

Something from nothing – bridging the gap between constraint-based and kinetic modelling

Kieran Smallbone^{1,2}, Evangelos Simeonidis^{1,3}, David S. Broomhead^{1,2} and Douglas B. Kell^{1,4}

1 Manchester Centre for Integrative Systems Biology, The University of Manchester, UK

2 School of Mathematics, The University of Manchester, UK

3 School of Chemical Engineering and Analytical Science, The University of Manchester, UK

4 School of Chemistry, The University of Manchester, UK

Keywords

flux balance analysis; linlog kinetics;
Saccharomyces cerevisiae

Correspondence

K. Smallbone, Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, 131 Princess Street, Manchester, M1 7 DN, UK
Fax: +44 161 30 65201
Tel: +44 161 30 65146
E-mail: kieran.smallbone@manchester.ac.uk
Website: <http://www.mcisb.org/>

(Received 29 June 2007, revised 17 August 2007, accepted 29 August 2007)

doi:10.1111/j.1742-4658.2007.06076.x

Two divergent modelling methodologies have been adopted to increase our understanding of metabolism and its regulation. Constraint-based modelling highlights the optimal path through a stoichiometric network within certain physicochemical constraints. Such an approach requires minimal biological data to make quantitative inferences about network behaviour; however, constraint-based modelling is unable to give an insight into cellular substrate concentrations. In contrast, kinetic modelling aims to characterize fully the mechanics of each enzymatic reaction. This approach suffers because parameterizing mechanistic models is both costly and time-consuming. In this paper, we outline a method for developing a kinetic model for a metabolic network, based solely on the knowledge of reaction stoichiometries. Fluxes through the system, estimated by flux balance analysis, are allowed to vary dynamically according to linlog kinetics. Elasticities are estimated from stoichiometric considerations. When compared to a popular branched model of yeast glycolysis, we observe an excellent agreement between the real and approximate models, despite the absence of (and indeed the requirement for) experimental data for kinetic constants. Moreover, using this particular methodology affords us analytical forms for steady state determination, stability analyses and studies of dynamical behaviour.

The emergent field of systems biology involves the study of the interactions between the components of a biological system, and how these interactions give rise to the function and behaviour of that system (for example, the enzymes and metabolites in a metabolic pathway). Nonlinear processes dominate such biological networks, and hence intuitive verbal reasoning approaches are insufficient to describe the resulting complex system dynamics [1–3]. Nor can such approaches keep pace with the large increases in -omics data (such as metabolomics and proteomics) and the accompanying advances in highthroughput experiments and bioinformatics. Rather, experience from other areas of science has taught us that quanti-

tative methods are needed to develop comprehensive theoretical models for interpretation, organization and integration of this data. Once viewed with scepticism, we now realize that mathematical models, continuously revised to incorporate new information, must be used to guide experimental design and interpretation.

We focus here on the development and analysis of mathematical models of cellular metabolism [4–6]. In recent years two major (and divergent) modelling methodologies have been adopted to increase our understanding of metabolism and its regulation. The first is constraint-based modelling [7,8], which uses physicochemical constraints such as mass balance, energy balance, and flux limitations to describe the

Abbreviations

BPG, 1,3-bisphosphoglycerate; ETOH, ethanol; FBA, flux balance analysis; PFK, phosphofructokinase.

potential behaviour of an organism. The biochemical structure of (at least the central) metabolic pathways is more or less well-known, and hence the stoichiometries of such a network may be deduced. In addition, the flux of each reaction through the system may be constrained through, for example, knowledge of its V_{\max} , or irreversibility considerations. From the steady state solution space of all possible fluxes, a number of techniques have been proposed to deduce network behaviour, including flux balance and extreme pathway or elementary mode analysis. In particular, flux balance analysis (FBA) [9] highlights the most effective and efficient paths through the network in order to achieve a particular objective function, such as the maximization of biomass or ATP production.

The key benefit of FBA lies in the minimal amount of biological knowledge and data required to make quantitative inferences about network behaviour. However, this apparent free lunch comes at a price – constraint-based modelling is concerned only with fluxes through the system and does not make any inferences nor any predictions about cellular metabolite concentrations. By contrast, kinetic modelling aims to characterize fully the mechanics of each enzymatic reaction, in terms of how changes in metabolite concentrations affect local reaction rates. However, a considerable amount of data is required to parameterize a mechanistic model; if complex reactions like phosphofructokinase are involved, an enzyme kinetic formula may have 10 or more kinetic parameters [6]. The determination of such parameters is costly and time-consuming, and moreover many may be difficult or impossible to determine experimentally. The *in vivo* molecular kinetics of some important processes like oxidative phosphorylation and many transport mechanisms are almost completely unknown, so that modelling assumptions about these metabolic processes are necessarily highly speculative.

In this paper, we define a novel method for the generation of kinetic models of cellular metabolism. Like constraint-based approaches, the modelling framework requires little experimental data regarding variables and no knowledge of the underlying mechanisms for each enzyme; nonetheless it allows inference of the dynamics of cellular metabolite concentrations. The fluxes found through FBA are allowed to vary dynamically according to linlog kinetics [10–12]. Linlog kinetics, which draws ideas from thermodynamics and metabolic control analysis, is known to be more appropriate for approximating hyperbolic enzyme kinetics than are other phenomenological relations such as power laws [13]. Indeed, when a version using linlog kinetics is compared with the original and mechanistic branched yeast glycolysis model of Teusink

et al. [14], we observe an excellent agreement between the real and approximate models. Moreover, we show that a model framed within the linlog format affords analytical forms for steady state determination, stability analyses and studies of dynamical behaviour. As such, it does not suffer from the usual [15] computational scalability problems, and could therefore be applied to existing genome scale models of metabolism [8,16–18]. Such a model has powerful predictive power in determining cellular responses to environmental changes, and may be considered a stepping-stone to a full kinetic model of cell metabolism: a ‘virtual cell’.

Results

The linlog approximation [10–12] is a method for simplifying reaction rate laws in metabolic networks. Drawing ideas from metabolic control analysis, it describes the effect of metabolite levels on flux as a linear sum of logarithmic terms (Eqn 2). By definition, it will provide a good approximation near a chosen reference state. Moreover, thermodynamic considerations show that we can expect a logarithmic response to changes in metabolite concentrations [10,13], and hence that the approximation may be valid some distance from the reference state. Indeed, linlog kinetics are known to be appropriate for approximating hyperbolic enzyme kinetics, and, in this case, are superior to other phenomenological relations such as power laws (including generalized mass action and S-systems) [13]. To illustrate this graphically, we compare in Fig. 1

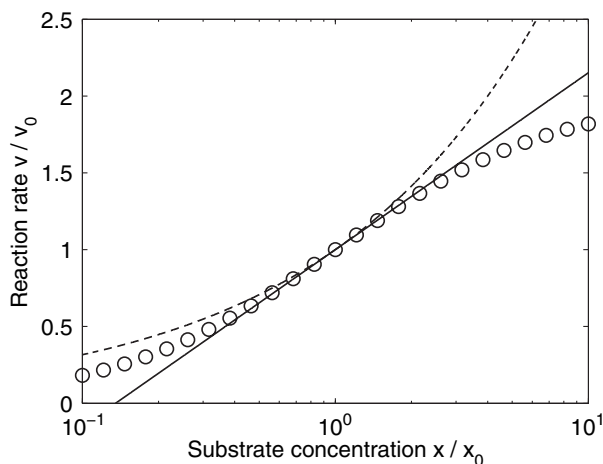


Fig. 1. A comparison of Michaelis–Menten kinetics $v(x) = Vx/(x + K_m)$ (o) to its linlog approximation $u(x) = v^* (1 + \varepsilon \log(x/x^*))$ (solid line) and its power law approximation $\omega(x) = v^*(x/x^*)^\varepsilon$ (dashed line), where $v^* = v(x^*)$ and $\varepsilon = K_m/(x^* + K_m)$. Parameter values used are $x^* = V = K_m = 1$.

typical irreversible Michaelis–Menten kinetics (o) with its linlog (solid line) counterpart. Notice that the abscissa is logarithmic, and Michaelis–Menten kinetics appears to be close to linear in this plotting regime. Thus linlog serves as an excellent approximation. Even an order of magnitude away from the reference state, the functions have comparable values. Also shown is the power law approximation (dashed line). We see that linlog provides a better approximation than power law for substrate concentrations greater than the reference state, whilst the two approximations are equally valid for concentrations less than the reference state.

Having described the validity of linlog kinetics at the single reaction level, we move on to apply the approximation to a full network: the branched model of yeast glycolysis of Teusink *et al.* [14], available in SBML format from JWS Online [19]. Taking the model's steady state as our reference state, elasticities may be calculated analytically from the kinetic equations using Eqn (3). Eqn (10) may then be used to predict changes in internal metabolite concentrations with external metabolite changes.

In the Teusink *et al.* model, there are three external effectors: ethanol, glucose and glycerol; in Fig. 2 we show, as an example, internal changes in response to changes in ethanol (ETOH). We see that both the zeroth and first derivatives of the linlog kinetics are correct around the reference state $[\text{ETOH}] = [\text{ETOH}]_0$, and hence the approximation is good in a region near this point. Moreover, we see that in many cases the approximation remains valid when the ethanol concentration is changed by an order of magnitude.

Linlog provides a good approximation to enzyme kinetics, and moreover (as we show in Eqns 10–13) affords analytical forms for steady state determination, stability analyses and temporal dynamics. However, the good fit in Fig. 2 was obtained through our exact knowledge of the underlying kinetic formulae. Phenomenological relations such as linlog are unlikely to be of such interest when all enzymatic mechanisms and corresponding parameters are known; rather the interest lies in their applicability when such information is not available and we require a best guess model of the

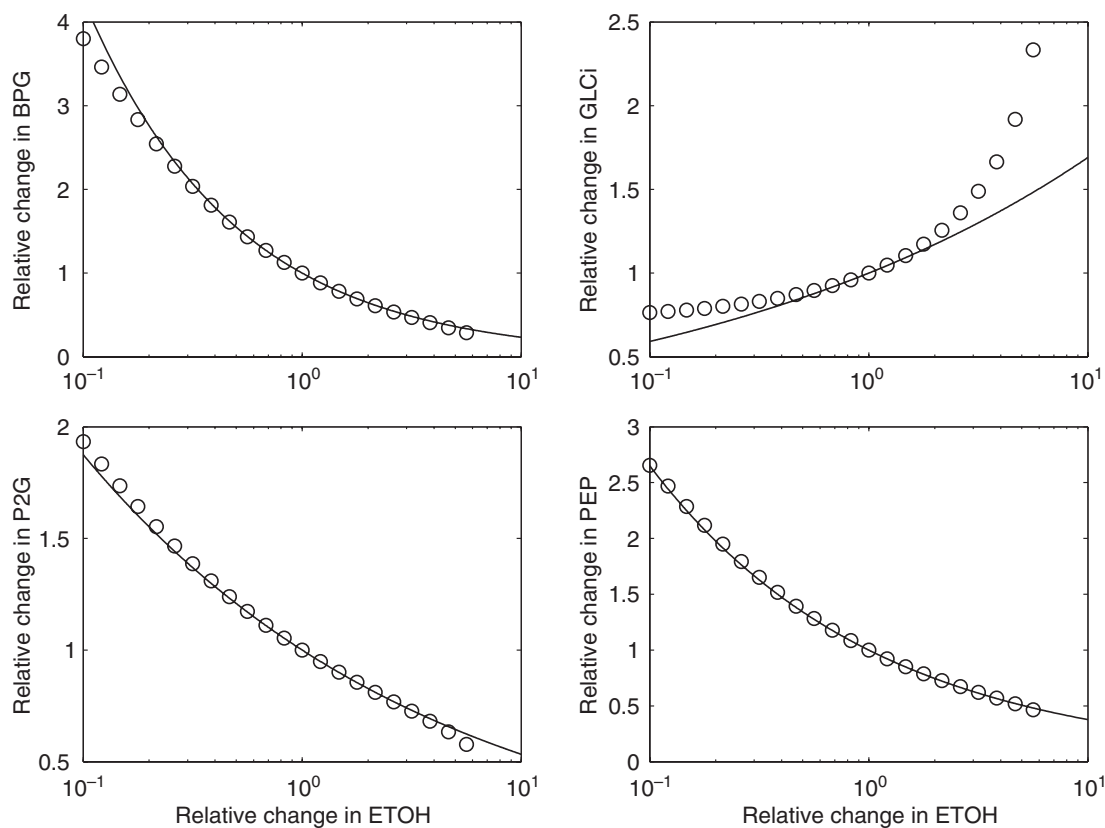


Fig. 2. From Eqn (10). Elected variations in steady state intracellular metabolite concentrations with changes in ethanol (ETOH) concentration in the branched model of yeast glycolysis of Teusink *et al.* [14]. Shown are the real model solutions (o) and the predictions of the linlog approximation (solid line). BPG, 1,3-bisphosphoglycerate; GLCi, glucose in cytosol; P2G, 2-phosphoglycerate; PEP, phosphoenolpyruvate.

underlying kinetics. Returning to Eqn (10), we see that to predict steady state behaviour in response to changes in external effectors, estimates are required for the reference system flux and the elasticities.

The first point, estimation of system fluxes with limited information, may be addressed through appealing to flux balance analysis [9]. This method allows us to identify the optimal path through the network to achieve a particular objective, such as biomass yield or ATP production. Biologically, this kind of objective function assumes that an organism has evolved over time to lie close to its maximal metabolic efficiency, within its underlying physicochemical, topobiological, environmental and regulatory constraints [8].

FBA (Eqn 15) is applied to the model of Teusink *et al.* defining the objective function as cellular ATP production; the results are presented in Table 1. We see that FBA does provide a good estimate to the real fluxes through the system as predicted by Teusink *et al.* The discrepancy between the real and FBA solution is due to FBA disregarding the branches of the pathway not involved in ATP production, namely glycerol, glycogen, succinate and trehalose synthesis. It is interesting to note that, in the full model, the fluxes through these branches are relatively small; the majority of flux is used to generate ATP as assumed by FBA.

It remains to estimate the elasticities. Of course these should ideally be measured explicitly using traditional enzyme assays, for example. In the absence of such

Table 1. Results from Eqn (15). A comparison between fluxes in Teusink *et al.* [14] and those predicted by FBA with ATP production maximization. For reaction abbreviation definitions, see supplementary Table S2.

Reaction	Flux ($\text{mM}\cdot\text{min}^{-1}$)	
	Teusink	FBA
ADH	129	176
ALD	77.3	88.1
ATP	84.5	176
ENO	136	176
G3PDH	18.1	0
GAPDH	136	176
GLK	88.1	88.1
GLYCO	6	0
PDC	136	176
PFK	77.3	88.1
PGI	77.3	88.1
PGK	136	176
PGM	136	176
PYK	136	176
SUC	3.64	0
Treha	2.4	0

information, assuming knowledge only of reaction stoichiometries, we follow the tendency modelling approach of Visser *et al.* [20] (see Materials and methods). The results of elasticity estimation when applied to Teusink *et al.*'s model are presented in Table 2. We see that this is a reasonable method for a first estimation of elasticities; in most cases the estimate falls within an order of magnitude of the true elasticity. It is interesting to observe that in one case, the estimate has the incorrect sign – the phosphofructokinase (PFK) reaction with respect to high energy phosphates. Whilst ATP is a substrate of PFK, at the reference state an increase in ATP leads to a decrease in reaction rate. Such a result is counter-intuitive and could not

Table 2. A comparison between elasticities in Teusink *et al.* and those estimated through stoichiometric considerations. For reaction and metabolite abbreviation definitions, see supplementary Tables S1–S2.

Reaction	Metabolite	Elasticity	
		Teusink	Estimate
ADH	ACE	3.20	1
	ETOH	− 2.95	− 1
	NAD	− 3.04	− 1
	NADH	3.20	1
ALD	F16P	1.89	1
	TRIO	− 3.08	− 2
ATP	P	1.80	1
ENO	P2G	0.826	1
	PEP	− 0.384	− 1
GAPDH	BPG	− 8.00 × 10 ^{−2}	− 1
	NAD	0.144	1
	NADH	− 9.14 × 10 ^{−2}	− 1
	TRIO	0.919	1
GLK	G6P	− 1.65 × 10 ^{−2}	1
	GLCi	0.458	− 1
	P	1.02	1
GLT	GLCi	− 7.20 × 10 ^{−2}	− 1
	GLC _O	2.54 × 10 ^{−2}	1
PDC	ACE	0	− 1
	PYR	0.423	1
PFK	F16P	− 0.402	− 1
	F6P	0.936	1
	P	− 3.21	1
PGI	F6P	− 0.709	− 1
	G6P	1.18	1
PGK	BPG	2.81	1
	P	− 9.47	− 1
	P3G	− 2.24	− 1
PGM	P2G	− 2.36	− 1
	P3G	2.94	1
PYK	P	− 1.82	− 1
	PEP	0.765	1
	PYR	− 0.243	− 1

be known without detailed knowledge of the underlying enzymatic mechanism.

Using the reference flux estimated in Table 1 and the elasticities estimated in Table 2, we again use

Eqn (10) to predict internal metabolite steady state concentrations for given external metabolite levels. In Fig. 3 we show the predicted internal variations in response to changes in ethanol concentration, using

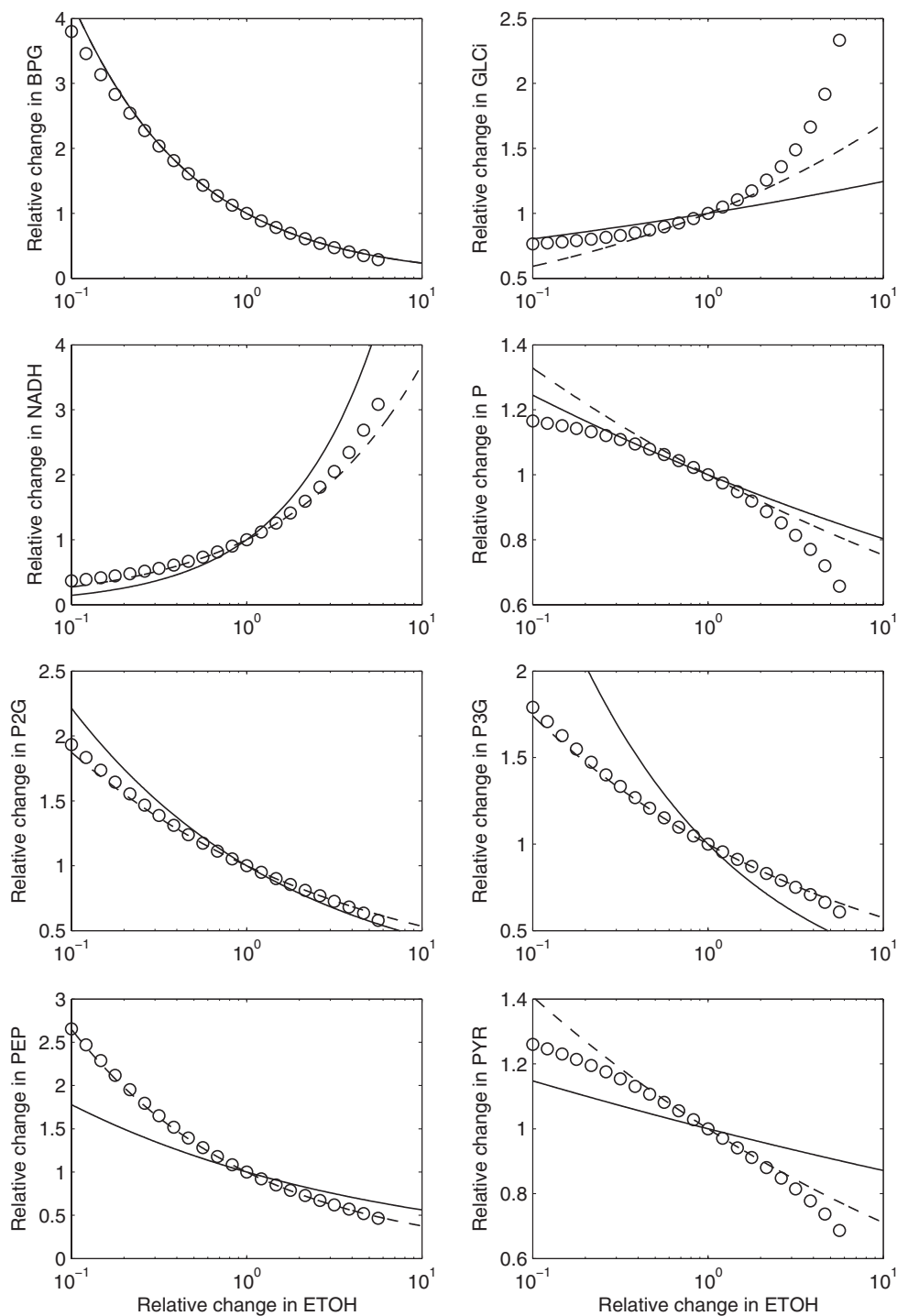


Fig. 3. Variations in steady state intracellular metabolite concentrations with changes in ethanol concentration. Shown are the real model solutions (o), and the predictions of the linlog model with both estimated (solid line) and correct (dashed line) fluxes and elasticities. ETOH, ethanol; P, high energy phosphates; P3G, 3-phosphoglycerate; PYR, pyruvate.

these estimated parameter values (solid line), alongside the real model solutions (o). As a comparison, we also present the predictions of the correctly parameterized linlog approximation (dashed line). Unsurprisingly, the version of the linlog model with estimated parameters does not reproduce real system dynamics as well as the linlog model with correct fluxes and elasticities. Somewhat surprisingly, however, given the limited information used, the estimated model still provides a reasonable approximation to the underlying kinetics. Indeed, in the case of 1,3-bisphosphoglycerate (BPG) steady state concentration, the two versions of linlog approximation are indistinguishable. For many of the remaining metabolites, there is a good agreement between the real and estimated model steady states, despite the differences in parameter values as set out in Tables 1 and 2. The result implies that system steady states are relatively insensitive to these parameters. To reinforce this point, in Table 3 we present the percentage error in steady state prediction by both the correctly parameterized and estimated linlog models, when ethanol concentration is halved from its reference value. The correctly parameterized linlog model provides an excellent approximation, with the majority of errors under 1%. The estimated model performs less well, but nonetheless the predicted concentration of all but one of the metabolites falls within 25% of its real value, which should be considered a success given the limited biological information used.

To complement the above results, in Fig. 4 we present the steady state fluxes as predicted by the linlog

model with both estimated (solid line) and real (dashed line) fluxes and elasticities. Again, both versions provide good approximations to the real model fluxes. The exception here is succinate synthesis; as we saw in Table 1, FBA disregards this branch of glycolysis as it is not involved in ATP production. Hence the fully estimated model assumes no succinate synthesis for any metabolite levels. However, such a result could easily be improved through incorporation of relevant biological information to FBA.

Discussion

Metabolism is arguably the best described network in the cell and there already exist various computational tools to model its behaviour. Kinetic modelling incorporates mechanistic rate equations for each reaction in the network and knowledge of kinetic parameters to accurately simulate system dynamics. However, it requires a large amount of data, which may not be always available for every reaction, and the model may become intractable as the size of the system under examination increases. Constraint-based approaches typically only consider stoichiometric information for the network, which is much more readily accessible, but the allowable solution space is much larger and cannot usually be reduced to a single point; furthermore a lot of biological information about the system may be disregarded, even when available, because there is no consideration of kinetics.

The goal of this paper has been to reconcile the separate methodologies of constraint-based modelling and mechanistic (kinetic) modelling. Like constraint-based methods, we begin from knowledge of only the stoichiometry of the network and cellular composition. Like mechanistic approaches, our estimated model provides at least an intimation of the kinetic nature and behaviour of the system. The proposed methodology was tested by applying it to the well-studied yeast glycolytic pathway, using the model proposed in Teusink *et al.* [14] as a starting point.

The results in Figs 3 and 4 demonstrate that our approach, even though admittedly not perfect in its predictions, is still capable of providing a very useful approximation for a metabolic network, in the absence of an accurate kinetic model and detailed kinetic rate equations for each reaction. The predictions of the proposed model agree well with the Teusink *et al.* model solutions and, therefore, when applied to a pathway that has not been as thoroughly studied, could yield invaluable information. To our knowledge, the approach presented in this paper is the first to provide a kinetic (albeit approximate) model for a

Table 3. Percentage errors in the predicted steady state concentrations when ethanol concentration is halved from its reference value. The table compares the percentage error between the real model steady states and those predicted by both the correctly parameterized linlog model and its fully estimated counterpart.

Metabolite	Linlog error (%)	
	Correctly parameterized	Fully estimated
ACE	9.98×10^{-2}	21.2
BPG	0.591	0.612
F16P	1.37×10^{-2}	19.0
F6P	1.68	36.7
G6P	7.58×10^{-2}	24.5
GLCi	3.00	6.34
NAD	0.251	0.805
NADH	2.52	19.5
P	1.54	0.418
P2G	0.38	4.75
P3G	0.275	14.5
PEP	0.535	11.8
PYR	1.03	4.98
TRIO	6.48×10^{-2}	13.4

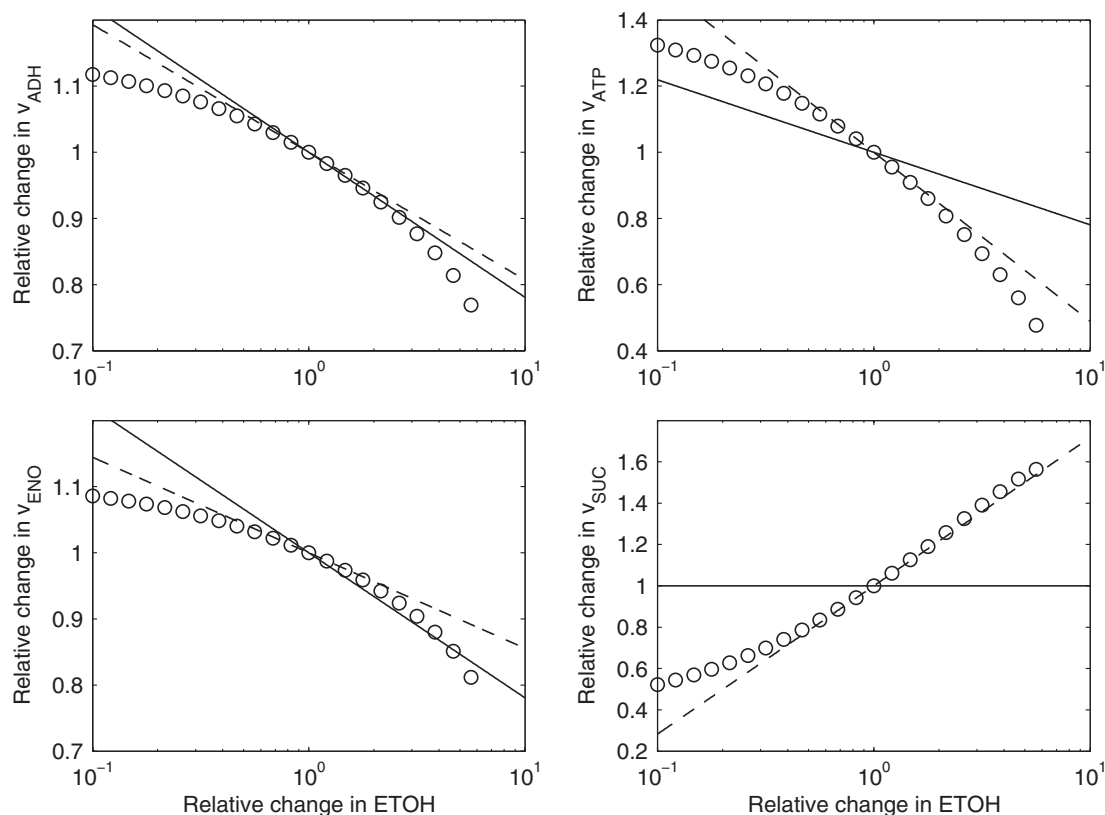


Fig. 4. Selected variations in steady state fluxes with changes in ethanol concentration. Shown are the real model solutions (o), and the predictions of the linlog model with both estimated (solid line) and correct (dashed line) fluxes and elasticities. ADH, alcohol dehydrogenase; ATP, ATPase activity; ENO, enolase; SUC, succinate synthesis.

metabolic network, based solely on the knowledge of the reaction stoichiometry and nutrient supply. Any known fluxes may be set as constraints to improve our FBA solution; similarly, any known elasticities may be used in place of our first approximations. Hence our modelling framework may be considered the first step in the deductive–inductive ‘cycle of knowledge’ [21] crucial for systems biology.

Materials and methods

Linlog kinetics

A generalized description of the temporal evolution of a metabolic network may be described in differential equation form as

$$\text{diag}(c) \frac{dx}{dt} = Nv(x, y) \quad (1)$$

where x is a vector of length m of internal metabolite concentrations, v is a vector of length n describing the functional form of reaction rates or fluxes, and N is a matrix of size $m \times n$ defining the stoichiometries of the system. Also

included is y , a vector of length m_y of external metabolite concentrations that affect flux, but whose temporal dynamics are not considered. Finally, c is a vector of length m whose elements c_i correspond to the volume of the compartment containing metabolite x_i .

We first define a reference state $(x, y) = (x^*, y^*)$. These are metabolite concentrations at a point of interest in the system; for example, y^* might represent the ‘normal’ external metabolite concentrations and x^* a steady state solution to Eqn (1) at $y = y^*$, if such a solution exists.

The linlog approximation [10–12] describes the effect of metabolite levels on flux v as a linear sum of logarithmic terms:

$$v(x, y) \approx u(x, y) := \text{diag}(v^*) \left(1_n + \varepsilon^x \log \frac{x}{x^*} + \varepsilon^y \log \frac{y}{y^*} \right) \quad (2)$$

where 1_n denotes a vector of length n with all components equal to unity, $v^* = v(x^*, y^*)$ is the reference flux, ε^x and ε^y are $n \times m$ and $n \times m_y$ elasticity matrices, and $\log(x/x^*)$ and $\log(y/y^*)$ are vectors with components $\log(x_i/x_i^*)$ and $\log(y_j/y_j^*)$, respectively. Implicit in the definition of linlog kinetics is the requirement that all components of the reference state (x^*, y^*) are nonzero. Through differentiation of Eqn (2), the elasticity matrices are defined by

$$\begin{aligned} (\varepsilon^x)_{ij} &= \varepsilon_{x_j}^{v_i} = \frac{\partial v_i(x, y)}{\partial x_j} \bigg|_{(x^*, y^*)} \frac{x_j^*}{v_i^*}, \\ (\varepsilon^y)_{ij} &= \varepsilon_{y_j}^{v_i} = \frac{\partial v_i(x, y)}{\partial y_j} \bigg|_{(x^*, y^*)} \frac{y_j^*}{v_i^*}. \end{aligned} \quad (3)$$

By definition, at the reference state $u = v$ and $J(u) = J(v)$, where J denotes the Jacobian matrix. Thus for each reaction, both the zeroth and first derivatives with respect to any metabolite are correct at the reference state, and hence u is a good approximation to v in a region near this point.

Substituting Eqn (2) in Eqn (1) we find

$$\text{diag}(c) \frac{dx}{dt} \approx N \text{diag}(v^*) \left(1_n + \varepsilon^x \log \frac{x}{x^*} + \varepsilon^y \log \frac{y}{y^*} \right) \quad (4)$$

$$= N \text{diag}(v^*) \left(\varepsilon^x \log \frac{x}{x^*} + \varepsilon^y \log \frac{y}{y^*} \right) \quad (5)$$

where the second equation holds as we choose the reference state (x^*, y^*) to be a steady state, so $N v^* = 0$.

In general, the rank $r(N \text{diag}(v^*) \varepsilon^x) = m_0 < m$ and the system defined above will display moiety conservations [22,23]—certain metabolites can be expressed as linear combinations of other metabolites in the system. Note that, within the linlog framework, the number of independent metabolites is not given simply by $r(N)$, as has been erroneously suggested [10]. The conservations may be removed through matrix decomposition, using an $m \times m_0$ link matrix L that relates the complete vector of internal metabolites to the vector of independent metabolites [23,24]. Following [10], we define

$$L = \text{diag}(x^*)^{-1} \text{diag}(c)^{-1} N \tilde{N}^+ \text{diag}(\bar{c}) \text{diag}(\bar{x}^*) \quad (6)$$

where \bar{x} denotes the independent metabolites, \bar{c} the corresponding compartments, \tilde{N} the corresponding rows of N and $^+$ the Moore-Penrose pseudoinverse [25]. Using the logarithmic approximation $\log(z) \approx z - 1$ for $z \approx 1$ we find

$$\log \frac{x}{x^*} \approx L \log \frac{\bar{x}}{\bar{x}^*}. \quad (7)$$

Now from Eqn (5):

$$\frac{d\chi}{dt} = \omega(-\chi) (\varepsilon_x \chi + \varepsilon_y \gamma) \quad (8)$$

where

$$\begin{aligned} \chi &= \log \frac{\bar{x}}{\bar{x}^*}, \varepsilon_x = \text{diag}(\bar{c})^{-1} \text{diag}(\bar{x}^*)^{-1} \tilde{N} \text{diag}(v^*) \varepsilon^x L, \\ \gamma &= \log \frac{y}{y^*}, \varepsilon_y = \text{diag}(\bar{c})^{-1} \text{diag}(\bar{x}^*)^{-1} \tilde{N} \text{diag}(v^*) \varepsilon^y, \end{aligned} \quad (9)$$

and $\omega(z) = \exp(\text{diag}(z))$ is a diagonal matrix with strictly positive diagonal elements e^{z_i} . Eqn (8) differs from previous linlog representations [11] in that we do not rely on the further approximation $\omega(-\chi) \approx I$, the identity matrix, and thereby maintain the nonlinearity of the system.

Using Eqn (8), for given fixed concentrations of external metabolites γ , we find that the steady state internal metabolite concentrations are given analytically by

$$\chi^* = -\varepsilon_x^{-1} \varepsilon_y \gamma = -(\tilde{N} \text{diag}(v^*) \varepsilon^x L)^{-1} (\tilde{N} \text{diag}(v^*) \varepsilon^y) \gamma \quad (10)$$

where invertibility is ensured through introduction of the link matrix. Linearizing the network about the steady state defined in Eqn (10), the stability matrix is given by

$$J = \omega(-\chi^*) \varepsilon_x = \omega(\varepsilon_x^{-1} \varepsilon_y \gamma) \varepsilon_x. \quad (11)$$

The steady state is linearly stable if and only if all eigenvalues of J have negative real parts [26].

Finally, assuming that $\chi \approx \chi^*$, i.e. that the system remains close its steady state, we may linearize Eqn (8) and hence [27] find the temporal solution

$$\chi(t) = e^{Jt} (\chi(0) - \chi^*) + \chi^*. \quad (12)$$

If the external metabolites γ are allowed to vary with time, we may instead approximate $\omega(-\chi) \approx I$ (following [11]), i.e. assume that we remain close to the reference state, when Eqn (8) has solution

$$\chi(t) = e^{\varepsilon_x t} \chi(0) + \int_0^t e^{\varepsilon_x(t-\tau)} \varepsilon_y \gamma(\tau) d\tau. \quad (13)$$

Flux balance analysis

Mathematically, flux balance analysis [9] is framed as a linear programming problem:

$$\begin{aligned} \text{Maximize } Z &= f^T v, \\ \text{subject to } Nv &= 0, \\ v^{\min} &\leq v \leq v^{\max}. \end{aligned} \quad (14)$$

That is, we define an objective function Z , a linear combination of the fluxes v_i , that we maximize over all possible steady state fluxes ($Nv = 0$) satisfying certain constraints. In many genome scale metabolic models a biomass production reaction is defined explicitly that may be taken as a natural form for the objective function; in other cases, a natural choice is to maximize the rate of cellular ATP production, $Z = v_{\text{ATP}}$.

Whilst the relation $Nv = 0$ constrains the possible fluxes to lie within the null space of the stoichiometric matrix N , upper and lower bounds may be further placed on the individual fluxes ($v_i^{\min} \leq v \leq v_i^{\max}$). For irreversible reactions, $v_i^{\min} = 0$. Specific upper bounds v_i^{\max} , based on enzyme capacity measurements, may be imposed on reactions; in the absence of any information these rates can be generally assumed unconstrained, i.e. $v_i^{\max} = \infty$, and $v_i^{\min} = -\infty$ for reversible reactions.

Special care must be taken when considering exchange fluxes connecting external effectors and internal metabolites, i.e. nutrient supply and waste product removal. If

all reactions are unconstrained, it is explicit from Eqn (14) that the objective function will be unconstrained. To circumvent this problem, we must constrain nutrient supplies to $v_i^{\max} < \infty$; we take these bounds to be the known reference state nutrient supplies. On a similar note, if waste removal reactions are assumed unconstrained and reversible, we may find that the cell can utilize the waste product to generate biomass or ATP, again leading to an unconstrained objective function. To circumvent this problem, we force net waste removal reactions to be irreversible, setting $v_i^{\min} = 0$.

The FBA problem is now well-defined, in that it leads to a unique, finite objective value $Z = Z^*$; however, in general there is degeneracy in the network, leading to an infinite number of flux distributions v with the same optimal value. It is a great focus of the FBA community to reduce the size of this optimal flux space, through imposing tighter limits on each flux based, for instance, on gene knockout experiments, and the measurement of intracellular fluxes with NMR. Such data are easily incorporated, but continuing the assumption that we have limited information, these techniques are not available to us. Instead we extract a biologically meaningful flux from the solution space by solving a secondary problem:

$$\begin{aligned} & \text{Minimize } \sum_i |v_i| \\ & \text{subject to } f^T v = Z^*, \\ & \quad Nv = 0, \\ & \quad v^{\min} \leq v \leq v^{\max}. \end{aligned} \quad (15)$$

That is, we make the sensible assumption that the cell will minimize the total flux required to produce the objective $Z = Z^*$, which by decomposing v into its positive and negative parts may be again viewed as a linear programming problem. Cells may be profligate with regard to flux [28], but one benefit of this approach is that internal cycles that can produce fluxes $v_i = \infty$ from Eqn (14) are removed. Whilst this secondary problem may still have alternative solutions v , we at least know that our nonunique solution will be sensible from a biological perspective.

Elasticity estimation

We follow the tendency modelling approach of Visser *et al.* [20], whereby the elasticities ε^x and ε^y are taken to be equal to the negative of their corresponding stoichiometric coefficient. For example, if two molecules of substrate are used in a reaction, its elasticity is estimated as $\varepsilon = 2$, whilst if one molecule of product is formed from a reaction, we estimate its elasticity as $\varepsilon = -1$. These elasticities are identical to those that would be found through the assumption of mass action kinetics. Whilst Visser *et al.* extended their generalized mass action approach to allow for allosteric

effectors, no such information may be derived from knowledge of the stoichiometric matrix alone.

Acknowledgements

This research was partially funded by the BBSRC/ EPSRC grant BB/C008219/1 'The Manchester Centre for Integrative Systems Biology (MCISB)'. We thank Nils Blüthgen for commenting on the manuscript.

References

- 1 Lazebnik Y (2002) Can a biologist fix a radio? – Or, what I learned while studying apoptosis. *Cancer Cell* **2**, 179–182.
- 2 Mendes P & Kell DB (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**, 869–883.
- 3 Szallasi Z, Stelling J & Periwal V (2006) *System Modeling in Cellular Biology: from Concepts to Nuts and Bolts*. MIT Press, Boston.
- 4 Klipp E, Herwig R, Kowald A, Wierling C & Lehrach H (2005) *Systems Biology in Practice: Concepts, Implementation and Application*. Wiley-VCH, Weinheim.
- 5 Palsson BØ (2006) *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge.
- 6 Wiechert W (2002) Modeling and simulation: tools for metabolic engineering. *J Biotechnol* **94**, 37–63.
- 7 Covert MW, Famili I & Palsson BØ (2003) Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? *Biotechnol Bioeng* **84**, 763–772.
- 8 Price ND, Reed JL & Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* **2**, 886–897.
- 9 Kauffman KJ, Prakash P & Edwards JS (2003) Advances in flux balance analysis. *Curr Opin Biotechnol* **14**, 491–496.
- 10 Visser D & Heijnen JJ (2003) Dynamic simulation and metabolic redesign of a branched pathway using linlog kinetics. *Metab Eng* **5**, 164–176.
- 11 Hatzimanikatis V & Bailey JE (1997) Effects of spatiotemporal variations on metabolic control: approximate analysis using (log) linear kinetic models. *Biotechnol Bioeng* **54**, 91–104.
- 12 Nielsen J (1997) Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates. *Biochem J* **321**, 133–138.
- 13 Heijnen JJ (2005) Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng* **9**, 534–545.

- 14 Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV *et al.* (2000) Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* **267**, 5313–5329.
- 15 Takahashi K, Yugi K, Hashimoto K, Yamada Y, Pickett CJF & Tomita M (2002) Computational challenges in cell simulation: a software engineering approach. *IEEE Intell Syst* **17**, 64–71.
- 16 Duarte NC, Herrgård MJ & Palsson BØ (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **14**, 1298–1309.
- 17 Förster J, Famili I, Fu P, Palsson BØ & Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* **13**, 244–253.
- 18 Reed JL, Famili I, Thiele I & Palsson BØ (2006) Towards multidimensional genome annotation. *Nature Rev Genet* **7**, 130–141.
- 19 Olivier BG & Snoep JL (2004) Web-based kinetic modelling using JWSONline. *Bioinformatics* **20**, 2143–2144. <http://jjj.biochem.sun.ac.za/>.
- 20 Visser D, van der Heijden R, Mauch K, Reuss M & Heijnen S (2000) Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. *Metab Eng* **2**, 252–275.
- 21 Kell DB (2004) Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **7**, 296–307.
- 22 Hofmeyr JHS, Kacser H & van der Merwe KJ (1986) Metabolic control analysis of moiety-conserved cycles. *Eur J Biochem* **155**, 631–641.
- 23 Reder C (1988) Metabolic control theory: a structural approach. *J Theor Biol* **135**, 175–201.
- 24 Ehldel M & Zacchi G (1996) A general formalism for metabolic control analysis. *Chem Eng Sci* **52**, 2599–2606.
- 25 Penrose R (1955) A generalized inverse for matrices. *Proc Cambridge Phil Soc* **51**, 406–413.
- 26 Murray JD (2002) *Mathematical Biology I. An Introduction*, 3rd edn. Springer-Verlag, New York, NY.
- 27 Golub GH & Van Loan CF (1996) *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore, MD.
- 28 Westerhoff HV, Hellingwerf KJ & van Dam K (1983) Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proc Natl Acad Sci USA* **80**, 305–309.

Supplementary material

The following supplementary material is available online:

Table S1. Metabolite abbreviations used in Teusink *et al.* [14].

Table S2. Reaction abbreviations used in Teusink *et al.* [14].

This material is available as part of the online article from <http://www.blackwell-synergy.com>

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.