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 (θ) 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 (θ) 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 (θ) in Equation (1), based on the measured data. The minimization of least square error between the measured (x^*) and modeled (x) concentrations, weighted by the variance in the experimental data σ^*_{kl} for each concentration at each time point, is used as an objective function (Equation 2a) for the optimization problem (Equation 2). The parameter values are determined within fixed upper (θ_u) and lower (θ_l) bounds (Equation 2b).

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

$$\theta_l \leq \theta \leq \theta_u \tag{2b}$$

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. The parameter vector (θ) is augmented with additional parameters that define this relationship. These additional parameters also need to be estimated.

$$\dot{x} = \mathbf{S}v \tag{3a}$$

$$v = f(x, y, \theta, u) \tag{3b}$$

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

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 θ, 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 θ 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 θ_1 and θ_2 satisfy the relationship in Equation (4) below.

$$\Phi(\theta_1, u) = \Phi(\theta_2, u) \iff \theta_1 = \theta_2 \tag{4}$$

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.

Methods and tools for structural identification of parameters based on differential algebra (Ljung AND Glad 1994; Audoly, ET AL. 2001; Bellu, ET AL. 2007) and profile likelihood (Raue, ET AL. 2009) are available. However, only the profile likelihood-based methods enable experimental design by facilitating practical identification of parameters. Nonetheless, this method still depend on solving a non-convex nonlinear least

squares problem (Equation 2) to get likelihood estimates of parameters, and hence still suffers from all the inherent difficulties associated with obtaining global optimal solutions for non-convex optimization problems. This also makes it un-scalable for experimental design and practical identifiability of parameters in kinetic models of large metabolic networks.

In this paper, we propose a scalable methodology to establish practical identifiability for parameters in kinetic models of metabolism using abundantly available steady state concentration and flux data. We present a computer algebra-based method that can facilitate experimental design through practical identifiability of parameters separately for each individual reaction within a metabolic network based on available steady state experimental data. We illustrate the utility of this method by applying it to 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.

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 x and the kinetic parameters θ (Figure 1a).

Let $\theta \in \mathbb{R}^p$ in Equation (3b) for each flux v_i in the network. For each experiment $j = 1, 2, ..., n$, 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 θ for each flux v_i (Figure 1b). This is done by solving a system of nonlinear algebraic

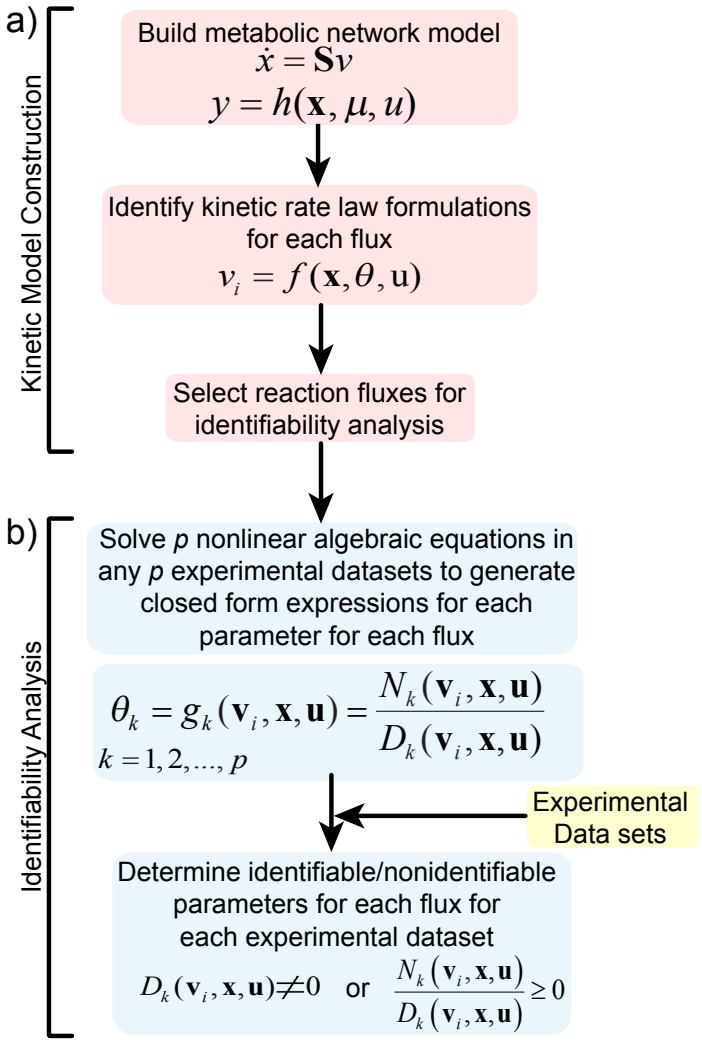

Figure 1: A flow diagram showing the methodology developed to establish practical identifiability of parameters in kinetic models of metabolism. a) The steps for the construction of a kinetic model of a metabolic network are shown. The choice of rate law formulations to describe metabolic fluxes influences 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.

equations in $\mathbb{R}^p$, shown in Equation (5).

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

Each equation in (5), indicated by the index k, corresponds to the kinetic rate law expression $f(x, \theta, u)$ 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 θ, where $N(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$ is the numerator of g, and $D(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$ is the denominator of g (Figure 1b).

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

The identifiability of parameter θ_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 θ_k is said to 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.

$$\frac{d}{dt} pep = v_1 - v_2 - v_4 \tag{7}$$

$$\frac{d}{dt} fdp = v_2 - v_3 \tag{8}$$

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

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.

$$v_1 = k_1^{cat} E \frac{ac}{ac + K_1^{ac}} \tag{10}$$

$$v_2 = V_2^{max} \frac{pep}{pep + K_2^{pep}} \tag{11}$$

$$v_3 = V_3^{max} \frac{f\tilde{d}p \left(1 + f\tilde{d}p\right)^3}{\left(1 + f\tilde{d}p\right)^4 + L_3 \left(1 + \frac{pep}{K_3^{pep}}\right)^{-4}} \tag{12}$$

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 $f\tilde{d}p$ 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).

$$v_4 = k_4^{cat}.pep \tag{13}$$

2.3 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_1 in the small model of gluconeogenesis described in Section 2.2. For the purposes of this demonstration, we assume that all relevant steady state metabolite concentrations and fluxes can be measured.

In flux v_1, the concentration of the enzyme E is used as a variable. If the enzyme concentration can be measured, then the expression for v_1 given in Equation (10) can be used for identifiability analysis of parameters k_1^{cat} and K_1^{ac}. However, if enzyme concentrations are not available, the expression in Equation (10) can be modified as given in Equation (14) below. This equation does not make use of enzyme concentrations as variables, and uses V_1^{max} and K_1^{ac} as parameters.

$$v_1 = V_1^{max} \frac{ac}{ac + K_1^{ac}} \tag{14}$$

We choose this expression for flux v_1 (Equation 14), expressed using Michaelis-Menten kinetics, to demonstrate our method for practical identifiability. Both V_1^{max} and K_1^{ac} 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 i.e., in Equation (5) $k = 1, 2$. We label the available

concentrations and fluxes as $ac^{(k)}$ and ${v_1}^{(k)}$, respectively. Accordingly, the nonlinear algebraic equations shown in Equation (5) can be formulated for v_1 as follows:

$$ {v_1}^{(k)} = V_1^{max} \frac{ac^{(k)}}{ac^{(k)} + K_1^{ac}} \qquad\qquad k = \{1, 2\} \tag{15} $$

Solving this simultaneous system of k equations using Mathematica (Wolfram Research, USA), a computer algebra system, we get $p = 2$ nonlinear algebraic equations in the parameters V_1^{max} and K_1^{ac} based on the form shown earlier in Equation (6).

$$ V_1^{max} = \frac{v_1^{(1)} v_1^{(2)} (ac^{(1)} - ac^{(2)})}{v_1^{(2)} ac^{(1)} - v_1^{(1)} ac^{(2)}} \tag{16a} $$

$$ K_1^{ac} = \frac{ac^{(1)} ac^{(2)} (v_1^{(1)} - v_1^{(2)})}{v_1^{(2)} ac^{(1)} - v_1^{(1)} ac^{(2)}} \tag{16b} $$

In Equation (16), the denominator of the right hand side expression is used to test the identifiability of parameters V_1^{max} (Equation 16a) and K_1^{ac} (Equation 16b) for different available experimental data set combinations. Since the enzyme binding constant (K_1^{ac}) and maximum reaction rate (V_1^{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 (Figure 1b).

2.4 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 (ac, V_1^{max}, V_2^{max} and V_3^{max}) are 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.

The minimum number of experiments from which data is required for identifying all the parameters of a given flux is determined by dimension $\mathbb{R}^p$ of the parameter space of a chosen flux v_i. For instance, as demonstrated above, data from two distinct experiments is required for identifying the two parameters of v_1. In this case $p = 2$. This also applies for identifying parameters in v_2. Hence, multiple data sets generated using a combination of any two experiments are used to test the identifiability of parameters for v_1 and v_2. The total number of such possible combinations is 306 (18 x 17) from the 18 different experiments mentioned earlier. A similar calculation for identifying the three parameters of v_3, where $p = 3$, leads to 4896 distinct data combinations (18 x 17 x 16) of three experiments.

2.5 Degee of Identifiability, a quantitative measure of practical identifiability of parameters:

We express the practical identifiability of kinetic parameters using a simple quantitative term called the degree of identifiability. We describe the degree of identifiability of any single parameter as the percentage of all data combinations (used to test practical identifiability) that can identify that parameter. As an example, if 90% of all the experimental data combinations used for testing can identify a parameter θ_i, then the degree of identifiability of θ_i is said to be 0.9 or 90%. On the other hand, if only 50% of the combinations can identify another parameter θ_j, then θ_j has a degree of identifiability of 0.5 or 50%. Furthermore, we can create a hierarchy of practically identifiable parameters using their degree of identifiabilities. In the above instance of the two parameters θ_i and θ_j that have degrees of identifiabilities of 90% and 50% respectively, θ_i is classified to be more identifiable than θ_j due to its relatively higher degree of identifiability. Determining this hierarchy of identifiable parameters can help in distinguishing parameters that can be identified by any type and any combination of experiments from parameters that can be identified by only a select type and combination of experiments, and design minimal sets of experiments that can practically identify all kinetic parameters used to model a metabolic network.

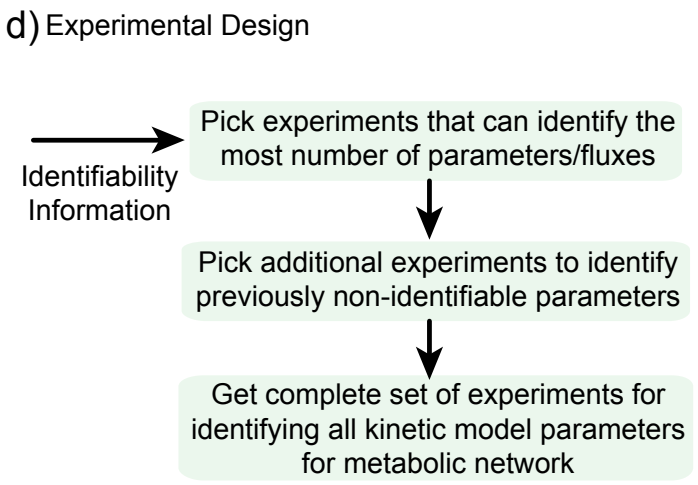

Figure 2: Flow diagram showing a method for experimental design that uses our methodology for practical identifiability of parameters to determine the number and type of experiments required to identify all fluxes within a given metabolic network.

2.6 Experimental design through practical parameter identifiability for kinetic models of metabolism:

Following the methodology described in Section 2.1, and demonstrated in Section 2.3 for a single flux using data from a combination of two different experiments, all distinct combinations found from Section 2.4 can be tested for their ability to practically identify parameters for any of the three fluxes in the small metabolic network. This step would help distinguish identifiable experiment combinations from combinations that do not practically identify any parameter in the model (Figure). Subsequently, it is possible to obtain a collection of experiments that make up all identifiable data combinations that can be performed to obtain the most minimal and informative set of experiments to identify as many parameters as possible (Figure 2). Consequently, the set of experiments can be used to estimate the identifiable parameters in the model. We explain this formally below.

The identifiability of each parameter based on each experiment indexed as $j = \{1, ..., n\}$ is established based on the methodology described previously in Section 2.1 and demonstrated in Section 2.3. 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 θ_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 $j = \{1, ..., n\}$ needs to be selected such that parameter θ_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.

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 3.1, we show that the ability to establish practical identifiability of kinetic parameters in metabolic network models using our methodology, described in Section 2.1, relies upon the nonlinearity of the kinetic rate law formulations used to describe the fluxes. We use the gluconeogenic model described in Section 2.3 as an example. Then, in Section 3.2 we show how the degree of identifiability of the maximum reaction rate parameter is always higher than the degree of identifiability of the corresponding enzyme binding parameter irrespective of enzyme kinetics rate law used to describe the flux. In Section 3.3 that follows, we discuss the the ability to determine experiments for parameter identifiability for a given flux based on the informativeness of a given experiment. In this section, we show how the informativeness of a given type of experiment to identify a specific flux can be deduced from its contribution towards the practical identifiability of the parameters for a given flux. In Section 3.4 we discuss the identifiability of parameters when data with additive noise is used with our methodology to test practical identifiability. Finally, in Section 3.5 we provide a motivation for the need for experimental design, especially to identify kinetic models metabolic networks.

3.1 Nonlinearity of enzyme kinetic rate law expression affects identifiability analysis:

Establishing parameter identifiability for fluxes in kinetic models of metabolism using the methodology we describe in the Methods section is governed by the nonlinear complexity of the enzyme kinetic rate law used to model a specific flux. We demonstrate one example of how the methodology works in Section 2.3 for parameters of flux v_1. The expression in Equation (16) is obtained by using a computer algebra system. To recall, in this specific case, a computer algebra system is used to solve for the parameters of a flux described using the Michaelis-Menten kinetic rate law when experimental data from two different experiments is available.

However, we find that the nonlinearity of the MWC kinetic rate law used to model the allosteric regulation of v_3 makes it computationally intractable for determining the closed form expressions of the three parameters V_3^{max}, K_3^{fdp} and K_3^{pep} using a computer algebra system (Mathematica and SymPy in Python). In order to overcome this computational obstacle, we model the reaction rate for v_3 using the convenience kinetic rate law formulation (Liebermeister AND Klipp 2006). The corresponding expression obtained for v_3 is given below (Equation 17).

$$v_3 = V_3^{max} \left(\frac{1}{1 + \frac{K_3^{pep}}{pep}} \right) \left(\frac{\frac{fdp}{K_3^{fdp}}}{1 + \frac{fdp}{K_3^{fdp}}} \right) \tag{17}$$

Using this expression for identifiability analysis, we find that each of the parameters V_3^{max}, K_3^{fdp} and K_3^{pep} have two different close form expressions owing to the presence of a square root term in their solutions. These distinct expressions are denoted by (1) and (2) following the respective parameter names throughout the rest of the document: $V_3^{max}(1)$, $K_3^{fdp}(1)$, $K_3^{pep}(1)$, and $V_3^{max}(2)$, $K_3^{fdp}(2)$, $K_3^{pep}(2)$.

The impact of nonlinearity of a kinetic rate law expression affecting the practical identifiability of parameters in the expression may not be an unique problem isolated to the system that we are investigating. Complicated expressions for describing fluxes have been extensively used to model observed experimental data for different fluxes in a variety of organisms (Chassagnole, ET AL. 2002; Peskov, Mogilevskaya, AND Demin 2012; Heerden, ET AL. 2014). However, the identifiability of the parameters used in these models

has never been truly examined. If these models are tractable, we believe our methodology can help in elucidating the identifiability of parameters in these models. If not, metabolic network fluxes can be expressed using alternative kinetic rate law models whose parameters can be tested for identifiability, and subsequently experiments can be designed for their estimation and model validation.

Next, we look some of the results we obtained for the practical identifiability of different parameters for all three fluxes v_1, v_2 and v_3 in the small gluconeogenic network model described in Section 2.2, with the flux for v_3 described by Equation (17).

3.2 Maximum reaction rates are more identifiable than enzyme binding constants:

In Figure 3 we show the number and percentage of combinations that are capable of identifying each parameter in each flux. Based on the definition given in Section 2.5, the percentages refer to the degree of identifiability of each parameter. The three panels in Figure 3 represent the degree of identifiability for parameters modeling the three different fluxes of the small network individually. Looking at these panels for v_1 (Figure 3a), v_2 (Figure 3b) and v_3 (Figure 3c), the degree of identifiability of the maximum reaction rates (V_i^{max}) in each of the three fluxes is higher than the degree of identifiability of the enzyme binding (K_i) constants and the allosteric activation constant (K_3^{pep}). We observe this trend irrespective of the fact that only v_1 and v_2 are represented by the same enzyme kinetic rate law (Michaelis-Menten), while the convenience kinetic rate law is used to model v_3. These observations lead us to conclude that the maximum reaction rate parameters are always more identifiable (as indicated by the higher degree of identifiability) than their enzyme binding constant counterparts irrespective of the enzyme kinetic rate law used to model the corresponding flux.

The most conspicuous of all observations from Figure 3 is the difference in the identifiability of parameters based on the flux in which these parameters are present. The variation in the degree of identifiability is not only dependent on the type of parameter being identified, but also depends on the flux whose parameters are being tested for identifiability. This can be tied to selecting experimental data that have the ability to satisfy conditions for practical identifiability for a given parameter. We discuss the dependence of the variation in

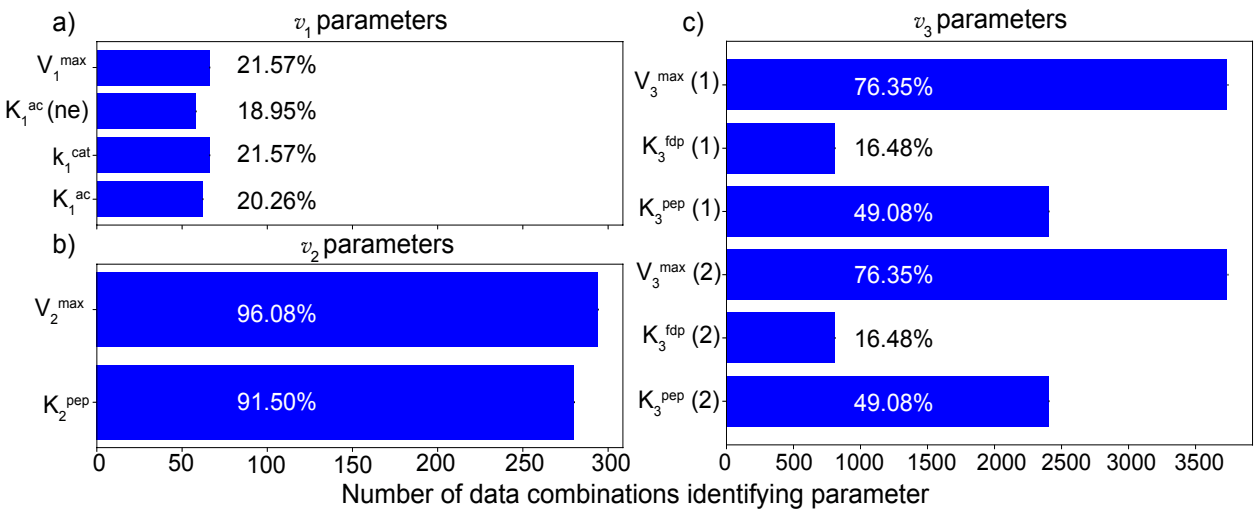

Figure 3: The number of data combination from 18 different in silico experiments that can practically identify each parameter in fluxes a) v_1, b) v_2 and c) v_3 when there is no noise in the input experimental data is shown. The percentage of total experimental data combinations that can identify each parameter is specified in/on the right hand side of the bar showing the total number of combinations identifying a given parameter. The total number of data combinations used to identify parameters in fluxes v_1 and v_2 is lower that the total number of combinations used to test identifiability of parameters for v_3 since both v_1 and v_2 have only two parameters while v_3 is modeled with three parameters. Hence, the total number of experiments used to identify parameters in v_1 and v_2 is two, while parameter identification in v_3 requires data from at least three different experiments.

the degree of identifiability for parameters of different fluxes from the point of view of the informativeness of the experiment from which data is collected for identifiability in the following section.

3.3 Experiments required for identifying parameters depends on the position of the flux in the metabolic network:

In system identification, any input signal should be rich or informative enough to guarantee full excitement of the dynamics of the system. Only information obtained from such changes in the input and the corresponding signals can be used to completely identify the system over its entire dynamic range. Input signals that guarantee to fully excite the dynamics of the system are termed as persistently excitable signals. In linear systems, persistence of excitation of any input signal is guaranteed. However, for nonlinear systems, like the metabolic network we deal with in this paper, persistence of excitation of input signals is not guaranteed and should be assessed on a case by case basis. The lack of persistence of excitation requires the design of experiments to satisfy the data needs for complete system identification.

In this section, we demonstrate a specific case of lack of persistence of excitation and the subsequent experimental design required to overcome this hurdle for parameter identifiability for a small nonlinear metabolic network model. The informativeness of experiments for identifying parameters in a given flux depends on its ability to satisfy the conditions determined for practical identifiability of that parameter. For example, we demonstrated in Section 2.3 that for a combination of any two experiments to be capable of identifying V_1^{max} and K_1^{ac} in v_1, the experiments must have distinct acetate concentrations as well as a different uptake flux v_1 between them. Thus, in this instance, the informativeness of the experiments (changes in the input acetate concentration) for identifying parameters of v_1 is determined by their ability to change the measured value of flux v_1. Similarly, the identification of parameters for v_2 requires the experiments to distinguish between values of both v_2 as well as *pep*, given both these variables are outputs of the model describing v_2.

In Figure 3a we show that only about 20% of all the experimental data sets satisfy these requirements for v_1 while more than 90% of the available experimental satisfies these requires for v_2 (Figure 3b). Thus, experimental design becomes crucial to determine the minimum number, as well as the nature of experiments

that can help identify parameters for both v_1 and v_2. Recall that we use experimental data from five different experiments to test the practical identifiability of parameters in the model (Section 2.4). In order to determine the type of experiments that help satisfy identifiability conditions for both parameters of v_1, we look at the distribution of the five different experiments in all data combinations that can identify these 2 parameters (Figure 4).

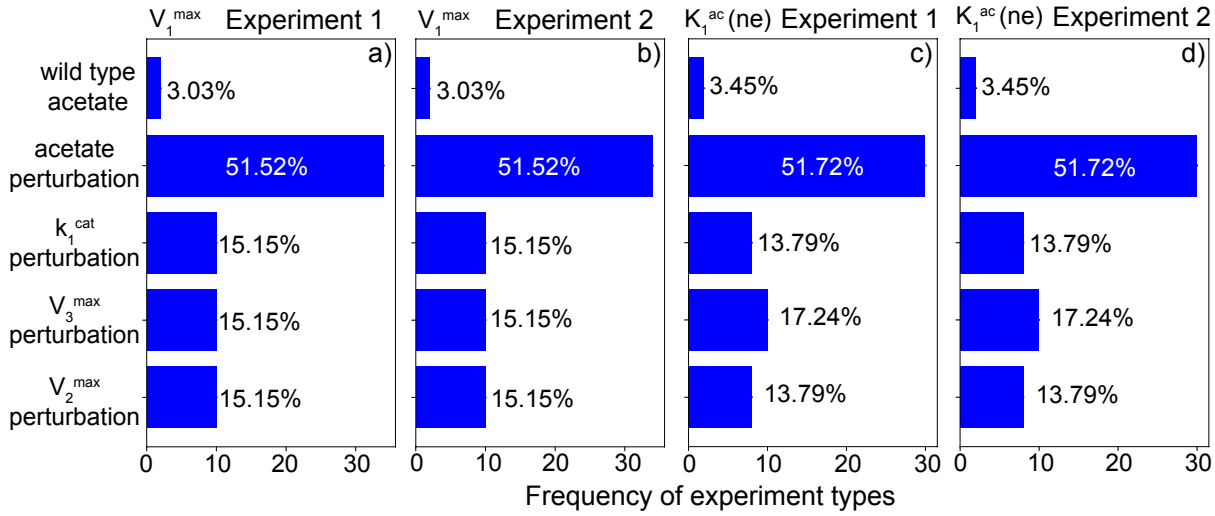

Figure 4: The contribution of different experiments types used in a combination of two experiments ($k = \{1, 2\}$) that can practically identify parameter V_1^{max} a) experiment 1 ($k = 1$), and b) experiment 2 ($k = 2$) in combination, and K_1^{ac} c) experiment 1 ($k = 1$), and d) experiment 2 ($k = 2$) in combination. The percentages reflect the fractional contribution of each experiment type towards all identifiable data combinations.

The contribution of experiments that involve changes in the acetate concentrations, which consequently bring about changes in the value of v_1 contribute to a significant part (¿ 50%) of the identifiable experimental data combinations in contrast to the other types of experiments. This matches with requirements for identifiability on the basis of the informativeness of experiments that was laid out earlier i.e., experiments must be able distinguish between changes in v_1 on the basis of changes in acetate. This informs why only about 20% of the 306 different experiment combinations can identify these parameters (Figure 3a). In can be gathered that only about 20% of all the data combinations used to test identifiability are informative enough to distinguish between both acetate as well as v_1, and most of these experiments require changes in the acetate concentration to be able to identify V_1^{max} and K_1^{ac} through changes in the input and the output

quantities acetate and v_1, respectively.

In contrast to v_1, we see that the different perturbation experiments have a higher contribution towards experimental data combinations that can identify parameters for v_2 (Figure). Also, it is important to note that since most experiment types can bring about changes in the dynamics of v_2 (changes to *pep* and value of v_2), the identifiability of parameters for v_2 is relatively higher than the identifiability of v_1 parameters (Figure 3b). We can further extend these arguments to also justify the observed contribution of experiments towards identifying parameters for v_3 (figure not shown).

Alternatively, both the degree of identifiability and the capability of different experiments to identify parameters of different fluxes based on their informativeness can be explained by the position of the flux in the metabolic network. We believe that the position of any given in the metabolic network determines the specific experiment that is persistently excitable enough to identify the parameters of that flux. This dependency of experiment informativeness on the position of the flux can be further elucidated using v_1 and v_2 as examples. We know from Equation (16) that identifiability of v_1 requires changes in both acetate and v_1. We also know, based on our knowledge of the Michaelis-Menten kinetic rate law that changes in the substrate concentration of a reaction can bring about a nonlinear change in the value of the corresponding reaction rate (flux). In this specific metabolic network, since the substrate is an input variable to the model and v_1 is the corresponding uptake flux, the substrate can be easily perturbed to create persistently excitable experiments to identify parameters in v_1. We can generalize this observation for the identification of all uptake fluxes in all metabolic networks, i.e., at a minimum, a change in the input substrate concentration may be necessary for an informative experiment to identify the uptake flux parameters.

On the other hand, the Michaelis-Menten model for v_2 also requires changes in *pep* and v_2 for persistently excitable experiments to identify v_2. However, since both of these are system outputs, satisfaction of this condition cannot be guaranteed without an analysis of the dynamics of the metabolic network and how changes in the input (acetate) bring about changes in the two requisite output quantities. Previously, a dynamical analysis of the current network has shown that the concentration of *pep*, and hence the flux v_2 are functions of not only the input acetate concentration, but also the levels of expression of the different enzymes within the network (Srinivasan et al., 2017). Hence, it is theoretically possible for any of the five

different experiment types to be persistently excitable to identify v_2. This is confirmed by the high degree of identifiability for both parameters of v_2, where in more than 90% of all data combinations can identify the parameters. Thus, this analysis informs us that the degree of identifiability and consequently, the type of experiments needed to identify different parameters varies widely depending the position of the flux in the metabolic network with respect to the inputs and the outputs of the network.

Nonetheless, all these explanations could only be verified due to the fact that the conditions for identifiability for both v_1 and v_2 are simple and readable due to the use of the Michaelis-Menten rate law to model these fluxes. Instead, in addition to presence of an allosteric interaction, when the convenience kinetic rate law is used to model this activation in v_3, we find that the complex expressions obtained for the three parameters preclude such an in-depth insight into the nature of persistently excitable experiments to identify v_3. Thus, the only avenue through which experiments can be designed for the identification of parameters of v_3 would be through our proposed methodology wherein manual curation of experimental combinations required for identifiability are not necessary.

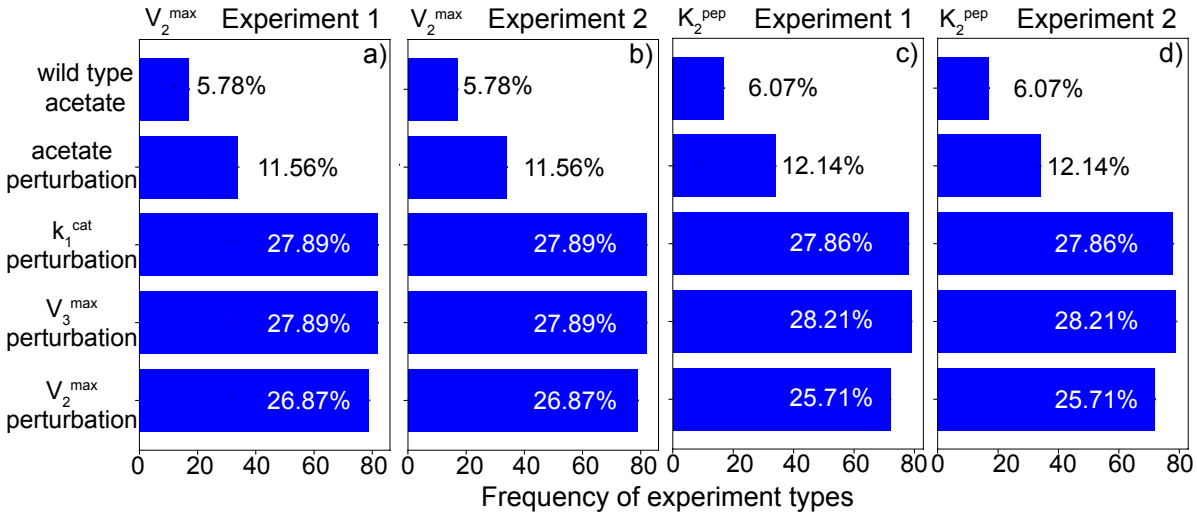

Figure 5: The contribution of different experiments types used in a combination of two experiments ($k = \{1, 2\}$) that can practically identify parameter V_2^{max} a) experiment 1 ($k = 1$), and b) experiment 2 ($k = 2$) in combination. Different experiments that contribute towards identifiability of K_2^{pep} c) experiment 1 ($k = 1$) and d) experiment 2 ($k = 2$) in combination.

3.4 Effect of noise on degree of identifiability is dependent on nonlinearity of enzyme kinetic rate law models:

The degree of identifiability for all the parameters shown in Figure 3 is obtained using noise-free in silico experimental data. However, actual experimental data from biological systems is usually noisy. Data gathered from experiments is usually associated with both biological noise attributable to the stochasticity of cellular function, as well as measurement noise associated with the techniques used to obtain concentration and flux measurements. In order to test the practical identifiability of parameters under circumstances where the presence of noise in the experimental data can affect the identifiability of the parameters, we apply our methodology on data (concentration and flux) with 5% additive noise. 50 different samples of noise with 0 mean and 0.05 standard deviation are drawn from a normal distribution and added to the steady state experimental data generated on the basis of experiments described in Section 2.4 to generate 50 different samples of noisy data. Subsequently, all 50 samples are tested for their ability to practically identify parameters for all three fluxes of the small network. The corresponding graphs generated for practical identifiability analysis using noisy experimental data are shown in the Appendix.

We hypothesize that the nonlinearity of the enzyme kinetic rate law used to model a flux, and consequently the complexity of the closed form expression obtained for the various parameters used in the model has an impact on the ability of noisy data to practically identify the parameters. The difference in the identifiability attributable to the aforementioned factors can be seen in the difference in the degree of identifiability for the parameters of v_1, v_2 and v_3 in the small metabolic network that we use to demonstrate our methodology.

Despite the presence of noise, the degree of identifiability of v_1 and v_2 (figure not shown) does not change from the case where no noise is present in the experimental data (Figure 3). However, the degree of identifiability for parameters of v_3 is affected by the presence of noise in the data (Figure). This effect is represented by the presence of a non-zero standard deviation and corresponding error bars on the graphs in the degrees of identifiability of parameters modeling v_3 (Figure). The standard deviation and the error bars represent the presence of data combinations that can identify a given parameter under certain values of noise, but not do so under different noise values. In the presence of 5% additive noise in the data, the conditions under which parameters for v_1 and v_2 can be identified are strictly satisfied by the same data combinations

that can identify these parameters in the absence of noise, i.e., the changes in the acetate concentration (for v_1) and *pep* (for v_2) are significant enough to enable identifiability. The contrinution of different experiment types towards identifiability also remains the same (figure not shown separately).

However, as stated in the previous section, the nonlinearity of the original kinetic rate law expression for v_3, and the complexity of the closed-form expression obtained for the parameters in the rate law model preclude us from testing the presence, and consequently the satisfaction of these conditions for parameters in v_3. Furthermore, the conditions for identifiability of v_3 parameters are not as simple as either requiring just a simple difference in *fdp* or *pep* or that of both concentrations. Thus, while under some noise values these identifiability conditions may be satisfied, it may not be possible to satisfy these conditions under other noise values. Accordingly, we end up with a distribution of experimental data combinations that can identify v_3 parameters depending on noise levels in the data. It is significant to note that the variability in the degree of identifiability of parameters is small (¡ 2%) when noise is present.

These observations, in conjunction with the estimation of the degree of identifiability of the different parameters (Figure 3), helps in determining not only the nature of experiments required for parameter identifiability (Figure 4), but also the order which the type of experiments necessary for identifying the entire metabolic network needs to be evaluated.

For instance, in the case of the example network that we use in this paper, given the high degree of identifiability for v_2 and large spread in the frequency of different experiment types that can contribute to identifiability for v_2, the focus should be on first choosing experiments for identifying v_1 followed by selecting experiments for identifying v_3. It is possible that the three or at most four experiments chosen to identify parameters of both v_1 and v_3 would suffice to identify v_2 as well without the need to perform additional experiments.

3.5 Utility of experimental data for practical parameter identification:

Apart from determining experiments that are informative enough to identify parameters for each flux in a kinetic model, experimental design can also be used to design and develop a minimal set of experiments that informative enough to estimate all parameters in a metabolic network model. However, there is not only a

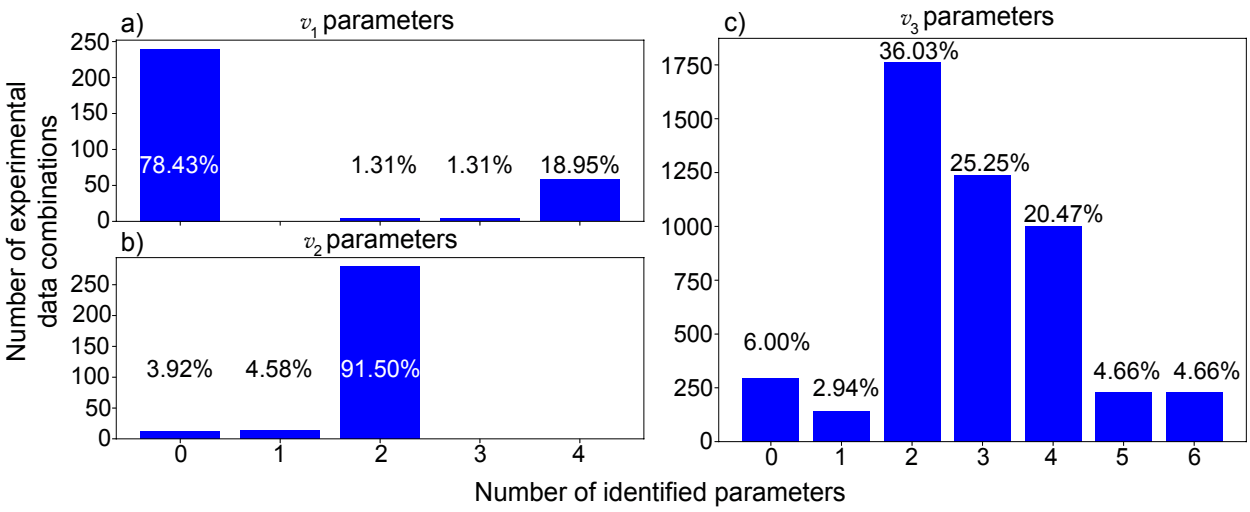

Figure 6: Utility of experimental data combinations on the basis of their ability to identify the most number of parameters. Information is shown for parameters modeling fluxes a) v_1, b) v_2 and c) v_3. The total possible number of parameters that can be identified by data from combinations of a) and b) two, and c) three experiments is shown in the horizontal axis. The vertical axis represents the total number of combinations that can identify the corresponding number of parameters. The percentages shown in the plots represent the fraction of the total combinations used to test identifiability of parameters for a given flux. A total of 306 data combinations are used for identifiability analysis for a) v_1 and b) v_2, and 4896 combinations are used to practical identifiability of parameters in c)v_3. Section 2.4 provides more details on how the combinations of experimental data are generated.

scarcity of experimental design methods in literature devoted to the purpose of designing experiments for parameter identifiability in kinetic models of metabolism, the motivation to develop such methods is not well understood. Hence, we found the need to stress the necessity for experimental design methodologies tailored specifically for metabolic network models. We do this by showing how useful the multiple combinations of experimental data sets are towards identifying the maximum number of parameters for each given flux (Figure 6).

In Figure 6, we show the percentage of all data combinations that can estimate a given number of parameters within a given flux. For flux v_1 (Figure 6a) we see that about 78% of all data combinations cannot identify a single parameter, while only about 18% of the data combinations can identify all the four parameters (two parameters each under different model setups) used in the model. This directly correlates with the degree of identifiability of the various parameters of v_1, shown earlier in Figure 3.

In contrast, given the high degree of identifiability of both parameters for v_2, most data combinations (¿90%) are not wasted, and can identify both parameters (Figure 6b). Furthermore, the data utility plots for v_3 (Figure 6c) show that about 36% of the 5000 data combinations can identify only two parameters while 6% of the experiment combinations cannot identify any data set. Thus, Figure 6 provides evidence to the idea that careful experimental design is necessary to minimize usage of resources devoted to performing experiments for parameter estimation. Choosing any of the experiment combinations for the 78% percent that does not identify any parameter in v_1, or the 6% that does not identify any parameters in v_3 can lead to waste of resources used to perform these non-informative experiments.

4 Discussions:

Methods have been extensively developed for the establishing structurally identifiability of parameters in nonlinear biological model based on differential algebra techniques (Ljung AND Glad 1994). However, very few methods exist for testing practical identifiability of parameters in nonlinear models and subsequently design experiments to estimate these parameters. One of the examples of a method for practical identifiability is the profile likelihood approach (Raue, ET AL. 2009). However, methods that can overcome some of the difficulties associated with implementing this method for models of large metabolic network are sorely

missing. Additionally, any test for identifiability is forgone in favor of direct parameter estimation using a nonlinear least squares optimization approach with powerful computers to overcome difficulties associated with getting global optimal solutions (). Deterministic parameter estimation techniques, based on optimization methods are paving way for probabilistic approaches based on Bayesian techniques (Lars Nielsen papers). The common thread that connects all these parameter estimation techniques is their failure to address the fundamental need for practically parameter identifiability before actual parameter estimation.

Parameter estimation for kinetic models has always focused on the ability to estimate parameters from existing data without the need for additional experiments, which might not be always possible if parameters are not identifiable from existing experimental data. Hence, these methods fail to address the issue of designing experiments to suit parameter estimation, and the need for ways to determine informative experiments remains.

In this paper we have developed a methodology to test the practical identifiability of parameters in kinetic models of metabolism. Our methodology enables practical parameter identification for each individual flux within a metabolic network under the assumption that all steady state substrate, product and allosteric effector concentrations and fluxes for a given reaction can be measured. Using a small metabolic network for gluconeogenesis, we have demonstrated that the identifiability of parameters for a given flux is dependent on the position of the flux within the metabolic network. We have also shown the ability to use our analysis to design the minimal number of experiments that are most informative for identifying all fluxes within a metabolic network.

Our work adds to this existing body of work wherein we develop a method for practical identifiability tailored for use with nonlinear enzyme kinetic rate laws that are typically used to model fluxes in metabolic networks. With our work we hope to change the status quo in the application of systems identification techniques for kinetic models of metabolic networks. Our methodology fills the niche gap of experimental design for parameter estimation by providing a way to design informative experiments to obtain data required for parameter estimation by spending the least amount of resources.

Based on the results we obtained in establishing the identifiability of parameters modeling all the fluxes of a small metabolic network, in the future we hope to extend this methodology to a larger metabolic network

with more reactions.

Appendix

In silico perturbation experiments for gluconeogenic model:

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

Experiment Type	Perturbed Parameter	Perturbed Values				
acetate	ac	0.1				
acetate perturbation	ac	0.5	1.0			
v_1 perturbation	k_1^{cat}	1.1	1.5	2.0	0.9	0.5
v_3 perturbation	V_3^{max}	1.1	1.5	2.0	0.9	0.5
v_2 perturbation	V_2^{max}	1.1	1.5	2.0	0.9	0.5

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.

Chassagnole, C., ET AL. (2002) Dynamic modeling of the central carbon metabolism of Escherichia coli, *Biotechnol. Bioeng.* 79.1, 53–73.

Costa, R. S., A. Hartmann, AND S. Vinga (2016) Kinetic modeling of cell metabolism for microbial production, *J. Biotechnol.* 219, 126–141.

Heerden, J. H. van, ET AL. (2014) Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* 343.6174, 1245114.

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.

Liebermeister, W. AND E. Klipp (2006) Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.* 3, 41.

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.

Peskov, K., E. Mogilevskaya, AND O. Demin (2012) Kinetic modelling of central carbon metabolism in Escherichia coli, *FEBS J.* 279.18, 3374–3385.

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.