1 Introduction:

The dynamic response of any system can be modeled as a set of ordinary differential equations (ode) (Equation 1a) in the state variables ($\mathbf{x}$). The changes in the state variables are expressed as a function of the state variables, the system parameters (p) and the inputs to the system (u).

$$\dot{\mathbf{x}} = g(\mathbf{x}, p, u) \tag{1a}$$

$$\mathbf{y} = h(\mathbf{x}, \mu, u) \tag{1b}$$

When not all states are observable, Equation (1a) is usually augmented with a system that defines the relationship between the observable output variables ($\mathbf{y}$) and the state variables ($\mathbf{x}$) as in Equation (1b). The parameters used to establish this relationship (Equation 1b), μ, may or may not include system parameters p defined in Equation (1a).

Typically, a nonlinear programming formulation with the objective of minimizing the least square error between the measured (y^*_{kl}) and modeled (y_{kl}) outputs over the time course ($l = 1, ..., d$) for which experimental data is collected (Equation 2) (Raue et al., and other parameter estimation/identifiability papers), is used to estimate parameters p and μ in Equation 1.

$$\chi^2(\theta) = \sum_{k=1}^{m} \sum_{l=1}^{d} \left(\frac{y^*_{kl} - y_{kl}}{\sigma^*_{kl}} \right)^2 \tag{2}$$

The difference between the data and the model estimate for each output, and at each time point, is weighted by the variance in the experimental data σ^*_{kl} for that corresponding variable and time point. The minimization problem is solved subject to constraints on the state and output variables expressed in Equation 1, as well as bounds on the estimated parameter values. The ability to determine a unique solution to the parameter estimation problem is however governed by the identifiability of the 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 1) is said to be structurally identifiable if, for an input-output mapping defined by $\mathbf{y} = \Phi(\mu, u)$ for at least one input function u, any two

values of parameters μ_1 and μ_2 satisfy the relationship in Equation (3) below.

$$\Phi(\mu_1, u) = \Phi(\mu_2, u) \iff \mu_1 = \mu_2 \tag{3}$$

Accordingly, any system that has an infinite number of solutions to the parameter estimation problem for all input functions is said to structurally non-identifiable. Thus, the structural identifiability of parameters in a dynamic model helps establish the presence or absence of a relationship between the unobservable system states and the observable system outputs. Accordingly, 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 (3) is referred to as practical identifiability. 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. On the other hand, establishing practical identifiability of parameters in any model helps design experiments that are minimal, informative and useful for parameter estimation.

The dynamics of metabolic networks can also be represented by Equation (1) wherein the metabolite concentrations, $\mathbf{x}$, are state variables. The changes in the metabolite concentrations ($\mathbf{x}$) are expressed as a function of reaction rates dependent on enzyme kinetics, which are in turn expressed as nonlinear functions of the states ($\mathbf{x}$), kinetic parameters (p) and other input variables (u).

Algorithms have been extensively developed for establishing structural identifiability of dynamic models of biological systems (IEEE Trans paper from 2007). Most of these algorithms use methods based on differential algebra (Glad and Ljung, 1994). More recent methods take a profile-likelihood-based approach (2012 Paper) to establish both structural and practical identifiability. However, these methods not only scale poorly

with increases in size of the modeled system, but also require dynamic time course data of the observable variables of the system. While computational burden due to poor scalability can be partly addressed with the current increases in computational power, the ability to obtain dynamic data for establishing identifiability of parameters in kinetic models of metabolism still remains a challenge.

In this paper, we propose a methodology to establish practical identifiability for parameters in kinetic models of metabolism. We present a computer algebra-based method that can facilitate experimental design for estimating parameters separately for each individual reaction within a metabolic network based on available steady state experimental data. This enables us to address the twin issues of scalability and data availability. We illustrate the utility of this method by applying it for a small network of gluconeogenesis in *E. coli* and demonstrating our ability to propose experiments that will facilitate parameter estimation for a kinetic model of this network. We also demonstrate the scalability of the proposed methodology to facilitate experimental design by applying it to a relatively larger metabolic network of the human red blood cell hepatocyte.

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)

1.1 Identifiability analysis: Definitions and Formulations

Any nonlinear dynamical system can be represented by a set of states $\mathbf{x}$, observables $\mathbf{y}$ that are dependent on the states, parameters μ, and inputs u as in Equation (4).

$$\dot{\mathbf{x}} = g(\mathbf{x}, \mu, u) \tag{4a}$$

$$\mathbf{y} = h(\mathbf{x}, \mu, u) \tag{4b}$$

Identifiability concerns with the ability to determine a unique solution to the problem of estimating parameters μ from given data on the system observables $\mathbf{y}$ for inputs u (McLean AND McAuley 2012). The identifiability of parameters in nonlinear models of physical processes can be classified into two categories: structural and practical identifiability.

2 Methods:

2.1 A method to establish posterior identifiability of metabolic network models:

This section details a method to establish the practical (posterior) identifiability of metabolic network models using the algebraic relationship between fluxes. Every flux, v, in a kinetic model of a metabolic network can be expressed as a nonlinear algebraic equation (Equation 5). The fluxes are expressed as a function of the metabolite concentrations x and the kinetic parameters θ in Equation (5).

$$v = f(\mathbf{x}, \theta) \tag{5}$$

Given the nonlinear nature of this model, the function f in Equation (5) can expressed, without loss of generality as,

$$v = \frac{N(\mathbf{x}, \theta)}{D(\mathbf{x}, \theta)} \tag{6}$$

where $N(\mathbf{x}, \theta)$ is the numerator of f, and $D(\mathbf{x}, \theta)$ is the denominator of f.

If $\theta \in \mathbb{R}^p$, given a set of experimental measurements for the metabolite concentrations $\mathbf{x}$ and the reaction fluxes $\mathbf{v}$, theoretically, it is possible to choose p sets of data from these measurements to solve for the p parameters in θ. However, if any these datasets do not satisfy the condition that $D(\mathbf{x}, \theta) \neq 0$, then the

number of experiments required to estimate the p parameters in θ can be established to be greater than p. An example is shown below.

This analysis can be performed for each flux in a metabolic network independent of all the other fluxes. This enables this method to be scalable to even genome-scale models. The following section demonstrates this methodology for one of the fluxes in the gluconeogenic model of Kotte et al., (**Kotte2014**).

2.2 Identifiability analysis of parameters in a kinetic model of gluconeogenesis:

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.

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

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

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

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).

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

We use flux v_2 to demonstrate the identifiability analysis method described in the previous section. Flux v_2 has two parameters, V_2^{max} and K_2^{pep} that need to be estimated from experimental data. Here, we assume that at least two different sets of experimental data for the concentrations and fluxes are available. Accordingly, we label these dataset as pep^1, ${v_2}^1$ and pep^2, ${v_2}^2$ respectively. Subsequently, these experimental datasets can be included in the model to form two simultaneous nonlinear algebraic equations in the parameters V_2^{max} and K_2^{pep} (Equation 14).

$$V_2^{max} = \frac{v_2^1 v_2^2 (pep^1 - pep^2)}{v_2^2 pep^1 - v_2^1 pep^2} \tag{14a}$$

$$K_2^{pep} = \frac{pep^1 (v_2^1 pep^2 - v_2^2 pep^2)}{v_2^2 pep^1 - v_2^1 pep^2} \tag{14b}$$

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

We use a profile likelihood-based approach (Raue, ET AL. 2009) to establish structural and practical identifiability of parameters in nonlinear kinetic models of metabolism. Briefly, the approach seeks to establish the existence/non-existence of bounds in confidence intervals for the estimates of parameters in nonlinear models. The profile likelihood is calculated based on Equation (15) for each parameter θ_i where $\chi^2(\theta_i)$ is given by Equation (**??**).

$$\chi^2_{PL}(\theta_i) = \min_{\theta_{j \neq i}} \left[\chi^2(\theta) \right] \tag{15}$$

In the minimization objective shown in Equation (**??**) for parameter estimation, y^*_{kl} is the available experimental time course data for each observable state k at each l time point. The difference between the data and the model estimates at these time points, y_{kl} is weighted by the variance in the experimental data σ^*_{kl}. An algorithm to calculate the profile likelihood, $\chi^2_{PL}(\theta_i)$, based on Equation 15 is given below.

The identifiability of parameters is established through the confidence intervals of their estimates, $\left[\sigma_i^-, \sigma_i^+\right]$. The likelihood-based confidence interval for any parameter whose profile likelihood is estimated can be written on the basis of a threshold Δ_α in the likelihood as in Equation (16).

$$\{\theta | \chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha\} \tag{16}$$

The threshold Δ_α in the likelihood is the 1-α quantile of the χ^2 distribution, represented as $\chi^2(\alpha, df)$. The confidence intervals obtained hold for df degrees of freedom. For a choice of df=1 the confidence intervals will hold for each parameter individually, and confidence intervals that hold jointly for all parameters can be obtained by choosing the number of parameters as df.

The visualization of structurally and practically non-identifiable parameters using the profile likelihood approach is illustrated in Figure 1. The points of intersection between the profile likelihood curves (solid line) with the one parameter likelihood threshold ($\Delta_\alpha = \chi^2(\alpha, 1)$, dashed line) provide the confidence intervals of the parameter θ_i. The confidence intervals of a structurally non-identifiable parameter are unbounded, i.e., $[-\infty, +\infty]$ (Figure 1a), while the confidence intervals of a practically non-identifiable parameter are unbounded in at least one direction, i.e., $\left[\sigma_i^-, \sigma_i^+\right]$ where either $\sigma_i^- = -\infty$ or $\sigma_i^+ = +\infty$ (Figure 1b). If a parameter's estimates have a finite confidence interval then the parameter is said to be identifiable (Figure 1c). Note that the horizontal dotted lines in Figure 1 represent the confidence interval thresholds (Δ_α) that are used to establish identifiability.

Due to the dependence of practical parameter identifiability on the experimental data, the profile likelihood approach can be used to design experiments in such a way that the observables that are derived from these experiments can improve the practical identifiability of the parameters. We show how experimental design can have a meaningful impact on parameter identification and estimation in Figure **??**. Assuming a parameter θ_i is practically non-identifiable (Figure **??**a), performing a profile-likelihood based identifiability analysis using simulated data can help determine the nature of experiments needed to make the parameter identifiable (Figure **??**b). In contrast, performing non-informative experiments without prior knowledge on their ability to change the identifiability of the parameter may provide data that cannot be used to estimate parameter θ_i (Figure **??**c).

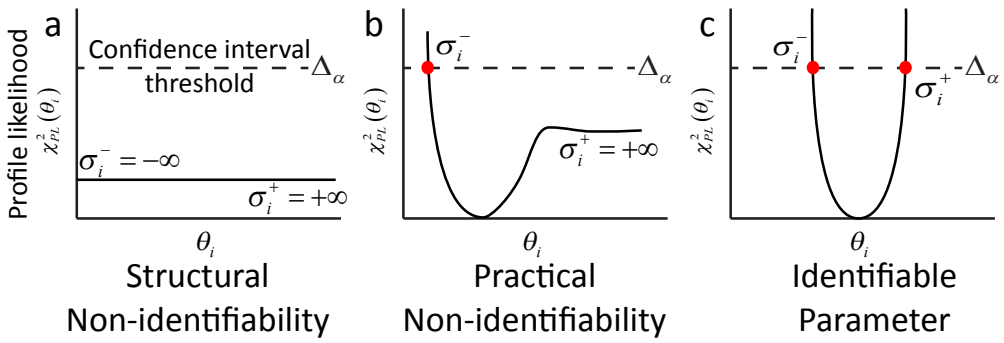

Figure 1: The profile likelihood estimates of a) a structurally non-identifiable, b) a practically non-identifiable and c) an identifiable parameter. The horizontal axis represents the changes in the value of the parameter (θ_i) whose identifiability is being determined and the profile likelihood ($\chi^2_{PL}(\theta_i)$) is shown in the vertical axis. The confidence interval threshold (Δ_α) used to determine the identifiability of the parameter is denoted by the horizontal dotted line. Identifiable parameters are distinguished from non-identifiable parameters by the presence of both upper and lower bounds on their confidence interval estimates $\left[\sigma_i^-, \sigma_i^+\right]$.

3 Results:

Outline:

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

Sections:

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

- identifiability in large rbc model

References

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.