Multiple high-throughput analyses monitor the response of *E. coli* to perturbations, Science 316 (5824) (2007) 593–597.
[51] B.D. Bennett, E.H. Kimball, M. Gao, R. Osterhout, S.J. Van Dien, J.D. Rabinowitz, Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*, Nat. Chem. Biol. 5 (8) (2009) 593–599.
[52] T. Gowers, J. Barrow-Green, I. Leader, The Princeton Companion to Mathematics, Princeton University Pres, Princeton, 2008.
[53] H. Richard, J.W. Foster, *Escherichia coli* glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential, J. Bacteriol. 186 (18) (2004) 6032–6041.
[54] R.A. Alberty, Standard transformed Gibbs energies of coenzyme A derivatives as functions of pH and ionic strength, Biophys. Chem. 104 (1) (2003) 327–334.
[55] A.R. Joyce, J.L. Reed, A. White, R. Edwards, A. Osterman, T. Baba, H. Mori, S.A. Lesely, B.Ø. Palsson, S. Agarwalla, Experimental and computational assessment of conditionally essential genes in *Escherichia coli*? J. Bacteriol. 188 (23) (2006) 8259–8271.
[56] J. Yuan, W.U. Fowler, E. Kimball, W. Lu, J.D. Rabinowitz, Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*, Nat. Chem. Biol. 2 (10) (2006) 529–530.
[57] H.E. Umbarger, Amino acid biosynthesis and its regulation, Annu. Rev. Biochem. 47 (1978) 532–606.
[58] Y.B. Tewari, R.N. Goldberg, J.D. Rozzell, Thermodynamics of reactions catalysed by branched-chain-amino-acid transaminase, J. Chem. Thermodyn. 32 (10) (2000) 1381–1398.
[59] A. Varma, B.Ø. Palsson, Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns, J. Theor. Biol. 165 (4) (1993) 503–522.
[60] E. Fischer, N. Zamboni, U. Sauer, High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13c constraints, Anal. Biochem. 325 (2) (2004) 308–316.
[61] J.D. Orth, R.M.T. Fleming, and Bernhard Ø. Palsson. *Escherichia coli* and Salmonella: Cellular and Molecular Biology, chapter Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide (in press No). ASM Press, 2009.

A Supplementary Material

Quantitative assignment of reaction directionality in constraint-based models of metabolism: Application to *Escherichia coli*

R.M.T. Fleming[1], I. Thiele, H.P. Nasheuer

A.1 Metabolite species standard Gibbs energy of formation

A.1.1 Measurement of apparent equilibrium constants

The standard Gibbs energy of formation of a metabolite species can, in principal, be calculated by measuring the change required to inorganically synthesize it from its elements in standard state. This is difficult or impossible for large complicated organic metabolites. Alternatively, if a reactant occurs in a reaction where only one reactant standard Gibbs energy of formation is unknown, it is possible to back-calculate this unknown indirectly by measuring the apparent equilibrium constant for the reaction. Consider the biochemical reaction catalyzed by glucose-6-phosphate isomerase, *PGI*.

$$F6P \rightleftharpoons G6P \tag{1}$$

This is an isomerisation between the reactants fructose-6-phosphate, F6P, and glucose-6-phosphate, G6P. At physiological pH, fructose-6-phosphate is distributed between two forms, $F6P^{-2}$ and $F6P^{-1}$, that differ by one hydrogen ion

$$F6P^{-1} \rightleftharpoons F6P^{-2} + H^{+}$$

Likewise, glucose-6-phosphate is distributed between two charged forms, $G6P^{-2}$ and $G6P^{-1}$. The apparent equilibrium constant for biochemical reaction 1 is given by

$$K' = \frac{[G6P]_{eq}}{[F6P]_{eq}} = \frac{\left[G6P^{-1}\right] + \left[G6P^{-2}\right]}{[F6P^{-1}] + [F6P^{-2}]}$$

which is a function of temperature, pressure, pH, ionic strength and the pKa of both weak acids. The apparent equilibrium constant of a reaction, K', is related to the standard transformed reaction Gibbs energy by

$$-RT \ln (K') = \Delta_r G'^{0} \tag{2}$$

Therefore, each time an apparent equilibrium constant is measured, one unknown reactant standard transformed Gibbs energy of formation can be back calculated. This back calculation process is discussed in detail by Alberty [4, 5]. Any error in the measurement of the apparent equilibrium constant is ameliorated in the estimation of reactant standard Gibbs energy of formation by virtue of the logarithmic function in Eq. 2.

A.1.2 Group contribution methodology

There are many reactants for which experimentally derived metabolite species standard Gibbs energies of formation are not available. However, it is possible to estimate standard metabolite species Gibbs energy of formation using a group contribution approach [22, 26]. Briefly, this method considers the thermodynamic properties of each metabolite species as the sum of the properties of its constituent structural subgroups. The standard Gibbs energy of formation of various structural subgroups are estimated by linear regression from experimentally known metabolite species structures, and their standard Gibbs energies of formation. To estimate the standard Gibbs energy of formation for an unknown metabolite species, the contributions of each of its structural subgroups are summed along with an origin term. Henry *et al.*[20, 21] have successfully used this approach to computationally estimate standard Gibbs energy of formation for the majority of metabolite species in *E. coli*. The data input to the algorithm which estimates the standard Gibbs energy of formation is structure files which hold information about the atoms, bonds, connectivity and coordinates of substrate and product molecules [22].

[1]Corresponding author. Address: Science Institute and Center for Systems Biology, University of Iceland, Reykjavik, Iceland. Tel.:+354 618 6245

A difficult task in estimating the group contribution of a metabolite species standard Gibbs energy of formation, is the uncertainty due to non-additivity of subgroup contributions. A group contribution is said to be additive if the magnitude of the contribution is doubled when the number of such a group in a metabolite species is doubled. Individually, the uncertainty associated with estimation of a particular subgroup standard Gibbs energy of formation is typically small relative to metabolite species standard Gibbs energy of formation. However, for complex metabolites with many subgroups, the combined uncertainty in metabolite species standard Gibbs energy of formation can be significant. Jankowski *et al.* [22] have recently reported a significant improvement to the group contribution methodology, that has reduced the uncertainty and has provided better estimates of metabolite species standard Gibbs energy of formation. They explicitly propagate the uncertainty in subgroup estimates, and provide a standard error, $SE_{\Delta_f G^0_{est,i}}$, for each estimated metabolite species standard Gibbs energy of formation, $\Delta_f G^0_{est,i}$, e.g., Supplementary Table 3.

We take into account the uncertainty in estimated metabolite species standard Gibbs energy of formation when making quantitative assignment of reaction directionality, by estimating minimum and maximum reaction standard Gibbs energy using

$$\begin{aligned} \Delta_r G^0_{min,k} &\equiv \inf\{\mathbf{S}_k^T \cdot (\Delta_f \mathbf{G}^0_{est} \pm \mathbf{SE}_{\Delta_f G^0_{est}})\} \\ \Delta_r G^0_{max,k} &\equiv \sup\{\mathbf{S}_k^T \cdot (\Delta_f \mathbf{G}^0_{est} \pm \mathbf{SE}_{\Delta_f G^0_{est}})\} \end{aligned}$$

where inf denotes the infimum and sup denotes the supremum. Group contribution estimates of metabolite species standard Gibbs energy of formation are based on a reference temperature of $298.15K$ [22]. Standard enthalpy of formation is not available for the majority of *E. coli* metabolite species, and therefore standard Gibbs energy of formation cannot be adjusted to $310.15K$ as described in Section A.5. However, the charge and number of hydrogen atoms in each metabolite species are known. Therefore, we can compute estimates for metabolite species standard transformed Gibbs energy of formation, $\Delta_f G'^0_{est}$, at *in vivo* ionic strength and pH. See Section A.3 and A.4, respectively. Similarly, we can estimate minimum and maximum reaction standard Gibbs energy, $\Delta_r G'^0_{min}$ and $\Delta_r G'^0_{max}$. The group contribution approach represents a valuable complementary approach to the estimation of metabolite species standard transformed Gibbs energy of formation, when such data is not available due to a dearth of experimentally measured equilibrium constants. This is the case for the majority of metabolite species in the *E. coli* genome scale model. Jankowski et al [22] continue to expand the number of different structural groups within the scope of the group contribution method, so we can expect the number of metabolite species without group contribution estimates to continue to shrink.

A.2 Reactant standard transformed Gibbs energy of formation

Chemical potential is a thermodynamic potential introduced by J. Willard Gibbs [17], and can be intuitively thought of as a measure of "the tendency of a substance to change its location, chemical composition or state of aggregation" [9]. In a chemical reaction, a difference in chemical potential between substrates and products represents a driving force for spontaneous change of location, chemical composition or state of aggregation. This is discussed further in Section **??**. First we discuss chemical potential, from which change in chemical potential can be derived.

Since it is not possible to measure an absolute value for the chemical potential of a metabolite species [30], we follow established nomenclature [1] and use the (molar) Gibbs energy of formation of a metabolite species, $\Delta_f G_j$. The chemical potential of a metabolite species is an intensive quantity, meaning that it does not depend on the absolute number of molecules of that metabolite in a system. In contrast, the Gibbs energy of a metabolite species is an extensive quantity which is given by the product of metabolite species chemical potential and the number of metabolite species. The molar Gibbs energy of a metabolite species is an intensive property since it is the Gibbs energy per mole of metabolite species. Depending on the strictness of convention, one may observe molar Gibbs energy and the chemical potential being used interchangeably.

A.3 *In vivo* ionic strength

In water, many metabolite species dissociate into anions$^-$ and cations$^+$. However, a charged metabolite species in solution does not behave ideally, even at very low concentrations in the physiological range of ionic strength [14]. This is mainly due to attraction of metabolite species with opposite charges, which has the effect of shielding the repulsion between metabolite species of like charge, see Chapt. 26 [10]. At non-zero ionic strength, metabolite

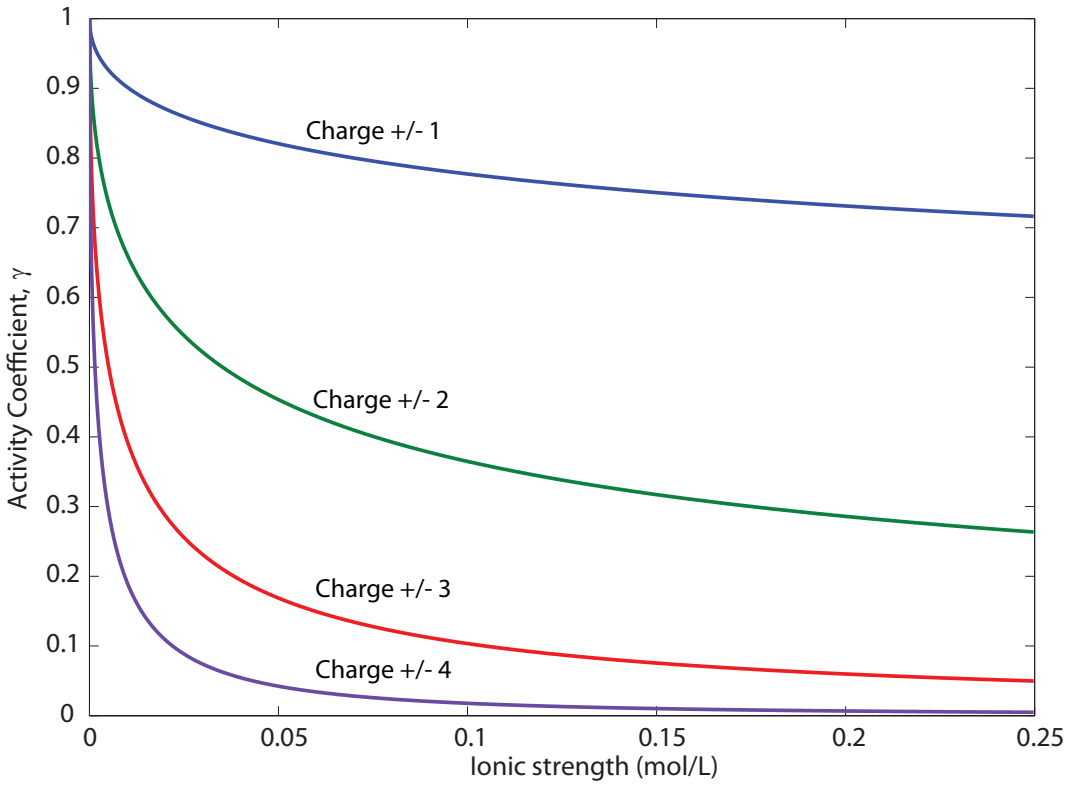

Figure 5: The dimensionless activity coefficient, γ, as a function of ionic strength, I, for metabolite species with different absolute charge, $|z|$. All else being equal, the activity coefficient is the same for a given positive or negative charge of the same magnitude, since z^2 appears in the function giving ionic strength. The activity coefficient of a metabolite species with a charge of either +4 or -4 may differ markedly from 1. This is significant for a number of reactants, such as for D-Fructose-1-6-bisphosphate, NADPH or ATP. e.g. at 0.25 M, and $310.15K$, the reactant D-Fructose-1-6-bisphosphate, is distributed between three dissociated forms, FDP^{4-}, $FDP^{3-}$and FDP^{2-}. However, the mole fraction of $FDP^{4-}$is greater than 0.99 with an activity coefficient of 4.8×10^{-3}. This means that a micromolar cytoplasmic concentration of FDP^{4-} actually corresponds to a nanomolar activity.

species molar Gibbs energy of formation, $\Delta_f G_j$, is given by

$$\Delta_f G_j = \Delta_f G_j^o + RT \ln (a_j) \tag{3}$$

where R is the gas constant, T is temperature, and a_j is metabolite species activity [14]. The activity of a metabolite species is related to its concentration by $a_j = \gamma_j x_j$, where the activity coefficient, γ_j, represents the non-ideal electrolyte behavior. Given the electric charge of a metabolite species, z_j, one can reasonably approximate the *in vivo* activity coefficient using the extended Debye-H*ü*ckel equation [14] relating charge, z_j, and ionic strength, I, to the activity coefficient

$$\ln (\gamma_j) = -\frac{A z_j^2 I^{\frac{1}{2}}}{1 + B I^{\frac{1}{2}}} \tag{4}$$

where A and B are empirically fitted parameters, which approximate experimentally measured activity coefficients [2]. The ionic strength, I, of a solution is defined by

$$I \equiv \frac{1}{2} \sum_{j=1}^{m} z_j^2 x_j$$

where z_j is the charge, and $x_j$is the concentration of metabolite species j. The *in vivo* ionic strength is thus a function of all charged metabolite concentrations, which are largely unknown [16].

Estimation of activity coefficient using extended Debye-H*ü*ckel equation is considered to be valid in the ionic strength range 0-0.35M [4]. By estimating the activity coefficient for each metabolite species, this term may then be absorbed into Eq. 3 for metabolite species Gibbs energy of formation

$$\begin{aligned} \Delta_f G_j &= \Delta_f G_j^o + RT \ln a_j \\ &= \Delta_f G_j^o + RT \ln (\gamma_j x_j) \\ &= \Delta_f G_j^o + RT \ln \gamma_j + RT \ln x_j \\ &\equiv \Delta_f G_j^0 + RT \ln x_j \end{aligned} \tag{5}$$

We shall still refer to $\Delta_f G_j^0$ as a metabolite species standard Gibbs energy of formation but the slight change in superscript, from ^{o} to ^{0}, is used to remind the reader that it is now a function of ionic strength. By absorbing the activity coefficient into the standard term we retain the convenience of dealing with experimentally measured concentrations and not activities [4]. Obviously, it would be of significant benefit to have experimental data on typical values of the actual *in vivo* ionic strength. Unless otherwise specified, we assume an ionic strength of 0.25 M.

A.4 *In vivo* pH

The second law of thermodynamics dictates that net spontaneous change must be in the direction of a drop in thermodynamic potential. When modeling the thermodynamics of any system, the set of variables that are held constant dictate the appropriate thermodynamic potential [14]. At constant temperature and pressure, the appropriate thermodynamic potential is Gibbs energy. In biochemical thermodynamics, we must also consider that the buffering capacity of intracellular proteins acts as a bath which maintains a constant pH. At constant temperature, pressure and pH, the appropriate thermodynamic potential is the transformed Gibbs energy [4]. The adjective refers to the Legendre transformation which is used to account for the fact that pH is a specified constant. This transformation can be made directly to the standard Gibbs energy of formation, giving rise to the definition of the standard transformed Gibbs energy of a metabolite species

$$\triangle_f G_j'^0 \equiv \triangle_f G_j^0 - N_j(H) \Delta_f G(H^+) \tag{6}$$

where $N_j(H)$ is the number of hydrogen atoms in metabolite species j and $\Delta_f G(H^+)$ is the hydrogen ion Gibbs energy of formation. It is important to realise that hydrogen ion Gibbs energy of formation is a function of ionic strength and temperature, just like any other metabolite species, see Eq. 4 and Eq. 5. For a comprehensive treatment of Legendre transforms in biochemical thermodynamics, see Alberty [2, 4, 5].

[2] $B = 1.6 L^{-\frac{1}{2}} mol^{-\frac{1}{2}}$. A is a function of temperature and pressure, at $298.15K$ and atmospheric pressure, then $A = 0.510651 L^{-\frac{1}{2}} mol^{-\frac{1}{2}}$.

A.5 *In vivo* temperature

The enthalpy, H, is a thermodynamic potential defined by a Legendre transformation of the internal energy

$$H \equiv U + PV \tag{7}$$

Only at constant pressure, the differential of enthalpy provides the appropriate criterion for spontaneous change. Metabolite species standard Gibbs energy of formation may be estimated as a function of temperature, $\triangle_f G_j^o(T)$, using

$$\triangle_f G_j^o(T) = \left(\frac{T}{298.15K}\right) \triangle_f G_j^o(T = 298.15K) + (1 - \frac{T}{298.15K}) \triangle_f H_j^o(T = 298.15K) \tag{8}$$

where $\triangle_f G_j^o(T = 298.15K)$ and $\triangle_f H_j^o(T = 298.15K)$ are metabolite species standard Gibbs energy of formation and metabolite species standard enthalpy of formation, respectively, measured at 298.15K [4]. Metabolite species standard enthalpy of formation may be obtained from calorimetric experiments [4] and it also may be a function of temperature but we assume here that this variation is small going from $298.15K \rightarrow 310.15K$.

Additional effects of temperature on Gibbs energy of formation arise due to the temperature dependence of the parameter A in the extended Debye-H*ü*ckle equation 4 used for predicting metabolite species activity coefficient. This is particularly important when estimating the activity of hydrogen ions at a specified pH. The effect of temperature on hydrogen ion activity appears as an adjustment to the hydrogen ion Gibbs energy of formation. This temperature effect on hydrogen ion Gibbs energy of formation will then be multiplied by the number of hydrogen atoms in each metabolite species when making a Legendre transformation for constant pH, as in Eq. 6.

A.6 Combined equation for ionic strength, pH and temperature adjustment

Amalgamation of all the adjustments to metabolite species standard Gibbs energy of formation, $\triangle_f G_j^o$, for specified temperature (Eq. 8), specified ionic strength (Eq. 5), and specified pH (Eq. 6) gives a metabolite species standard transformed Gibbs energy of formation as

$$\Delta_f G_j'^0 = \left(\frac{T}{298.15}\right) \triangle_f G_j^o + (1 - \frac{T}{298.15}) \triangle_f H_j^o - \frac{RTAz_j^2 I^{\frac{1}{2}}}{1 + BI^{\frac{1}{2}}} - N_j(H)\Delta_f G(H^+)$$

This equation highlights that the metabolite species standard transformed Gibbs energy of formation depends on properties that are specific to that metabolite species.

A.7 Concentrations of water and dissolved gases

In the treatment of biochemical reactions, as in this study, it is typical to assume an activity of water equal to one [3]. In addition, we use the dissolved oxygen (reactant) concentration range, $0.1 - 8.6 \times 10^{-6}$ M . In aqueous phase the reactant carbon dioxide is distributed between a number of metabolite species, some involving water $[CO_2] = [CO_2(aq)] + [CO_3^{2-}] + [HCO_3^-] + [H_2CO_3]$. We assumed a carbon dioxide (reactant) concentration of $[CO_2] = 1$ mM [20]. The details follow the special thermodynamic treatment of carbon dioxide as described by Alberty, (Section 8.7 in [4]).

A.8 Qualitative assignment of reaction directionality

During reconstruction of a biochemical network, qualitative assignment of reaction directionality is made on a reaction-by-reaction basis using various forms of evidence from biochemical characterization experiments of specific enzymes [37]. If, at physiological concentrations of substrates and products, a reaction is experimentally observed to favor the production of products then it may be classified as irreversible in a forward direction. If biochemical characterization is not available, a qualitative assignment of reaction irreversibility can be made if a reaction is co-factor coupled, such as to the hydrolysis of ATP, since ATP driven reactions tend to be effectively irreversible.

Reaction	E.C. / Gene	Eq. and qualitative direction	$\Delta_r G'^0_{low}$	$\Delta_r G'^0_{high}$	$\Delta_r G'_{min}$	$\Delta_r G'_{max}$
D-Erythrose-4-phosphate dehydrogenase*	1.2.1.72	**e4p** + h2o + nad ⇌ **4per** + 2 h + nadh	-66.4	-51.1	-105.6	-11.9
galactokinase	2.7.1.6	atp + gal ⇌ adp + gal1p + h	-49.0	-49.0	-88.2	-9.8
γ-glutamyl γ-aminobutyric acid dehydrogenase	puuC	**ggbutal** + h2o + nadp ⇌ **gg4abut** + 2 h + nadph	-71.8	-43.5	-111.0	-4.3
glutathione oxidoreductase	1.8.1.7	gthox + h + nadph ⇌ 2 gthrd + nadp	-914.0	-914.0	-982.9	-884.9
IMP cyclohydrolase*	3.5.4.10	h2o + imp ⇌ **fprica**	-60.1	-26.4	-79.7	-6.7
phosphoribosylpyrophosphate synthetase	2.7.6.1	atp + r5p ⇌amp + h + **prpp**	-43.3	-43.3	-82.5	-4.1
phosphate reversible transport via symport periplasm	pitA or pitB	h[p] + pi[p] ⇌ h + pi	8.5	8.5	6.7	26.3
glycogen phosphorylase (GLCP)	2.4.1.1	**glycogen** + pi → g1p	693.0	794.0	685.3	825.5
glycogen phosphorylase (GLCP2)	2.4.1.1	**bglycogen** + pi → g1p	693.0	794.0	685.3	825.5
2-C-methyl-D-erythritol-2-4-cyclodiphosphate synthase*	4.6.1.12	**2p4c2me** → **2mecdp** + **cmp**	63.8	167.2	14.5	176.7
methionine adenosyltransferase*	2.5.1.6	atp + h2o + met-L → **amet** + pi + ppi	114.6	176.1	63.5	203.5
phosphoribosylaminoimidazole synthase*	6.3.3.1	atp + **fpram** → adp + **air** + 2 h + pi	84.7	156.2	33.7	183.5

Table 2: Reactions, as given in the iAF1260 reconstruction, with quantitatively assigned reaction directions that are more constrained, or reversed, in comparison with qualitative assignments. The * denotes reactions that, contrary to quantitative assignment, are essential in the forward direction, for aerobic growth on glucose minimal medium. The key to reactant abbreviations is given in Table 3. The **bold** reactant abbreviations denote the use of group contribution estimates where data back calculated from equilibrium constants was not available. Where appropriate, the low and high standard transformed reaction Gibbs energy, $\Delta_r G'^0_{low}$ and $\Delta_r G'^0_{high}$, account for a one standard deviation uncertainty in reactant group contribution estimates. The minimum and maximum transformed reaction Gibbs energy, $\Delta_r G'_{min}$ and $\Delta_r G'_{max}$, also assume a $0.01-20\,mM$ physiological reactant concentration range, except: $[CO_2] = 10^{-4}\ M$, $[H_2O] = 1\ M$, and $[O_2] = 8.2 \times 10^{-8} - 8.2 \times 10^{-6}\ M$. Data given for a temperature of $310.15\ K$, ionic strength of $0.25\ M$, and pH of 7.7. See Supplementary Table 3 for metabolite legend.

A.9 Quantitative tightening of qualitatively assigned reaction directionality: reconciliation with experimental literature

A.9.1 Qualitatively reversible yet quantitatively forward

Erythrose 4 phosphate dehydrogenase (E.C. 1.2.1.72) D-Erythrose-4-phosphate dehydrogenase catalyzes the first step in the biosynthesis of essential coenzyme pyridoxal 5'-phosphate (active vitamin B6), from the pentose phosphate precursor, D-erythrose-4-phosphate [43]

$$\text{D-Erythrose-4-phosphate+NA}D^{+}\text{+}H_2\text{O} \rightleftharpoons \text{4-Phosphoerythronate+2h+NADH}$$

In silico, the forward oxidation reaction direction is essential for growth on glucose minimal medium. Biochemical characterization experiments established that, at pH 7, the oxidation of 1 mM D-Erythrose-4-phosphate and reduction of 1 mM NAD^+ proceeded to completion, upon addition of D-Erythrose-4-phosphate dehydrogenase as measured by NADPH production [11]. This suggests that indeed the D-Erythrose-4-phosphate dehydrogenase reaction is strongly thermodynamically favorable, supporting the quantitative assignment.

Abbr	Metabolite	$\pm SE_{\Delta_f G^0_{est}}$
2mecdp	2-C-methyl-D-erythritol-2-4-cyclodiphosphate	18.4
2p4c2me	2-phospho-4-cytidine-5-diphospho-2-C-methyl-D-erythritol	17.3
4per	4-Phospho-D-erythronate	3.8
air	5-amino-1-5-phospho-D-ribosyl-imidazole	16.4
amet	S-Adenosyl-L-methionine	30.8
bglycogen	branching-glycogen	50.5
cmp	Cytidine-5'-monophosphate	16.0
e4p	D-Erythrose-4-phosphate	3.9
fpram	2-Formamido-N1-5-phospho-D-ribosyl-acetamidine	19.4
fprica	5-Formamido-1-5-phospho-D-ribosyl-imidazole-4-carboxamide	16.9
g1p	D-Glucose-1-phosphate	-
gal1p	alpha-D-Galactose-1-phosphate	-
gal	D-Galactose	-
gg4abut	gamma-glutamyl-gamma-aminobutyric-acid	7.1
ggbutal	gamma-glutamyl-gamma-butyraldehyde	7.1
glycogen	glycogen	50.5
gthox	Oxidized-glutathione	-
gthrd	Reduced-glutathione	-
imp	Inosine monophosphate	-
met-L	L-Methionine	-
nad	Nicotinamide-adenine-dinucleotide	-
nadh	Nicotinamide-adenine-dinucleotide–reduced	-
nadp	Nicotinamide-adenine-dinucleotide-phosphate	-
nadph	Nicotinamide-adenine-dinucleotide-phosphate–reduced	-
nmn	Nicotinamide ribonucleotide	-
pi	Phosphate	-
ppi	Diphosphate	-
prpp	5-Phospho-alpha-D-ribose-1-diphosphate	-
r5p	alpha-D-Ribose-5-phosphate	-

Table 3: Reactants involved in reactions with quantitatively assigned directions that are more constrained, or reversed, as compared with qualitatively assigned direction (Reactions given in Supplementary Table 2). The reactants where group contribution estimates were used are indicated with the associated standard error in standard Gibbs energy of formation, $SE_{\Delta_f G^0_{est}}$, [22].

Galactokinase (E.C. 2.7.1.6) Galactokinase is a phosphotransferase that catalyzes the phosphorylation of D-galactose to D-galactose-1-phosphate . At a pH of 7 and temperature of 298 K the apparent equilibrium constant was experimentally measured to be 25 ± 9 [8]. The corresponding standard Gibbs energy change in these conditions, $-7.9 \pm 5.4\ kJ\ mol^{-1}$, is considerably greater than the $-49\ kJ\ mol^{-1}$ standard transformed reaction Gibbs energy back calculated from experimentally measures equilibrium constants, at a temperature of 310.15 K, ionic strength of 0.25 M, and pH of 7.7. The standard transformed reaction Gibbs energy was calculated using the following reactant standard transformed Gibbs energy of formation: ATP = -2188.79, D-Galactose = -336.02, ADP = -1326.02 and alpha-D-Galactose-1-phosphate = -1247.75 , H = 0 ($kJ\ mol^{-1}$). These standard transformed reactant Gibbs energies do not account for the 10 mM concentration of Magnesium ions in the solution when the equilibrium constant of galactokinase was measured. Since ATP is know to bind magnesium ions, a further investigation of this reaction is warranted at constant 10 mM magnesium concentration [4].

γ-glutamyl γ-aminobutyric acid dehydrogenase (puuC) γ-glutamyl γ-aminobutyric acid dehydrogenase catalyzes a key step in the recently discovered putrescine utilization pathway in *E. coli* [25]. Putrescine is an organic polyamine that used as essential cofactor in bacterial ribosomes and is a precursor for spermidine and spermine which are also essential cofactors of ribosomes. Putrescine is also produced by the breakdown of amino acids in living and dead organisms [25]. *E. coli* can catabolise putrescine as the sole nitrogen source in minimal medium, and physiologically this function is used in conditions of nutrient starvation where other more preferable sources of nitrogen are not present. The carbon backbone of putrescine also enters the tricarboxylic acid cycle as succinate. To our knowledge, the equilibrium constant of the γ-glutamyl γ-aminobutyric acid dehydrogenase reaction has not been measured. However, in the unbranched putrescine utilization pathway, the distal reaction, gamma glutamyl putrescine oxidase, is both qualitatively and quantitatively assigned to be irreversible in the forward direction. Therefore the quantitative assignment is consistent with the physiological use of this reaction as an alternate source of nitrogen and carbon.

Glutathione oxidoreductase (E.C. 1.8.1.7) Glutathione oxidoreductase (synonym glutathione reductase), is a member of a flavoproteins family that transfers electrons from a pyridine nucleotide to a specific disulfide containing substrate, in order to maintain the sulfhydryl redox potential in the cytoplasm. Specifically, glutathione oxidoreductase catalyzes the transfer of an electron from NADPH to glutathione disulphide (synonym oxidised glutathione) [23]. In iAF1260, the stoichiometry of this reaction is given as

$$\text{glutathione(oxidized)+h+NADPH} \rightleftharpoons 2 \times \text{glutathione(reduced)+NADP}$$

With this stoichiometry, at 310.15 K, ionic strength 0.25 M, and pH of 7.7, the standard transformed Gibbs energy of the reaction is $-914.0\ kJ\ mol^{-1}$. Although this reaction is experimentally observed to be thermodynamically favourable in the forward direction, this standard transformed Gibbs energy seems excessively large in magnitude. Since Tewari and Goldberg [36] estimated that the apparent equilibrium constant of the reaction is $\simeq 0.013$, giving a standard transformed Gibbs energy of $11.33\ kJ\ mol^{-1}$ at 311.15 K, ionic strength 0.25 M, and pH of 7. The large discrepancy between experimental and back-calculated value seems to originate from the large difference in standard *transformed* Gibbs energy of formation between oxidised and reduced gluthathione (p413 of [5]).

Inosine monophosphate cyclohydrolase (E.C. 3.5.4.10) Inosine monophosphate (IMP) is a precursor in the purine nucleotide *de novo* synthesis pathway, which can either be converted into AMP or GMP. These nucleoside monophosphates are then phosphorylated to produce ATP and GTP. IMP cyclohydrolase catalyzes the cyclization of 5-formamido-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide, eliminating water, to form the imidazole ring of IMP. In iAF1260, this qualitatively reversible reaction was written with IMP as the substrate. With this convention, and using group contribution estimates of reactant standard transformed Gibbs energy, this reaction seems thermodynamically irreversible in the direction written. However, experimentally it was observed that the reverse reaction, synthesis of IMP, is highly favored [12]. At 298 K and pH 7.5, the unidirectional rate constants for the catalytic step were, slower than the observational time for the forward direction $k_f < 1\ s^{-1}$, and much fast for the reverse direction, $40 s^{-1}$, suggesting that the overall reaction is more favorable in the reverse direction, i.e., the production of IMP [12]. Group contribution estimates are derived by estimating a standard Gibbs energy of formation for each of the structural subgroups in a metabolite species. IMP presents a challenge to this method since non-additive effects between subgroups can be expected to be increased in due to the imidazole ring, which is not present in 5-formamido-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide, see Figure 6.

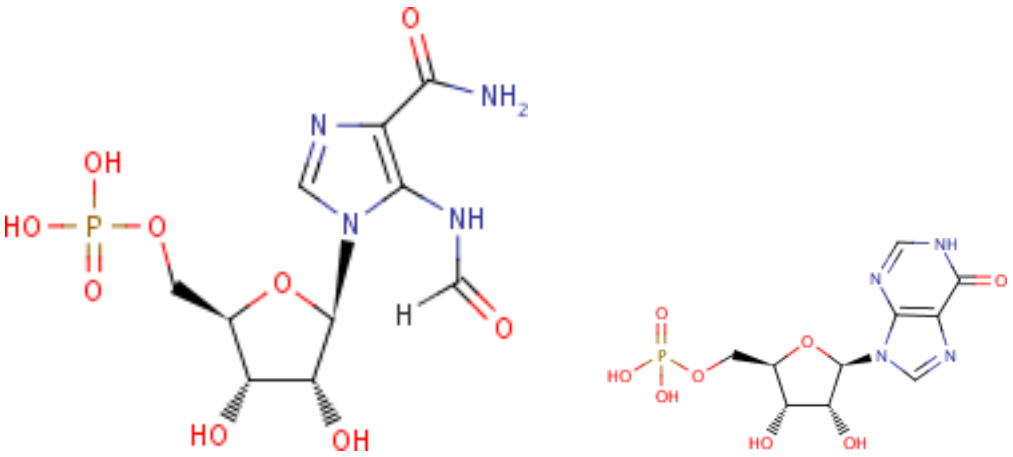

Figure 6: 5-formamido-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide (left) and Inosine monophosphate with the imidazole ring (right) [13].

Phosphoribosylpyrophosphate synthetase (E.C. 2.7.6.1) The biosynthesis of 5-Phospho-alpha-D-ribose-1-diphosphate from alpha-D-Ribose-5-phosphate and ATP is the point at which carbons are drained from the oxidative pentose cycle and utilized for biosynthesis of macromolecular precursors [41]. Thus, this is the first step of a biosynthesis leading to pyrimidine, purine and pyridine as well as histidine. The experimentally measured apparent equilibrium constant was 28.6 [35], at 310 K, pH 7.5 and 5 mM $MgCl_2$. This corresponds to a standard transformed Gibbs energy of -8.64 $kJ\ mol^{-1}$, which is significantly different than the estimate of -43.3 $kJ\ mol^{-1}$ at a temperature of 310.15 K, ionic strength of 0.25 M, and pH of 7.7. The 5-Phospho-alpha-D-ribose-1-diphosphate standard Gibbs energy of formation was estimated using the group contribution method, its structure is illustrated in Figure 7.

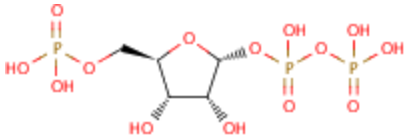

Figure 7: 5-Phospho-alpha-D-ribose-1-diphosphate [13].

A.9.2 Qualitatively reversible yet quantitatively reverse

Phosphate 'reversible' transport via symport periplasm (pitA or pitB) There are two transport mechanisms in *E. coli* for uptake of inorganic phosphate[39]. When periplasmic inorganic phosphate, $pi[p]$, is plentiful, *E. coli* transports it into the cell by a low-affinity, high velocity transport system. The pitA or pitB genes [18, 42] encode this transport system, termed 'Phosphate reversible transport via symport periplasm' in iAF1260. The stoichiometry of this proton symport reaction is assumed to be

$$\text{h[p]+pi[p]} \rightleftharpoons \text{h+pi}$$

and the reaction is assigned to be reversible. At a temperature of 310.15 K, pH of 7.7, and ionic strength of 0.25 M, which are assumed to be identical for both cytoplasm and periplasm, the standard transformed reaction Gibbs energy is 8.5 $kJ\ mol^{-1}$. The standard transformed Gibbs energy of the reactants are, $h[p] = 8.68$, $pi[p] = -1062.17$, $h = 0$, $pi = -1045$ $kJ\ mol^{-1}$.

The 90 mV electrical potential difference between the cytoplasm and periplasm [32] is responsible for the differences in standard transformed reactant Gibbs energy for these charged metabolites. Various regulatory mechanisms maintain a $8 - 10$ mM cytoplasmic phosphate concentration in various growth conditions [7, 38]. Quantitatively, it seems as if the current 1:1 proton:phosphate stoichiometry is insufficient to facilitate the transport of inorganic

phosphate into the cell, except when periplasmic inorganic phosphate is present at a high concentration, $\gtrsim 0.26\ M$. However, it is known that when the concentration of periplasmic inorganic phosphate is lower than 20 mM, a different high-affinity, low velocity transport system, encoded by pstA/pstB, is induced [31]. Therefore, we predict that either there is a third transport mechanism for uptake of inorganic phosphate, or the proton:phosphate stoichiometry of the low-affinity, high velocity pitA or pitB transporter is 2:1. The latter stoichiometry would give rise to a transformed reactant Gibbs energy of $\sim -2\ kJmol^{-1}$ at a periplasmic inorganic phosphate concentration of 20 mM. However, the caveat is that quantiative prediction of stoichiometry for transport reactions might be altered when non-identical pH and ionic strength for cytoplasm and periplasm are considered.

A.9.3 Qualitatively forward yet quantitatively reverse

Glycogen phosphorylase (E.C. 2.4.1.1) Glycogen is a polysaccharide dendrimer consisting of 12-14 glucose linked in a chain by $\alpha - 1,4$ bonds, chains linked together by $\alpha - 1,6$ branch bonds, see Figure 8. Glycogen phosphorylase, encoded by glgP, is responsible for glyocogenolysis. During extended periods of substrate deprivation, it cleaves $\alpha - 1,4$ bonds releasing one glucose-6-phosphate leaves a chain of the glycogen dendrimer one unit shorter [6]. In iAF1260 this process is represented by the irreversible reaction (GCLP) with the stoichiometry

$$\text{glycogen+pi} \rightarrow \text{D-Glucose1-phosphate}$$

where glycogen represents a single unit in a linear chain of glucose linked by $\alpha - 1,4$ bonds. Implicit in this representation is the fact that in the real reaction the glycogen dendrimer is one unit shorter. In most species, glycogen phosphorylase, cannot cleave a chain with less than 5 glucose units [38]. There is some evidence that this is also the case in *E.coli* [6], therefore the reaction abbreviated GLCP2 involving branched glycogen (bglycogen) seems superfluous.

The equilibrium constant of the glycogenolysis reaction cannot be measured by conventional means since substrate glycogen is indistinguishable from product glycogen [38]. Therefore, it is necessary to measured the reverse reaction, since one can quantify the formation of inorganic phosphate. At pH 7 and 298 K the apparent equilibrium constant for the reaction is 0.28 [38]. This corresponds to a standard transformed Gibbs energy of 3.15 $kJ\ mol^{-1}$. This is markedly different from the quantitative estimate, which uses a group contribution estimate for glycogen. Because the glycogen dendrimer has a variable structure it cannot be accurately represented by a single structure file, complicating group contribution estimation. It has been observed *in vivo* that the glycogen phosphorylase reaction (GLCP) proceeds towards the production of D-Glucose 1-phosphate due to the high concentration of inorganic phosphate, $8 - 10\ mM$, versus D-Glucose 1-phosphate, 170 μM [38]. Therefore, the qualitatively assigned forward direction is correct.

2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (E.C. 4.6.1.12) Isoprenoids are a large group of organic lipids, which are are all assembled from the universal precursors dimethylallyl diphosphate and isopentenyl diphosphate [19]. To synthesize these precursors, *E. coli* uses an alternate pathway to the mammalian mevalonate pathway, and 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase catalyzes a step in this alternate pathway. The pathway begins with 1-deoxy-D-xylulose 5-phosphate which is assembled from pyruvate and D-glyceraldehyde 3-phosphate. When isotopically labeled 1-deoxy-D-xylulose was incubated with a recombinant strain engineered for hyperexpression of genes in the non mevalonate pathway, upto and including ispF, coding for 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, the ^{13}C nuclear magnetic resonance spectrum of the cell extract was dominated by the known signals of 2C-methyl-D-erythritol 2,4-cyclodiphosphate [19]. While, to our knowledge, the equilibrium constant of the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase reaction has not been measured, the aforementioned study indicates that this reaction is correctly qualitatively assigned to be forward, rather than the quantitative assignment of reverse. The substrates and products of the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase reaction are significantly structurally distinct which perhaps is a reason for the discrepancy between the quantitative assignment and qualitative assignment of reaction directionality, see Figure 9.

Methionine adenosyltransferase (E.C. 2.5.1.6) Methylation has a myriad of functions in *E. coli* including regulation of gene expression, protein function, and RNA metabolism. Methionine adenosyltransferase catalyzes the synthesis of the most widely uses methy donor, S-adenosylmethionine. The reaction occurs in a two-step reaction, in which the complete tripolyphosphate chain is cleaved from ATP as S-adenosylmethionine is formed, and the tripolyphosphate is further hydrolyzed to pyrophosphate (PPi) and orthophosphate (Pi) [24]. At 298.15 K

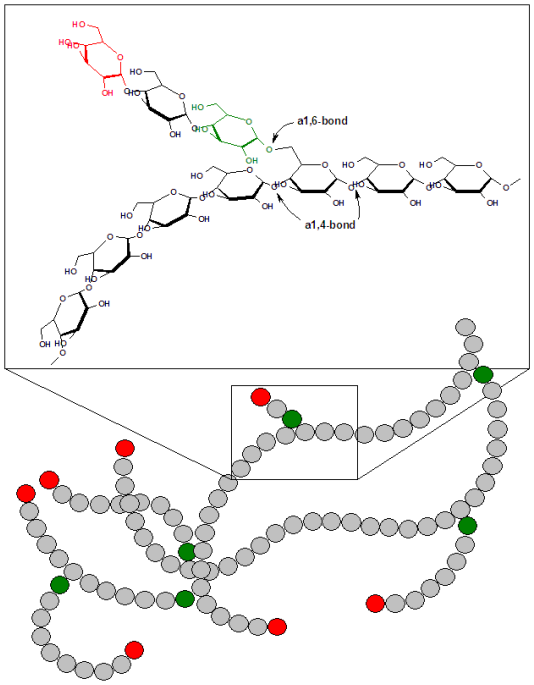

Figure 8: Glycogen

and pH 8, the experimentally measured apparent equilibrium constant was approximately 1×10^4 [27], which corresponds to a standard transformed Gibbs energy of $-23\ kJ\ mol^{-1}$. This conflicts with the positive estimated reaction standard transformed Gibbs energy. At 298.15 K and pH 8, the Gibbs energy profile for the methionine adenosyltransferase reaction was constructed assuming *in vivo* concentrations of ATP (2.6 mM), methionine (0.1 mM), S-adenosylmethionine ($0.08mM$), pyrophosphate (0.5 mM), and orthophosphate (9 mM) [27]. The overall transformed Gibbs energy was $-42\ kJ\ mol^{-1}$ supporting the qualitative assignment of a forward direction for the methionine adenosyltransferase reaction [27].

Phosphoribosylaminoimidazole synthase (E.C. 6.3.3.1) This enzyme catalyzes an intermediate step in de novo purine synthesis, the conversion of 2-Formamido-N1-5-phospho-D-ribosyl-acetamidine (fpram) to 5-amino-1-5-phospho-D-ribosyl-imidazole (air), driven by ATP hydrolysis. At 20 μM ATP and 5 μM 2-Formamido-N1-5-phospho-D-ribosyl-acetamidine, the reaction was experimentally observed to be effectively irreversible in the forward direction [34]. The conversion of 2-Formamido-N1-5-phospho-D-ribosyl-acetamidine to 5-amino-1-5-phospho-D-ribosyl-imidazole, as the name suggests, requires a significant structural rearrangement to form the imidazole ring. Perhaps this is responsible for the erroneous quantitative assignment in this case.

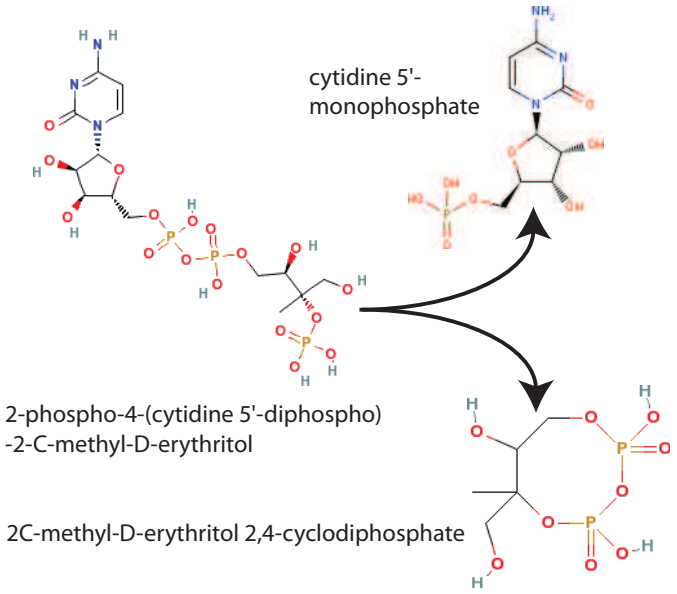

Figure 9: The reaction catalyzed by 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, (E.C. 4.6.1.12) [13].

Supplementary Figures

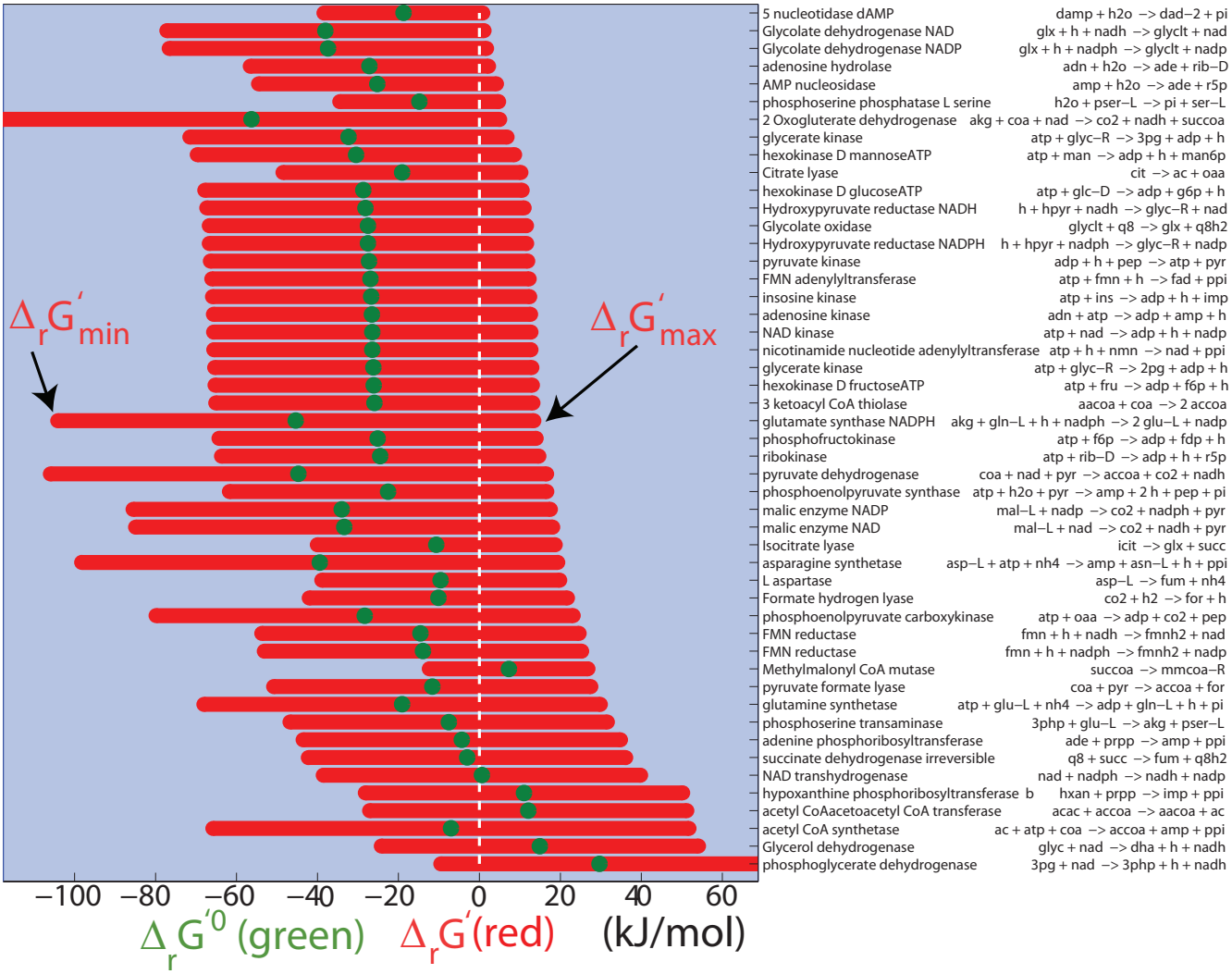

Figure 10: Non transport reactions in iAF1260 which were qualitatively assigned to be forward, yet are quantitatively assigned to be reversible using standard transformed reactant Gibbs energy of formation back calculated from experimentally measured equilibrium constants [5]. Reactions are ordered by ascending maximum transformed reaction Gibbs energy, $\Delta_r G'_{max}$. The reactions towards the bottom are increasingly dependent on higher concentrations of substrates, compared to products, in order to remain quantitatively thermodynamically favorable in the forward direction. We assume a temperature of 310.15 K, ionic strength of 0.25 M, and pH of 7.7 and a physiological range of reactant concentration, $0.01 - 20$ mM, for all reactants, except $[CO_2] = 10^{-4}$ M, $[H_2O] = 1$ M, and $[O_2] = 1 \times 10^{-8} - 8.2 \times 10^{-6}$ M.

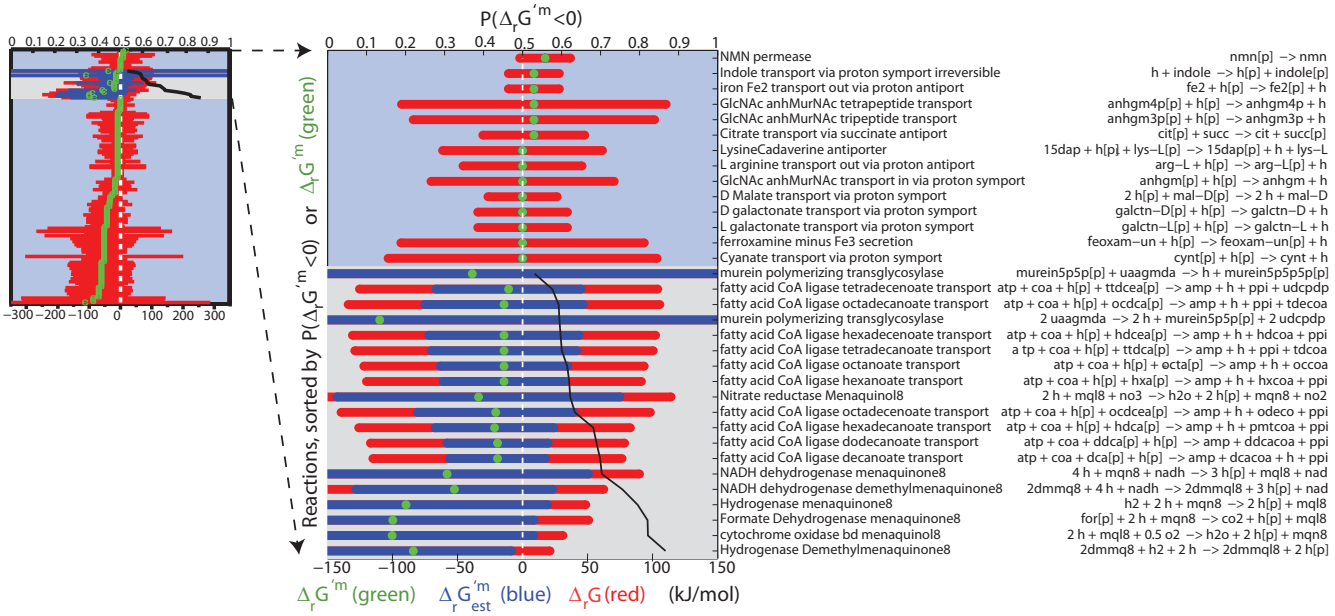

Figure 11: Overview (top left) and zoom (right) of transport reactions which are qualitatively forward, quantitatively reverse, yet required to operate in a forward direction to avoid net shuttling of protons into the periplasmic compartment. Without these reactions qualitatively set to forward, the increased reflux of periplasm protons through the ATP synthase reaction would generate excess ATP, thereby obviating the requirement for oxidative phosphorylation and artificially raising the growth rate. The abbreviatied reaction equations are given for non trivial reactions. The suffix [p] denotes a periplasmic reactant and the remainder of reactants are cytoplasmic. The fatty acid ligase reactions occur less in the periplasm than in the cytoplasmic side of the periplasmic membrane. Representing such spatial details will become more significant when non-identical pH and ioni strength is assumed for cytoplasmic and periplasmic 'compartments'.

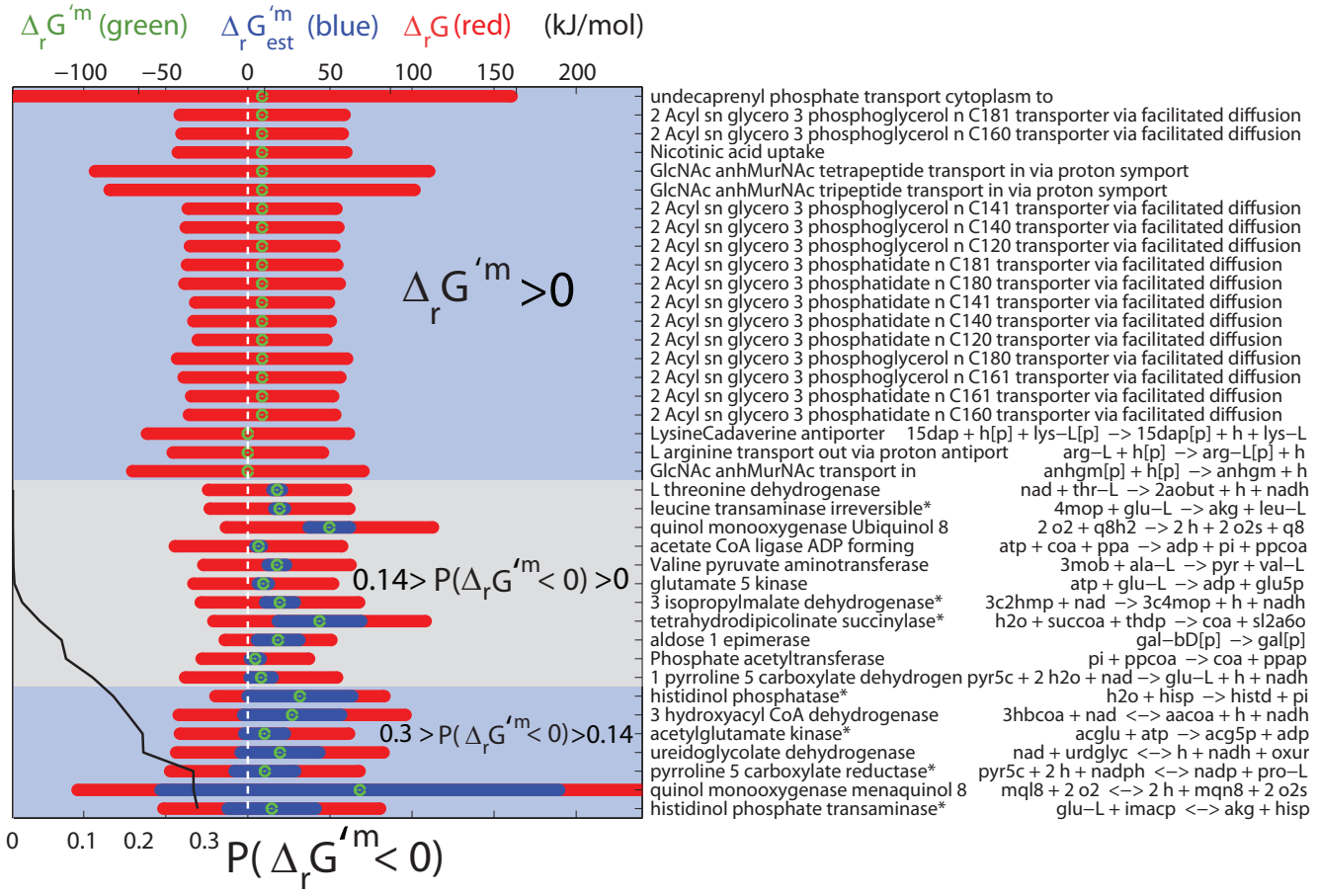

Figure 12: Qualitatively forward reactions that were quantitatively assigned to be reverse in the genome scale *E. coli* model, iAF1260. All reactions either have a positive physiological standard transformed reaction Gibbs energy, $0 < \Delta_r G'^m$, or are below the lower cutoff for cumulative probability of forward physiological standard transformed reaction Gibbs energy, $0.3 \geq P(\Delta_r G'^m < 0)$. Consider the 'nicotinic acid uptake' reaction, which is not coupled to any other thermodynamically favorable reaction in iAF1260. Nicotinic acid is a precursor of the essential coenzymes NAD and NADP. At pH 7, nicotinic acid has a charge of -2, therefore translocation from periplasm to cytoplasm, and accumulation within the cell is predicted to require an energy-dependent process, as has been observed experimentally [33]. The abbreviated reaction Equations are given for non trivial reactions. The suffix [p] denotes a periplasmic reactant and the remainder of reactants are cytoplasmic. It is possible that in vivo reactant concentration may drive any of these reactions in a forward direction. The forward direction of the 7 reactions denoted $*$ are essential for biomass production in glucose minimal medium and were qualitatively set accordingly.

1. R. A. Alberty. IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN). Recommendations for nomenclature and tables in biochemical thermodynamics. Recommendations 1994. *Eur J Biochem*, 240(1):1–14, Aug 1996.

2. R. A. Alberty. Legendre transforms in chemical thermodynamics. *J Chem Thermodyn*, 29(5):501–516, 1997.

3. R. A. Alberty. The role of water in the thermodynamics of dilute aqueous solutions. *Biophys Chem*, 100(1-3):183–192, 2003.

4. R. A. Alberty. *Thermodynamics of Biochemical Reactions*. Wiley-Interscience, Hoboken, NJ, 2003.

5. R. A. Alberty. *Biochemical Thermodynamics: Applications of Mathematica*. Wiley-Interscience, Hoboken, NJ, 2006.

6. N. Alonso-Casajús, D. Dauvillée, A. M. Viale, F. J. Muñoz, E. Baroja-Fernández, M. T. Morán-Zorzano, G. Eydallin, S. Ball, and J. Pozueta-Romero. Glycogen phosphorylase, the product of the glgp gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*. *J Bacteriol*, 188(14):5266–5272, Jul 2006.

7. N. Amin and A. Peterkofsky. A dual mechanism for regulating cAMP levels in *Escherichia coli*. *J Biol Chem*, 270(20):11803–11805, May 1995.

8. M. R. Atkinson, E. Johnson, and R. K. Morton. Equilibrium constant of the galactokinase reaction and free energy of hydrolysis of adenosine triphosphate. *Nature*, 184:1925–1927, Dec 1959.

9. R. Baierlein. The elusive chemical potential. *Am J Phys*, 69(4):423–434, 2001.

10. S. R. Berry, S. A Rice, and J. Ross. *Physical Chemistry*. Oxford University Press, Oxford, 2nd edition, 2000.

11. S. Boschi-Muller, S. Azza, D. Pollastro, C. Corbier, and G. Branlant. Comparative enzymatic properties of gapb-encoded erythrose-4-phosphate dehydrogenase of escherichia coli and phosphorylating glyceraldehyde-3-phosphate dehydrogenase. *J Biol Chem*, 272(24):15106–15112, Jun 1997.

12. K. G. Bulock, G. P. Beardsley, and K. S. Anderson. The kinetic mechanism of the human bifunctional enzyme atic (5-amino-4-imidazolecarboxamide ribonucleotide transformylase/inosine 5'-monophosphate cyclohydrolase). a surprising lack of substrate channeling. *J Biol Chem*, 277(25):22168–22174, Jun 2002.

13. K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, and M. Ashburner. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res*, 36(Database issue):D344–D350, Jan 2008.

14. K. A. Dill and S. Bromberg. *Molecular driving forces: Statistical thermodynamics in Chemistry and Biology*. Garland Science, London, 2003.

15. A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*, 3(1):e121, 2007.

16. O. Fiehn. Metabolomics: the link between genotypes and phenotypes. *Plant Mol Biol*, 48(1-2):155–171, 2002.

17. J. W. Gibbs. *Elementary principles in statistical mechanics, developed with especial reference to the rational foundation of thermodynamics*. Dover Publications, New York, 1902.

18. R. M. Harris, D. C. Webb, S. M. Howitt, and G. B. Cox. Characterization of PitA and PitB from *Escherichia coli*. *J Bacteriol*, 183(17):5008–5014, Sep 2001.

19. S. Hecht, W. Eisenreich, P. Adam, S. Amslinger, K. Kis, A. Bacher, D. Arigoni, and F. Rohdich. Studies on the nonmevalonate pathway to terpenes: the role of the gcpe (ispg) protein. *Proc Natl Acad Sci U S A*, 98(26):14837–14842, Dec 2001.

20. C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophys J*, 92(5):1792–1805, 2007.

21. C. S. Henry, M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys J*, 90(4):1453–1461, 2006.

22. M. D. Jankowski, C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys J*, 95(3):1487–1499, Aug 2008.

23. K.G. Knapp and J. R. Swartz. Cell-free production of active *E. coli* thioredoxin reductase and glutathione reductase. *FEBS Lett*, 559(1-3):66–70, Feb 2004.

24. J. Komoto, T. Yamada, Y. Takata, G. D. Markham, and F. Takusagawa. Crystal structure of the s-adenosylmethionine synthetase ternary complex: a novel catalytic mechanism of s-adenosylmethionine synthesis from atp and met. *Biochemistry*, 43(7):1821–1831, Feb 2004.

25. S. Kurihara, S. Oda, K. Kato, H. G. Kim, T. Koyanagi, H. Kumagai, and H. Suzuki. A novel putrescine utilization pathway involves gamma-glutamylated intermediates of *Escherichia coli* k-12. *J Biol Chem*, 280(6):4602–4608, Feb 2005.

26. M. L. Mavrovouniotis. Estimation of standard Gibbs energy changes of biotransformations. *J Biol Chem*, 266(22):14440–14445, 1991.

27. M. S. McQueney, K. S. Anderson, and G. D. Markham. Energetics of s-adenosylmethionine synthetase catalysis. *Biochemistry*, 39(15):4443–4454, Apr 2000.

28. J.D. Orth, R.M.T. Fleming, and Bernhard Ø. Palsson. Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*, chapter Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide (in press). ASM Press, 2009.

29. E. Padan and S. Schuldiner. Intracellular pH and membrane potential as regulators in the prokaryotic cell. *J Membr Biol*, 95(3):189–198, 1987.

30. M. Planck. *Treatise on Thermodynamics*. Courier Dover Publications, Chelmsford, MA, 1945.

31. N. N. Rao and A. Torriani. Molecular aspects of phosphate transport in *Escherichia coli*. *Mol Microbiol*, 4(7):1083–1090, Jul 1990.

32. H. Richard and J. W. Foster. *Escherichia coli* glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential. *J Bacteriol*, 186(18):6032–6041, Sep 2004.

33. J. J. Rowe, R. D. Lemmon, and G. J. Tritz. Nicotinic acid transport in *Escherichia coli*. *Microbios*, 44(179-180):169–184, 1985.

34. J. L. Schrimsher, F. J. Schendel, J. Stubbe, and J. M. Smith. Purification and characterization of aminoimidazole ribonucleotide synthetase from *Escherichia coli*. *Biochemistry*, 25(15):4366–4371, Jul 1986.

35. R. L. Switzer. Regulation and mechanism of phosphoribosylpyrophosphate synthetase. i. purification and properties of the enzyme from *Salmonella typhimurium*. *J Biol Chem*, 244(11):2854–2863, Jun 1969.

36. Y.B. Tewari and R.N. Goldberg. Thermodynamics of the oxidation–reduction reaction {2 glutathionered (aq)+ NADPox (aq)= glutathioneox (aq)+ NADPred (aq)}. *J Chem Thermodyn*, 35(8):1361–1381, 2003.

37. I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, (in press), 2009.

38. T. Traut. *Allosteric Regulatory Enzymes*. Springer Verlag, 2007.

39. A. K. White and W. W. Metcalf. Microbial metabolism of reduced phosphorus compounds. *Annu Rev Microbiol*, 61:379–400, 2007.

40. J.C. Wilks and J. L. Slonczewski. pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol*, 189(15):5601–5607, Aug 2007.

41. M. Willemos and B. Hove-Jensen. Binding of divalent magnesium by *Escherichia coli* phosphoribosyl diphosphate synthetase. *Biochemistry*, 36(16):5078–5083, Apr 1997.

42. G. R. Willsky and M. H. Malamy. Characterization of two genetically separable inorganic phosphate transport systems in *Escherichia coli*. *J Bacteriol*, 144(1):356–365, Oct 1980.

43. G. Zhao, A. J. Pease, N. Bharani, and M. E. Winkler. Biochemical characterization of gapb-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* k-12 and its possible role in pyridoxal 5'-phosphate biosynthesis. *J Bacteriol*, 177(10):2804–2812, May 1995.