

## Outline:

- Importance of parameter values for estimating in vivo response using kinetic models of metabolism and consequently for metabolic network design using kinetic models of metabolism
- The need for parameter identifiability to determine unique and true parameter values from observed data
- Types and purpose of identifiability for parameters
- Methods for structural identifiability and existing methods for practical identifiability
- Lack of methods for practical identifiability and consequently experimental design (not covered?)
- work done in this paper for practical identifiability
- scalability of computer algebra-based methods for structural identifiability (using CRNT to reduce networks to make structural identifiability scalable) (move to discussion - may be in discussion)

## 1 Introduction:

Kinetic models of metabolism can be used to study the dynamic characteristics of metabolic networks. In these models, ordinary differential equations (ode) are used to express the rate of change of metabolite concentrations ( $x$ ) as a function of the reaction fluxes ( $v$ ) in the metabolic network (Equation 1). The matrix  $\mathbf{S}$  in Equation (1a) defines the stoichiometric relationship between the fluxes and the concentrations of the metabolic network.

$$\dot{x} = \mathbf{S}v \tag{1a}$$

$$v = f(x, \theta, u) \tag{1b}$$

The expression for the nonlinear function ( $f$ ) used to describe each reaction flux  $v_i$  in a kinetic model (Equation 1b) is dependent on the enzyme kinetic mechanism that is used to model the reaction (Heijnen 2005; Link, Christodoulou, AND Sauer 2014; Machado, ET AL. 2011; Srinivasan, Cluett, AND Mahadevan

2015). Accordingly,  $f$  is a nonlinear function of the metabolite concentrations, enzyme kinetic parameters ( $\theta$ ) and other input concentrations ( $u$ ).

The ability to predict the steady state and dynamic responses of metabolic networks, under in vivo conditions, to different perturbations is dependent on the numerical values of the enzyme kinetic parameter values ( $\theta$ ) in Equation (1). Analyzing the ability of a metabolic network to exhibit dynamic characteristics like multiple steady states and oscillations, irrespective of the structure of the network, is one example where parameter values might play a crucial role (Srinivasan, Cluett, AND Mahadevan 2015; Vital-Lopez, Maranas, AND Armaou 2006)(Srinivasan et al., 2017?). The use of in vitro, or unreliable in vivo parameter estimates, reduces confidence in the model predicted behaviour. Consequently, the reduction in confidence hampers the use of these models to gain insight into the functioning of metabolic networks (Tran, Rizk, AND Liao 2008; Chakrabarti, ET AL. 2013). The insights gained from the use of kinetic models are subsequently used to design changes to these metabolic networks to achieve various goals. These goals could either be to increase metabolite production for biosynthesis of different chemicals (Almquist, ET AL. 2014; Khodayari, ET AL. 2016; Costa, Hartmann, AND Vinga 2016; Andreozzi, ET AL. 2016)(Srinivasan et al., 2017) or to find therapeutic targets to cure ailments (Apaolaza, ET AL. 2017). Hence, an increase in uncertainty in model predicted responses is also an obstacle for using the predicted responses as a basis for designing the metabolic networks to achieve these goals.

If all intracellular metabolite concentrations can be measured over a time course, a nonlinear programming problem can be formulated to estimate the enzyme kinetic parameters ( $\theta$ ) in Equation (1), based on the measured data. The minimization of least square error between the measured ( $x^*$ ) and modeled ( $x$ ) concentrations can be used as an objective function (Equation 2a) for the optimization problem (Equation 2), with fixed upper ( $\theta_u$ ) and lower ( $\theta_l$ ) bounds for the parameter values (Equation 2b). The difference between the data and the model estimate for each output and at each time point in the objective function is weighted by the variance in the experimental data  $\sigma_{kl}^*$  for that corresponding variable and time point.

$$\min_{\theta} \sum_{k=1}^m \sum_{l=1}^d \left( \frac{x_{kl}^* - x_{kl}}{\sigma_{kl}^*} \right)^2 \quad (2a)$$

$$\theta_l \leq \theta \leq \theta_u \quad (2b)$$

However, not all metabolite concentrations used in the model (Equation 1) can be measured. Additionally, measurable fluxes in the metabolic network also need to be included as part of the parameter estimation problem. In such scenarios, the parameter estimation problem is modified to suit a new system of equations shown below (Equation 3). The new system of equations is obtained by augmenting the original system (Equation 1) with Equation (3c) that models the relationship between the measurable metabolite concentrations and fluxes ( $y$ ) and the unmeasured concentrations ( $x$ ) that are used in the original model (Equation 1) above. Equation (3) now has additional parameters  $\mu$  that define this relationship and also need to be estimated.

$$\dot{x} = \mathbf{S}v \quad (3a)$$

$$v = f(x, y, \theta, u) \quad (3b)$$

$$\dot{y} = h(x, y, \theta, u) \quad (3c)$$

In systems identification, the measured concentrations and fluxes ( $y$ ) are called output or observed variables, and the unmeasured concentrations ( $x$ ) are called the state variables. For estimating  $\theta$ , the metabolite concentrations  $x$  in the optimization problem (Equation 2) are substituted with the output variables  $y$ .

However, the ability to determine unique solutions to parameters  $\theta$  is governed by the identifiability of these parameters in the model (McLean AND McAuley 2012). The identifiability of parameters in nonlinear models can be classified into two categories: structural (or a priori) and practical (or posterior) identifiability. Any system (Equation 3) is said to be structurally identifiable if, for an input-output mapping defined by  $y = \Phi(\theta, u)$  for at least one input function  $u$ , any two values of parameters  $\theta_1$  and  $\theta_2$  satisfy the relationship in Equation (4) below.

$$\Phi(\theta_1, u) = \Phi(\theta_2, u) \iff \theta_1 = \theta_2 \quad (4)$$

Accordingly, the system can have a unique solution, a finite number of non-unique solutions or an infinite number of solutions for all input functions, and is said to be structurally globally identifiable, locally identifiable or non-identifiable, respectively. So, the structural identifiability of parameters in a dynamic model helps establish the presence or absence of a relationship between the unobservable state variables and the observable output variables. Consequently, the effect of model structure and parameterization on the ability

to infer true parameter values from experimental data is determined by the structural identifiability of the parameter.

Experimental data from many physical systems is usually noisy, and when parameters are estimated on the basis of noisy data, the ability to estimate unique parameter values to satisfy Equation (4) is referred to as practical identifiability. So, the effect of the available experimental data on the ability to estimate unique parameter values is determined by the practical identifiability of the parameter. Accordingly, practical identifiability of a parameter is contingent upon the nature, quality and quantity of data available to estimate the parameter as opposed to the structure and parameterization of the model.

Thus, on the one hand, establishing the structural identifiability of parameters enables one to propose models that are not only appropriate representations of physical processes, but also are parameterized in such a way that the value of these parameters can be estimated from measurable data. On the other hand, establishing practical identifiability of parameters in any model helps design experiments that are minimal, informative and useful for parameter estimation.

Algorithms have been extensively developed for establishing structural identifiability of dynamic models of biological systems (Audoly, ET AL. 2001)(IEEE Trans paper from 2007). Most of these algorithms use methods based on differential algebra (Ljung AND Glad 1994). More recent methods take a profile-likelihood-based approach to establish both structural and practical identifiability (Raue, ET AL. 2009). All of these methods are heavily dependent on the availability of dynamic time course data for all the output variables of the system and scale poorly with increase in size of the modeled system (Bellu, ET AL. 2007).

Computational burden due to poor scalability can be partly addressed with the current increases in computational power. Although, improvements in measurement technologies can enable one to obtain high frequency dynamic data, no scalable methodologies exist to utilize the abundantly available steady state data for parameter estimation and identifiability for large metabolic networks.

In this paper, we propose a methodology to establish practical identifiability for parameters in kinetic models of metabolism. We present a computer algebra-based method that can facilitate practical identifiability as well as experimental design for estimating parameters separately for each individual reaction within a metabolic network based on available steady state experimental data. This enables us to address the

twin issues of scalability and data availability. We illustrate the utility of this method by applying it for a small network of gluconeogenesis in *E. coli* and demonstrating our ability to propose experiments that will facilitate parameter estimation for a kinetic model of this network. We also demonstrate the scalability of the proposed methodology to facilitate experimental design by applying it to a relatively larger metabolic network of the human red blood cell hepatocyte.

## 2 Methods:

### 2.1 A method for practical identifiability of kinetic models of metabolism:

In this section, we show how practical identifiability of kinetic parameters in a dynamic model of metabolism can be established. A summary of the methodology in the form of a flow diagram is shown in Figure 1. In a kinetic model, the value of every flux  $v_i$  is expressed using one of the many available enzyme kinetic formulations (Equation 3b). Without loss of generality, all of these kinetic formulations can be expressed as nonlinear algebraic equations. The fluxes are expressed as a function of the metabolite concentrations  $\mathbf{x}$  and the kinetic parameters  $\theta$  (Figure 1a).

Let  $\theta \in \mathbb{R}^p$  in Equation (3b) for each flux  $v_i$  in the network. For each experiment  $j = 1, 2, \dots, n$ , we assume that all metabolite concentrations  $x$  and reaction fluxes  $v$  are measurable. The pertinent information for each experiment is available as a vector of concentrations and fluxes,  $\mathbf{x}_j$  and  $\mathbf{v}_j$ , respectively (Figure 1b).

In order to establish the identifiability of kinetic parameters for each flux  $v_i$ , we describe a computer algebra-based method. The primary use of the computer algebra system is to obtain closed form expressions for each parameter in  $\theta$  for each flux  $v_i$  (Figure 1b). This is done by solving a system of nonlinear algebraic equations in  $\mathbb{R}^p$ , shown in Equation (5).

$$v_{i,k} = f_k(\mathbf{x}_k, \theta, u_k) \quad \forall k = \{1, 2, \dots, p\} \subset \{1, 2, \dots, n\} \quad (5)$$

Each equation in (5), indicated by the index  $k$ , corresponds to the kinetic rate law expression  $g(x, \theta, u)$  for  $v_i$ , described earlier in Equation (3b), written for concentrations and fluxes obtained from experiment  $k$ . Solving the system in Equation (5) results in  $\mathbb{R}^p$  nonlinear expressions for parameters in  $\theta$ , where  $N(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$

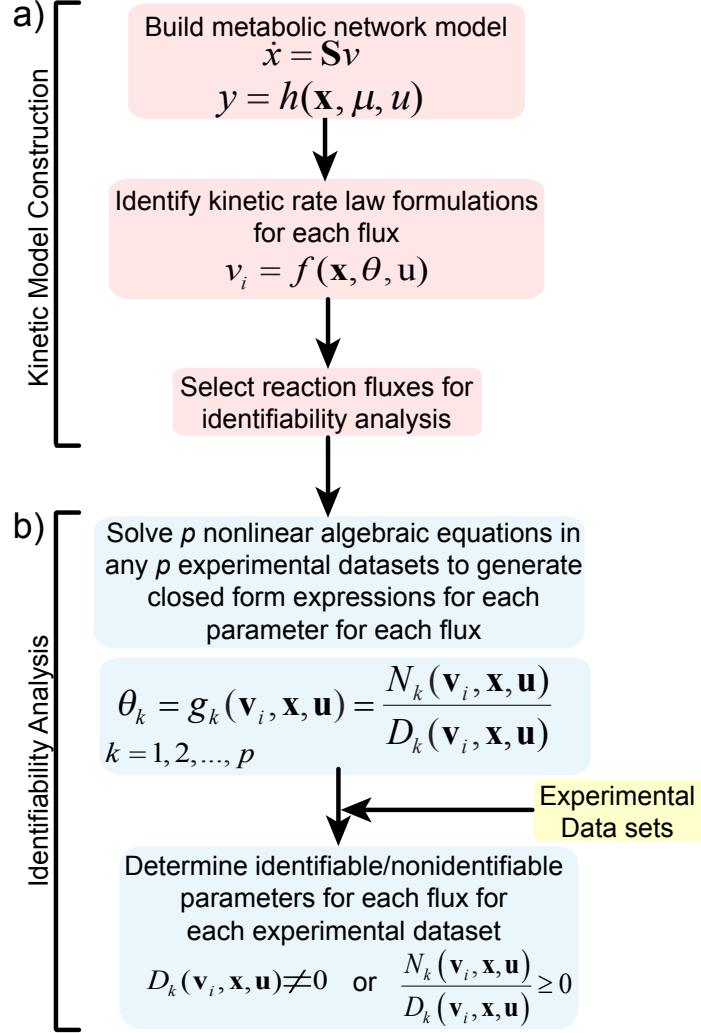


Figure 1: A flow diagram showing the methodology developed to establish practical identifiability of parameters in kinetic models of metabolism. a) The steps for the construction of a kinetic model of a metabolic network are shown. The choice of rate law formulations to describe metabolic fluxes influences the identification methodology. The identifiability of parameters for each flux can be established independently. b) The steps for identifiability analysis for parameters of a single flux are shown.

is the numerator of  $g$ , and  $D(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$  is the denominator of  $g$  (Figure 1b).

$$\theta_k = g_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) = \frac{N_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})}{D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})} \quad (6)$$

The identifiability of parameter  $\theta_k$  for flux  $v_i$  can be established by determining the value of  $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$  (Figure 1b): any parameter  $\theta_k$  is said to be practically identifiable (practically non-identifiable) if  $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) \neq 0$  ( $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) = 0$ ). Furthermore, the physical properties of the kinetic parameter values can be used to distinguish between identifiable and non-identifiable parameter values by designating only parameters with a non-negative value of  $g_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$  as identifiable (Figure 1b).

In the following sections we provide a previously published kinetic model of a small gluconeogenic network, followed by a demonstration of our methodology to establish practical identifiability for one of the fluxes in this network.

## 2.2 Kinetic model of gluconeogenesis in *E. coli*:

The proposed model for acetate consumption through gluconeogenesis and its corresponding kinetic model is used as a case study to illustrate the utility of identifiability analysis for the design of experiments for estimating parameters in kinetic models of metabolism. The kinetic model is described below.

$$\frac{d}{dt} pep = v_1 - v_2 - v_4 \quad (7)$$

$$\frac{d}{dt} fdp = v_2 - v_3 \quad (8)$$

$$\frac{d}{dt} E = v_{e,max} \left( \frac{1}{1 + \left( \frac{fdp}{K_e^{fdp}} \right)^{n_e}} \right) - dE \quad (9)$$

The kinetic expressions for fluxes  $v_1$  through  $v_4$  are given below. The consumption of acetate through  $v_1$  and conversion of  $pep$  through  $v_2$  are expressed in Equations (10) and (11) respectively using Michaelis-Menten kinetics. The acetate flux through  $v_1$  is also governed by the quantity of available enzyme  $E$ .

$$v_1 = k_1^{cat} E \frac{acetate}{acetate + K_1^{acetate}} \quad (10)$$

$$v_2 = V_2^{max} \frac{pep}{pep + K_2^{pep}} \quad (11)$$

$$v_3 = V_3^{max} \frac{fdp (1 + fdp)^3}{(1 + fdp)^4 + L_3 \left( 1 + \frac{pep}{K_3^{pep}} \right)^{-4}} \quad (12)$$

The allosterically regulated flux  $v_3$  for the consumption of  $fdp$  is expressed in Equation (12) using the Monod-Wyman-Changeux (MWC) model for allosterically regulated enzymes, where  $\tilde{f}dp$  refers to the ratio of  $fdp$  with respect to its allosteric binding constant  $K_3^{fdp}$ . The added flux  $v_4$  for the export of  $pep$  is expressed as a linear equation dependent on  $pep$  in Equation (13).

$$v_4 = k_4^{cat} \cdot pep \quad (13)$$

### 2.3 Data for establishing parameter identifiability in kinetic model of gluconeogenesis:

Steady state metabolomics and fluxomics data can be gathered under different physiological conditions by either perturbing the expression levels for different enzymes within a metabolic network, or by changing the substrate concentrations under which the cells grow. The aforementioned model of gluconeogenesis has three different fluxes ( $v_1$ ,  $v_2$  and  $v_3$ ) whose enzyme expression parameters ( $V_1^{max}$ ,  $V_2^{max}$  and  $V_3^{max}$ ) can be perturbed to simulate the repression and over expression of the corresponding enzymes. Furthermore, the acetate concentration, that determines the acetate uptake flux  $v_1$ , can also be perturbed to measure cellular response to changes in the substrate concentration. We use the in silico data generated by 18 different experiments wherein these four model parameters (*acetate*,  $V_1^{max}$ ,  $V_2^{max}$  and  $V_3^{max}$ ) were perturbed to demonstrate practical parameter identification with our methodology. The experiments and the perturbed values of each of the four the parameters are given in the Appendix.

### 2.4 Identifiability of parameters in a kinetic model of gluconeogenesis:

Here, we demonstrate the use of our computer algebra-based methodology to establish practical identifiability of parameters for flux  $v_2$  in the small model of gluconeogenesis described in Section 2.2. We choose flux  $v_2$ , expressed using Michaelis-Menten kinetics for this demonstration.  $v_2$  has two parameters,  $V_2^{max}$  and  $K_2^{pep}$  that need to be identifiable so that they can be estimated from experimental data. Here, we assume that data (for the concentrations and fluxes) from at least two different sets of experiments is available. We label the available concentrations as  $pep^1$  and  $pep^2$ , and the fluxes as  $v_2^1$  and  $v_2^2$  respectively, from each experiment. Accordingly, the nonlinear algebraic equations shown in Equation (5) can be formulated for flux



$v_2$  as follows:

$$v_2^1 = V_2^{max} \frac{pep^1}{pep^1 + K_2^{pep}} \quad (14a)$$

$$v_2^2 = V_2^{max} \frac{pep^2}{pep^2 + K_2^{pep}} \quad (14b)$$

Solving this simultaneous system of equations using Mathematica (Wolfram Research, USA), a computer algebra system, we get two nonlinear algebraic equations in the parameters  $V_2^{max}$  and  $K_2^{pep}$  based on the form shown earlier in Equation (6).

$$V_2^{max} = \frac{v_2^1 v_2^2 (pep^1 - pep^2)}{v_2^2 pep^1 - v_2^1 pep^2} \quad (15a)$$

$$K_2^{pep} = \frac{pep^1 (v_2^1 pep^2 - v_2^2 pep^1)}{v_2^2 pep^1 - v_2^1 pep^2} \quad (15b)$$

In Equation (15), the denominator of the right hand side expression is used to test the identifiability of parameters  $V_2^{max}$  (Equation 15a) and  $K_2^{pep}$  (Equation 15b) for different available experimental data set combinations. Since the enzyme binding constant ( $K_2^{pep}$ ) and maximum reaction rate ( $V_2^{max}$ ) cannot be negative, we can further reduce the criteria for identifiability for both these parameters by saying that the evaluated expressions should be non-negative.

## 2.5 Experimental design for parameter estimation in kinetic models of metabolism:

The identifiability of each parameter based on each experiment indexed as  $j = \{1, \dots, n\}$  is established based on the methodology described previously in Section 2.1 and demonstrated in Section 2.4. Subsequently, for any flux  $v_i$ , if for any  $p$  combinations of indices  $j$ , the experimental concentrations ( $\mathbf{x}_j$ ) and fluxes ( $\mathbf{v}_j$ ) do not satisfy the condition for identifiability for any parameter in  $\theta \in \mathbb{R}^p$ , i.e.,  $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) = 0$  for any  $k$ , then at least one of the  $p$  experiments needs to be changed to make a parameter  $\theta_k$  identifiable. Consequently, the corresponding experiment cannot be used for parameter estimation and needs to be discarded from the set of all necessary experiments. Furthermore, another experiment from  $j = \{1, \dots, n\}$  needs to be selected such that parameter  $\theta_k$  is identifiable. This process has to be repeated until all parameters in  $\theta \in \mathbb{R}^p$  are identifiable for flux  $v_i$ . In doing so, we can arrive at a set of  $p$  experiments that will always result in practically identifiable parameters for flux  $v_i$ . This analysis can be performed for each flux in a metabolic

network independent of all the other fluxes. Hence, our method is theoretically scalable even to genome-scale models. We show the application of this design methodology for flux  $v_2$  in the gluconeogenic model in Section.

Note that if none of the  $n$  pre-selected experiments satisfy the identifiability condition, then we can design an  $(n+1)^{th}$  experiment that can replace one of the experiments that causes practical non-identifiability using our methodology.

### 3 Results:

First, in Section 2.4 we demonstrate how our methodology can be applied for one of the fluxes in the gluconeogenic model. In Section 3.1 that follows, we discuss the ability to apply our methods to different nonlinear kinetic rate law formulations within the context of the gluconeogenic model. Then, in Section we discuss the different experimental design strategies that were found through our identifiability analysis to enable kinetic parameter estimation for the each flux in the model using steady state data. Finally, in Section we expand the application of our method to a relatively large (39 reactions) kinetic model of the red blood cell metabolism.

#### 3.1 Getting closed-form expressions for each flux in Kotte model:

The complexity of the equations in our specific scenario is determined by the complexity of enzyme-metabolite interaction models used describe fluxes in metabolic networks. Although computer algebra systems (CAS) are capable of handling complex symbolic calculations, sometimes, the complexity of getting closed form expressions for all kinetic parameters of certain rate law formulations is too much for the CAS to overcome. We encountered this scenario in the case of the gluconeogenic model for flux  $v_3$  where the kinetics of the allosterically activated reaction are described by the MWC model for allosteric regulation. In order to overcome this computational difficulty, we used a convenience kinetic rate law formulation to describe the allosteric interaction in  $v_3$ . We give this formulation in Section.

We believe this problem of computational tractability to occur in other complex kinetic rate law formulations as well. Examples?

### 3.2 Experiments that can establish identifiability for each flux in Kotte model:

As indicated previously in Section, we performed different experiments by perturbing the parameters for fluxes  $v_1$ ,  $v_2$  and  $v_3$  in the model. We also perturbed the input substrate concentration to collect data on the concentrations and fluxes within the network. We describe the results of using this data to establish the identifiability of various parameters in the kinetic model of the small metabolic network.

### 3.3 Combinations of experiments that will enable identification of all model parameters:

### 3.4 Expanding methodology to RBC model:

If any data generated from a perturbation experiment  $i$  results in nonidentifiability, we eliminate experiment  $i$  from the list of experiments that need to be performed for parameter estimation.

Table 1: Table showing the perturbed values of all fluxes used for parameter estimation.

Designation	Perturbed Fluxes	Perturbed Values
P1	$v_1$	2
P2	$v_2$	0.2
P3	$v_3$	0.5

#### Outline:

- parameter estimation is a well developed field typically using minimization of least square error to estimate model parameters from available experimental data
- if parameters are structurally identifiable, it does not guarantee practical identifiability from noisy experimental data
- identifiability dependent on whether given datasets (outputs) for estimation can sufficiently distinguish between different parameter values

## Sections:

- datasets required for parameter estimation in kinetic models of metabolism (methods?)
- identifiability in kotte model - scalability, number of experiments required, requirements for time course data(? in the intro)
- identifiability in large rbc model

## References

- Almquist, J., ET AL. (2014) Kinetic models in industrial biotechnology - Improving cell factory performance, *Metab. Eng.* 24, 38–60.
- Andreozzi, S., ET AL. (2016) Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models, *Metab. Eng.* 35, 148–159.
- Apaolaza, I., ET AL. (2017) An in-silico approach to predict and exploit synthetic lethality in cancer metabolism, *Nat. Commun.* 8.1, 459.
- Audoly, S., ET AL. (2001) Global identifiability of nonlinear models of biological systems, *IEEE Trans. Biomed. Eng.* 48.1, 55–65.
- Bellu, G., ET AL. (2007) DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* 88.1, 52–61.
- Chakrabarti, A., ET AL. (2013) Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints, *Biotechnol. J.* 8.9, 1043–1057.
- Costa, R. S., A. Hartmann, AND S. Vinga (2016) Kinetic modeling of cell metabolism for microbial production, *J. Biotechnol.* 219, 126–141.
- Heijnen, J. J. (2005) Approximative kinetic formats used in metabolic network modeling, *Biotechnol. Bioeng.* 91.5, 534–545.
- Khodayari, A., ET AL. (2016) A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains, *Nat. Commun.* 7, 13806.

- Link, H., D. Christodoulou, AND U. Sauer (2014) Advancing metabolic models with kinetic information, *Curr. Opin. Biotechnol.* 29.1, 8–14.
- Ljung, L. AND T. Glad (1994) On global identifiability for arbitrary model parametrizations, *Automatica* 30.2, 265–276.
- Machado, D., ET AL. (2011) Modeling formalisms in systems biology, *AMB Express* 1.45.
- McLean, K. A. P. AND K. B. McAuley (2012) Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures, *Can. J. Chem. Eng.* 90.2, 351–366.
- Raue, A., ET AL. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, *Bioinformatics* 25.15, 1923–1929.
- Srinivasan, S., W. R. Cluett, AND R. Mahadevan (2015) Constructing kinetic models of metabolism at genome-scales: A review. *Biotechnol. J.* 10.9, 1345–59.
- Tran, L. M., M. L. Rizk, AND J. C. Liao (2008) Ensemble modeling of metabolic networks. *Biophys. J.* 95.12, 5606–5617.
- Vital-Lopez, F., C. Maranas, AND A. A. Armaou (2006) Bifurcation analysis of metabolism of E. coli at optimal enzyme levels, in: *Proc. 2006 Am. Control Conf.* IEEE, Minnesota, 3439–3444.