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 (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. Addition-

ally, 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. 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 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 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 $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$

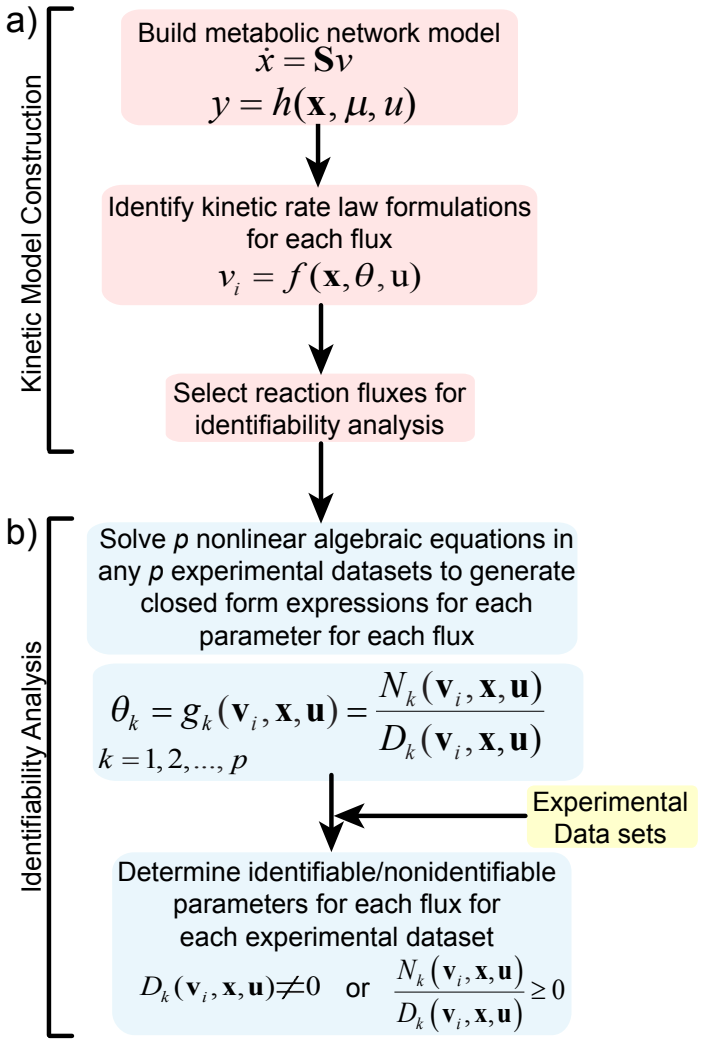

Figure 1: A flow diagram showing the methodology developed to establish practical identifiability of parameters in kinetic models of metabolism. a) The steps for the construction of a kinetic model of a metabolic network are shown. The choice of rate law formulations to describe metabolic fluxes influences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

(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.

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

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

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

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 $f\tilde{dp}$ 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).

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

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.

Looking at the three fluxes whose parameters have to be identified, v_3 has the maximum number of

parameters at three. Hence, we choose to select multiple distinct data sets, each with three different experiments to establish the practical identifiability of all the parameters in the model. Accordingly, we can select 4896 (18 x 17 x 16) distinct combinations of three experiments from the 18 different experiments mentioned earlier and elaborated in the Appendix.

2.5 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. Consequently, the set of experiments can be used to estimate the identifiable parameters in the model.

3 Results:

First in Section 3.1, we discuss the ability to apply our methods to different nonlinear kinetic rate law formulations within the context of the gluconeogenic model.

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

As demonstrated for identifying parameters of v_1 earlier in Section 2.3, the computational effort required to establish identifiability is governed by the nonlinear complexity of the enzyme kinetic rate law used to model a specific flux. Although computer algebra systems 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 computer algebra system to overcome. We encounter this scenario when expression for the the allosterically activated reaction v_3 is described by the MWC model for allosteric

regulation. We find that the complexity of the MWC kinetic rate law precludes its use to determine closed form expressions for the parameters in v_3 using either Mathematica or 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 (reference for convenience kinetics). 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.

We believe that this is not a 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, other yeast and e. coli kinetic modeling papers). However, the identifiability of the parameters used in these models has never been truly examined. 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 (). This is however paving way for probabilistic approaches based on Bayesian techniques (Lars Nielsen papers). Furthermore none of these methods address the issue of designing experiments to suit 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.

We believe this problem of computational tractability will occur in other complex kinetic rate law formulations as well. Example(s)?

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

We use the percentage of data combinations that can identify a parameter as a simple measure of the degree of identifiability of 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%, and θ_i has a higher degree of identifiability than θ_j. We can use this criteria to distinguish 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.

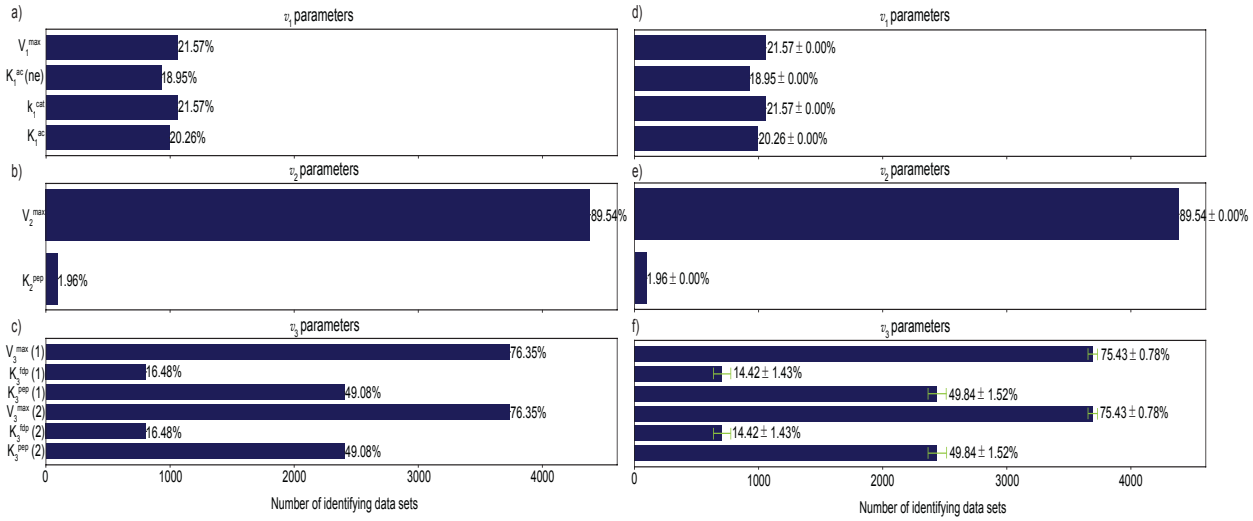

Figure 2: The number of data sets, made from a combination of 18 different in silico experiments that can practically identify each parameter in fluxes a) v_1, b) v_2 and c) v_3 when there is no noise in the input experimental data. In the presence of additive noise in the flux and concentration data used for identifiability, the number of data sets practically identifying parameters in fluxes d) v_1, e) v_2 and f) v_3 are shown. 25 samples of additive noise ϵ $N(0, 0.05)$ were used to generate noisy data for figures in the right hand side panel. The average and standard deviation values shown for identifiability of parameters with data in the presence of noise were calculated on the basis of the number of data combinations that can identify the respective parameter for each of the 25 noisy data samples.

In Figure 2 we show the number and percentage of combinations that are capable of identifying each parameter in each flux. In the left panels (Figure 2a-c), the data is shown when noise-free data is used for identification. The identifiability information for parameters using data with additive noise is shown on the right panel (Figure 2d-f). The data shown in the right hand side panel was generated using 25 different samples of random additive noise selected from a normal distribution with 0 mean 0.05 standard deviation. The value used for standard deviation reflects a 5% noise level. The average and standard deviation values shown for identifiability of parameters with data in the presence of noise were calculated on the basis of the number of data combinations that can identify the respective parameter for each of the 25 noisy data samples. As seen in these figures, irrespective of the presence or absence of noise in the data, the number of combinations that can identify each parameter, and concordantly, the degree of identifiability of each parameter varies widely. Next, we look at some of the trends in the degree of identifiability of each parameter in each flux.

Since both fluxes v_1 and v_2 are represented by the same enzyme kinetic rate law, we compare the number of data combinations that are capable of identifying the different parameters used to model these fluxes. The number of data combinations identifying the maximum reaction rates (V_i^{max}) for both v_1 (Figure 2a and d) and v_2 (Figure 2b and e) is always greater than the number of data combinations that identify the enzyme binding (K_i) constants. This trend is observed for data combinations with and without noise. Even for parameters in v_3 the percentage of data combinations identifying $V_3^{max}(1)$ and $V_3^{max}(2)$ are greater than the percentage of combinations that can identify $K_3^{fdp}(1)$ and $K_3^{fdp}(2)$ (Figure 2c and f). Hence, all maximum reaction rate parameters in the network have a higher degree of identifiability that their corresponding enzyme binding constants.

The difference between the degrees of identifiability of each parameter within a single flux varies widely depending on the reaction flux. For instance, the difference between the percentage of combinations that can identify the parameters used to model v_1 is less than 2% (Figure 2a and d). In contrast, a high number of combinations (89.5%) can identify V_2^{max} in v_2, while less than 2% of the combinations are able to identify K_2^{pep} (Figure 2b and e). This variability in the degree of identifiability for various parameters modeling the same flux is also seen for v_3 (Figure 2c and f): $V_3^{max}(1)$ and $V_3^{max}(2)$ have an average degree of 75% in

comparison to the 15% identifiability for both $K_3^{fdp}(1)$ and $K_3^{fdp}(2)$. Again, this observation is common for both data with and without added noise.

Also notable is the difference in the degree of identifiability between the maximum reaction rates of the uptake flux v_1 and the intracellular fluxes v_2 and v_3. We discuss this difference and its implications for experimental design in the following section.

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

Only about 20% of the 4896 different data combinations can identify either the maximum reaction rate (V_1^{max}) or the enzyme turnover rate (k_1^{cat}) in v_1 (Figure 2a and d), in contrast to the high number of data combinations capable of identifying V_2^{max} (89.5%) (Figure 2b and e), $V_3^{max}(1)$ or $V_3^{max}(2)$ (>75%) (Figure 2c and f).

This 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. The dependency can be further elucidated using the closed form solutions obtained for parameters of v_1 in Section 2.3. The identifiability expressions for both parameters in v_1 (V_1^{max} and K_1^{ac} in Equation 16) are dependent on changes in the both the uptake flux v_1 as well as the acetate concentration ac in both the experiments used for identification. If either of these values cannot be distinguished between the two experiments whose data is used for identification, then the corresponding combination of experiments can be guaranteed to result in practical non-identifiability of the parameters in v_1. Furthermore, the distinct values should satisfy the requirements for practical identifiability mentioned earlier for each parameter (Figure 1b).

As the uptake flux for the metabolic network that is modeled to be dependent on the substrate concentration, the only way meaningful changes in v_1 can be observed are through experiments involving changes either in the substrate concentration or changes in the enzyme concentration for v_1. As the change in v_1 is directly proportional to changes in acetate concentrations, one can except the conditions for practical identifiability of parameters of v_1 (Equation 16) to be satisfied with relative certainty when experimental

data gathered under different acetate concentrations is used for identification. This is further illustrated in Figure 3 where the different experiments in each combination identifying both V_1^{max} and k_1^{cat} are shown.

However, in certain cases unlike v_1, satisfaction of this conditions may not be as easy as one might except due to the nonlinearity in the relationship between different parts of the network.

We see that experiments involving changes to the substrate concentration figure prominently in the data combinations (as experiment 1) used to identify V_1^{max} or k_1^{cat}, and the corresponding enzyme binding constant K_1^{ac} (see Supplementary Figure S1). Experiments involving perturbations to other parameters appear in higher percentages as experiment 2 in the combination identifying these parameters. This can be attributed to the level of change in both v_1 as well a

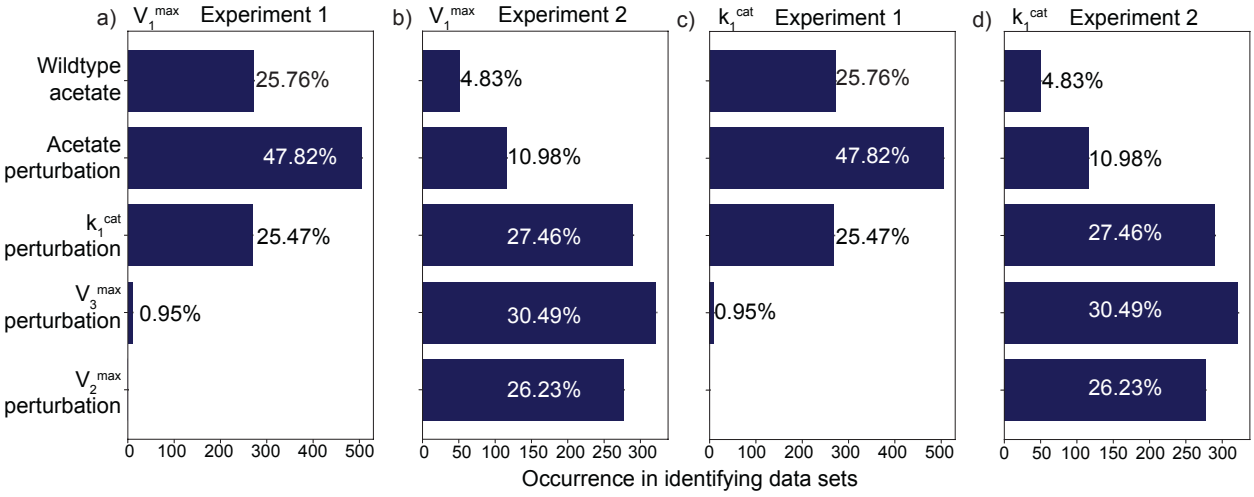

Figure 3: The contribution of different experiments types used in a combination of two experiments ($k = \{1, 2\}$) that can practically identify parameter V_1^{max} a) experiment 1 ($k = 1$), and b) experiment 2 ($k = 2$) in combination. Different experiments that contribute towards identifiability of k_1^{cat} c) experiment 1 ($k = 1$) and d) experiment 2 ($k = 2$) in combination.

In contrast, any experiment that can distinguish between different values of *pep* and v_2 can be used to identify V_2^{max}.

When the experimental data has no measurement noise associated with it, then V_2^{max} is the only parameter that can be identified by the most number of experimental combinations (89.5%). In contrast to the high identifiability of this parameter, K_2^{pep}, the other parameter for v_2 is identified by less than 2% of the

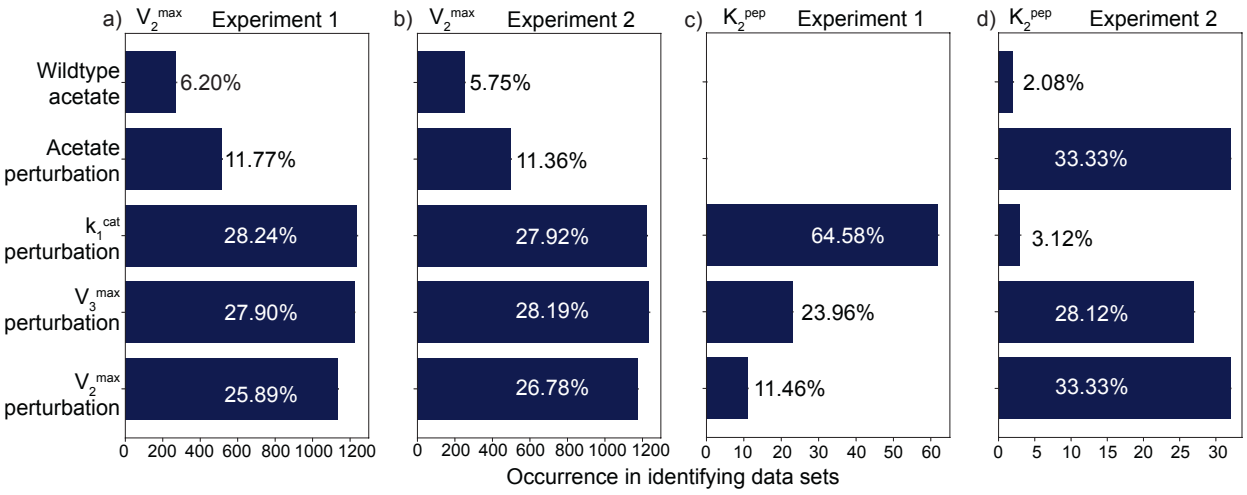

Figure 4: The contribution of different experiments types used in a combination of two experiments ($k = \{1, 2\}$) that can practically identify parameter V_2^{max} a) experiment 1 ($k = 1$), and b) experiment 2 ($k = 2$) in combination. Different experiments that contribute towards identifiability of K_2^{pep} c) experiment 1 ($k = 1$) and d) experiment 2 ($k = 2$) in combination.

available data combinations.

This is followed by V_3^{max} that is capable of being identified by 76.3% of the available data combinations.

In the small model that we use to elucidate our methodology, we find that only

3.4 Utility of data for identifiability purposes:

3.5 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.

3.6 Experimental design for parameter estimation in kinetic models of metabolism:

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. 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.

3.7 Combinations of experiments that will enable identification of all model parameters:

3.8 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

Appendix

In silico perturbation experiments for gluconeogenic 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.