Modular Dynamic Biomolecular Modelling with Bond Graphs: The Unification of Stoichiometry, Thermodynamics, Kinetics and Data

Peter J. Gawthrop*[1], Michael Pan[1,2], and Edmund J. Crampin[1,2]

[1] Systems Biology Laboratory, School of Mathematics and Statistics, and Department of Biomedical Engineering, University of Melbourne, Victoria 3010, Australia.
[2]ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, School of Chemical and Biomedical Engineering, University of Melbourne, Victoria 3010, Australia

June 7, 2021

Abstract

Renewed interest in dynamic simulation models of biomolecular systems has arisen from advances in genome-wide measurement and applications of such models in biotechnology and synthetic biology. In particular, genome-scale models of cellular metabolism beyond the steady state are required in order to represent transient and dynamic regulatory properties of the system. Development of such whole-cell models requires new modelling approaches. Here we propose the energy-based bond graph methodology, which integrates stoichiometric models with thermodynamic principles and kinetic modelling. We demonstrate how the bond graph approach intrinsically enforces thermodynamic constraints, provides a modular approach to modelling, and gives a basis for estimation of model parameters leading to dynamic models of biomolecular systems. The approach is illustrated using a well-established stoichiometric model of *Escherichia coli* (*E. coli*) and published experimental data.

*Corresponding author. **peter.gawthrop@unimelb.edu.au**

Contents

1 Introduction **3**

2 Bond Graphs **4**
2.1 Basic components 4
2.2 Chemostats, Flowstats and Pathways 7

3 Bond Graphs Integrate Stoichiometry and Energy **8**
3.1 Generating a bond graph from a stoichiometric matrix 8
3.2 Modularity 9
3.3 Energy Balance Analysis (EBA) in a bond graph context 11

4 Application to the *E. coli* Core Model **11**
4.1 Glycolysis & Pentose Phosphate pathway 13
4.2 Respiration 14
4.2.1 Analysis of individual modules 14
4.2.2 Analysis of combined network 14

5 Dynamic Modelling and Parameter Estimation **16**
5.1 Species potentials 16
5.2 Pathway flows 17
5.3 Reaction constants 18
5.4 Dynamical parameters 19
5.5 Parameters for the Glycolysis & Pentose Phosphate model 19
5.6 Simulation 21

6 Conclusion **21**

A EBA examples **32**
A.1 Example: Parallel reactions 32
A.2 Example: three-reaction cycle 33

B Pathways: Illustrative example **34**

C Glycolysis & Pentose Phosphate Pathways **35**

D Modular representation of Metabolism: Reactions **36**
D.1 Glycolysis 37
D.2 TCA cycle 37
D.3 Electron Transport Chain 38
D.4 ATP Synthase 38

E Enzyme-catalysed Reaction **38**

1 Introduction

The recent explosion of omics data has generated an interest in developing dynamic whole-cell models that account for the function of every gene and biomolecule over time. Such models have the potential to "predict phenotype from genotype" [1–3] and hence to "transform bioscience and medicine" [4]. Critical to understanding the large-scale metabolism within cells is the stoichiometric approach [5–8], which has had notable successes including the genome-scale reconstruction of the metabolism of *Escherichia coli* (*E. coli*) [9–11] and Neocallimastigomycota Fungus [12].

The stoichiometric approach can give rise to constraint-based models such as Flux Balance Analysis (FBA)[13], which predict metabolic fluxes at steady state. However, most implementations of such constraint-based models do not explicitly consider energy. This can lead to mass flows that are not thermodynamically possible because they violate the second law of thermodynamics. Such non-physical flows can be detected and eliminated by adding additional thermodynamic constraints, as in *Thermodynamics-based Metabolic Flux Analysis* (TFA)[14, 15], *Energy Balance Analysis* (EBA) and *Expression, Thermodynamics-enabled FLux models* (ETFL) [16–20] and loopless FBA [21]. Whereas constraint-based models provide metabolic fluxes, they generally do not explicitly account for metabolite concentrations, or how fluxes vary over time, both of which are required for dynamic whole-cell modelling. However, the stoichiometric approach can help to bridge towards dynamic models capable of satisfying these requirements. In this context, there has been work into developing two types of large-scale dynamic models: fully detailed mass action stoichiometric simulation (MASS) models [7, 22, 23] and simplified network models that use non-mass action rate laws such as lin-log laws or modular rate laws [24, 25]. Although mass-action approaches seem restrictive, we note that models of enzyme kinetics can be built from elementary mass-action reactions [26].

MASS models are parameterised by reaction rate constants which are subject to thermodynamic constraints such as the Wegscheider conditions [27] (Wegscheider conditions are a formulation of *detailed balance* conditions which avoid models which are inconsistent with thermodynamic laws [26, § 1.5]). This paper focuses on the mass-action formulation and introduces an alternative to MASS which explicitly incorporates thermodynamics. Specifically, the approach uses an alternative parameterisation related to that of Thermodynamic-Kinetic Modelling (TKM) [27, 28]. TKM explicitly divides parameters into those associated with capacities and resistances by analogy with electrical systems; this approach gives thermodynamic consistency without invoking additional constraints such as the Wegscheider conditions [27, 28]. Mason and Covert[29] developed a similar approach for a non-mass-action rate law.

Recently, the bond graph approach from engineering [30–33] has been adapted to biochemistry [34–37]. Bond graphs are close in spirit and application to TKM in that they produce ordinary differential equations for dynamic simulation [37] and that their parameters satisfy thermodynamic consistency without the need to invoke Wegscheider conditions [34–37]. However, bond graph models are endowed with several additional features:

1. Bond graphs can be easily generalised to model multi-physics systems and thus readily incorporate the physics of electrically charged species into an integrated model combining both chemical and electrical potential [38–42].

2. Bond graphs are modular [43, 44], a key requirement of any large-scale modelling endeavour [45].

3. Bond graph models can be systematically modified to give simpler bond graph models which remain compatible with thermodynamic laws [37, 46, 47].

The stoichiometric matrix of a biomolecular network can be derived from the corresponding bond graph [37, 43]. Similarly, as shown herein, a bond graph model can be constructed from a stoichiometric matrix. Thus, the large repository of models of biomolecular systems available in stoichiometric form are available as templates for developing bond graph models; we provide a methodology for this later in the paper. Furthermore, once rate laws such as mass action are added, such templates provide a basis for complete dynamic models of metabolic systems.

A key challenge in the development of dynamic models is the fitting of parameters to experimental data, especially when thermodynamic constraints need to be satisfied [48, 49]. For large-scale biomolecular models such as whole cell models, applying these constraints is particularly challenging [50]. In this paper, we use the thermodynamically safe parameterisation provided by bond graphs to resolve this issue. As in the TKM [27, 28] approach, the bond graph approach uses an alternative parameterisation which satisfies thermodynamic constraints *as long as the parameters are positive*; such inequality constraints are easier to handle than non-linear constraints. We illustrate this approach by generating a dynamic bond graph model of *E. coli* metabolism, using a well-established stoichiometric model [51] as a template and show that the use of thermodynamic parameters can significantly streamline the process of parameter estimation.

In summary, this paper proposes the fusion of the stoichiometric and bond graph approaches to modelling biological systems and illustrates its potential for the unification of stoichiometry, thermodynamics, kinetics and data.

§ 2 summarises the bond graph background to the rest of the paper. § 3 shows how bond graph models can be extracted from stoichiometric information, used to create modular models and analysed in terms of pathways; the relationship of the approach to Energy Balance Analysis is also discussed. § 4 applies these concepts to two subsystems within the *E. coli* core model – a well-documented [8, 51] and readily-available [52] stoichiometric model of a biomolecular system. § 5 shows how thermodynamically-consistent bond graph parameters can be extracted from experimental data and gives a dynamic simulation of the parameterised model. § 6 concludes the paper and gives directions for future work.

2 Bond Graphs

This section gives a brief introduction to the bond graph approach to modelling biomolecular systems based on the seminal work of Oster et al. [34, 35] as extended by Gawthrop and Crampin [37, 44, 53].

2.1 Basic components

Bond graphs represent the energetic connections between components of a system. The $\rightharpoonup$ symbol is used to indicate an energetic connection, or 'bond', between components; the half-arrow indicates the direction corresponding to positive energy flow. In the biomolecular context, each bond is associated with two covariables: chemical potential μ $(\mathrm{J\,mol^{-1}})$ and flow v $(\mathrm{mol\,s^{-1}})$. The key point is that the product of μ and v is power $p = \mu v$ (W). This ensures that models are consistent with the laws of thermodynamics, as energy flow is explicitly accounted for. In the context of cellular metabolism, and in line with the measurement of redox potentials, it is convenient to scale these co-variables by *Faraday's constant* $F \approx 96\,485\,\mathrm{C\,mol^{-1}}$ to give

$$\phi = \frac{1}{F}\mu \ \ (\mathrm{V}) \qquad\qquad f = Fv \ \ (\mathrm{A}) \tag{1}$$

where $(\mathrm{J\,C^{-1}})$ has been replaced by the more convenient unit volt (V) and $(\mathrm{C\,s^{-1}})$ has been replaced by the more convenient unit ampere (A) [38]. As a useful rule-of-thumb, μ $(\mathrm{kJ\,mol^{-1}})$ can be converted to ϕ (mV) by dividing by $\frac{10^6}{F} \approx 10$.

Bonds transmit, but do not store or dissipate energy. Within this context, the bonds connect four distinct types of component:

0 & 1 Junctions provide a method of connecting two or more bonds, and therefore creating a network. Analogous to electrical systems, there are two types of junction, denoted **0** and **1** . The bonds impinging on a **0** junction share a common effort (chemical potential); the bonds impinging on a **1** junction share a common flow. Both **0** and **1** junctions transmit, but do not store or dissipate energy. As discussed previously [37], the arrangement of bonds and junctions represents the stoichiometry of the corresponding biomolecular system and thus the relationship both between reaction and species flows and between species potentials and reaction forward and reverse potentials. Furthermore, the reverse is also true: the stoichiometric matrix of a biomolecular system uniquely determines the bond graph, as will be discussed further below.

Ce represents biochemical *species*. Thus species A is represented by **Ce**:**A** with the equations:

$$x_A(t) = \int_0^t f_A(t')dt' + x_A(0) \tag{2}$$

$$\phi_A = \phi_A^{\ominus} + \phi_N \ln \frac{x_A}{x_A^{\ominus}} = \phi_N \ln K_A x_A \tag{3}$$

$$\text{where } \phi_N = \frac{RT}{F} \approx 26.7 \text{ mV at } T = 310 \text{ K} \tag{4}$$

$$\text{and } K_A = \frac{1}{x_A^{\ominus}} \exp \frac{\phi_A^{\ominus}}{\phi_N} \tag{5}$$

Equation (2) accumulates the flow f_A of species A. Equation (3) generates chemical potential ϕ_A in terms of the reference potential $\phi_A^{\ominus}$ at reference conditions $x_A^{\ominus}$. **Ce** components thus store, but do not dissipate, energy. An equivalent parameterisation that we use in this paper is to express the chemical potential in terms of ϕ_N and the species constant K_A, as defined in Equation (5).

Re represents *reactions*. The flow f associated with each reaction is given by the *Marcelin – de Donder* formula [37, 54]:

$$f = \kappa \left(\exp \frac{\Phi^f}{\phi_N} - \exp \frac{\Phi^r}{\phi_N} \right) \tag{6}$$

where Φ^f and Φ^r are the forward and reverse reaction potentials (or affinities), defined as the sums of the chemical potentials of the reactants and products respectively. If κ is constant, this represents the mass-action formula.

In general, κ is a function of Φ^f, Φ^r and enzyme concentration [37]; for example, a reversible Michaelis-Menten formulation used in Gawthrop et al. [43] is:

$$\kappa = \frac{f_{max}}{K_f + (1-\rho) \exp \frac{\Phi^f}{\phi_N} + \rho \exp \frac{\Phi^r}{\phi_N}} \tag{7}$$

where the three constants f_{max}, K_f and ρ define the kinetics. As discussed elsewhere [26, 37], enzyme kinetics can be modelled using the pair of reactions with mass-action kinetics

$$\begin{aligned} \text{A + E} &\overset{\text{r}_1}{\rightleftharpoons} \text{C} \\ \text{C} &\overset{\text{r}_2}{\rightleftharpoons} \text{B + E} \end{aligned} \tag{8}$$

where A, B, E and C are the substrate, product, enzyme and complex of substrate and enzyme respectively; the bond graph representation is given in Appendix E. Equation (7) arises from the steady-state analysis of this model [37]. In particular:

$$\kappa_1 = \frac{\bar{\kappa}}{\rho}; \qquad \kappa_2 = \frac{\bar{\kappa}}{1-\rho} \text{ where } \bar{\kappa} = \frac{f_{max}}{K_c e_0} \text{ and } K_f = \frac{K_C}{K_E} \tag{9}$$

K_C and K_E correspond to Equation (5) for the complex and enzyme respectively and e_0 is the total amount of enzyme (unbound and bound within the complex).

Re components dissipate, but do not store, energy. In general

$$f = f(\Phi, \phi) \tag{10}$$

where $\Phi = \Phi^f - \Phi^r$ and ϕ is a vector containing the chemical potentials of every species. Since f always has the same sign as Φ, $f()$ is dissipative in Φ for all ϕ:

$$f\Phi > 0 \tag{11}$$

The key *stoichiometric* equations arising from bond graph analysis are [37]:

$$\dot{x} = Nf \tag{12}$$

$$\Phi = -N^T\phi \tag{13}$$

where x, f, Φ and ϕ are the species amounts, reaction fluxes, reaction potentials and species potentials respectively, all represented as vector quantities. N is the stoichiometric matrix of the network. Combining Equations (12) and (13):

$$\phi^T\dot{x} = \phi^T N f = -\Phi^T f \tag{14}$$

$\phi^T\dot{x}$ is the rate of energy into the species (which must be negative or zero for closed systems) and $\Phi^T f$ is the rate of energy dissipated by the reactions. Since $\phi^T\dot{x} + \Phi^T f = 0$, it follows that the network of bonds and junctions transmits, but does not dissipate or store, energy [37].

Moreover, the stoichiometric matrix N can be decomposed as [37]:

$$N = N^r - N^f \tag{15}$$

where N^r corresponds to the positive entries of N and N^f to the negative entries. The forward and reverse reaction potentials Φ^f and Φ^r are given by:

$$\Phi^f = N^f\phi, \quad \Phi^r = N^r\phi \tag{16}$$

In other words, the stoichiometric matrix N can be derived from the system bond graph. § 3 shows that, conversely, the system bond graph can be derived from the stoichiometric matrix N.

2.2 Chemostats, Flowstats and Pathways

Modularity implies the interconnection of subsystems; thus such subsystems must be thermodynamically open. As discussed previously [38, 44], the notion of a *chemostat* [55] is useful in creating an open system from a closed system. The chemostat has a number of interpretations [38]:

1. one or more species are fixed to give a constant concentration [43]; this implies that an appropriate external flow is applied to balance the internal flow of the species.
2. as a **Ce** component with a fixed state.
3. as an external *port* of a module which allows connection to other modules.

In the context of stoichiometric analysis, the chemostat concept provides a flexible alternative to the primary and currency exchange reactions [6, 8, 56].

Alternatively, reaction flows can be fixed using the dual concept of *flowstats* [44], which has a number of interpretations:

1. as an **Re** component with a fixed flow.
2. as an external *port* of a module which allows connection to other modules.

In the context of this paper, we use flowstats to isolate parts of a network by setting the flows of certain reactions to zero. Such zero flow flowstats can also be interpreted as removing the corresponding enzyme via gene knockout.

In terms of stoichiometric analysis, the closed system equations (12) and (13) are replaced by:

$$\dot{x} = N^{cd} f \tag{17}$$

$$\Phi = -N^T \phi \tag{18}$$

where N^{cd} is created from the stoichiometric matrix N by setting *rows* corresponding to chemostats species and *columns* corresponding to flowstatted reactions to zero [44]. As discussed by Gawthrop and Crampin [44], system pathways corresponding to Equation (17) are defined by the right-null space of N^{cd}, that is, the columns of a matrix K_p satisfying the equation $N^{cd} K_p = 0$. At steady state, the flows through these pathways are defined by:

$$f = K_p f_p \tag{19}$$

where f_p is the pathway flow. It follows from Equation (17) that Equation (19) implies that $\dot{x} = 0$. The *pathway* stoichiometric matrix N_p is defined as [53]:

$$N_p = N K_p \tag{20}$$

In a similar fashion to equation (18), the pathway reaction potentials Φ_p are given by

$$\Phi_p = -N_p^T \phi \tag{21}$$

In the same way as the stoichiometric matrix N relates reaction flows to species and thus represents a set of reactions, the pathway stoichiometric matrix N_p also represents a set of reactions: these reactions will be called the *pathway reactions*.

Pathways can be divided into three mutually exclusive types [56] according to the species corresponding to the non zero elements in the relevant column of the *pathway* stoichiometric matrix N_p:

Type I The species include primary metabolites; these pathways are of functional interest.

Type II The species include currency metabolites only; these pathways dissipate energy without creating or consuming primary metabolites. Such pathways are sometimes called *futile cycles*; however they have an important role to play in regulating metabolite flow [57–62].

Type III There are no species. These may arise when the same reaction is catalysed by different isoforms of the same enzyme.

Pathway reactions for type I pathways contain both primary and currency metabolites; pathway reactions for type II pathways contain currency metabolites only; pathway reactions for type III pathways are empty. The concept of pathways is applied to a simple example in Appendix B and to a biomolecular example (the Pentose Phosphate Pathway) in § 4.1.

3 Bond Graphs Integrate Stoichiometry and Energy

As discussed in the previous section, the stoichiometric matrix can be directly derived from the bond graph; this section shows that the converse is true and thus bond graphs can be automatically derived from preexisting stoichiometric representations thereby allowing bond graph energy based analysis and modularity to be applied to such models.

3.1 Generating a bond graph from a stoichiometric matrix

A bond graph can be constructed from a stoichiometric matrix by using the following procedure:

1. For each *species* create a **Ce** component with appropriate name and a **0** junction; connect a bond from the **0** junction to the **Ce** component.
2. For each *reaction* create an **Re** component with appropriate name and two **1** junctions; connect a bond from one **1** junction to the forward port of the **Re** component and a bond from the reverse port of the **Re** component to the other **1** junction.
3. For each *negative* entry N_{ij} in the stoichiometric matrix, connect $-N_{ij}$ bonds from the zero junction connected to the ith species to the the **1** junction connected to the forward port of the jth reaction.
4. For each *positive* entry N_{ij} in the stoichiometric matrix, connect N_{ij} bonds from the one junction connected to the reverse port of the jth reaction to the zero junction connected to the ith species.
5. If an Michaelis-Menten formulation is required, each **Re** component is replaced by a bond graph module (§ 3.2) corresponding to the enzyme catalysed reaction pair (8) and Appendix E.

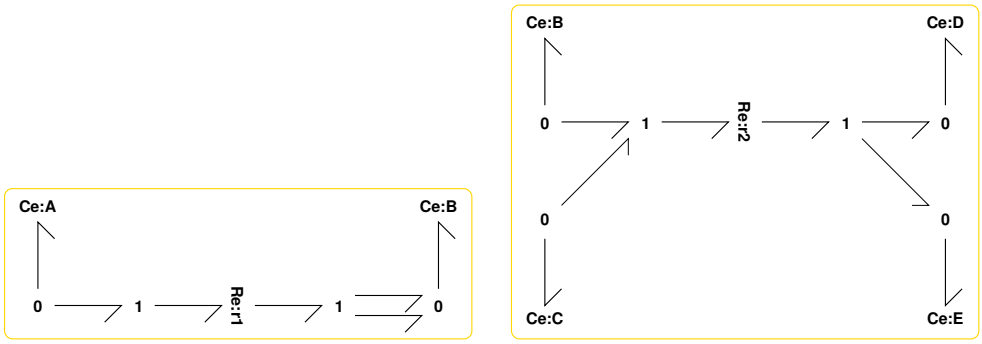
For example, the reaction $\mathrm{A} \overset{\mathrm{r_1}}{\rightleftharpoons} 2\,\mathrm{B}$ has the stoichiometric matrix

$$N = \begin{pmatrix} -1 & 2 \end{pmatrix}^T \tag{22}$$

and the bond graph of Figure 1(a). The reaction B + C $\overset{r_2}{\rightleftharpoons}$ D + E has the stoichiometric matrix

$$N = \begin{pmatrix} -1 & -1 & 1 & 1 \end{pmatrix}^T \tag{23}$$

and has the bond graph of Figure 1(b).



(a) A $\overset{r_1}{\rightleftharpoons}$ 2 B (b) B + C $\overset{r_2}{\rightleftharpoons}$ D + E

Figure 1: Bond graphs of simple reactions. (a) and (b) are used as modules M1 and M2 in § 3.2.

Bond graphs provide a graphical representation of a system. While this provides an intuitive and clear visual representation when dealing with small systems such as the ones shown above, such visualisation becomes cumbersome for large systems. We employ two approaches to overcome this issue for the large-scale systems considered in this paper: modularity and a non-graphical (or programmatic) representation. In particular, we use a recent concept of bond graph modularity [38] in § 3.2 and the recently developed BondGraphTools package[63] (`https://pypi.org/project/BondGraphTools/`) as a non-graphical representation that allows large-scale systems to be constructed in a scalable and automated manner. This is discussed further below.

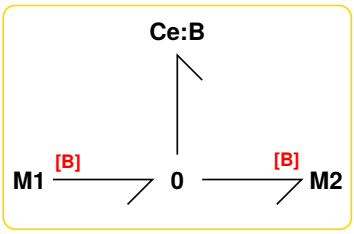
3.2 Modularity

Two related but distinct concepts of modularity [44] are *computational modularity*, where physical correctness is retained, and *behavioural modularity*, where module behaviour (such as ultrasensitivity) is retained. Here we discuss computational modularity. In particular, it is shown how the concept of external flows, as discussed in § 2.2, is key to bond graph modularity.

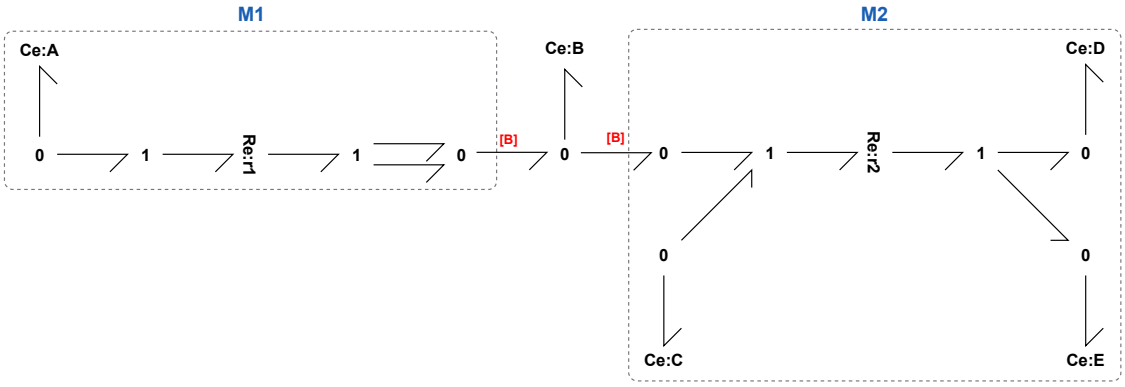
Modular bond graphs provide a way of decomposing complex biomolecular systems into manageable subsystems [43, 44, 53]. This paper combines the modularity concepts of Neal et. al. [64–66] with the bond graph approach to give a more flexible approach to modularity. The basic idea is simple [38] . Modules are self-contained and have no explicit ports, but any species represented by a **Ce** component has the potential to become a port available for external connection. Thus, if two modules share the same species, the corresponding **Ce** component in each module is replaced by a port (labelled with the same name), and the species is explicitly represented as a **Ce** component in the parent model. This approach allows each module to be individually tested prior to being integrated into a larger model.

We use the following algorithm to merge bond graph models of stoichiometric networks:

1. Within each module, each **Ce** component corresponding to a common species is *exposed*, that is, replaced by a *port*, or external connection.
2. For each common species, create a **Ce** component connected to a **0** component.
3. Connect all module ports associated with each species to the **0** junction associated with the species; all instances of **Ce** components corresponding to each species are thus *unified* into the same component.



(a) Modular bond graph



(b) Equivalent bond graph

Figure 2: Modularity. Modules M1 and M2 correspond to Figures 1(a) & 1(b) respectively. The common species B is exposed as a port in each module and connected to the new **Ce**:B component via a **0** junction. (a) shows the compact modular form and (b) contains equivalent bond graph when the contents of the modules are expanded.

For example, let modules M1 and M2 correspond to Figures 1(a) & 1(b) respectively. The composition of these modules requires the common species B to be exposed in both modules. This is illustrated in Figure2, where both modules are connected to the new **Ce**:B component via a **0** junction. The composite system contains the two coupled reactions:

$$\mathrm{A} \underset{}{\overset{\mathrm{r_1}}{\rightleftharpoons}} 2\,\mathrm{B} \tag{24}$$

$$\mathrm{C} + \mathrm{B} \overset{\mathrm{r_2}}{\rightleftharpoons} \mathrm{D} + \mathrm{E} \tag{25}$$

§ 4 gives examples of modular decomposition of a metabolic system and § 4.2 gives an example of how such modules can be combined using the methods of this section. The pathway analysis of § 2.2

can be applied to modules themselves, and to systems built of modules, to give insight into the overall behaviour of complex systems; this is illustrated in § 4.2.

The concept of modularity can be extended to include common **Re** (reaction) components [67]; but this concept is not pursued in this paper.

3.3 Energy Balance Analysis (EBA) in a bond graph context

FBA [13] uses the linear equation (19) within a constrained linear optimisation to compute pathway flows. EBA [16] adds two sorts of nonlinear constraint arising from thermodynamics. This section shows that the bond graph approach automatically includes the EBA constraint equations by considering Inequality (11) and Equation (18). In particular:

1. Inequality (11) corresponds to Equation 8 of Beard et al. [16]. This inequality can be re-expressed as:

$$\Phi_i = r_i(\phi) f_i \text{ where } r_i(\phi) > 0 \tag{26}$$

   r_i corresponds to the "flux resistances" on p.83 of Beard and Qian [16].

2. If K is the right nullspace matrix of N, it follows from Equation (18) that

$$K^T \Phi = 0 \tag{27}$$

   This corresponds to Equation 7 of Beard et al [16]. Note that K defines the pathways of the closed system system, with no chemostats.

Moreover, the pathways of the open system as defined by K^{cd} can be considered by defining $R = \text{diag } r_i$ and using Equation (19):

$$\boxed{K^T R K_p f_p = 0} \tag{28}$$

Equation (28) and inequality (26) constrain the pathway flows f_p. This is illustrated in Appendix A.

4 Application to the *E. coli* Core Model

The *E. coli* Core Model [8, 51] is a well-documented and readily-available stoichiometric model of a biomolecular system; species, reactions and stoichiometric matrix were extracted from the CobraPy model: "textbook". Using the methods of § 3.1, the corresponding bond graph model was created which, as discussed in the Introduction, automatically satisfies thermodynamic constraints.

To illustrate the concepts developed above, we analyse two subsets of reactions within this model:

1. § 4.1 uses the methods of § 2.2 to examine possible pathways within the system formed from the combined Glycolysis & Pentose Phosphate pathway (which produces precursors to the synthesis of nucleotides).

2. § 4.2 uses the modularity approach of § 3.2 to build a modular model of respiration using Glycolysis, the TCA cycle, the Electron Transport Chain and ATP synthase as modules. Furthermore, the methods of § 2.2 are applied to examine the pathway properties of an individual module (the TCA cycle) as well as the overall system.

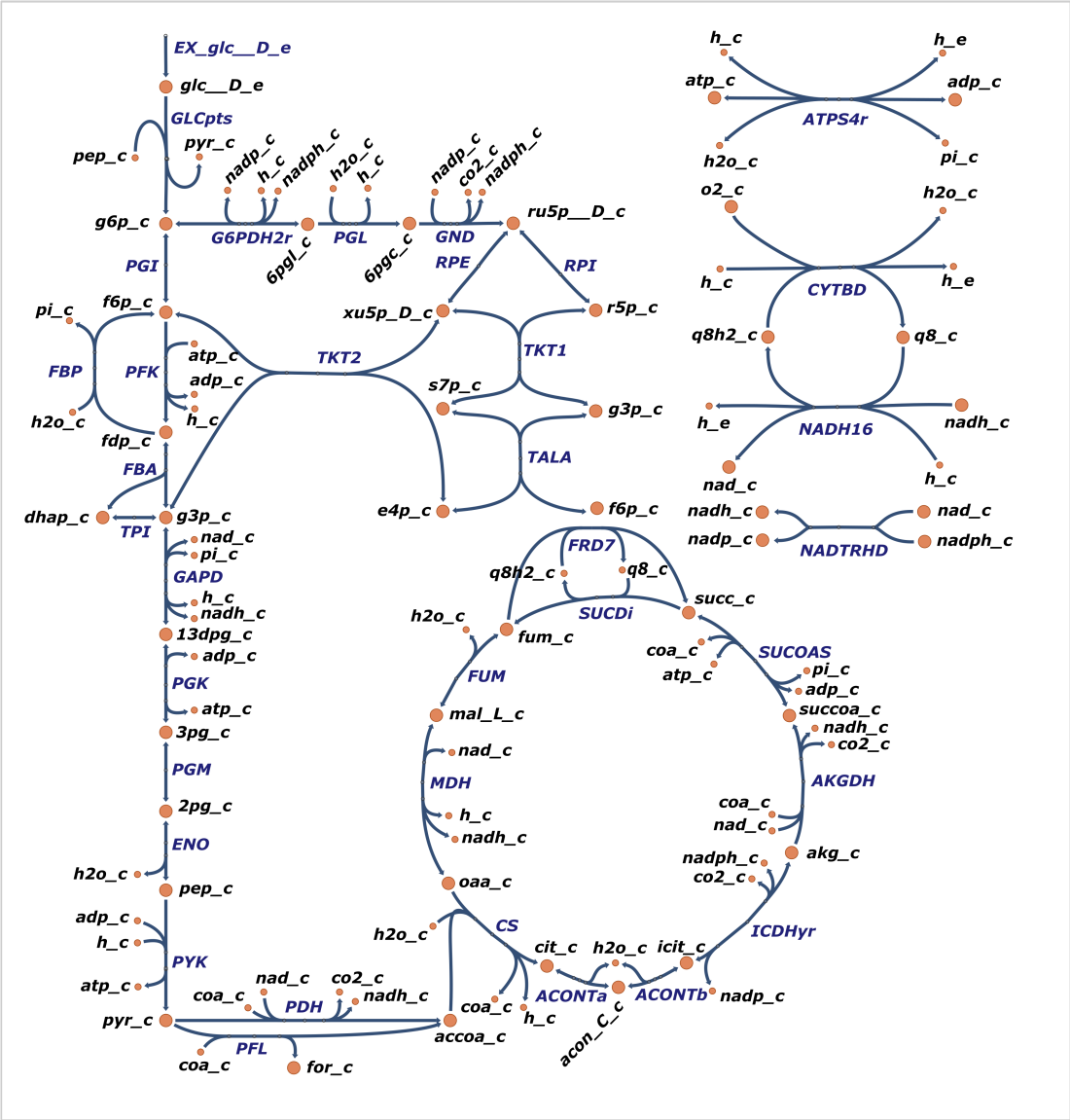


Figure 3: *E. coli* Core Model. The extracted reactions corresponding to the Glycolysis, Pentose-Phosphate Pathways and TCA cycle parts of the model are shown; a complete list of reactions is given in Appendix D. The diagram was created using Escher [68].

4.1 Glycolysis & Pentose Phosphate pathway

The combination of the Glycolysis & Pentose Phosphate networks provides a number of different products from the metabolism of glucose. This flexibility is adopted by proliferating cells, such as those associated with cancer, to adapt to changing requirements of biomass and energy production [69, 70].

We construct a stoichiometric model of these pathways, consisting of the upper reactions of glycolysis and the pentose phosphate pathway. The full reaction network is given in Appendix C, and a bond graph is constructed using the methods of § 3.1.

As discussed in the textbooks [61, 71], it is illuminating to pick out individual paths through the network to see how these may be utilised to provide a variety of products. This is reproduced here by choosing appropriate chemostats and flowstats (§ 2.2) to give the results listed by Garrett and Grisham [61] § 22.6d. In each case, the corresponding pathway reaction potential is given. For consistency with Garrett and Grisham [61] § 22.6d, each pathway starts with Glucose 6-phosphate (G_6P).

We use the following list of chemostats (together with additional chemostats) for the pathway analysis below: {ADP, ATP, CO_2, G_6P, H, H_2O, NAD, NADH, NADP, NADPH, PI, PYR}. The pathways are generated using the methods of § 2.2.

1. R_5P & NADPH generation

   Chemostats: RP5

   Flowstats: PGI, TKT2

   Pathway: G6PDH2R + PGL + GND + RPI

   Reaction: $G_6P + H_2O + 2\,NADP \rightleftharpoons CO_2 + 2\,H + 2\,NADPH + R_5P$

2. R_5P generation

   Chemostats: RP5

   Flowstats: GAPD, G6PDH2R

   Pathway: - 5 PGI - PFK - FBA - TPI - 4 RPI + 2 TKT2 + 2 TALA + 2 TKT1 + 4 RPE

   Reaction: $ADP + H + 6\,R_5P \rightleftharpoons ATP + 5\,G_6P$

3. NADPH generation

   Chemostats: None

   Flowstats: GAPD

   Pathway: - 5 PGI - PFK - FBA - TPI + 6 G6PDH2R + 6 PGL + 6 GND + 2 RPI + 2 TKT2 + 2 TALA + 2 TKT1 + 4RPE

   Reaction: $ADP + G_6P + 6\,H_2O + 12\,NADP \rightleftharpoons ATP + 6\,CO_2 + 11\,H + 12\,NADPH$

In § 5, we use the model of the Glycolysis & Pentose Phosphate pathways as a basis for inferring parameters from experimental data. Once the parameters have been identified (§ 5.4), dynamic simulations of these pathways can be run. This is shown later in § 5.6.

4.2 Respiration

To illustrate the utility of using bond graphs for the modular construction of stoichiometric models, we construct a model of respiration by combining the subsystems of Glycolysis, TCA cycle, Electron Transport Chain and ATP Synthase. Reactions for each of these subnetworks were extracted from the CobraPy model; these reactions are listed in Appendix D. For simplicity, reactions PDH and PFL (converting PYR to ACCOA) and reaction NADTRHD (converting NADP/NADPH to NAD/NADH) were included in the TCA cycle module. Once these are converted into bond graphs, the algorithm in § 3.2 was used to combine these models together into a model of respiration.

4.2.1 Analysis of individual modules

An advantage of considering subsystems as separate modules is that these modules can be analysed individually. For example, the TCA cycle module can be analysed using the set of chemostats (see § 2.2):

$$\{\mathrm{PYR}, \mathrm{CO_2}, \mathrm{ADP}, \mathrm{ATP}, \mathrm{H_2O}, \mathrm{NAD}, \mathrm{NADH}, \mathrm{PI}, \mathrm{NADP}, \mathrm{NADPH}, \mathrm{H}, \mathrm{Q_8}, \mathrm{Q_8H_2}, \mathrm{FOR}\}$$

Three pathways result from this analysis:

1. **- FRD7 + SUCDI**
   This is a type III pathway with no overall reaction.

2. **CS + ACONTA + ACONTB + ICDHYR + AKGDH + SUCOAS + FRD7 + FUM + MDH + PDH**
   This is a type I pathway with the reaction

$$\begin{aligned} &\mathrm{ADP} + 2\,\mathrm{H_2O} + 3\,\mathrm{NAD} + \mathrm{NADP} + \mathrm{PI} + \mathrm{PYR} + \mathrm{Q_8} \\ &\quad \rightleftharpoons \mathrm{ATP} + 3\,\mathrm{CO_2} + 2\,\mathrm{H} + 3\,\mathrm{NADH} + \mathrm{NADPH} + \mathrm{Q_8H_2} \end{aligned}$$

3. **CS + ACONTA + ACONTB + ICDHYR + AKGDH + SUCOAS + FRD7 + FUM + MDH + PFL**
   This is a type I pathway with the reaction

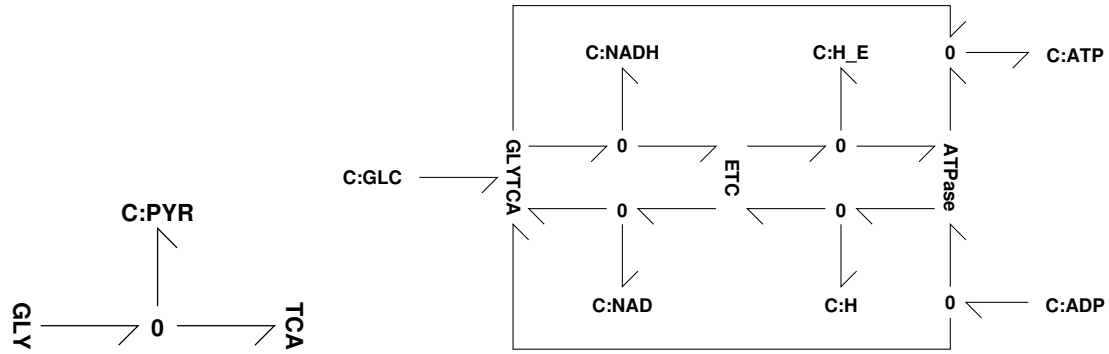
$$\begin{aligned} &\mathrm{ADP} + 2\,\mathrm{H_2O} + 2\,\mathrm{NAD} + \mathrm{NADP} + \mathrm{PI} + \mathrm{PYR} + \mathrm{Q_8} \\ &\quad \rightleftharpoons \mathrm{ATP} + 2\,\mathrm{CO_2} + \mathrm{FOR} + 2\,\mathrm{H} + 2\,\mathrm{NADH} + \mathrm{NADPH} + \mathrm{Q_8H_2} \end{aligned}$$

Pathways 2 and 3 utilise the potential of PYR to generate NADH, NADHP, ATP and Q_8H_2 whilst releasing CO_2 and H.

4.2.2 Analysis of combined network

The bond graph approach provides a method for easily combining stoichiometric models using the methods of § 3.2. Here we demonstrate this by constructing a model of respiration from the individual modules Glycolysis, TCA cycle, Electron Transport Chain and ATP Synthase. We begin by first combining the Glycolysis and TCA modules, as indicated in Figure 4(a). As well as the common species PYR (pyruvate) explicitly shown, the set of species

$$\{\mathrm{ATP}, \mathrm{ADP}, \mathrm{PI}, \mathrm{H}, \mathrm{NAD}, \mathrm{NADH}, \mathrm{H_2O}\}$$



(a) Glycolysis-TCA module (GLYTCA)

(b) Modular Metabolism

Figure 4: Modularity. (a) The two modules **GLY** (Glycolysis) and **TCA** (TCA cycle) each contain a bond graph representation of the relevant reactions. As discussed in § 3.2, they are combined into a single module by combining common species; in this case PYR is shown explicitly – other common species are {ATP, ADP, PI, H, NAD, NADH, H_2O}. (b) The three modules **GLYTCA** (containing the two modules **GLY** and **TCA**), **ETC** and **ATP Synthase** are combined by unifying common species. This is shown for principle common species and emphasises that **ETC** is powered by NADH from **GLYTCA**, **ATP Synthase** is powered by the external protons H_E and both **GLYTCA** and **ATP Synthase** generate ATP from ADP. Common species not explicitly shown are { PI, H_2O, Q_8, Q_8H_2}.

were also declared to be common.

The full model of respiration is then constructed by combining the Glycolysis+TCA cycle module with the Electron Transport Chain and ATP Synthase modules, as indicated in Figure 4(b). In addition to the common species explicitly shown

$$\{\mathrm{ATP}, \mathrm{ADP}, \mathrm{H}, \mathrm{H_E}, \mathrm{NAD}, \mathrm{NADH}\}$$

the set of species

$$\{\mathrm{PI}, \mathrm{H_2O}, \mathrm{Q_8}, \mathrm{Q_8H_2}\}$$

were also declared to be common.

To analyse this overall module, the chemostats were chosen to be:

$$\{\mathrm{GLCD_E}, \mathrm{CO_2}, \mathrm{O_2}, \mathrm{ADP}, \mathrm{ATP}, \mathrm{H_2O}, \mathrm{PI}, \mathrm{H}\}$$

Using the methods of § 2.2, the three pathways in this network are

1. **PFK + FBP**
   This is a type II pathway with the overall reaction

   $$\mathrm{ATP} + \mathrm{H_2O} \overset{\mathrm{P_1}}{\rightleftharpoons} \mathrm{ADP} + \mathrm{PI} + \mathrm{H}$$

   This Futile Cycle has regulatory implications[62].

2. **-FRD7 + SUCDI**
   This is a type III pathway with no overall reaction.

3. **2 GLCPTS + 2 PGI + 2 PFK + 2 FBA + 2 TPI + 4 GAPD + 4 PGK - 4 PGM + 4 ENO + 2 PYK + 4 PDH + 4 CS + 4 ACONTA + 4 ACONTB + 4 ICDHYR + 4 AKGDH + 4 SUCOAS + 4 FRD7 + 4 FUM + 4 MDH + 4 NADTRHD + 20 NADH16 + 12 CYTBD + 27 ATPS4R**
   This is a type I pathway with the reaction

$$2\,\mathrm{GLCD_E} + 12\,\mathrm{O_2} + 35\,\mathrm{ADP} + 35\,\mathrm{PI} + 35\,\mathrm{H} \xleftrightarrow{\mathrm{P_3}} 12\,\mathrm{CO_2} + 35\,\mathrm{ATP} + 47\,\mathrm{H_2O}$$

Pathway 3 corresponds to the metabolic generation of ATP using the free energy of $\mathrm{GLCD_E}$. The ratio of ATP to $\mathrm{GLCD_E}$ is 17.5; this is the value quoted by Palsson [8] § 19.2.

5 Dynamic Modelling and Parameter Estimation

Dynamic models of biochemical networks have the potential to aid the understanding of how subprocesses change over time, and can potentially elucidate important control structures within these networks [72]. However, due to their nonlinear nature, parameter estimation is one of the most challenging aspects of developing models of biomolecular systems [73].

Parameter estimation depends on both the form of the model and the type of data available. This section assumes a bond graph model with the mass-action kinetics of Equation (6) and that the following data are available for a single steady-state condition:

1. Reaction potentials Φ (equivalent to reaction Gibbs free energy).
2. Reaction flows f.
3. Species concentration c.

If data at three or more steady-state conditions were available, more complex kinetics such as the reversible Michaelis-Menten formulation (7) could be used but this is not pursued in this paper.

In recent times, such data are becoming more readily available; species concentrations can be obtained from metabolomics data, and tracer experiments involving ^{13}C and ^{2}H have been used to infer both fluxomics data for reaction flows [74, 75] and thermodynamic data for reaction potentials [74, 76, 77]. In the following examples, we make use of the dataset obtained by Park et al. [74] to infer the thermodynamic parameters using a relatively fast quadratic programming (QP) algorithm.

Because bond graph models are thermodynamically consistent, the estimated parameters have physical meaning and the resultant estimated model, though not necessarily correct, is physically plausible [47]. Moreover, physical constraints imply parametric constraints thus reducing the parameter search space.

5.1 Species potentials

Because of the energetic constraints implied by the bond graph, the reaction potentials Φ are related to the species potentials ϕ by Equation (13). Since some reaction potentials may be unavailable, we rearrange and partition Φ and the stoichiometric matrix N so that

$$N = \begin{bmatrix} N_0 & N_1 \end{bmatrix} \text{ and } \Phi = \begin{bmatrix} \Phi_0 \\ \Phi_1 \end{bmatrix} \tag{29}$$

where Φ_0 and Φ_1 contain the known and unknown values of Φ respectively.

Given the measured value of Φ_0 and the estimated species potentials $\hat{\phi}$, the *estimation error* ϵ is defined as:

$$\epsilon = \hat{\Phi}_0 - \Phi_0 = -N_0^T \hat{\phi} - \Phi_0 \tag{30}$$

$$\text{hence } \epsilon^2 = \hat{\phi}^T N_0 N_0^T \hat{\phi} + 2\Phi_0^T N_0^T \phi + \Phi_0^T \Phi_0 \tag{31}$$

where the hat notation denotes estimated quantities. Although Φ_1 is unknown, it is subject to the physical inequality (11). In this case, all of the measured flows are positive, hence inequality (11) can be combined with Equation (13) and rewritten as:

$$-N_1^T \phi > 0 \tag{32}$$

Equation (30) and inequality (32) can be embedded in a *quadratic program* (QP) [78]:

$$\text{minimise } \frac{1}{2}\hat{\phi}^T P \hat{\phi} + q^T \phi \text{ subject to } N_1^T \hat{\phi} < 0 \tag{33}$$

$$\text{where } P = N_0 N_0^T + \lambda I \text{ and } q = N_0 \Phi_0 \tag{34}$$

I is the $n_\phi \times n_\phi$ unit matrix and $\lambda > 0$ a small positive number. In some cases, there are more species than reactions and so the stoichiometric matrix N has more rows than columns. As a result, the number of species potentials ϕ is greater than the number of reaction potentials Φ and so Equation (13) has no unique solution for ϕ given Φ. Thus we use the λI term to which is required to turn a non-unique solution for ϕ into a minimum norm solution.

Having deduced a set of estimated species potentials $\hat{\phi}$ using the QP, the corresponding reaction potentials $\hat{\Phi}_0$ and $\hat{\Phi}$ can be obtained from Equation (13) rewritten as:

$$\hat{\Phi}_0 = -N_0^T \hat{\phi} \text{ and } \hat{\Phi}_1 = -N_1^T \hat{\phi} \tag{35}$$

Once again, $\hat{\Phi}_0 = \Phi_0$ and the other values of Φ can be deduced from (35); because of the inequality constraint in the QP, these values are positive and thus physically plausible.

QP also handles equality constraints [78]; this provides a potential mechanism for incorporating known parameters into the procedure.

5.2 Pathway flows

From basic stoichiometric analysis, steady-state flows f can be written in terms of the pathway matrix K_p and pathway flows f_p by Equation (19) repeated here as

$$f = K_p f_p \tag{36}$$

Note that, as discussed in § 2.2, the *pathway matrix* K_p is dependent on the choice of chemostats. In general, K_p has more rows than columns and thus the pathway flow f_p is over-determined by the reaction flows f. Hence, given a set of experimental flows f, an estimate $\hat{f}_p$ of f_p can be obtained from the *least-squares* formula:

$$(K_p^T K_p)\hat{f}_p = K_p^T f \tag{37}$$

Note that:

1. $(K_p^T K_p)$ is a square $n_p \times n_p$ matrix where n_p is the number of pathways.

2. If some flows are not measured, the corresponding rows of K_p are deleted.
3. The reaction flows (including the missing ones) can be estimated from $\hat{f} = K_p \hat{f}_p$.
4. From Equation (12), the estimated chemostat flows are given by the non-zero elements of

$$\hat{x} = N\hat{f} \tag{38}$$

5.3 Reaction constants

In terms of estimated quantities, the reaction flow of Equation (6) can be rewritten as:

$$\hat{f} = \hat{\kappa}\hat{f}_0 \text{ where } \hat{f}_0 = \hat{f}_0^+ - \hat{f}_0^- \tag{39}$$

$$\hat{f}_0^+ = \exp\frac{\hat{\Phi}^f}{\phi_N};\ \hat{f}_0^- = \exp\frac{\hat{\Phi}^r}{\phi_N} \text{ and } \hat{\Phi}^f = N^f\hat{\phi};\ \hat{\Phi}^r = N^r\hat{\phi} \tag{40}$$

For each reaction, the estimated reaction constant $\hat{\kappa}$ is then given by Equation (39).

Similarly, reversible Michaelis-Menten reaction kinetics can be written in terms of estimated quantities and three estimated parameters $\hat{f}_{max}$, $\hat{K}_f$ and $\hat{\rho}$ from Equation (7)

$$\hat{f} = \frac{\hat{f}_{max}\hat{f}_0}{\hat{K}_f + (1-\hat{\rho})\hat{f}_0^+ + \hat{\rho}\hat{f}_0^-} \tag{41}$$

This can be rearranged as:

$$\hat{f}_{max} + \hat{f}\ \hat{\rho} - \frac{\hat{f}}{\hat{f}_0}\ \hat{K}_f = \frac{\hat{f}}{\hat{f}_0}\hat{f}_0^+ \tag{42}$$

and can be in rewritten in linear-in-the-parameters form [79] as:

$$y = X\theta \tag{43}$$

$$\text{where } X = \begin{pmatrix} 1 & \hat{f} & -\frac{\hat{f}}{\hat{f}_0} \end{pmatrix} \tag{44}$$

$$\theta = \begin{pmatrix} f_{max} & \rho & K_f \end{pmatrix}^T \tag{45}$$

$$\text{and } y = \frac{\hat{f}}{\hat{f}_0}\hat{f}_0^+ \tag{46}$$

Given an estimate $\hat{\theta}$ of θ, the *estimation error* ϵ' is

$$\epsilon' = y - X\hat{\theta} \tag{47}$$

Because there are three unknown parameters ($\hat{f}_{max}$, $\hat{K}_f$ and $\hat{\rho}$), at least three different sets of steady-state data are required to uniquely determine the parameters; this case is not considered here. Alternatively, these unknown parameters can be determined using measured constants from the literature [29]. Such known parameters can be included using an *equality* constraint of the form $A_{ec}\hat{\theta} = b_{ec}$ – an example appears in § 5.5. Noting that all elements of θ are positive, $\hat{\theta}$ also has the *inequality constraint* $\hat{\theta} > 0$, the error equation (47) together with the constraints can be embedded *quadratic program* (QP) [78]:

$$\text{minimise } \frac{1}{2}\hat{\theta}^T P\hat{\theta} + q^T\hat{\theta} \text{ subject to } \hat{\theta} > 0 \text{ and } A_{ec}\hat{\theta} = b_{ec} \tag{48}$$

$$\text{where } P = X^T X + \lambda I;\ q = X^T y$$

I is the $n_\theta \times n_\theta$ unit matrix and $\lambda > 0$ a small number. The parameters of the equivalent bond graph model can be deduced using Equation (9).

More general reaction kinetics [29] can be incorporated in a straightforward manner but, however, would require non-linear fitting procedures to determine parameters.

5.4 Dynamical parameters

The parameter K of the species components (**Ce**) determines the time course of species amounts and reaction flows when there is a deviation from steady-state. Using Equation (5), this can be determined from the species potential estimate $\hat{\phi}$ and the amount of species $x^{\ominus}$ at the steady-state conditions. Expressing amounts per unit volume, it follows that $x^{\ominus} = c$, the species concentration at the steady-state conditions.

5.5 Parameters for the Glycolysis & Pentose Phosphate model

The bond graph of the Glycolysis & Pentose Phosphate model (§ 4.1) was parameterised to fit *E. coli* experimental data [74] using the approach described in this section. Table 2 [74] gives experimentally measured values of the reaction Gibbs energy ΔG for all of the reactions in the model except for G6PDH2R and PGL. The known values of ΔG were converted to reaction potentials $\Phi_0\ (\mathrm{mV})$. The unknown potentials Φ_1 were constrained to be greater than $1\,\mathrm{mV}$. The first column of Table 1(c) gives the experimental values of reaction potential Φ with the unknown values indicated by –; the second column gives the corresponding estimates $\hat{\Phi}\ (\mathrm{mV})$. The estimated and known values are identical; of the two estimated unknown values, that for PGL lies on the constraint – unconstrained optimisation gives physically impossible negative value.

As discussed in § 2.2, pathways are determined by chemostats. In this case it was assumed that the set of chemostats was: {ADP, ATP, CO2, G3P, G6P, H, H2O, NADP, NADPH, R5P}. Using the methods of § 2.2, there were three pathways:

1. PGI + PFK + FBA + TPI
2. G6PDH2R + PGL + GND + RPI
3. - 2 PGI + 2 G6PDH2R + 2 PGL + 2 GND + TKT2 + TALA + TKT1 + 2 RPE

with pathway matrix K_p given by

$$K_p^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 2 & 2 & 2 & 0 & 1 & 1 & 1 & 2 \end{pmatrix} \tag{49}$$

and corresponding reactions:

$$\mathrm{ATP} + \mathrm{G_6P} \xrightleftharpoons{\mathrm{PPP_1}} \mathrm{ADP} + 2\,\mathrm{G_3P} + \mathrm{H} \tag{50}$$

$$\mathrm{G_6P} + \mathrm{H_2O} + 2\,\mathrm{NADP} \xrightleftharpoons{\mathrm{PPP_2}} \mathrm{CO_2} + 2\,\mathrm{H} + 2\,\mathrm{NADPH} + \mathrm{R_5P} \tag{51}$$

$$2\,\mathrm{H_2O} + 4\,\mathrm{NADP} + \mathrm{R_5P} \xrightleftharpoons{\mathrm{PPP_3}} 2\,\mathrm{CO_2} + \mathrm{G_3P} + 4\,\mathrm{H} + 4\,\mathrm{NADPH} \tag{52}$$

(a) Species parameters			
Species	$\hat{\phi}\, mV$	c/c_0	$\hat{K}$
6PGC	29	0.4784	6.2335
ADP	-27	0.0704	5.1546
ATP	27	1.2221	2.2539
CO2	-30	0.0095	33.7942
DHAP	-10	0.3883	1.7790
E4P	-27	0.0062	57.9353
F6P	-21	0.3198	1.4140
FDP	-8	1.9289	0.3880
G3P	-18	0.0344	14.9020
G6P	-5	1.0000	0.8377
NADP	30	0.0003	11747.0633
NADPH	-30	0.0154	21.0027
R5P	5	0.0999	12.2419
RU5PD	5	0.0142	86.1551
S7P	24	0.1119	21.7513
XU5PD	5	0.0230	51.6829

(b) Chemostat flows	
Chemostat	flow
ADP	63.12
ATP	-63.12
CO2	11.58
G3P	128.03
G6P	-71.10
H	86.27
H2O	-11.58
NADP	-23.16
NADPH	23.16
R5P	6.19

(c) Pathway flows	
Pathway	$\hat{f}_p$
PPP1	63.12
PPP2	7.98
PPP3	1.80

(d) Reaction flows and Parameters							
Reaction	Φ mV	$\hat{\Phi}$ mV	f/f_0	$\hat{f}/f_0$	$\hat{\kappa}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$
PGI	16.48	16.48	60.00	59.52	154.39	66.44	16.61
PFK	68.82	68.82	62.62	63.12	54.85	30.59	7.65
FBA	20.00	20.00	63.43	63.12	160.08	61.59	15.40
TPI	7.98	7.98	62.82	63.12	353.93	133.64	33.41
G6PDH2R	–	82.84	–	11.58	4.67	5.14	1.28
PGL	–	1.00	–	11.58	291.64	171.06	42.77
GND	114.53	114.53	11.70	11.58	1.27	4.78	1.19
RPI	0.04	0.04	7.87	7.98	4206.98	2785.35	696.34
TKT2	16.38	16.38	0.91	1.80	9.17	2.24	0.56
TALA	54.41	54.41	–	1.80	1.66	0.94	0.23
TKT1	4.04	4.04	2.92	1.80	8.82	6.66	1.67
RPE	0.83	0.83	3.83	3.59	96.07	63.27	15.82

Table 1: Estimated flows and parameters; flows and concentration normalised by f_0 and c_0 (53). Missing data indicated by –. (a) Estimated species potentials $\hat{\phi}$ (§ 5.1), normalised concentration and species constants (§ 5.4). (b) Estimated chemostat flows (§ 5.2). (c) Estimated pathway flows (§ 5.2). (d) The estimated reaction potentials $\hat{\Phi}$; these are identical to the measured reaction potentials Φ where known (§ 5.1). The estimated reaction flows $\hat{f}$ are close to the measured reaction flows where known (§ 5.2). The estimated mass-action reaction constants $\hat{\kappa}$ and the estimated Michaelis-Menten equivalent parameters $\hat{\kappa}_1$ and $\hat{\kappa}_2$ using (9) (§ 5.3) with $K_f = 0.1$ and $\rho = 0.2$.

Data normalisation is important in the context of parameter identification in Systems Biology [80]. Here, the experimental concentration and flow data [74] was normalised with respect to the concentration of G_6P and flow of PGI (given in mM/min) by defining:

$$c_0 = c_{G6P} = 7.88\,\mathrm{mM} \qquad f_0 = f_{PGI}/60 = 0.992\,\mathrm{mM/s} \qquad t_0 = \frac{c_0}{f_0} = 7.95\,\mathrm{s} \tag{53}$$

where t_0 is the corresponding time unit.

Using the pathway decomposition and the method of § 5.2, the three pathway flows were deduced to be those of Table 1(d). The estimated reaction flows $\hat{f}$ are then deduced from Equation (36) and given in the fifth column of Table 1(c). The chemostat flows are given in Table 1(b). The concentrations given in Table 3 [74] were used to derive the species parameters of Table 1(a).

The reaction constants κ of the mass action formulation are given in Table 1(d) together with the reaction constants κ_1 and κ_2 of the Michaelis-Menten formulation derived using the QP of (48). These parameters are used to perform a dynamical simulation in § 5.6.

5.6 Simulation

The parameters of Table 1(a)&(d) were used with the bond graph model of the Glycolysis & Pentose Phosphate pathway (§ 4.1) to run simulations. In § 4.1, we derived three pathways within this system; these are now simulated separately here. In particular, chemostats and flowstats (as defined in § 4.1) were implemented for the three cases and the initial concentrations were set to those in Table 1(a) where known and to unit values where unknown.

The simulation was performed separately for two cases: the mass-action formulation using the κ parameters and the Michaelis-Menten formulation using the $\hat{\kappa}_1$ and $\hat{\kappa}_2$ parameters.

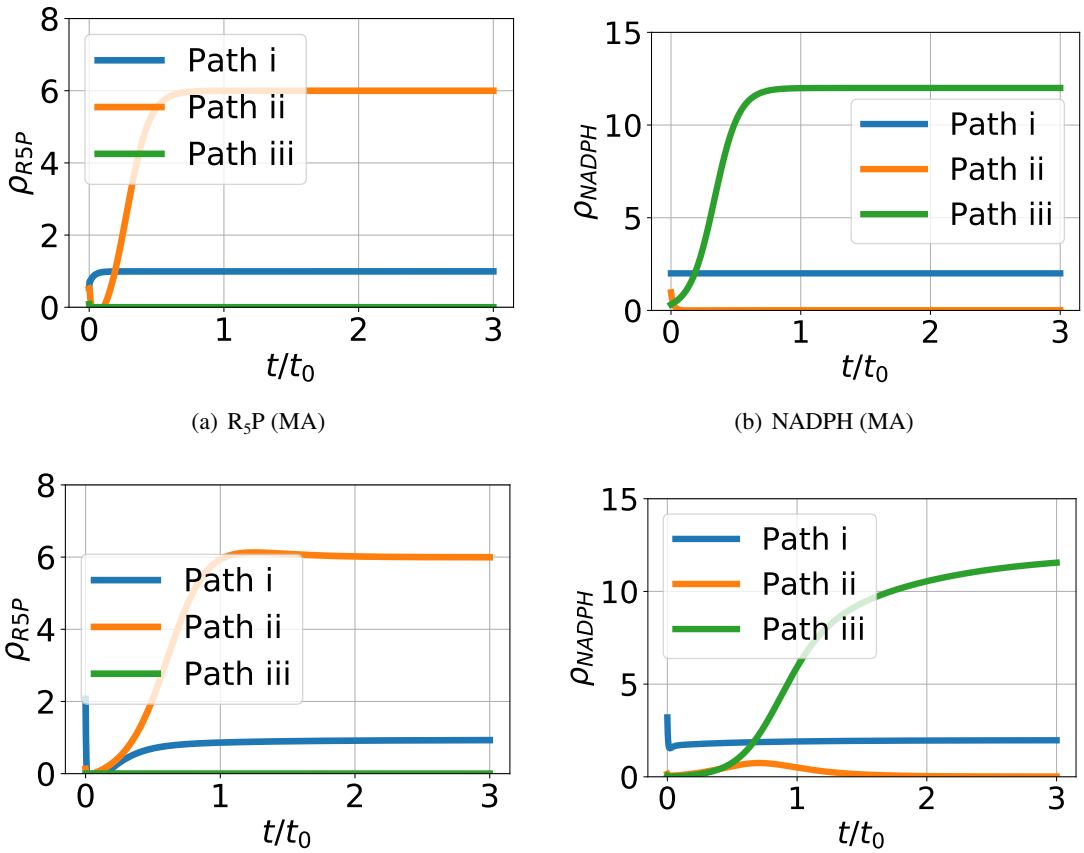
Figure 5 shows the ratios $\rho_{R5P} = \frac{f_{R5P}}{f_{G6P}}$ and $\rho_{NADPH} = \frac{f_{NADPH}}{f_{G6P}}$ of the chemostat flows corresponding to the products R_5P and NADH to the chemostat flow corresponding to the substrate G_6P. At steady state, these ratios correspond to the stoichiometry of the three pathways of § 2.2. In particular, pathway i yields both products, pathway ii yields more R_5P at the expense of NADPH and pathway iii yields more NADPH at the expense of R_5P. Figures 5(a) and 5(b) correspond to the mass-action formulation and Figures 5(c) and 5(d) correspond to the Michaelis-Menten formulation.

Because the two-reaction Michaelis-Menten formulation of enzyme catalysed reactions (8) explicitly includes the enzyme, such models can be used to examine system behaviour as enzyme levels change.

6 Conclusion

The formulation of dynamic simulation models for large-scale biological systems remains a key challenge in systems biology. With the advent of genome-scale simulation and whole-cell modelling, there is increasing recognition of the need for a modular approach in which model components can be formulated, tested and validated independently, and then seamlessly integrated together to form a model of the whole system. However, a dynamic modelling framework which is modular and which can in principle describe the broad range of biochemical and biophysical cellular processes has been elusive.

Several authors have acknowledged the need for energetic considerations to be integrated into modelling approaches, both to ensure that models are consistent with basic thermodynamic principles, and to enable calculation of energy flows and related concepts such as efficiency [67]. Here we have shown that thermodynamically compliant dynamic models of metabolism can be generated



(a) R_5P (MA)

(b) NADPH (MA)

(c) R_5P (MM)

(d) NADPH (MM)

Figure 5: Pathway simulation. Ratios (ρ) of product (R_5P & NADPH) chemostat flow to substrate (G_6P) chemostat flow, plotted against time normalised by t_0 (53), for each of the three pathways of § 2.2. The results are given for two cases: using the estimated mass-action (MA) parameter $\hat{\kappa}$ and using the estimated Michaelis-Menten equivalent parameters (MM) $\hat{\kappa}_1$ and $\hat{\kappa}_2$ from Table 1. As discussed in § 2.2, pathway i yields both products, pathway ii yields more R_5P at the expense of NADPH and pathway iii yields more NADPH at the expense of R_5P. In this case both MA and MM give the same steady-state values but with differing dynamic response.

using the bond graph modelling approach, with the stoichiometric matrix as the starting point. Bond graphs, first advocated in the context of biological network thermodynamics by Oster et al. [34], represent both energy and mass flow through the biochemical network. Bond graphs separate the system connectivity from energy-dissipating processes (reactions), and thus are a very natural fit to network-based modelling in systems biology. As a port-based modelling approach, bond graphs are also inherently modular. Furthermore, application of bond graph modelling principles automatically endows models with a number of necessary features for large-scale modelling including modularity, thermodynamically distinguished parameters (wherein system-wide thermodynamic parameters relating to biochemical species are distinguished from reaction-specific parameters) and hence, as noted by Mason and Covert [29], improved opportunity for parameter identification from data.

Energy-based modelling of biochemical reaction networks using bond graphs naturally encompasses the EBA approach [16], where we have shown that the key equations of EBA are implicit in the system bond graph. This is a powerful advantage as it means that no additional steps are required in order to satisfy thermodynamic constraints. Any model formulated as a bond graph implicitly satisfies these constraints; it is not possible to impose, or infer from data, parameters which break these constraints.

A further benefit for large-scale modelling is that bond graphs naturally lend themselves to model reduction, for example through generation of reduced-order models using pathway analysis [53, 81]: any such simplified model will also satisfy the same thermodynamic constraints. This enables a hierarchical approach to modelling, and it is not necessary to model all aspects of the system at the same level of detail. Different levels of representation can be used as required, for example reflecting available knowledge and data about different parts of the system.

As noted in the Introduction, a key challenge in the development of dynamic models is the fitting of parameters to experimental data. We have shown that both mass-action kinetics and (reversible) Michaelis-Menten kinetics fall within the bond graph framework and therefore have a thermodynamically safe parameterisation; moreover, it is shown this parameterisation leads to a linear-in-the parameters estimation problem. Bond graphs separate the constitutive relations describing the reactions from the connectivity of the model; it is therefore possible to incorporate more complex kinetic schemesc̃itepCor12, including inhibition, allosteric modulation and cooperativity within the bond graph approach thus retaining thermodynamically safe parameterisation. However, the resultant parameter estimation problem will not, in general, be linear-in-the parameters and will therefore require an optimisation approach such as that used by K-FIT [82]. Optimisation approaches such as K-FIT do not use a set of parameters that is thermodynamically safe by design, hence they need to derive additional constraints to incorporate thermodynamic consistency. Future work will examine how the thermodynamically safe parameterisation induced by the bond graph approach can be used to simplify such optimisation when applied to large systems and data sets.

According to Noor et al. [83] in the context of obtaining biological insights though omics data integration: "To maximize predictive power and mechanistic insights on the molecular level, ODE simulations based on physical models of binding and catalysis remain the gold standard." The illustrative example of this paper shows how data involving flows, concentrations and chemical potentials can be integrated using the physical model structure provided by combining stoichiometric and bond graph approaches. It is believed that this provides a basis for integrating the larger and more varied omics data becoming available. Moreover, the physical basis of the approach can be used to indicate what additional data should be gathered to fully parameterise the model.

Here we have demonstrated that thermodynamically compliant dynamic models can be constructed starting from the stoichiometric matrix. The plethora of existing stoichiometric models for metabolic networks provides a natural starting point for this endeavour. However, while metabolic

models are of central importance in a number of contexts, models of cellular physiology in general, and whole-cell models in particular, require a framework that can incorporate a much broader range of cellular processes, feedback and regulation. As a general tool for physically plausible systems modelling, bond graphs can naturally include energy compliant connections to other physical domains and processes, including transport [84], electrochemical transduction [38, 39], membrane potential dynamics [41], mechanochemical transduction and photosynthesis. Furthermore, through incorporation of control-theoretic concepts, enzyme modulation and feedback control can be represented in a coherent manner [62]. There remain however several key domains of cellular biology where to our knowledge there are as yet no examples of bond graph modelling, including transcription and translation [20, 85, 86]. These will need to be demonstrated in order to provide a complete road map for construction of modular and thermodynamically compliant whole-cell models using bond graphs.

Data Accessibility

The figures and tables in this paper were generated using the Jupyter notebooks and Python code available at `https://github.com/gawthrop/GawPanCra21`.

Acknowledgements

PJG would like to thank the Faculty of Engineering and Information Technology, University of Melbourne, for its support via a Professorial Fellowship. This research was in part conducted and funded by the Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology (project number CE140100036). The authors would like to thank the referees for their constructive comments.

References

[1] Jonathan R. Karr, Jayodita C. Sanghvi, Derek N. Macklin, Miriam V. Gutschow, Jared M. Jacobs, Benjamin Bolival Jr., Nacyra Assad-Garcia, John I. Glass, and Markus W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389 – 401, 2012. ISSN 0092-8674. doi:10.1016/j.cell.2012.05.044.

[2] Markus W. Covert. *Fundamentals of Systems Biology From Synthetic Circuits to Whole-cell Models*. CRC Press, 2015. doi:10.4324/9781315222615.

[3] Xin Fang, Colton J. Lloyd, and Bernhard O. Palsson. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology*, 18(12): 731–743, Dec 2020. ISSN 1740-1534. doi:10.1038/s41579-020-00440-4. URL `https://doi.org/10.1038/s41579-020-00440-4`.

[4] Balázs Szigeti, Yosef D. Roth, John A.P. Sekar, Arthur P. Goldberg, Saahith C. Pochiraju, and Jonathan R. Karr. A blueprint for human whole-cell modeling. *Current Opinion in Systems Biology*, 7:8 – 15, 2018. ISSN 2452-3100. doi:10.1016/j.coisb.2017.10.005.

[5] Reinhart Heinrich and Stefan Schuster. *The regulation of cellular systems*. Chapman & Hall New York, 1996.

[6] Bernhard Palsson. *Systems biology: properties of reconstructed networks*. Cambridge University Press, 2006. ISBN 0521859034.

[7] Bernhard Palsson. *Systems Biology: Simulation of Dynamic Network States*. Cambridge University Press, 2011.

[8] Bernhard Palsson. *Systems Biology: Constraint-Based Reconstruction and Analysis*. Cambridge University Press, Cambridge, 2015.

[9] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism—2011. *Molecular Systems Biology*, 7(1), 2011. ISSN 1744-4292. doi:10.1038/msb.2011.65.

[10] Ines Thiele, Neil Swainston, Ronan M. T. Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K. Aurich, Hulda Haraldsdottir, Monica L. Mo, Ottar Rolfsson, Miranda D. Stobbe, Stefan G. Thorleifsson, Rasmus Agren, Christian Bolling, Sergio Bordel, Arvind K. Chavali, Paul Dobson, Warwick B. Dunn, Lukas Endler, David Hala, Michael Hucka, Duncan Hull, Daniel Jameson, Neema Jamshidi, Jon J. Jonsson, Nick Juty, Sarah Keating, Intawat Nookaew, Nicolas Le Novere, Naglis Malys, Alexander Mazein, Jason A. Papin, Nathan D. Price, Evgeni Selkov Sr, Martin I. Sigurdsson, Evangelos Simeonidis, Nikolaus Sonnenschein, Kieran Smallbone, Anatoly Sorokin, Johannes H. G. M. van Beek, Dieter Weichart, Igor Goryanin, Jens Nielsen, Hans V. Westerhoff, Douglas B. Kell, Pedro Mendes, and Bernhard O. Palsson. A community-driven global reconstruction of human metabolism. *Nat Biotech*, 31:419–425, May 2013. ISSN 1087-0156. doi:10.1038/nbt.2488.

[11] Neil Swainston, Kieran Smallbone, Hooman Hefzi, Paul D. Dobson, Judy Brewer, Michael Hanscho, Daniel C. Zielinski, Kok Siong Ang, Natalie J. Gardiner, Jahir M. Gutierrez, Sarantos Kyriakopoulos, Meiyappan Lakshmanan, Shangzhong Li, Joanne K. Liu, Veronica S. Martínez, Camila A. Orellana, Lake-Ee Quek, Alex Thomas, Juergen Zanghellini, Nicole Borth, Dong-Yup Lee, Lars K. Nielsen, Douglas B. Kell, Nathan E. Lewis, and Pedro Mendes. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, 12(7):109, Jun 2016. ISSN 1573-3890. doi:10.1007/s11306-016-1051-4.

[12] St. Elmo Wilken, Jonathan M. Monk, Patrick A. Leggieri, Christopher E. Lawson, Thomas S. Lankiewicz, Susanna Seppälä, Chris G. Daum, Jerry Jenkins, Anna M. Lipzen, Stephen J. Mondo, Kerrie W. Barry, Igor V. Grigoriev, John K. Henske, Michael K. Theodorou, Bernhard O. Palsson, Linda R. Petzold, and Michelle A. O'Malley. Experimentally validated reconstruction and analysis of a genome-scale metabolic model of an anaerobic neocallimastigomycota fungus. *mSystems*, 6(1), 2021. doi:10.1128/mSystems.00002-21.

[13] Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nat Biotech*, 28:245–248, March 2010. ISSN 1087-0156. doi:10.1038/nbt.1614.

[14] Christopher S. Henry, Linda J. Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical Journal*, 92(5):1792 – 1805, 2007. ISSN 0006-3495. doi:10.1529/biophysj.106.093138.

[15] Maria Masid, Meric Ataman, and Vassily Hatzimanikatis. Analysis of human metabolism by reducing the complexity of the genome-scale models using redhuman. *Nature Communications*, 11(1):2821, Jun 2020. ISSN 2041-1723. doi:10.1038/s41467-020-16549-2.

[16] Daniel A. Beard, Shoudan Liang, and Hong Qian. Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83(1):79 – 86, 2002. ISSN 0006-3495. doi:10.1016/S0006-3495(02)75150-3.

[17] Hong Qian, Daniel A. Beard, and Shou-dan Liang. Stoichiometric network theory for nonequilibrium biochemical systems. *European Journal of Biochemistry*, 270(3):415–421, 2003. ISSN 1432-1033. doi:10.1046/j.1432-1033.2003.03357.x.

[18] Elad Noor, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLOS Computational Biology*, 10(2):1–12, 02 2014. doi:10.1371/journal.pcbi.1003483.

[19] Elad Noor. Removing both Internal and Unrealistic Energy-Generating Cycles in Flux Balance Analysis. *arXiv e-prints*, art. arXiv:1803.04999, Mar 2018.

[20] Pierre Salvy and Vassily Hatzimanikatis. The etfl formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nature Communications*, 11 (1):30, Jan 2020. ISSN 2041-1723. doi:10.1038/s41467-019-13818-7.

[21] Jan Schellenberger, Nathan E. Lewis, and Bernhard Ø Palsson. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, 100(3):544–553, 2011. ISSN 0006-3495. doi:10.1016/j.bpj.2010.12.3707.

[22] Neema Jamshidi and Bernhard Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98(2):175 – 185, 2010. ISSN 0006-3495. doi:10.1016/j.bpj.2009.09.064.

[23] Zachary B. Haiman, Daniel C. Zielinski, Yuko Koike, James T. Yurkovich, and Bernhard O. Palsson. Masspy: Building, simulating, and visualizing dynamic biological models in python using mass action kinetics. *PLOS Computational Biology*, 17(1):1–20, 01 2021. doi:10.1371/journal.pcbi.1008208.

[24] Natalie J. Stanford, Timo Lubitz, Kieran Smallbone, Edda Klipp, Pedro Mendes, and Wolfram Liebermeister. Systematic construction of kinetic models from genome-scale metabolic networks. *PLoS ONE*, 8(11):e79195, 11 2013. doi:10.1371/journal.pone.0079195.

[25] Kieran Smallbone and Pedro Mendes. Large-scale metabolic models: From reconstruction to differential equations. *Industrial Biotechnology*, 9(4):179–184, 2013. doi:10.1089/ind.2013.0003.

[26] James P Keener and James Sneyd. *Mathematical Physiology: I: Cellular Physiology*, volume 1. Springer, New York, 2nd edition, 2009.

[27] Michael Ederer and Ernst Dieter Gilles. Thermodynamically feasible kinetic models of reaction networks. *Biophysical Journal*, 92(6):1846 – 1857, 2007. ISSN 0006-3495. doi:10.1529/biophysj.106.094094.

[28] M. Ederer and E.D. Gilles. Thermodynamic constraints in kinetic modeling: Thermodynamic-kinetic modeling in comparison to other approaches. *Engineering in Life Sciences*, 8(5):467–476, 2008. doi:10.1002/elsc.200800040.

[29] John C. Mason and Markus W. Covert. An energetic reformulation of kinetic rate laws enables scalable parameter estimation for biochemical networks. *Journal of Theoretical Biology*, 461: 145 – 156, 2019. ISSN 0022-5193. doi:10.1016/j.jtbi.2018.10.041.

[30] H. M. Paynter. *Analysis and Design of Engineering Systems*. MIT Press, Cambridge, Mass., 1961.

[31] P. J. Gawthrop and L. P. S. Smith. *Metamodelling: Bond Graphs and Dynamic Systems*. Prentice Hall, Hemel Hempstead, Herts, England., 1996. ISBN 0-13-489824-9.

[32] Peter J Gawthrop and Geraint P Bevan. Bond-graph modeling: A tutorial introduction for control engineers. *IEEE Control Systems Magazine*, 27(2):24–45, April 2007. doi:10.1109/MCS.2007.338279.

[33] Dean C Karnopp, Donald L Margolis, and Ronald C Rosenberg. *System Dynamics: Modeling, Simulation, and Control of Mechatronic Systems*. John Wiley & Sons, Hoboken, New Jersey, 5th edition, 2012. ISBN 978-0470889084.

[34] George Oster, Alan Perelson, and Aharon Katchalsky. Network thermodynamics. *Nature*, 234: 393–399, December 1971. doi:10.1038/234393a0.

[35] George F. Oster, Alan S. Perelson, and Aharon Katchalsky. Network thermodynamics: dynamic modelling of biophysical systems. *Quarterly Reviews of Biophysics*, 6(01):1–134, 1973. doi:10.1017/S0033583500000081.

[36] A.S. Perelson. Network thermodynamics. an overview. *Biophysical Journal*, 15(7):667 – 685, 1975. ISSN 0006-3495. doi:10.1016/S0006-3495(75)85847-4.

[37] Peter J. Gawthrop and Edmund J. Crampin. Energy-based analysis of biochemical cycles using bond graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 470(2171):1–25, 2014. doi:10.1098/rspa.2014.0459. Available at arXiv:1406.2447.

[38] P. J. Gawthrop. Bond graph modeling of chemiosmotic biomolecular energy transduction. *IEEE Transactions on NanoBioscience*, 16(3):177–188, April 2017. ISSN 1536-1241. doi:10.1109/TNB.2017.2674683. Available at arXiv:1611.04264.

[39] P. J. Gawthrop, I. Siekmann, T. Kameneva, S. Saha, M. R. Ibbotson, and E. J. Crampin. Bond graph modelling of chemoelectrical energy transduction. *IET Systems Biology*, 11(5):127–138, 2017. ISSN 1751-8849. doi:10.1049/iet-syb.2017.0006. Available at arXiv:1512.00956.

[40] Michael Pan, Peter J. Gawthrop, Kenneth Tran, Joseph Cursons, and Edmund J. Crampin. A thermodynamic framework for modelling membrane transporters. *Journal of Theoretical Biology*, 2018. ISSN 0022-5193. doi:10.1016/j.jtbi.2018.09.034.

[41] Michael Pan, Peter J. Gawthrop, Kenneth Tran, Joseph Cursons, and Edmund J. Crampin. Bond graph modelling of the cardiac action potential: implications for drift and non-unique steady states. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 474(2214), 2018. ISSN 1364-5021. doi:10.1098/rspa.2018.0106. Available at arXiv:1802.04548.

[42] Peter J. Gawthrop and Michael Pan. Network thermodynamical modeling of bioelectrical systems: A bond graph approach. *Bioelectricity*, 2020. doi:10.1089/bioe.2020.0042. Published Online: 18 Dec 2020.

[43] Peter J. Gawthrop, Joseph Cursons, and Edmund J. Crampin. Hierarchical bond graph modelling of biochemical networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2184):1–23, 2015. ISSN 1364-5021. doi:10.1098/rspa.2015.0642. Available at arXiv:1503.01814.

[44] P. J. Gawthrop and E. J. Crampin. Modular bond-graph modelling and analysis of biomolecular systems. *IET Systems Biology*, 10(5):187–201, October 2016. ISSN 1751-8849. doi:10.1049/iet-syb.2015.0083. Available at arXiv:1511.06482.

[45] Peter Hunter. The virtual physiological human: The physiome project aims to develop reproducible, multiscale models for clinical practice. *IEEE Pulse*, 7(4):36–42, July 2016. ISSN 2154-2287. doi:10.1109/MPUL.2016.2563841.

[46] Michael Pan, Peter J. Gawthrop, Joseph Cursons, Kenneth Tran, and Edmund J. Crampin. The cardiac Na+/K+ ATPase: An updated, thermodynamically consistent model. *Physiome*, 8 2020. doi:10.36903/physiome.12871070.v1. URL `https://physiome.figshare.com/articles/journal_contribution/The_cardiac_Na_K_ATPase_An_updated_thermodynamically_consistent_model/12871070`.

[47] Peter J. Gawthrop, Peter Cudmore, and Edmund J. Crampin. Physically-plausible modelling of biomolecular systems: A simplified, energy-based model of the mitochondrial electron transport chain. *Journal of Theoretical Biology*, 493:110223, 2020. ISSN 0022-5193. doi:10.1016/j.jtbi.2020.110223.

[48] Melissa J. Lambeth and Martin J. Kushmerick. A computational model for glycogenolysis in skeletal muscle. *Annals of Biomedical Engineering*, 30(6):808–827, 2002. ISSN 0090-6964. doi:10.1114/1.1492813.

[49] Lucie Parent, Stéphane Supplisson, Donald D. F. Loo, and Ernest M. Wright. Electrogenic properties of the cloned Na+/glucose cotransporter: I. voltage-clamp studies. *The Journal of Membrane Biology*, 125(1):49–62, 1992. ISSN 1432-1424. doi:10.1007/BF00235797.

[50] Pedro A. Saa and Lars K. Nielsen. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances*, 35(8):981 – 1003, 2017. ISSN 0734-9750. doi:10.1016/j.biotechadv.2017.09.005. Metabolic engineering frontiers emerging with advanced tools and methodologies.

[51] J. Orth, R. Fleming, and B. Palsson. Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. *EcoSal Plus*, 2010. doi:10.1128/ecosalplus.10.2.1.

[52] Ali Ebrahim, Joshua A. Lerman, Bernhard O. Palsson, and Daniel R. Hyduke. Cobrapy: Constraints-based reconstruction and analysis for python. *BMC Systems Biology*, 7(1):74, Aug 2013. ISSN 1752-0509. doi:10.1186/1752-0509-7-74.

[53] Peter J. Gawthrop and Edmund J. Crampin. Energy-based analysis of biomolecular pathways. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 473(2202), 2017. ISSN 1364-5021. doi:10.1098/rspa.2016.0825. Available at arXiv:1611.02332.

[54] Pierre Van Rysselberghe. Reaction rates and affinities. *The Journal of Chemical Physics*, 29(3): 640–642, 1958. doi:10.1063/1.1744552.

[55] Matteo Polettini and Massimiliano Esposito. Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws. *The Journal of Chemical Physics*, 141(2):024117, 2014. doi:10.1063/1.4886396.

[56] Christophe H. Schilling, David Letscher, and Bernhard Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203(3):229 – 248, 2000. ISSN 0022-5193. doi:10.1006/jtbi.2000.1073.

[57] E.A. Newsholme, R.A.J. Challiss, and B. Crabtree. Substrate cycles: their role in improving sensitivity in metabolic control. *Trends in Biochemical Sciences*, 9(6):277 – 280, 1984. ISSN 0968-0004. doi:10.1016/0968-0004(84)90165-8.

[58] H. Qian and D. A. Beard. Metabolic futile cycles and their functions: a systems analysis of energy and control. *IEE Proceedings - Systems Biology*, 153(4):192–200, July 2006. ISSN 1741-2471. doi:10.1049/ip-syb:20050086.

[59] Athel Cornish-Bowden. *Fundamentals of enzyme kinetics*. Wiley-Blackwell, London, 4th edition, 2013. ISBN 978-3-527-33074-4.

[60] Yo-Cheng Chang, Judith P Armitage, Antonis Papachristodoulou, and George H Wadhams. A single phosphatase can convert a robust step response into a graded, tunable or adaptive response. *Microbiology*, 159:1276–1285, 2013. doi:10.1099/mic.0.066324-0.

[61] Reginald H. Garrett and Charles M. Grisham. *Biochemistry*. Cengage Learning, Boston, MA, 6th edition, 2017.

[62] P. J. Gawthrop. Energy-based modeling of the feedback control of biomolecular systems with cyclic flow modulation. *IEEE Transactions on NanoBioscience*, 20(2):183–192, April 2021. ISSN 1558-2639. doi:10.1109/TNB.2021.3058440.

[63] Peter Cudmore, Peter J. Gawthrop, Michael Pan, and Edmund J. Crampin. Computer-aided modelling of complex physical systems with BondGraphTools. Available at arXiv:1906.10799, Jun 2019.

[64] Maxwell L. Neal, Michael T. Cooling, Lucian P. Smith, Christopher T. Thompson, Herbert M. Sauro, Brian E. Carlson, Daniel L. Cook, and John H. Gennari. A reappraisal of how to build modular, reusable models of biological systems. *PLoS Comput Biol*, 10(10):e1003849, 10 2014. doi:10.1371/journal.pcbi.1003849.

[65] Maxwell L. Neal, Brian E. Carlson, Christopher T. Thompson, Ryan C. James, Karam G. Kim, Kenneth Tran, Edmund J. Crampin, Daniel L. Cook, and John H. Gennari. Semantics-based composition of integrated cardiomyocyte models motivated by real-world use cases. *PLoS ONE*, 10(12):1–18, 12 2016. doi:10.1371/journal.pone.0145621.

[66] Maxwell L Neal, Christopher T Thompson, Karam G Kim, Ryan C James, Daniel L Cook, Brian E Carlson, and John H Gennari. SemGen: a tool for semantics-based annotation and composition of biosimulation models. *Bioinformatics*, 35(9):1600–1602, 09 2018. ISSN 1367-4803. doi:10.1093/bioinformatics/bty829.

[67] Peter J. Gawthrop and Edmund J. Crampin. Biomolecular system energetics. In *Proceedings of the 13th International Conference on Bond Graph Modeling (ICBGM'18)*, Bordeaux, 2018. Society for Computer Simulation. Available at arXiv:1803.09231.

[68] Zachary A. King, Andreas Dräger, Ali Ebrahim, Nikolaus Sonnenschein, Nathan E. Lewis, and Bernhard O. Palsson. Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLOS Computational Biology*, 11(8):e1004321, Aug 2015. doi:10.1371/journal.pcbi.1004321.

[69] Matthew G. Vander Heiden, Lewis C. Cantley, and Craig B. Thompson. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science*, 324(5930):1029–1033, 2009. ISSN 0036-8075. doi:10.1126/science.1160809.

[70] Krushna C. Patra and Nissim Hay. The pentose phosphate pathway and cancer. *Trends in Biochemical Sciences*, 39(8):347 – 354, 2014. ISSN 0968-0004. doi:10.1016/j.tibs.2014.06.005.

[71] Jeremy M. Berg, John L. Tymoczko, Gregory J. Gatto, and Lubert Stryer. *Biochemistry*. W.H. Freeman, ninth edition, 2019. ISBN 978-1-319-11465-7.

[72] Mathieu Cloutier and Peter Wellstead. The control systems structures of energy metabolism. *Journal of The Royal Society Interface*, 7(45):651–665, 2010. doi:10.1098/rsif.2009.0371.

[73] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A. Kaandorp, and Joke G. Blom. Systems biology: parameter estimation for biochemical models. *FEBS Journal*, 276:886–902, 2009. ISSN 1742-4658. doi:10.1111/j.1742-4658.2008.06844.x.

[74] Junyoung O. Park, Sara A. Rubin, Yi-Fan Xu, Daniel Amador-Noguez, Jing Fan, Tomer Shlomi, and Joshua D. Rabinowitz. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol*, 12(7):482–489, Jul 2016. ISSN 1552-4450. doi:10.1038/nchembio.2077.

[75] Charles J. Foster, Saratram Gopalakrishnan, Maciek R. Antoniewicz, and Costas D. Maranas. From escherichia coli mutant 13c labeling data to a core kinetic model: A kinetic model parameterization pipeline. *PLOS Computational Biology*, 15(9):e1007319, Sep 2019. doi:10.1371/journal.pcbi.1007319.

[76] Junyoung O. Park, Lukas B. Tanner, Monica H. Wei, Daven B. Khana, Tyler B. Jacobson, Zheyun Zhang, Sara A. Rubin, Sophia Hsin-Jung Li, Meytal B. Higgins, David M. Stevenson, Daniel Amador-Noguez, and Joshua D. Rabinowitz. Near-equilibrium glycolysis supports metabolic homeostasis and energy yield. *Nature Chemical Biology*, 15(10):1001–1008, Oct 2019. ISSN 1552-4469. doi:10.1038/s41589-019-0364-9.

[77] Tyler B. Jacobson, Travis K. Korosh, David M. Stevenson, Charles Foster, Costas Maranas, Daniel G. Olson, Lee R. Lynd, and Daniel Amador-Noguez. In vivo thermodynamic analysis of glycolysis in clostridium thermocellum and thermoanaerobacterium saccharolyticum using 13c and 2h tracers. *mSystems*, 5(2), 2020. doi:10.1128/mSystems.00736-19.

[78] R. Fletcher. *Practical Methods of Optimization. 2nd Edition*. Wiley, Chichester, 1987.

[79] Lennart Ljung. *System Identification: Theory for the User*. Information and Systems Science. Prentice-Hall, Englewood Cliffs, New Jersey, 2nd edition, 1999.

[80] Andrea Degasperi, Dirk Fey, and Boris N. Kholodenko. Performance of objective functions and optimisation procedures for parameter estimation in system biology models. *Systems Biology and Applications*, 3(1):20, Aug 2017. ISSN 2056-7189. doi:10.1038/s41540-017-0023-2.

[81] Peter J. Gawthrop, Peter Cudmore, and Edmund J. Crampin. Physically-Plausible Modelling of Biomolecular Systems: A Simplified, Energy-Based Model of the Mitochondrial Electron Transport Chain. Available at arXiv:1905.12958, May 2019.

[82] Saratram Gopalakrishnan, Satyakam Dash, and Costas Maranas. K-fit: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. *Metabolic Engineering*, 61:197–205, 2020. ISSN 1096-7176. doi:https://doi.org/10.1016/j.ymben.2020.03.001.

[83] Elad Noor, Sarah Cherkaoui, and Uwe Sauer. Biological insights through omics data integration. *Current Opinion in Systems Biology*, 15:39 – 47, 2019. ISSN 2452-3100. doi:10.1016/j.coisb.2019.03.007. Gene regulation.

[84] Michael Pan, Peter J. Gawthrop, Kenneth Tran, Joseph Cursons, and Edmund J. Crampin. A thermodynamic framework for modelling membrane transporters. *Journal of Theoretical Biology*, 481:10 – 23, 2019. ISSN 0022-5193. doi:10.1016/j.jtbi.2018.09.034. Available at arXiv:1806.04341.

[85] Colton J. Lloyd, Ali Ebrahim, Laurence Yang, Zachary A. King, Edward Catoiu, Edward J. O'Brien, Joanne K. Liu, and Bernhard O. Palsson. Cobrame: A computational framework for genome-scale models of metabolism and gene expression. *PLOS Computational Biology*, 14(7): 1–14, 07 2018. doi:10.1371/journal.pcbi.1006302.

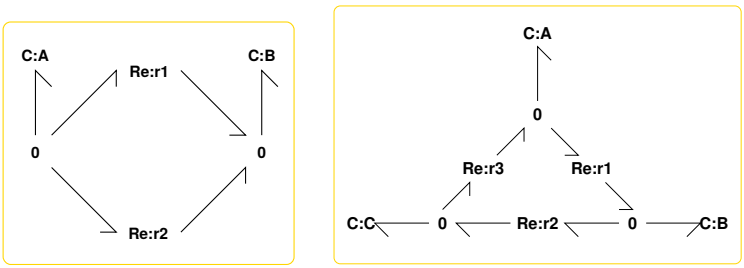
[86] Bin Du, Laurence Yang, Colton J. Lloyd, Xin Fang, and Bernhard O. Palsson. Genome-scale model of metabolism and gene expression provides a multi-scale description of acid stress responses in escherichia coli. *PLOS Computational Biology*, 15(12):1–21, 12 2019. doi:10.1371/journal.pcbi.1007525. URL `https://doi.org/10.1371/journal.pcbi.1007525`.

[87] P. Gawthrop. Computing biomolecular system steady-states. *IEEE Transactions on NanoBioscience*, 17(1):36–43, March 2018. ISSN 1536-1241. doi:10.1109/TNB.2017.2787486. Published online 25th December 2017.

A EBA examples

These examples refer to § 3.3 and drawn from Beard et al [16].

A.1 Example: Parallel reactions



(a) Example: Parallel reaction. (b) Example: three-reaction cycle.

Figure 6: Bond graphs corresponding to examples from Beard et al [16] (**1** junctions are not shown for clarity)). (a) Beard et al [16, Fig. 2], (b) Beard et al [16, Fig. 3]

Beard et al [16] motivate EBA using the example of two resistors in parallel. Figure 6(a) shows the bond graph of the analogous reaction system: the species A and B are joined by two reactions:

$$\mathrm{A} \underset{}{\overset{\mathrm{r_1}}{\rightleftharpoons}} \mathrm{B} \tag{54}$$

$$\mathrm{A} \overset{\mathrm{r_2}}{\rightleftharpoons} \mathrm{B} \tag{55}$$

The stoichiometric matrix is:

$$N = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} \tag{56}$$

and the null space matrix K is

$$K = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{57}$$

corresponding to the pathway: $-r_1 + r_2$.

Setting A and B as chemostats:

$$N^{cd} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{58}$$

$$K_p = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{59}$$

Equation (28) then becomes:

$$-r_1 v_1 + r_2 v_2 = 0 \tag{60}$$

As $r_i > 0$, it follows that v_1 and v_2 must either be zero or have the same sign.

A.2 Example: three-reaction cycle

Beard et al [16] give the example of a three-reaction cycle. Figure 6(b) shows the corresponding bond graph. The species A, B and C are joined by three reactions:

$$\mathrm{A} \overset{\mathrm{r_1}}{\rightleftharpoons} \mathrm{B} \tag{61}$$

$$\mathrm{B} \overset{\mathrm{r_2}}{\rightleftharpoons} \mathrm{C} \tag{62}$$

$$\mathrm{C} \overset{\mathrm{r_3}}{\rightleftharpoons} \mathrm{A} \tag{63}$$

The stoichiometric matrix is:

$$N = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \tag{64}$$

and the null space matrix K is

$$K = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{65}$$

corresponding to the pathway: $r_1 + r_2 + r_3$.

Setting A and B as chemostats:

$$N^{cd} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \tag{66}$$

$$K_p = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \tag{67}$$

Equation (28) then becomes:

$$r_1 v_1 + r_2 v_2 + r_3 v_2 = r_1 v_1 + (r_2 + r_3)\, v_2 = 0 \tag{68}$$

As $r_i > 0$, it follows that v_1 and v_2 must either be zero or have the opposite sign.

Alternatively, setting A, B and C as chemostats:

$$N^{cd} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{69}$$

$$K_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{70}$$

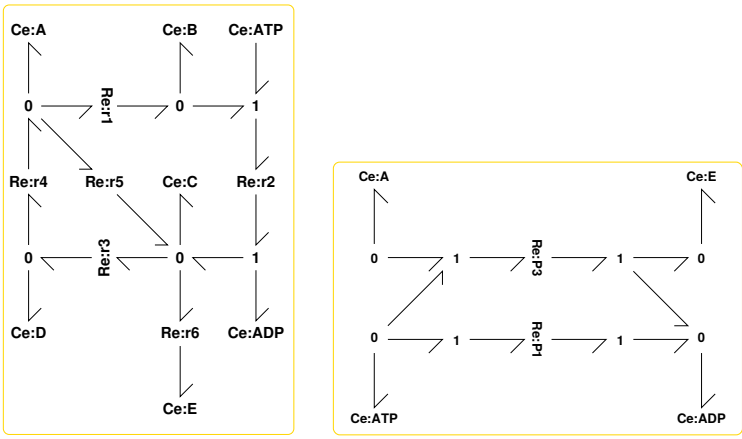
Equation (28) then becomes:

$$r_1 v_1 + r_2 v_2 + r_3 v_3 = 0 \tag{71}$$

As $r_i > 0$, there are three possibilities: all flows are zero; one of the three pathway flows must have one sign and the other two flows the opposite sign; or one flow is zero and the other two have opposite signs.

B Pathways: Illustrative example

This example refers to § 2.2. Noor [19] gives a simple illustrative example of the three types of



(a) Bond graph (b) Pathway bond graph

Figure 7: Bond graphs for illustrative example [19]

pathway; Figure 7(a) gives the corresponding bond graph. the reactions are:

$$\mathrm{A} \overset{\mathrm{r}_1}{\rightleftharpoons} \mathrm{B} \tag{72}$$

$$\mathrm{ATP} + \mathrm{B} \overset{\mathrm{r}_2}{\rightleftharpoons} \mathrm{ADP} + \mathrm{C} \tag{73}$$

$$\mathrm{C} \overset{\mathrm{r}_3}{\rightleftharpoons} \mathrm{D} \tag{74}$$

$$\mathrm{D} \overset{\mathrm{r}_4}{\rightleftharpoons} \mathrm{A} \tag{75}$$

$$\mathrm{A} \overset{\mathrm{r}_5}{\rightleftharpoons} \mathrm{C} \tag{76}$$

$$\mathrm{C} \overset{\mathrm{r}_6}{\rightleftharpoons} \mathrm{E} \tag{77}$$

The there are seven species and six reactions giving states x and flows v:

$$x = \begin{pmatrix} x_A \\ x_{ADP} \\ x_{ATP} \\ x_B \\ x_C \\ x_D \\ x_E \end{pmatrix} \qquad v = \begin{pmatrix} v_{r1} \\ v_{r2} \\ v_{r3} \\ v_{r4} \\ v_{r5} \\ v_{r6} \end{pmatrix} \tag{78}$$

The stoichiometric matrix is:

$$N = \begin{pmatrix} -1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{79}$$

Setting A, E, ATP and ADP as chemostats, N^{cd} is constructed by setting the corresponding rows of N to zero. The corresponding null space is three dimensional and corresponds to the three pathways:

1. r1 + r2 + r3 + r4
2. r3 + r4 + r5
3. r1 + r2 + r6

Using (20), the pathway stoichiometric matrix N_p is:

$$N_p = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{80}$$

The three pathway reactions are:

$$\text{ATP} \xrightleftharpoons{\text{P}_1} \text{ADP} \tag{81}$$

$$\xrightleftharpoons{\text{P}_2} \tag{82}$$

$$\text{A + ATP} \xrightleftharpoons{\text{P}_3} \text{ADP + E} \tag{83}$$

Pathway reaction P1 corresponds to a type II pathway, pathway reaction P2 to a type III pathway and pathway reaction P3 to a type I pathway where A is converted to E driven by the conversion of ATP to ADP. The example is extended by assigning a set of nominal chemical potentials $\phi^{\oslash}$ to the species: ${\phi^{\oslash}}_A = 1$, $\phi_{ATP} = 0$, $\phi_{ADP} = 3$, ${\phi^{\oslash}}_B = 1$, ${\phi^{\oslash}}_C = 1$, ${\phi^{\oslash}}_D = 1$, ${\phi^{\oslash}}_E = 0$. The pathway reaction potentials are then computed using (21) as $\Phi_{P1} = -2$, $\Phi_{P2} = 0$, $\Phi_{P3} = -1$. As the potential for each pathway only depends on the species appearing in the pathway reactions, the potential of non-chemostatted species are irrelevant for this computation. In fact the potentials of the species will correspond to the steady-state values of concentrations of the non-chemostatted species arising from the flow patterns corresponding to the chemostat potentials [87]. The pathway bond graph appears in Figure 7(b).

C Glycolysis & Pentose Phosphate Pathways

This section contains the reactions used in § 4.1 to generate the three pathways arising from the upper reactions of glycolysis and the pentose phosphate pathway. The reactions are extracted as discussed in § 4.

The reactions are:

$$G_6P \xrightleftharpoons{PGI} F_6P \tag{84}$$

$$ATP + F_6P \xrightleftharpoons{PFK} ADP + FDP + H \tag{85}$$

$$FDP \xrightleftharpoons{FBA} DHAP + G_3P \tag{86}$$

$$DHAP \xrightleftharpoons{TPI} G_3P \tag{87}$$

$$G_6P + NADP \xrightleftharpoons{G_6PDH_2R} {}_6PGL + H + NADPH \tag{88}$$

$${}_6PGL + H_2O \xrightleftharpoons{PGL} {}_6PGC + H \tag{89}$$

$${}_6PGC + NADP \xrightleftharpoons{GND} CO_2 + NADPH + RU_5PD \tag{90}$$

$$RU_5PD \xrightleftharpoons{RPI} R_5P \tag{91}$$

$$E_4P + XU_5PD \xrightleftharpoons{TKT_2} F_6P + G_3P \tag{92}$$

$$G_3P + S_7P \xrightleftharpoons{TALA} E_4P + F_6P \tag{93}$$

$$R_5P + XU_5PD \xrightleftharpoons{TKT_1} G_3P + S_7P \tag{94}$$

$$RU_5PD \xrightleftharpoons{RPE} XU_5PD \tag{95}$$

D Modular representation of Metabolism: Reactions

This section contains the reactions used in § 4.2 which illustrates the utility of using bond graphs for the modular construction of stoichiometric models by constructing a model of respiration by combining the modular subsystems: Glycolysis, TCA cycle, Electron Transport Chain and ATP Synthase.

The reaction CYTBD (containing $\frac{1}{2}O_2$) was multiplied by 2 to give integer stoichiometry and, for clarity, the reactions RPI, PGK, PGM, SUCOAS and FRD7 were reversed to give the conventional direction.

D.1 Glycolysis

The reactions extracted are:

$$\mathrm{GLCD_E + PEP} \xrightleftharpoons{\mathrm{GLCPTS}} \mathrm{G_6P + PYR} \tag{96}$$

$$\mathrm{G_6P} \xrightleftharpoons{\mathrm{PGI}} \mathrm{F_6P} \tag{97}$$

$$\mathrm{ATP + F_6P} \xrightleftharpoons{\mathrm{PFK}} \mathrm{ADP + FDP + H} \tag{98}$$

$$\mathrm{FDP + H_2O} \xrightleftharpoons{\mathrm{FBP}} \mathrm{F_6P + PI} \tag{99}$$

$$\mathrm{FDP} \xrightleftharpoons{\mathrm{FBA}} \mathrm{DHAP + G_3P} \tag{100}$$

$$\mathrm{DHAP} \xrightleftharpoons{\mathrm{TPI}} \mathrm{G_3P} \tag{101}$$

$$\mathrm{G_3P + NAD + PI} \xrightleftharpoons{\mathrm{GAPD}} \mathrm{{}_{13}DPG + H + NADH} \tag{102}$$

$$\mathrm{{}_{13}DPG + ADP} \xrightleftharpoons{\mathrm{PGK}} \mathrm{{}_3PG + ATP} \tag{103}$$

$$\mathrm{{}_3PG} \xrightleftharpoons{\mathrm{PGM}} \mathrm{{}_2PG} \tag{104}$$

$$\mathrm{{}_2PG} \xrightleftharpoons{\mathrm{ENO}} \mathrm{H_2O + PEP} \tag{105}$$

$$\mathrm{ADP + H + PEP} \xrightleftharpoons{\mathrm{PYK}} \mathrm{ATP + PYR} \tag{106}$$

D.2 TCA cycle

As well as the TCA cycle itself, this module includes:

1. the pyruvate (PYR) connection reactions: PDH and PFL and
2. the NAD/NADP interconversion reaction NADTRHD.

The reactions extracted are:

$$\mathrm{ACCOA + H_2O + OAA} \xrightleftharpoons{\mathrm{CS}} \mathrm{CIT + COA + H} \tag{107}$$

$$\mathrm{CIT} \xrightleftharpoons{\mathrm{ACONTA}} \mathrm{ACON + H_2O} \tag{108}$$

$$\mathrm{ACON + H_2O} \xrightleftharpoons{\mathrm{ACONTB}} \mathrm{ICIT} \tag{109}$$

$$\mathrm{ICIT + NADP} \xrightleftharpoons{\mathrm{ICDHYR}} \mathrm{AKG + CO_2 + NADPH} \tag{110}$$

$$\mathrm{AKG + COA + NAD} \xrightleftharpoons{\mathrm{AKGDH}} \mathrm{CO_2 + NADH + SUCCOA} \tag{111}$$

$$\mathrm{ADP + PI + SUCCOA} \xrightleftharpoons{\mathrm{SUCOAS}} \mathrm{ATP + COA + SUCC} \tag{112}$$

$$\mathrm{Q_8 + SUCC} \xrightleftharpoons{\mathrm{FRD_7}} \mathrm{FUM + Q_8H_2} \tag{113}$$

$$\mathrm{Q_8 + SUCC} \xrightleftharpoons{\mathrm{SUCDI}} \mathrm{FUM + Q_8H_2} \tag{114}$$

$$\mathrm{FUM + H_2O} \xrightleftharpoons{\mathrm{FUM}} \mathrm{MALL} \tag{115}$$

$$\mathrm{MALL + NAD} \xrightleftharpoons{\mathrm{MDH}} \mathrm{H + NADH + OAA} \tag{116}$$

$$\mathrm{NAD + NADPH} \xrightleftharpoons{\mathrm{NADTRHD}} \mathrm{NADH + NADP} \tag{117}$$

$$\mathrm{COA + NAD + PYR} \xrightleftharpoons{\mathrm{PDH}} \mathrm{ACCOA + CO_2 + NADH} \tag{118}$$

$$\mathrm{COA + PYR} \xrightleftharpoons{\mathrm{PFL}} \mathrm{ACCOA + FOR} \tag{119}$$

D.3 Electron Transport Chain

The reactions extracted are:

$$4\,H + NADH + Q_8 \xrightleftharpoons{NADH_{16}} 3\,H_E + NAD + Q_8H_2 \tag{120}$$

$$4\,H + O_2 + 2\,Q_8H_2 \xrightleftharpoons{CYTBD} 2\,H_2O + 4\,H_E + 2\,Q_8 \tag{121}$$

D.4 ATP Synthase

The reaction extracted is:

$$ADP + 4\,H_E + PI \xrightleftharpoons{ATPS_4R} ATP + 3\,H + H_2O \tag{122}$$

E Enzyme-catalysed Reaction

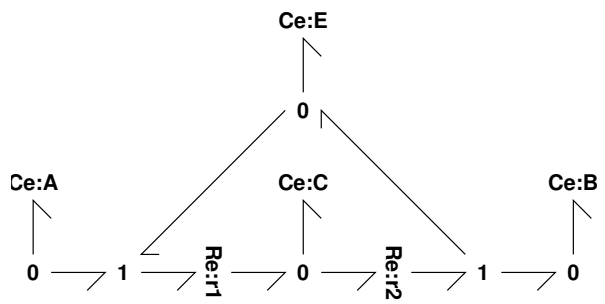


Figure 8: Bond graph representation of an Enzyme-catalysed Reaction