Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism

R.M.T. Fleming*
Department of Biochemistry,
National University of Ireland, Galway

I. Thiele
Bioinformatics Program,
University of California, San Diego

G. Provan
Department of Computer Science,
University College Cork

H.P. Nasheuer
Department Biochemistry,
National University of Ireland, Galway

26th August 2009

*Corresponding author. Present address: Science Institute, University of Iceland, Reykjavik, Iceland. Tel: +354 618 6246

Abstract

The quantitative analysis of biochemical reactions and metabolites is at frontier of biological sciences. The recent availability of high-throughput technology data sets in biology has paved the way for new modeling approaches at various levels of complexity including the metabolome of a cell or an organism. Understanding the metabolism of a single cell and multi-cell organism will provide the knowledge for the rational design of growth conditions to produce commercially valuable reagents in biotechnology. Here, we demonstrate how equations representing steady state mass conservation, energy conservation, the second law of thermodynamics, and reversible enzyme kinetics can be formulated as a single system of linear equalities and inequalities, in addition to linear equalities on exponential variables. Even though the feasible set is non-convex, the reformulation is exact and amenable to large-scale numerical analysis based on linear algebra, a prerequisite for computationally feasible genome scale modeling. Integrating flux, concentration and kinetic variables in a unified constraint-based formulation is aimed at increasing the quantitative predictive capacity of flux balance analysis. Incorporation of experimental and theoretical bounds on thermodynamic and kinetic variables ensures that the predicted steady state fluxes are both thermodynamically and biochemically feasible. The theoretical approach outlined is implemented for a genome scale model of *E. coli* metabolism and the current numerical challenges are discussed.

Key words: systems biology; constraint-based modelling; linear polytope; exponential polytope; algebraic geometry;

1 Introduction

It has long been appreciated that biological systems are subject to physico-chemical laws. These laws impose certain constraints on the feasible set of enzymatic reaction rates and metabolite concentrations. Within the physico-chemically feasible set, the actual values of evolved kinetic parameters will determine the functional state of the cell. In the absence of experimentally determined kinetic parameters, it is still possible to mathematically formulate and computationally implement systems of equations, the solutions to which represent a physico-chemically feasible set of functional states. Ideally, one seeks to incorporate all relevant physico-chemical constraints in a single framework. This brings a model closer to our conception of reality, reduces the size of the feasible set, and is aimed at increasing the predictive capacity of such models. However, it is not always computationally tractable to apply all known physical laws since the resulting equations are difficult to solve. This is especially so for genome scale models where there are typically a large number of equations and variables.

Flux balance analysis is a standard constraint-based modelling approach which is based the assumption that mass is conserved in a biochemical system of reactions (37). In addition, within ones timescale of interest, often one can reasonably assume that the concentrations of metabolites are time invariant, that is, the system is in a steady state. These two constraints define a convex polytope of feasible reaction fluxes and metabolite concentrations are not explicitly represented. Herein we derive a formalism where reaction flux to

obeys mass conservation, energy conservation, the second law of thermodynamics, and mass action kinetics. The resulting constraint equations define a non-convex feasible set, but now metabolite concentration and kinetic parameters are explicitly represented. We also demonstrate the application of this modelling approach with an example using the genome scale *E. coli* metabolic model. In addition, theoretically we illustrate how experimental knowledge of physiologically feasible ranges and theoretical biochemical predictions of feasible ranges of other pertinent variables may be incorporated into this new modelling framework (21). Natural selection has to operate within the constraints enforced by physical laws. By integration of genome scale stoichiometric, thermodynamic and kinetic constraints, in a framework amenable to numerical analysis, we aim to investigate *in silico*, the mechanisms responsible for the results of adaptive evolution experiments. This is the ultimate goal, but there are significant mathematical and computational modelling challenges which need to be addressed before these biologically more pertinent issues can be addressed.

Stoichiometric models of metabolism continue to grow in size due to new models of higher organisms (46) and incremental extension of existing models (20). As the dimensions of the stoichiometric matrix increases, computational complexity theory dictates that some computational analysis techniques no longer become practical due to memory or time constraints (48). This problem is exemplified by vertex enumeration of convex polytope where no known polynomial time algorithm exists (3). Vertex enumeration is the key step in elementary mode, or extreme pathway analysis (37). Unless significant advances are made in reducing the complexity of vertex enumeration, one cannot hope to utilise this analysis method for the latest genome scale models (20). An additional problem is that existing methods of low complexity may be numerically unstable for large matrices with coefficients spread over many orders of magnitude. The application of less conventional linear algebra to design a more numerically stable, yet low complexity algorithm can help to alleviate such problems (43).

In tandem with growth of stoichiometric matrices, there are efforts to increase their predictive capacity by incorporating additional constraints, beyond mass conservation with a steady state assumption. The addition of Boolean genetic regulatory constraints, to flux balance models of *E. coli* metabolism, increases their qualitative predictive capacity when compared to experimental data of mutant growth capability on various media (14). Predictions of qualitative growth capability or quantitative growth rate (28) are interpretations of the numerical values of exchange fluxes from flux balance analysis of stoichiometric models. It is not possible to reliably predict internal fluxes in genome scale metabolic models without additional thermodynamic constraints. The simple fact is that stoichiometrically balanced loops are invariably present in such networks and, with only stoichiometric constraints, they exhibit cyclic net fluxes entirely independent of exchange fluxes. Energy conservation, and the second law of thermodynamics, can be combined with stoichiometric constraints to retain only mass balanced and thermodynamically feasible fluxes (6). However, although such fluxes obey mass conservation, energy conservation and the second law of thermodynamics, additional constraints are required to make them biochemically feasible.

Standard chemical potential forms a bridge between metabolite concentration and metabolic flux in thermodynamically constrained stoichiometric models. Experimentally, one can measure standard chemical potential and combine this data with stoichiometric models. This approach has shown promise in constraining viable fluxes and assisting interpretation of metabolome data (32). Theoretically, one can estimate standard chemical potential using

group contribution methods (34, 24, 31). The group contribution method uses knowledge of the chemical energy within recurring chemical bonds and enumerates the prevalence of these bonds, within a compound of known structure, to estimate the entire standard chemical potential. Using this method, it is possible to estimate the standard chemical potential of almost all compounds in *E. coli* (20). These estimated results compare favourably with experimental measures, provided that the associated uncertainty in estimation is acknowledged (23, 31). Experimental and theoretical approaches are desirable and complimentary, since it is laborious and difficult to measure all the values experimentally whereas estimates need experimental confirmation.

Using standard chemical potential and metabolite concentration, one can place upper and lower bounds on the *in vivo* change in chemical potential of almost all reactions in genome scale stoichiometric models of metabolism (20). Given a stoichiometric model where all reactions were initially reversible, quantitative bounds on change in chemical potential, in combination with mass conservation, energy conservation and second law of thermodynamics, constrain the feasible steady state fluxes to thermodynamically feasible and physiologically relevant solutions. Without such bounds on chemical potential, the set of thermodynamically feasible solutions could include degradation of biomass into external nutrient metabolites, the opposite of the more probable biomass synthesis direction. Maximising bacterial growth, constrained by mass balance, energy conservation and the second law of thermodynamics, integrated with standard chemical potential and metabolite concentration estimates, has recently been shown to significantly improve the correspondence between computed and experimental internal flux rates in *E. coli* (26).

With the size of stoichiometric models increasing and simultaneously the scope of constraints expanding a pressing problem is to find a mathematical formalism which encompasses all current constraints in a concise form yet lends itself to robust, scalable numerical analysis. In addition, it is desirable to use a form which is invariant to the addition of new, possibly non-linear, constraints. In particular, efforts are under way to collate the kinetic equations and parameters of enzymes in model organisms (38), including *E. coli*. Building on earlier work, herein we present a mathematical formulation which includes steady state mass conservation, energy conservation, the second law of thermodynamics, and extends to encompass a system of reversible kinetic equations where bounds on standard chemical potential, metabolite or enzyme concentration and kinetic variables can be incorporated. We refer to kinetic variables rather than kinetic parameters to emphasise that this is a constraint-based approach to steady state kinetics where the variables are not exactly specified, but constrained to a bounded interval. The resulting system of simultaneous linear equations in linear variables, combined with linear equations in logarithmic variables, shows a strong mathematical similarity with the system of equations studied in chemical reaction equilibrium analysis (45, 42). The outline of the mathematical description below mimics the order of the aforementioned constraints. During the description we highlight salient mathematical features and use simple toy metabolic examples for illustration of key concepts. Initial efforts to apply this modelling approach to a genome scale model of *E. coli* metabolism is also illustrated.

2 Mass conservation at steady state

The set of chemical reactions in a metabolic network can be represented as a set of chemical equations. Embedded in these chemical equations is information about reaction stoichiometry. Stoichiometry represents the number of molecules of reactant consumed, and product produced, in a single chemical reaction. Since elements, Carbon, Nitrogen, Oxygen etc, are neither created nor destroyed in a biochemical reaction, the total number of each element, or moiety, is conserved even though their relative arrangements may change as old chemical bonds are broken and new ones are formed. All of the stoichiometric information about a metabolic network can be represented within a matrix, the stoichiometric matrix, $\mathbf{S} \in \mathbb{Z}^{m,n}$, $i \in 1 \dots m$, $j \in 1 \dots n$, where m is the number of metabolites and n is the number of reactions (e.g. Supplementary Material A). Each elementary reversible chemical transformation is represented by a pair of conjugate columns of the stoichiometric matrix, one forward elementary reaction and one reverse elementary reaction, odd columns, j, and even columns, $j+1$, respectively. Each row represents the participation of a single metabolite in all possible reactions. An integer coefficient in a column of a stoichiometric matrix gives the number of molecules of a single metabolite consumed or produced in that reaction. The stoichiometric matrix linearly transforms a flux vector, $\mathbf{v} \in \mathbb{R}^n_{>0}$, into a vector of metabolite concentration time derivatives, $\frac{d\mathbf{m}}{dt}$, giving the fundamental equation of dynamic mass conservation (27), which, with the assumption of steady state concentration, gives

$$\mathbf{S} \cdot \mathbf{v} = \frac{d\mathbf{m}}{dt} \equiv \mathbf{0} \tag{1}$$

The steady state assumption asserts that the concentration of metabolites is time invariant. Since Eq. 1 is generally under-determined, $rank(\mathbf{S}) < m < n$, there are many flux vectors which satisfy the steady state mass conservation constraint.

3 Energy conservation and the second law of thermodynamics

Theoretical application of thermodynamic principles to chemical networks in terms of differential geometry, algebraic topology and circuit theory were considered by Oster and Perelson in the 1970's (35, 36). This work remains to fully impact systems biology, most likely because of a dearth of advanced mathematical training amongst the systems biology community. When applied to biochemical networks, energy conservation, and the second law of thermodynamics, prevent net flux around stoichiometrically balanced cycles within metabolic networks, as emphasised by Beard, Liang & Qian (6). A finer point to note here is the implicit assumption that thermodynamically unfavourable biochemical transformations cannot be driven by a gradient of temperature, this is may not be the case for endothermic chemotrophes (33). Given the entire set of elementary reactions in a metabolic network, a stoichiometrically balanced cycle is a subset of contiguous reactions that form a subnetwork of chemical transformations that are perfectly mass balanced, without a member of the subset being an exchange reaction. An example of such a stoichiometrically balanced cycle is given in Figure 1. By the second law of thermodynamics, net reaction flux from reactants to

products requires a concomitant thermodynamic driving force, provided by a drop in chemical potential from reactants to products. In Figure 1, if A has a greater chemical potential than B, and B greater than C, then by the principal of conservation of energy, C must have lower chemical potential than A. By the second law of thermodynamics, the *net flux* is in the direction of decreasing chemical potential, therefore there must be net flux in the direction $A \rightarrow C$. Therefore, there is no *net flux* around the cycle $A \rightarrow B \rightarrow C \rightarrow A$. Many more complicated stoichiometrically balanced cycles exist in genome scale models, to the extent that they become unidentifiable by sight alone and innumerable using vertex enumeration algorithms (3) in the latest genome scale models for *E. coli* (20).

Quantitative thermodynamic constraints prevent net flux around stoichiometrically balanced loops but it is possible to have no net flux around a loop yet still violate thermodynamic constraints. Consider Figure 1, if A is in equilibrium with B and B is in equilibrium with C then by the second law of thermodynamics, there is no net flux $A \rightarrow B \rightarrow C$ and hence no net flux around the cycle. In flux balance analysis models, net flux between $A \rightarrow C$ is still possible even if there is no net flux $A \rightarrow B \rightarrow C$. However, net flux $A \rightarrow C$ violates energy conservation since A must be in equilibrium with C, when A is in equilibrium with B and B is in equilibrium with C. Conservation principles and model consistency from different perspectives are two sides of the same coin.

Let us consider the first $\widetilde{n}$ columns of the stoichiometric matrix corresponding exclusively to internal reactions, denoted $\widetilde{\mathbf{S}} \in \mathbb{Z}^{\mathbf{m},\tilde{\mathbf{n}}}$. If one computes a rational basis, $\mathbf{K} \in \mathbb{Z}^{\tilde{\mathbf{n}},\mathbf{k}}$, for the null space of $\widetilde{\mathbf{S}}$, then $\mathbf{K}$ spans the vector space of stoichiometrically balanced loops at steady state and, by the definition of a null space, $\widetilde{\mathbf{S}} \cdot \mathbf{K} \equiv \mathbf{0}$. Multiplying across by a vector of metabolite chemical potentials, $\mu \in \mathbb{R}^{1,m}$, gives $\mu \cdot \widetilde{\mathbf{S}} \cdot \mathbf{K} = \mathbf{0}$. Since $\mu \cdot \tilde{\mathbf{S}} = \triangle\widetilde{\mu}$, the vector of change in chemical potential for internal reactions, by substitution then transposing, one arrives at

$$\mathbf{K}^{\mathbf{T}} \cdot \triangle\widetilde{\mu}^{\mathbf{T}} = \mathbf{0} \tag{2}$$

This is an invariant, first derived by Beard and Qian (6), on total change in chemical potential at steady state for each stoichiometrically balanced loop and linear combinations thereof. Biochemically it enforces conservation of chemical potential energy by internal reactions of a stoichiometric network. By the second law of thermodynamics, net flux is with the direction of drop in chemical potential so this constraint is required in addition to energy conservation, in order to eliminate flux around these loops.

The net steady state change in chemical potential with respect to the forward reaction, $\Delta\widetilde{\mu}_j$, is given by the logarithm of the ratio of reverse divided by forward flux, scaled by temperature, T, and the gas constant, R,

$$\Delta\widetilde{\mu}_j = RT \ln\left(\frac{v_{j+1}}{v_j}\right) = -\Delta\widetilde{\mu}_{j+1} \tag{3}$$

Here we assign the odd index j to the forward reaction, and even index $j+1$ to the reverse reaction. The change in chemical potential with respect to the forward direction is equal to the change in chemical potential with respect to the reverse reaction if one changes its sign. Eq. 3 ensures that net flux is in the direction of a drop in chemical potential energy, in accordance with the second law of thermodynamics. Assuming the existence of a chemical potential for each metabolite, mass action kinetics, constant temperature and pressure,

within a finite volume element, this thermodynamic relationship applies to elementary reversible reactions arbitrarily far from equilibrium (39). Obviously, at equilibrium, the rate of the forward and reverse reactions is equal so there is no change in chemical potential. The connection between Eq. 3 and the work of Hill on catalytic cycle fluxes (25), Ussing, Hodgkin and Huxley on ion transport, and Crooks on entropy production in microscopically reversible systems (15) has recently been discussed (7). It is fair to say that the exact validity of this relation, beyond the assumption of an elementary reaction, is still an open issue.

By considering all chemical transformations to be reversible, rather than approximating certain steps with 'irreversible' reactions, with Eq. 3, we see that assigning zero flux to a reverse reaction dictates that the change in chemical potential, for the 'irreversible' forward reaction, should be negative infinity. Such a situation is not physically possible as all changes in chemical potential are finite. This is a difference between the kinetic studies which routinely assume irreversibility and thermodynamic studies which can not. Given Eq. 2, by substituting an appropriately signed vector of equations 3 in place of $\triangle\tilde{\mu}$, after some algebraic manipulation we derived the novel equation

$$\mathbf{P} \cdot \ln(\widetilde{\mathbf{v}}) = \mathbf{0} \tag{4}$$

where $\mathbf{P} \in \mathbb{Z}^{\mathbf{k},\tilde{\mathbf{n}}}$ and $\ln(\tilde{\mathbf{v}})$ denotes the component-wise natural logarithm of each flux in the vector of fluxes. (See Supplementary Material B for details an example). Eq. 4 simultaneously applies energy conservation and the second law of thermodynamics as constraints on logarithmic elementary flux. One advantage of this novel reformulation is that it eliminates thermodynamically infeasible fluxes without dependence on *a priori* knowledge of change in chemical potential, cf. (6). This is important since standard chemical potential is unknown experimentally for certain metabolites and accurate theoretical estimates are difficult because the metabolites contain substructures for which no group contribution estimate is known (31). Another advantage is that it explicitly considers elementary forward and reverse flux, rather than net flux, which is important in isotopomer analysis (44).

4 Kinetic constraints

The vast majority of kinetic parameters available in the literature are inappropriate for models of *in vivo* metabolism because they are byproducts of enzyme mechanistic studies carried out at temperature and pH far from *in vivo* conditions (13). However, mechanistic enzyme classification also indicates a specific mathematical form of kinetic equation for an enzyme in accordance with its mechanism. Different enzyme classes, with distinct kinetic equations, specify different potential relationships between their variables independent of actual numerical values for kinetic variables. For example, consider two bisubstrate reactions, one with a sequential ordered mechanism versus one with a ping pong mechanism (11). Since their rate equations are different, the relationship between the possible values of their variables are different. Just as the stoichiometric matrix constrains flux distributions, kinetic equations constrain distributions of kinetic variables, metabolite and enzyme concentrations at steady state. Since flux is a function of kinetic variables, metabolite and enzyme concentration, then kinetic equations implicitly constrain flux distributions. The kinetic equations for many *E. coli* enzymes are available in the literature and, as prerequisite for genome scale modelling,

are in the process of being collated in a curated database (38). Even in cases where the rate equation is unknown, one can approximate the kinetics with a simple mass action rate law.

4.1 Phenomenological versus elementary kinetic variables

Reversible kinetic equations typically appear in one of two mathematical forms depending on whether phenomenological or elementary kinetic variables are used. Phenomenological variables are microscopic properties of enzymes (18) but are more easily measured in macroscopic experiments with ensembles of enzymes (12), whereas elementary kinetic variables correspond to discrete approximations of major conformational transitions and are much more difficult to measure experimentally. Consider one of the isomerase reactions from Figure 1 modelled as a three step kinetic mechanism

$$A + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} EA \underset{k_4}{\overset{k_3}{\rightleftharpoons}} EB \underset{k_6}{\overset{k_5}{\rightleftharpoons}} E + B \tag{5}$$

with associated elementary reaction kinetic variables, k. Reactions are modelled stoichiometrically as a one step mechanism as it is the net stoichiometry of the entire reaction which acts as a constraint. The phenomenological, reversible, Michaelis-Menten kinetic equation for mechanism 5 is

$$v_{net} = v_j - v_{j+1} = \frac{k_A e_{AB} a - k_P e_{AB} b}{1 + \frac{a}{K_{mA}} + \frac{b}{K_{mB}}} \tag{6}$$

where v_{net} is the net flux through the reaction, a and b are substrate and product concentrations, and e_{AB} is the total enzyme concentration. Specificity constants, k_A and k_B, reflect the specificity of the enzyme for A or B in the presence of competing substrates. Michaelis constants, K_{mA} and K_{mB}, are apparent dissociation constants. Phenomenological kinetic variables are functions of overlapping sets of elementary kinetic variables so they are not independent. In the case of reaction 5

$$\begin{array}{ll} k_A = \frac{k_1 k_3 k_5}{k_2 k_4 + k_2 k_5 + k_3 k_5} & K_{mA} = \frac{k_2 k_4 + k_2 k_5 + k_3 k_5}{k_1 (k_{-2} + k_2 + k_3)} \\ k_P = \frac{k_2 k_4 k_6}{k_2 k_4 + k_2 k_5 + k_3 k_5} & K_{mP} = \frac{k_2 k_4 + k_2 k_5 + k_3 k_5}{(k_2 + k_4 + k_3) k_6} \end{array} \tag{7}$$

These functions can be viewed as implicit polynomial equations relating experimentally measurable quantities to quantitatively modelled kinetic variables, see Supplementary Material B. This facilitates the incorporation of appropriate experimental kinetic data (38) or theoretical predictions on optimal enzyme characteristics (21).

4.2 Thermodynamic Haldane equation

For a reversible reactions, with ordered binding and dissociation of substrates and products, the thermodynamic Haldane equation (11) relates standard change in chemical potential, $\Delta\mu_j^o$, to a ratio of forward and reverse elementary rate constants

$$\Delta\mu^o = -RT \ln \left(\frac{\prod k_+}{\prod k_-} \right) \tag{8}$$

For reversible reactions where the kinetic mechanism is not ordered, such as the 'Ping Pong Bi Bi' mechanism (11), or random, such as the 'Rapid Equilibrium Random Uni Bi Bi' mechanism, the thermodynamic Haldane is no longer simply the ratio of two monomials formed from the forward and reverse rate constants (11). Instead, the thermodynamic Haldane is a more complicated polynomial. In such a situation one may use a transformation from polynomial to linear, linear-logarithmic equations in order to incorporate the thermodynamic Haldane equation into the system of constraint equations. The transformation from polynomial to linear, linear-logarithmic equations is discussed in section 4.3.

For ordered kinetic mechanisms, given the logarithmic identity

$$\ln(ab) = \ln(a) + \ln(b) \tag{9}$$

the relation between standard change in chemical potential for an elementary reaction, and a single row of the stoichiometric matrix, $\mu^o \cdot \mathbf{S}_j = \Delta\mu_j^o$, one can rewrite Eq. 8 in linear homogeneous form in terms of logarithmic elementary kinetic parameters

$$\frac{\mu^o \cdot \mathbf{S}_j}{RT} + \sum \ln(k_+) - \sum \ln(k_-) = 0 \tag{10}$$

Thus, standard chemical potential, $0 < \mu^o \in \mathbb{R}^{1,m}$, can be used to linearly constrain the relationship between logarithmic elementary kinetic constants. Generalised to a system of reactions, in matrix form, we have

$$\left[\begin{array}{ccc} \widetilde{\mathbf{S}^T} & \mathbf{I} & -\mathbf{I} \end{array} \right] \cdot \left[\begin{array}{c} \ln(\mathbf{u}^o) \\ \ln(\mathbf{k}_+) \\ \ln(\mathbf{k}_-) \end{array} \right] = \mathbf{0} \tag{11}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{u}_i^o \equiv \exp(\frac{\mu_i^o}{RT})$, $0 < \mathbf{u}^o \in \mathbb{R}^m$, and $\mathbf{k}_+ \in \mathbb{R}^{\widetilde{n}}$ and $\mathbf{k}_- \in \mathbb{R}^{\widetilde{n}}$ are vectors of elementary kinetic constants, ordered appropriately according to each internal reaction. An identity matrix is a square matrix with all zero coefficients, except for the diagonal which consists of unitary coefficients. Eq. 11 constrains the kinetic parameters of the model to be consistent with thermodynamics, a constraint often overlooked in kinetic models (17) and likewise in numerical sampling of candidate values for kinetic parameters (19).

4.3 Kinetic rate equations

For reversible reactions, elementary kinetic variable rate equations express net flux as a function of elementary kinetic variables, metabolite concentrations and enzyme concentration. By splitting a reversible rate equation, into separate unidirectional forward and reverse flux equations, we have two polynomial equations relating elementary flux, metabolite and enzyme concentrations. For an entire metabolic network, all elementary unidirectional rate equations, and the relations between phenomenological and elementary kinetic variables, form a system of implicit homogeneous polynomials. We demonstrate how any system of polynomial equations can be algebraically transformed into a conjugate set of linear equalities, logarithmic equalities and logarithmic inequalities. This conjugate linear-logarithmic system

has the same mathematical form as the aforementioned linear-logarithmic flux constraints representing mass balance and thermodynamic feasibility.

Consider a d dimensional system of m homogeneous polynomials, with n distinct monomials. Here m is the number of metabolites + enzymes in a system, whilst d is the number of unidirectional reactions + kinetic constants + metabolites + enzymes. A distinct monomial is a single term in a polynomial ignoring coefficients. e.g. ignoring coefficients, $6x^2y^3$ and $-4x^2y^3$ are indistinct monomials. The number of distinct monomials depends on the type of polynomial kinetic expression used to model each reaction (e.g. Supplementary Material D). Any system of polynomials, $i \in 1...m$, can be represented, without algebraic manipulation, in the form

$$\begin{aligned} \sum_{j=1}^{n} \mathbf{A}_{i,j} \left(\prod_{r=1}^{d} \mathbf{y}_r^{\mathbf{Q}_{j,r}} \right) &= 0 \\ \vdots &= 0 \\ \sum_{j=1}^{n} \mathbf{A}_{m,j} \left(\prod_{r=1}^{d} \mathbf{y}_r^{\mathbf{Q}_{j,r}} \right) &= 0 \end{aligned} \tag{12}$$

where $\mathbf{A}_{i,j}$ is the coefficient of monomial j in polynomial i, $\mathbf{y}_r$ is a variable and $\mathbf{Q}_{j,r}$ is a an exponent of variable $\mathbf{y}_r$ within distinct monomial j. The coefficients are integers in the matrix $\mathbf{A} \in \mathbb{Z}^{m,n}$. $0 < \mathbf{y} \in \mathbb{R}^d$ since it is non-physical to consider variables or concentrations of zero or less. Each row of $\mathbf{Q} \in \mathbb{N}^{n,d}$ corresponds to one distinct monomial and each column of $\mathbf{Q}$ corresponds to one variable.

Now we arrive at the first key algebraic step. Let $\mathbf{y}_r \equiv \exp(\mathbf{z}_r)$, then, by the exponential identities, $(e^a)^b = e^{ab}$ and $e^a e^b = e^{a+b}$, we have

$$\prod_{r=1}^{d} \mathbf{y}_r^{\mathbf{Q}_{j,r}} = \prod_{r=1}^{d} \exp(\mathbf{z}_r)^{\mathbf{Q}_{j,r}} = \prod_{r=1}^{d} \exp(\mathbf{Q}_{j,r}\mathbf{z}_r) = \exp(\mathbf{Q}_j \cdot \mathbf{z})$$

If we consider $\exp(\mathbf{Q}_j \cdot \mathbf{z})$, $j \in 1...n$, as one component of an n dimensional vector then Eq. 12 condenses to become

$$\mathbf{A} \cdot \exp(\mathbf{Q} \cdot \mathbf{z}) = \mathbf{0} \tag{13}$$

where $z \in \mathbb{R}^{\mathbf{d}}$. This prepares the way for the second key step.

In the field of algebraic geometry, in particular, the study of algebraic varieties, which are geometric manifestations of solution sets of systems of polynomial equations, the matrix $\mathbf{Q}$ composed of exponents of distinct monomials, leads to the study of mathematical structures known as Newton polytopes (22). A polytope is a generalisation of a polygon to arbitrary dimension. To construct a Newton polytope, one considers the exponents of a single distinct monomial, a row in $\mathbf{Q}$, as a vertex of this polytope. A vertex is a generalisation of a corner, to arbitrary dimension. The set of rows $j \in 1...n$ of $\mathbf{Q}$ form such a set of vertices that define a polytope. It is a fundamental result from convex geometry that there are two equivalent ways to represent a polytope, in terms of a set of vertices, a vertex representation, or terms of a set of linear equations and inequalities, a half-space representation. Facet enumeration, using a convex hull algorithm, can transform a vertex representation into a half space representation.

The Newton polytope of a polynomial system is the convex hull of $\mathbf{Q}$ when one considers each row as a vertex. We invite the interest of the algebraic geometry community at this stage as we depart from the usual perspective and instead consider each column of $\mathbf{Q}$ as the vertex of a polytope and use a vertex enumeration algorithm to compute the 'dual Newton polytope'.

Since the polynomials we consider are homogeneous, no entire row of $\mathbf{Q}$ is devoid of nonzero coefficients. $\mathbf{Q}_{j,r} = 0$ when a monomial indexed j has a variable indexed r with a zero exponent since $\mathbf{y}_r^0 = 1$. Since $\mathbf{Q}_{j,r} \in \mathbb{N}$ we consider each column of $\mathbf{Q}$, which is formed from the exponents of a single variable $\mathbf{y}_r$ across all n distinct monomials of the polynomial system, as a vertex of a polytope in the positive quadrant. Using facet enumeration we can compute the convex hull of $\mathbf{Q}$ and derive an equivalent half-space representation of the dual Newton polytope as a system of affinely independent inhomogeneous linear equalities and inequalities with new matrices $\mathbf{B} \in \mathbb{Z}^{b,n}$, $\mathbf{C} \in \mathbb{Z}^{c,n}$, and vectors $\mathbf{b} \in \mathbb{R}^b$, $\mathbf{c} \in \mathbb{R}^c$, which by the definition of $\mathbf{Q}$ as a set of vertices of this 'dual Newton polytope' satisfy the equations

$$\mathbf{B} \cdot \mathbf{Q} \cdot \mathbf{z} \equiv \mathbf{b} \tag{14}$$

$$\mathbf{C} \cdot \mathbf{Q} \cdot \mathbf{z} > \mathbf{c} \tag{15}$$

i.e. the new matrices $\mathbf{B}$ and $\mathbf{C}$, with vectors $\mathbf{b}$ and $\mathbf{c}$ are defined by the halfspace representation of the same polytope with vertex representation defined by the set of vertices in$\mathbf{Q}$. Since the matrix $\mathbf{Q}$ is composed of exponents of distinct monomials arising from distinctNow, let $\mathbf{Q} \cdot \mathbf{z} \equiv \ln(\mathbf{x})$, then Eq. 13 becomes

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{0} \tag{16}$$

and equations 14 and 15 become

$$\mathbf{B} \cdot \ln(\mathbf{x}) = \mathbf{b} \tag{17}$$

$$\mathbf{C} \cdot \ln(\mathbf{x}) > \mathbf{c} \tag{18}$$

Intuitively, if one considers Eq. 16 as a polytope in linear scale, and Eqs. 17 and 18 as a polytope in logarithmic scale, then the simultaneous solution to equations 16, 17 and 18 is the intersection of a polytope in linear scale with the exponential of a logarithmic scale polytope. This concept is illustrated in Figure 2.

Since we let $\mathbf{y}_r \equiv \exp(\mathbf{z}_r)$ and $\mathbf{Q} \cdot \mathbf{z} \equiv \ln(\mathbf{x})$, it is consistent to relate the new vector $\mathbf{x}$ to the original vector $\mathbf{y}$ with $\mathbf{Q} \cdot \ln(\mathbf{y}) = \ln(\mathbf{x})$. This relation can be equivalently expressed as

$$\begin{bmatrix} \mathbf{I} & -\mathbf{Q} \end{bmatrix} \cdot \begin{bmatrix} \ln(\mathbf{x}) \\ \ln(\mathbf{y}) \end{bmatrix} = \mathbf{0} \tag{19}$$

where $\mathbf{I} \in \mathbb{Z}^{n,n}$ is an identity matrix. If the number of different monomials is greater than the number of dimensions of the polynomial system, $\mathbf{Q} \in \mathbb{N}^{\mathbf{n},\mathbf{d}}$, $n > d$, then $\mathbf{Q} \cdot \ln(\mathbf{y}) = \ln(\mathbf{x})$ is overdetermined. In this case, not every $\mathbf{x}$ will necessarily correspond to a solution $\mathbf{y}$. However, since Eq. 19 is the of same mathematical form as 17 we can incorporate it into the system of constraints to guarantee a solution to $\mathbf{y}$ for each $\mathbf{x}$. The output of the transformation from a system of polynomial equations in $\mathbf{y}$, using exponential identities and facet enumeration, is a set of linear equations in linear variables, plus a set of equalities and inequalities in $\ln(\mathbf{x})$ and $\ln(\mathbf{y})$. In the case of a polynomial system of kinetic equations, $\mathbf{y}$

encompasses external fluxes, $\overline{\mathbf{v}}$, internal fluxes, $\widetilde{\mathbf{v}}$, forward and reverse elementary kinetic variables, $\mathbf{k}_+$ and $\mathbf{k}_-$, metabolite concentrations, $\mathbf{m}$, and enzyme concentrations, $\mathbf{e}$. i.e. $\mathbf{y} = [\ \overline{\mathbf{v}},\ \widetilde{\mathbf{v}},\ \mathbf{k}_+,\ \mathbf{k}_-,\ \mathbf{m},\ \mathbf{e}\ ]^\top \in \mathbb{R}^{\frac{5n}{2}+m}$. Here we assume a unique enzyme for each reaction, so the dimension of $\mathbf{y}$ will be lower if the same enzyme catalyses multiple reactions. Thus we have the combined system of m linear equalities on linear variables, $b+n$ linear equalities on logarithmic variables, and c linear inequalities on logarithmic variables.

$$\mathbf{A}\cdot\mathbf{x} = \mathbf{0} \tag{20}$$

$$\begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I} & \mathbf{Q}_{\overline{v}} & \mathbf{Q}_{\widetilde{v}} & \mathbf{Q}_{k_+} & \mathbf{Q}_{k_-} & \mathbf{Q}_m & \mathbf{Q}_e \end{bmatrix} \cdot \begin{bmatrix} \ln(\mathbf{x}) \\ \ln(\overline{\mathbf{v}}) \\ \ln(\widetilde{\mathbf{v}}) \\ \ln(\mathbf{k}_+) \\ \ln(\mathbf{k}_-) \\ \ln(\mathbf{m}) \\ \ln(\mathbf{e}) \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \tag{21}$$

$$\mathbf{C}\cdot\ln(\mathbf{x}) \geq \mathbf{c} \tag{22}$$

where the various subscripts to $\mathbf{Q}$ denote the columns appropriate to the variables indicated. A detailed example of the transformation is given in Supplementary Material D. In the case of simple mass action kinetics, where flux is a function of a single monomial, then vertex enumeration is not required to exactly linearise the equations in terms of logarithmic variables, see Supplementary Material E.

5 Thermodynamic and kinetic inequalities

Broad bounds on metabolite or enzyme concentrations, fluxes or phenomenological kinetic variables, are available from theoretical consideration of intracellular kinetics (21) guided in specific cases by experimental data. In all cases, whether due to lack of knowledge or inherent errors in experimental determination, such data results in a pair of bounding inequalities on the variable of interest. For example, since estimates of standard chemical potential come with a margin of uncertainty (23), these constraints are represented as inequalities on the allowed range of standard chemical potential

$$\mu^o_{min} \leq \mu^o \leq \mu^o_{max} \tag{23}$$

Likewise, broad estimates on physiologically realistic intracellular metabolite concentrations (20) lead to similar bounding inequalities. When experimental metabolite concentrations are available this data leads to tighter bounds on the model. Model representation of inter-compartmental concentration gradients is essential to permit net diffusive flux of transport reactions. Such considerations, beyond the realm of flux balance analysis, are vital to permit *in silico* growth of more realistic models which include potential gradients, in addition to fluxes. Of particular importance is the difference in concentration of H^+ ions between separate compartments of the electrochemical chain (23). This electron-motive force can be represented in compartmentalised, charge balanced, constraint based models (20) when experimental data on intra- and periplasmic pH is included.

Establishing tight bounds on genome wide enzyme concentrations is a significant problem. Given the high level of noise in both microarray (29) and quantitative proteomics data (16) it is difficult to make broad conclusions on the relationship between transcript and protein abundance. If one assumes enzyme concentration is proportional to transcript abundance in global gene expression studies(14), then enzyme concentration can be quantitatively estimated through protein molecular weight, estimates of total dry weight of protein per cell and thermodynamically grounded preprocessing of raw microarray data (9). At best, such effort with experimental data gives rise to loose bounds on intracellular enzyme concentration for a particular nutrient medium appropriate to the microarray data (8). However, this may be important if the upper bound is so close to zero as to render flux through that particular enzyme inconsequential with respect to the flux of active pathways. As such, detecting the presence versus absence of a transcript is a crucial output of microarray preprocessing (10). As experimental systems biology continues its rapid progress one can expect tightening of these inequalities.

6 Integrated stoichiometric, thermodynamic and kinetic constraint-based modelling

The rationale for the reformulation of constraints becomes apparent when one compares the mathematical form of the resulting equations. Eq. 1, the steady state mass conservation invariant and Eq. 20, the linear portion of the reformulated kinetic constraints are of the same mathematical form and together can be represented by

$$\begin{bmatrix} A & 0 & 0 \\ \mathbf{0} & \overline{S} & \widetilde{S} \end{bmatrix} \cdot \begin{bmatrix} x \\ \overline{v} \\ \widetilde{v} \end{bmatrix} = \mathbf{0} \tag{24}$$

where $\overline{\mathbf{S}}$ and $\widetilde{\mathbf{S}}$ denote the external and internal columns of the stoichiometric matrix. Equation 4, the novel compact reformulation of energy conservation and the second law of thermodynamics, Eq. 11, the linear reformulation of the thermodynamic Haldane in terms of logarithmic kinetic variables and Eq. 21, the logarithmic equality portion of the reformulated kinetic constraints, are all linear forms in logarithms and together can be represented by

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{P} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \widetilde{\mathbf{S}^T} \\ \mathbf{B} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I} & \mathbf{Q}_{\overline{v}} & \mathbf{Q}_{\widetilde{v}} & \mathbf{Q}_{k_+} & \mathbf{Q}_{k_-} & \mathbf{Q}_m & \mathbf{Q}_e & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \ln(\mathbf{x}) \\ \ln(\overline{\mathbf{v}}) \\ \ln(\widetilde{\mathbf{v}}) \\ \ln(\mathbf{k}_+) \\ \ln(\mathbf{k}_-) \\ \ln(\mathbf{m}) \\ \ln(\mathbf{e}) \\ \ln(\mathbf{u}^o) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ b \\ 0 \end{bmatrix} \tag{25}$$

Experimental data or theoretical estimates of kinetic or thermodynamic variables give rise to bounding inequalities such as 23. By taking the logarithm, splitting paired bounding inequalities into two inequalities and reversing the sign of the upper bound inequality, we

have the same form as Eq. 22, the logarithmic inequality portion of the reformulated kinetic constraints. Combined they are of the form

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \ln(\mathbf{x}) \\ \ln(\mathbf{u}) \end{bmatrix} > \begin{bmatrix} \mathbf{c} \\ \ln\left(\mathbf{u}^o_{min}\right) \\ -\ln\left(\mathbf{u}^o_{max}\right) \end{bmatrix} \tag{26}$$

Observing equations 24, 25 and 26, biologically, they represent the combination of many different constraints, but mathematically, they are in the same form as equations 20, 21 and 22, derived from a system of polynomial equations. If we let $\mathbf{w} \equiv \ln(\mathbf{x})$ then equations 20, 21 and 22 are mathematically equivalent to

$$\begin{aligned} \mathbf{A} \cdot \exp(\mathbf{w}) &= \mathbf{0} \\ \mathbf{B} \cdot \mathbf{w} &= \mathbf{b} \\ \mathbf{C} \cdot \mathbf{w} &> \mathbf{c} \end{aligned} \tag{27}$$

but this has the algorithmic advantage that a wide dynamic range of variable is facilitated in logarithmic scale. In fact, all constraints formulated thus far could be expressed as a system of polynomial equations by applying the inverse of the transformation sequence outlined in subsection 4.3.

7 Application to *E. coli*

We tested the numerical scalability of our approach by implementation of mass conservation, energy conservation and the second law of thermodynamics for the genome scale model of *E. coli* metabolism, (iAF1260, $m = 1668$ metabolites, $\overline{n} = 4152$ one way internal reactions, $\tilde{n} = 306$ net exchange reactions) (20). We assumed that each overall reaction followed pseudo-elementary mass action kinetics and applied thermodynamic Haldane constraints to the ratio of forward and reverse pseudo-elementary rate constants. The equality constraints took the general form

$$\begin{bmatrix} \overline{\mathbf{S}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{v}} \\ \mathbf{k}_+ \\ \mathbf{k}_- \\ \mathbf{m} \\ \mathbf{u}^o \end{bmatrix} = \mathbf{a} \tag{28}$$

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & -\mathbf{I} & \mathbf{0} & \widetilde{\mathbf{S}^T} \\ -\mathbf{I} & \mathbf{Q}_{k_+} & \mathbf{Q}_{k_-} & \mathbf{Q}_m & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ln\left(\widetilde{\mathbf{v}}\right) \\ \ln\left(\mathbf{k}_+\right) \\ \ln\left(\mathbf{k}_-\right) \\ \ln\left(\mathbf{m}\right) \\ \ln\left(\mathbf{u}^o\right) \end{bmatrix} = \mathbf{0} \tag{29}$$

where the boundary condition on the mass conservation constraints, $\mathbf{a} \equiv -\overline{\mathbf{S}} \cdot \overline{\mathbf{v}}$, was derived from the optimal exchange fluxes obtained using flux balance analysis. For the latter, we

fixed certain exchange fluxes using glucose, oxygen, ethanol, acetate, D-lactate, succinate, pyruvate and formate uptake/secretion rates for aerobic growth in glucose minimal medium, as reported experimentally (30) (Sample ID: GR04). In addition, the growth rate was fixed to that measured for the same experiment (30). A unique optimal net flux was then obtained by minimizing the Euclidean norm of all remaining free exchange and internal net fluxes.

We successfully implemented a solver to satisfy the non-convex feasibility problem posed by Eq. 28 and 29, to a tolerance less than 10^{-12} for all constraints. The matrix in Eq. 29 had $8,788$ rows and $11,640$ columns for the genome scale *E. coli* model. The domain of the logarithm implies that all variables are non-negative. However, without additional inequality constraints, it is not possible to represent the qualitatively assigned local reaction directionality which accompany the iAF1260 model (47). Such constraints seem essential for a reliable prediction using flux balance analysis but their inclusion into the current formulation presents a significant numerical analysis challenge. The main difficulty is that local reaction directionality constraints may give rise to a feasible flux balance analysis problem, but an infeasible problem when additional thermodynamic constraints are required to be satisfied. Moreover, when a solver fails to converge to the solution of a non-convex feasibility problem, it is presently difficult to differentiate between a limitation in the 'solving' capability of the algorithm versus an infeasible problem to begin with. This problem is an inherent feature of many non-convex feasibility problems. We continue efforts to add the capability to reliably solve such problems with inequality constraints as reaction directionality constraints are necessary, though not sufficient for accurate prediction of fluxomic data (30) (see Supplementary Figure 3).

8 Discussion

We have demonstrated how the mathematical constraints representing steady state mass conservation, energy conservation, the second law of thermodynamics, reversible polynomial kinetics, experimental and theoretical estimates of thermodynamic and kinetic variables can all be integrated into the composite mathematical form of equations 27. This reformulation is mathematically exact in the sense that it uses no approximations around arbitrary reference states, as with the 'linlog kinetics' approach (41). The thermodynamic reformulation relies on a fundamental equation in non-equilibrium steady state thermodynamics (5) complementing existing work of Beard et. al. (6). The reformulation of reversible polynomial kinetic equations relies on judicious application of exponential identities, their logarithmic analogues, and a novel application of a well known principle from convex geometry. This principle is the representational equivalence of a polytope, as a set of vertices, a vertex representation, or a set of linear equations and inequalities, a half-space representation. The mathematical incorporation of numerical bounds on experimental and theoretical estimates of thermodynamic and kinetic variables, detailed in table 1 follows trivially in this constraint-based approach.

The mathematical generality of the integrated system of linear, linear exponential constraint equations 27, lies in the possibility to reformulate any system of polynomial equations in such a form. For constraint-based modelling of biological networks, this means that, theoretically, any new nonlinear constraints, formalised as an arbitrary system of implicit polynomials, can be integrated with the existing constraint-based modelling of steady state metabolism.

Practically, computational complexity may limit the size of polynomial system that can be transformed in reasonable time. In systems biology, variants of vertex enumeration algorithms are used to compute a convex basis for the stoichiometric matrix (40). These convex bases correspond directly to stoichiometrically balanced pathways of contiguous reactions. Facet enumeration is dual to vertex enumeration so the worst case computational complexity of vertex enumeration is identical to facet enumeration. Given $\mathbf{Q} \in \mathbb{N}^{\mathbf{n},\mathbf{d}}$, the latest algorithms (4) find all facets in worst case time of $\mathcal{O}(nd^2)$ per facet (3).

The number of facets contributes to the size of the constraint satisfaction problem through the matrix $\mathbf{A}$ in Eq. 24 and matrix $\mathbf{B}$ in Eq. 25. To our knowledge, there is no closed form for the number of facets corresponding to a given a set of vertices. However, it is possible to estimate the number of facets and hence the computation time, prior to facet enumeration (4). Assuming mass action kinetics for each reaction, the integration of kinetic constraints does not require facet enumeration, see Supplementary Material E. In this case the numerical constraint feasibility problem requires a simultaneously solving m linear mass conservation constraints (one for each metabolite) Eq. 24, $\frac{5n}{2} + m$ linear thermodynamic & kinetic equalities on logarithmic variables (for $\frac{n}{2}$ reversible elementary reactions) Eq. 25, and m linear inequalities on the estimated range of *in vivo* standard chemical potential.

The motivation for this work arose out of a requirement to integrate various constraints into a concise mathematical form, yet preserve desirable characteristics of the equations such that they are amenable to numerical analysis. Even though the feasible set is non-convex, and therefore non-linear, the linear structure of equations 27, even with a combination of linear and logarithmic variables, indicates potential for scalable numerical analysis. This is confirmed with numerical experiments on a genome scale model of *E. coli* metabolism. The current numerical challenge lies in solving this non-convex feasibility problem when there are non-trivial inequality constraints. It Practice, the current difficulty is distinguishing between failure of the current solver to converge due to (i) problem infeasibility due to over tight inequality constraints, versus (ii) failure to converge due to a shortcoming in algorithm design. Further work needs to be done to either, develop an algorithmic test for infeasibility of linear, linear-logarithmic equations with non-trivial inequalities, or, develop an algorithmic approach which is guaranteed to converge to a solution, if a solution exists. It is interesting to note that systems of linear, linear-logarithmic equations also arise in the classical problem of predicting metabolite concentrations at equilibrium (45, 42). There, the requirement for equilibrium concentration to be positive requires a trivial inequality constraint and the algorithms for solving such systems can be proven to converge(45). It remains to be seen if the sucessful algorithmic approaches from chemical reaction equilibrium analysis can be adapted to solve our novel non-equilibrium formulation with non-trivial inequality constraints.

Acknowledgements

The authors would like to thank Neema Jamshidi & Hong Qian for critical reading of the manuscript. R.M.T.F. was supported by a National University of Ireland, Galway, Science Faculty Fellowship. I.T. was supported by NIH grant Grant 5R01GM057089-11.

References

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

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

3. Avis, D., 2000. Polytopes - Combinatorics and Computation, Birkhauser-Verlag, chapter lrs: A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm, 177–198.

4. Avis, D., 1998. Computational experience with the reverse search vertex enumeration algorithm. *Optimization Methods and Software* 10:107–124.

5. Beard, D. A., E. Babson, E. Curtis, and H. Qian, 2004. Thermodynamic constraints for biochemical networks. *J Theor Biol* 228:327–333.

6. Beard, D. A., S. Liang, and H. Qian, 2002. Energy balance for analysis of complex metabolic networks. *Biophys J* 83:79–86.

7. Beard, D. A., and H. Qian, 2007. Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS ONE* 2:e144.

8. Beyer, A., J. Hollunder, H. P. Nasheuer, and T. Wilhelm, 2004. Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. *Molecular & cellular proteomics* 3:1083–1092.

9. Binder, H., 2006. Thermodynamics of competitive surface adsorption on DNA microarrays. *Journal of Physics: Condensed Matter* 18:S491–S523.

10. Binder, H., and S. Preibisch, 2006. GeneChip microarrays-signal intensities, RNA concentrations and probe sequences. *Journal of Physics: Condensed Matter* 18:S537–S566.

11. Cook, P. F., and W. W. Cleland, 2007. Enzyme Kinetics and Mechanism. Taylor & Francis Group, London.

12. Cornish-Bowden, A., 2004. Fundamentals of Enzyme Kinetics (3rd edition). Portland Press, London.

13. Cornish-Bowden, A., and J. H. S. Hofmeyr, 2005. Enzymes in context: Kinetic characterization of enzymes for systems biology. *The Biochemist* 27:11–13.

14. Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard, and B. Ø. Palsson, 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429:92–96.

15. Crooks, G. E., 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E* 60:2721–2726.

16. Deeds, E. J., O. Ashenberg, and E. I. Shakhnovich, 2006. A simple physical model for scaling in protein-protein interaction networks. *Proc Natl Acad Sci U S A* 103:311–316.

17. Ederer, M., and E. Gilles, 2007. Thermodynamically feasible kinetic models of reaction networks. *Biophys J* 92:1846–1857.

18. English, B. P., W. Min, A. M. van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, and X. S. Xie, 2006. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat Chem Biol* 2:87–94.

19. Famili, I., R. Mahadevan, and B. Ø. Palsson, 2005. k-Cone analysis: Determining all candidate values for kinetic parameters on a network scale. *Biophys J* 88:1616–1625.

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

21. Fersht, A., 1999. Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. W. H. Freeman, New York.

22. Gel'fand, I., and A. Zelevinskij, 1994. Discriminants, Resultants, and Multidimensional Determinants. Birkhauser, New York.

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

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

25. Hill, T. L., 1977. Free Energy Transduction in Biology. Academic Press, New York.

26. Hoppe, A., S. Hoffmann, and H. G. Holzhutter, 2007. Including metabolite concentrations into flux balance analysis: Thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BioMed Central Systems Biology* 1:23.

27. Horn, F., and R. Jackson, 1972. General mass action kinetics. *Archive for Rational Mechanics and Analysis* 47:81–116.

28. Ibarra, R. U., J. S. Edwards, and B. Ø. Palsson, 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420:186–189.

29. Irizarry, R. A., B. Hobbs, F. Collin, Y. D. . Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.

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

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

32. Kümmel, A., S. Panke, and M. Heinemann, 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* 2:e34.

33. Liu, J. S., I. W. Marison, and U. von Stockar, 2001. Microbial growth by a net heat up-take: a calorimetric and thermodynamic study on acetotrophic methanogenesis by *Methanosarcina barkeri*. *Biotechnol Bioeng* 75:170–180.

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

35. Oster, G. F., and A. S. Perelson, 1974. Chemical reaction dynamics. *Archive for Rational Mechanics and Analysis* 55:230–274.

36. Oster, G. F., and A. S. Perelson, 1974. Chemical reaction dynamics, part II: Reaction networks. *Archive for Rational Mechanics and Analysis* 57:31–98.

37. Palsson, B. Ø., 2006. Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, Cambridge.

38. Rojas, I., M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, and U. Wittig, 2007. Storing and annotating of kinetic data. *In Silico Biol* 7:37–44.

39. Ross, J., 2008. Thermodynamics and fluctuations far from equilibrium, volume 80 of *Springer Series in Chemical Physics*. Springer, New York.

40. Schilling, C. H., D. Letscher, and B. Ø. Palsson, 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 203:229–248.

41. Smallbone, K., E. Simeonidis, D. S. Broomhead, and D. B. Kell, 2007. Something from nothing: Bridging the gap between constraint-based and kinetic modelling. *Federation of European Biochemical Societies Journal* 274:5576–5585.

42. Smith, W. R., and R. W. Missen, 1982. Chemical Reaction Equilibrium Analysis: Theory and Algorithms. Wiley, New York.

43. Vallabhajosyula, R. R., V. Chickarmane, and H. M. Sauro, 2006. Conservation analysis of large biochemical networks. *Bioinformatics* 22:346–353.

44. Wiechert, W., 2007. The thermodynamic meaning of metabolic exchange fluxes. *Biophys J* 93:2255.

45. Zeleznik, F., and S. Gordon, 1968. Calculation of complex chemical equilibria. *Industrial & Engineering Chemistry* 60:27–57.

46. Duarte, N. C., S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ø. Palsson, 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104:1777–1782.

47. Fleming, R. M. T., I. Thiele, and H. P. Nasheuer. Quantitative assignment of reaction directionality in constraint-based models of metabolism: Application to *Escherichia coli*. *(accepted), Biophysical Chemistry* .

48. Yeung, M., I. Thiele, and B. Palsson, 2007. Estimation of the number of extreme pathways for metabolic networks. *BioMed Central Bioinformatics* 8:363.

Table

Data	Comments
Stoichiometry	Essential
Standard Chemical Potentials	Experimentally measured(2, 1) or group contribution estimates(24, 23).
Kinetic Rate Equations	Mass action assumption or enzyme specific polynomial kinetic rate laws(11)
Fluxes	Optionally, upper & lower bounds from experiments
Concentrations	Optionally, upper & lower bounds from experiments

Table 1: Input data for integrated modelling. Reaction stoichiometry is essential input data whereas experimentally measured fluxes or concentrations are optional. Experimentally derived or computationally estimated bounds on standard chemical potential are not essential but serve to quantitatively dictate the direction of net flux for kinetically irreversible reactions in the absence of manually assigned metabolic reconstruction reaction directions. Since standard chemical potential is on a logarithmic scale, with respect to concentration, bounds on the former are more significant for determining the feasible set of reaction directions. In an iterative model development approach, initially, one may assume mass action kinetics for all reactions then subsequently incorporate more specific polynomial rate laws, such as Michaelis-Menten kinetics, as the model becomes more sophisticated(38). Kinetic rate laws dictate the magnitude of net reaction flux given a particular metabolite concentration, assuming the reaction is thermodynamically feasible.

Figure Legends

Figure 1.

An artificial metabolic network consisting of one source exchange reaction, one sink exchange reaction and a three reversible isomerisations in a stoichiometrically balanced cycle, $\{v_1, v_3, v_5\}$or equivalently $\{v_2, v_4, v_6\}$. The oval represents the boundary between the internal organism and external environment.

Figure 2.

The linear polytope (blue) is a set of solutions to $\begin{bmatrix} 1 & 1 & -2 \end{bmatrix} \cdot \begin{bmatrix} x & y & z \end{bmatrix}^{\mathbf{T}} = 0$. The exponential of a logarithmic scale polytope (red) is a set of solutions to $\begin{bmatrix} 1 & -2 & 0 \end{bmatrix} \cdot \begin{bmatrix} \ln(x) & \ln(y) & \ln(z) \end{bmatrix}^{\mathbf{T}} = 0$. The intersection (green) is a two dimensional curve sitting in a three dimensional space, and is defined by the polynomial $y^2 + y - 2z = 0$. For illustration, there are arbitrary bounds on the variables.

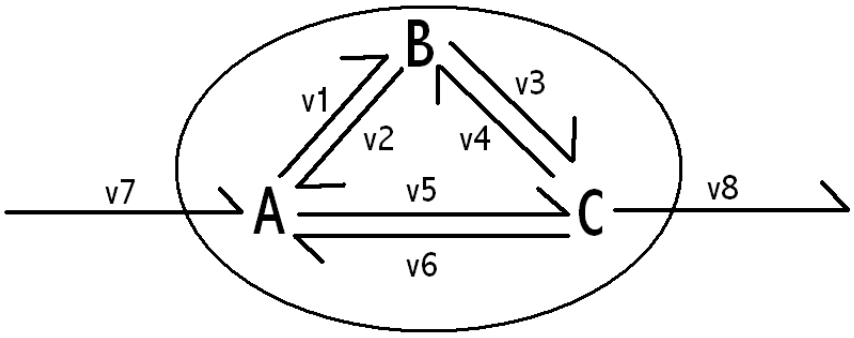

Figure 1: A toy metabolic network

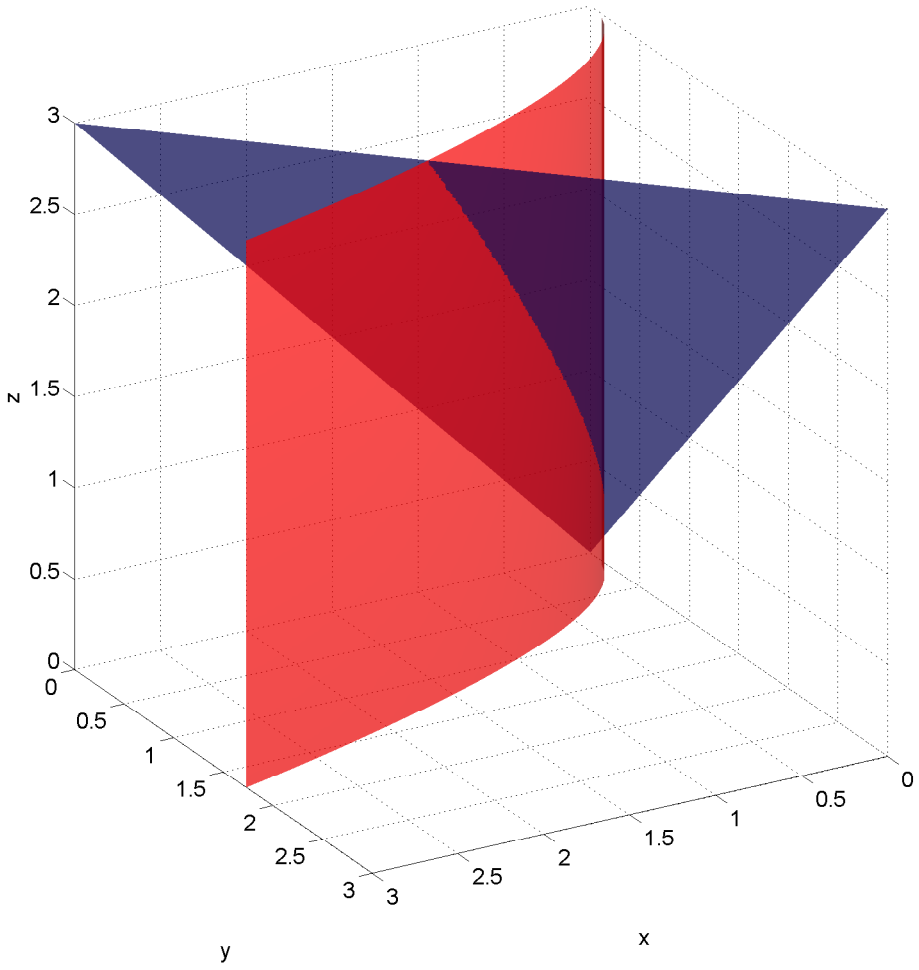

Figure 2: Linear, logarithmic polytope intersection.

Supplementary Material

Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism R.M.T. Fleming, I. Thiele, G. Provan, H.P. Nasheuer

A The stoichiometric matrix

Consider the highly artificial metabolic network in Figure 1 its stoichiometric matrix $\mathbf{S}$ appears in the dynamic equation for steady state mass conservation as

$$\mathbf{S}\cdot\mathbf{v} = \left[\begin{array}{cccccc|cc} -1 & 1 & 0 & 0 & 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & -1 \end{array}\right] \cdot \left[\begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ \hline v_7 \\ v_8 \end{array}\right] = \frac{d\mathbf{m}}{dt} \equiv \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right] \qquad (30)$$

where the metabolites A, B, C are denoted by rows $1, 2, 3$. The first column represents an elementary chemical reaction where one molecule of A is consumed and one molecule of B is produced. The matrix is artificially divided into six elementary internal reactions and two elementary exchange reactions. Internal reactions represent chemical transformations occurring within the organism. All elementary fluxes, or reaction rates, are non-negative and an internal chemical transformation is assumed to be reversible, so each internal reversible reaction is represented by two columns. The net flux through an internal reaction is the difference between the forward and reverse elementary reaction flux. Exchange reactions, on the organism boundary, provide a chemical potential gradient across the internal metabolic network, representing open non-equilibrium living conditions. Source exchange reactions pump metabolites into the organism and sink exchange reactions soak up waste metabolites from the organism (second and last column respectively). Strictly speaking, exchange reactions do not conserve mass since metabolites are input from infinite sources and output into infinite sinks (3). Most metabolites participate in a small number of reactions, thus the stoichiometric matrix is sparse. There is a heavy tailed distribution of non-zero coefficients with respect to the rows of stoichiometric matrices. Typically, very few rows have $> 10\%$ non-zero coefficients and most rows have less than 1% non-zero coefficients.

B Derivation of thermodynamic constraints

Given the equation for steady state energy conservation

$$\mathbf{K}^T.\Delta\widetilde{\mu}^T = \mathbf{0}$$

and the relationship between change in chemical potential and elementary fluxes

$$\Delta\widetilde{\mu}_j = RT\ln\left(\frac{v_{j+1}}{v_j}\right) = -\Delta\widetilde{\mu}_{j+1}$$

if one substitutes for the vector of change in chemical potential, for each row, $\mathbf{K}_i^T$, of $\mathbf{K^T} \in \mathbb{R}^{\mathbf{k},\widetilde{\mathbf{n}}}$, we have

$$RT\left(\mathbf{K}_{i,1}\ln\left(\frac{v_2}{v_1}\right)+\mathbf{K}_{i,2}\ln\left(\frac{v_1}{v_2}\right)+\ldots+\mathbf{K}_{i,\widetilde{n}-1}\ln\left(\frac{v_{\widetilde{n}}}{v_{\widetilde{n}-1}}\right)+\mathbf{K}_{i,\widetilde{n}}\ln\left(\frac{v_{\widetilde{n}-1}}{v_{\widetilde{n}}}\right)\right)=0 \tag{31}$$

By dividing across by the scalar constant, RT, expanding the logarithmic ratios and grouping the coefficients we arrive at

$$\begin{aligned}(\mathbf{K}_{i,2}-\mathbf{K}_{i,1})\ln(v_1)+(\mathbf{K}_{i,1}-\mathbf{K}_{i,2})\ln(v_2)+\cdots\\ +(\mathbf{K}_{i,\widetilde{n}}-\mathbf{K}_{i,\widetilde{n}-1})\ln(v_{\widetilde{n}-1})+(\mathbf{K}_{i,\widetilde{n}-1}-\mathbf{K}_{i,\widetilde{n}})\ln(v_{\widetilde{n}}) &= 0\end{aligned}$$

This equation can be represented as the multiplication of a row of a new matrix, denoted $\mathbf{P}_i$, by a vector of logarithmic fluxes

$$\mathbf{P}_i\cdot\ln(\widetilde{\mathbf{v}})=0$$

where the coefficients in the odd columns are given by

$$\mathbf{P}_{i,j}=(\mathbf{K}_{i,j+1}-\mathbf{K}_{i,j}) \tag{32}$$

and the coefficients in the even columns are given by

$$\mathbf{P}_{i,j+1}=(\mathbf{K}_{i,j}-\mathbf{K}_{i,j+1}) \tag{33}$$

We can now collate all rows into a dot product of a matrix and logarithmic flux

$$\mathbf{P}\cdot\ln(\widetilde{\mathbf{v}})=\mathbf{0}$$

where , $\mathbf{P}\in\mathbb{Z}^{k,\widetilde{n}}$.

C Reformulation of thermodynamic constraints

Consider the stoichiometric matrix 30 as an example. Computing a rational basis for the null space of 30 we find

$$\mathbf{K}^T=\begin{bmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\1&0&1&0&1&0\\-1&0&-1&0&0&1\end{bmatrix}$$

so applying equations 32 and 33 we find

$$\mathbf{P}=\begin{bmatrix}-1&1&-1&1&-1&1\end{bmatrix}$$

$\mathbf{P}$ has only one row since two rows can be considered as equivalent constraints if they are identical except for a sign change. In addition two rows contain all zero coefficients since the integer coefficients of $\mathbf{K}^T$ cancel exactly. Thus, the constraints of energy conservation and

the second law of thermodynamics can be condensed into linear equations on logarithmic flux

$$\mathbf{P} \cdot \ln(\widetilde{\mathbf{v}}) = \begin{bmatrix} -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \ln(v_1) \\ \ln(v_2) \\ \ln(v_3) \\ \ln(v_4) \\ \ln(v_5) \\ \ln(v_6) \end{bmatrix} = 0 \tag{34}$$

D Reformulation of kinetic constraints

To illustrate the mathematical transformation of polynomial kinetic equations into linear logarithmic constraints, for brevity, we consider the isomerisation $A \rightleftharpoons B$, modelled kinetically with the three step reversible Michaelis-Menten Eq. 6. Relating phenomenological to elementary kinetic parameters using equations 7, we have the elementary kinetic equation

$$v_{net} = v_1 - v_2 = \frac{k_1 k_3 k_5 e_{AB} a - k_2 k_4 k_6 e_{AB} b}{k_2 k_5 + k_2 k_4 + k_3 k_5 + k_1(k_3 + k_4 + k_5)a + k_6(k_2 + k_3 + k_4)b} \tag{35}$$

Multiplied out, Eq. 35 becomes

$$\begin{aligned} k_2 k_4 k_6 e_{AB} b - k_1 k_3 k_5 e_{AB} a + v_1 k_2 k_5 + v_1 k_2 k_4 + v_1 k_3 k_5 + v_1 k_1 k_3 a + v_1 k_1 k_4 a + \\ v_1 k_1 k_5 a + v_1 k_2 k_6 b + v_1 k_3 k_6 b + v_1 k_4 k_6 b - v_2 k_2 k_5 - v_2 k_2 k_4 - v_2 k_3 k_5 - \\ - v_2 k_1 k_3 a - v_2 k_1 k_4 a - v_2 k_1 k_5 a - v_2 k_2 k_6 b - v_2 k_3 k_6 b - v_2 k_4 k_6 b = 0 \end{aligned}$$

Expressing this implicit polynomial, in the form of Eq. 12, gives rise to the matrices $\mathbf{A}$ and $\mathbf{Q}$ where each column $\mathbf{A} \in \mathbb{Z}^{1,20}$ corresponds to a row of $\mathbf{Q} \in \mathbb{N}^{20,11}$ hence the equations

$$\mathbf{A} \cdot \mathbf{x} = \begin{bmatrix} 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \cdot \mathbf{x} = \mathbf{0}$$

$$\mathbf{Q} \cdot \ln(\mathbf{y}) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \ln(v_1) \\ \ln(v_2) \\ \ln(k_1) \\ \ln(k_2) \\ \ln(k_3) \\ \ln(k_4) \\ \ln(k_5) \\ \ln(k_6) \\ \ln(a) \\ \ln(b) \\ \ln(e_{AB}) \end{bmatrix} = \ln(\mathbf{x})$$

where each column of $\mathbf{Q}$ corresponds to a variable in Eq. 35. The 'dual Newton polytope', obtained by vertex enumeration of $\mathbf{Q}$, is defined by equations 17 and 18. In this case $\mathbf{b} = \mathbf{0}$ and the matrices $\mathbf{B} \in \mathbb{Z}^{12,20}$ and $\mathbf{C} \in \mathbb{Z}^{19,20}$ are

$$\mathbf{B} \cdot \ln(\mathbf{x}) = \begin{bmatrix} 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & -1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \ln(\mathbf{x}) = \mathbf{0}$$

$$\mathbf{C} \cdot \ln(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 1 & -3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 0 & -3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -2 & 1 & -4 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 2 & 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 2 & 1 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & -2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & -2 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \ln(\mathbf{x}) \geq \mathbf{0}$$

E Linearisation of mass action rate laws

If each implicit polynomial is a binomial, then facet enumeration results in the trivial constraint $\mathbf{x} > 0$, and $\mathbf{Q} \cdot \ln(\mathbf{y}) = \ln(\mathbf{x})$ is under-determined so the constraints can be summarised

$$\begin{aligned} \mathbf{A} \cdot \mathbf{x} &= \mathbf{a} \\ \begin{bmatrix} \mathbf{I} & -\mathbf{Q} \end{bmatrix} \cdot \begin{bmatrix} \ln(\mathbf{x}) \\ \ln(\mathbf{y}) \end{bmatrix} &= \mathbf{0} \\ \mathbf{x} &> 0 \end{aligned}$$

If, in addition, the coefficients of both monomials have a magnitude of 1 then application of the logarithm to both sides of an explicit equation linearises the equation in terms of logarithmic variables. This is the case if one approximates the kinetics of an elementary

reactions with a simple mass action mechanism and rate law

$$B + E \underset{k_8}{\overset{k_7}{\rightleftharpoons}} E + C$$

$$\begin{aligned} v_3 &= k_7 b e_{BC} \\ v_4 &= k_8 c e_{BC} \end{aligned} \tag{36}$$

Taking the logarithmic of both sides of equations 36 and using the identity 9, we have an implicit linear equation on a logarithmic scale

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \ln(v_3) \\ \ln(v_4) \\ \ln(k_7) \\ \ln(k_8) \\ \ln(b) \\ \ln(c) \\ \ln(e_{BC}) \end{bmatrix} = \mathbf{0}$$

Developing algorithms which solve systems of binomial equations using the logarithm in this way is part of geometric programming (1).

Supplementary Figure 3

Comparison of *in vivo* fluxomic data with *in silico* prediction by flux balance analysis and the equality constrained linear, linear logarithmic approach, as discussed in section 7. The *in vivo* data is mainly limited to reactions in central metabolism (Fluxome (4)). Numerical challenges preclude the addition of inequality constraints representing reaction directionality to the latter constraint satisfaction problem. As a result, even though additional thermodynamic constraints are satisfied, many of the directions of predicted fluxes agree no better with *in vivo* fluxomic data, than those predicted by flux balance analysis (FBA). Reaction names and abbreviations are as given with the genome scale reconstruction of *E. coli* metabolism (2).

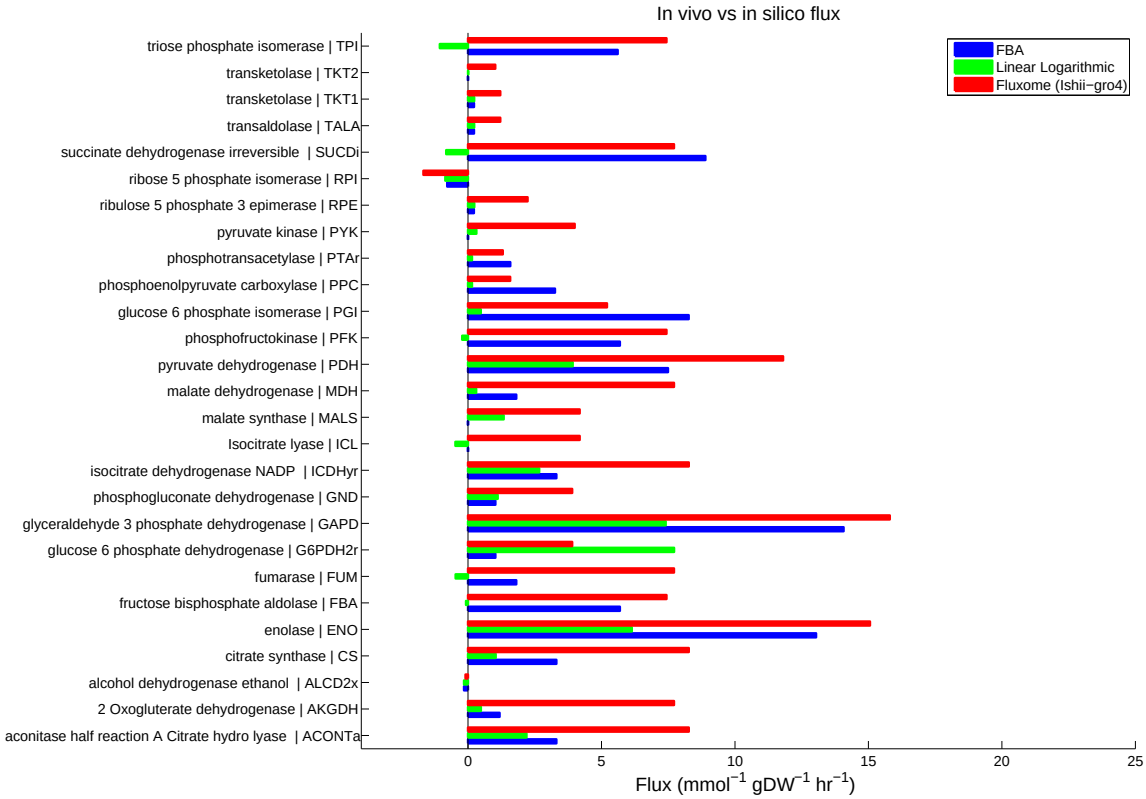

Figure 3: Comparison of *in silico* and experimental fluxes.

Supplementary References

1. Boyd, S., S. J. Kim, L. Vandenberghe, and A. Hassibi, 2007. A tutorial on geometric programming. *Optimization and Engineering* 8:67–127.

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

3. Horn, F., and R. Jackson, 1972. General mass action kinetics. *Archive for Rational Mechanics and Analysis* 47:81–116.

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