

¹ Practical Identification and Experimental Design for Parameter Estimation
² in Kinetic Models of Metabolism

³ Shyam Srinivasan^a, William R. Cluett^a and Radhakrishnan Mahadevan^{*,a,b}

⁴ a - Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON,
⁵ Canada.

⁶ b - Institute for Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada.

⁷ * Corresponding author

⁸ **Abstract**

⁹ **1 Introduction:**

¹⁰ The use of metabolic engineering spans a wide variety of applications. Some notable examples include the
¹¹ design of microorganisms for the biosynthesis of commodity and specialty chemicals (Andreozzi, Chakrabarti,
¹² ET AL. 2016), engineering mammalian cells as therapeutic targets for cures to some ailments affecting hu-
¹³ mans (Di Filippo, ET AL. 2016; Apaolaza, ET AL. 2017), and changing the constituents of the human gut
¹⁴ microbial community to cure related diseases (Zerfaß, Chen, AND Soyer 2018). These applications require
¹⁵ us to understand the numerous complex interaction^S, their roles in cell function, and sometimes even the
¹⁶ mechanisms behind these interactions. Computational models offer a systematic way to integrate available
¹⁷ experimental data, and to study and understand these interactions through mathematical representations
¹⁸ of the biological systems in which these interactions occur (Bordbar, Monk, ET AL. 2014; Saa AND Nielsen
¹⁹ 2017). They are also used to predict changes in cell function based on changes in the type and nature of
²⁰ the modeled interactions (Andreozzi, Chakrabarti, ET AL. 2016), or aid in the identification of therapeutic
²¹ targets for drug discovery and development (Bordbar, McCloskey, ET AL. 2015; Chandrasekaran, ET AL.
²² 2017)

²³ Constraint-based models (CBMs) of metabolism are used to improve our understanding of metabolism by
²⁴ representing it as a stoichiometric network of reactions that operate under a pseudo-steady state assumption
²⁵ (Bordbar, Monk, ET AL. 2014). The ability of CBMs to shine light on the nonintuitive interactions that

26 govern cellular metabolism is leveraged to engineer and asses the impact of designs that alter the ability of
27 a cell to grow, or produce a desired metabolite (Maia, M. Rocha, AND I. Rocha 2016). However, in CBMs,
28 metabolism is assumed to operate under a pseudo steady state. Consequently, the metabolite concentrations
29 within the metabolic network are assumed to be constant, and changes in metabolite concentrations are not
30 modeled. Furthermore, since CBMs represent metabolism using only the stoichiometry of its constituent
31 reactions, they do not account for the various non-catalytic regulatory interactions that are also responsible
32 for? of metabolic function. These shortcomings prevent CBMs from being used to fully understand the steady
33 state as well as the dynamic characteristics of metabolic networks.

34 In contrast, the effects of regulatory interactions and changes in metabolite concentrations on dif-
35 ferent characteristics of metabolism can be studied using kinetic models of metabolism (Saa AND Nielsen
36 2017). These models account for changes in metabolite concentrations subject to thermodynamic and reg-
37 ulatory constraints that underly metabolic networks in addition to its stoichiometry (Link, Christodoulou,
38 AND Sauer 2014). Kinetic models can not only help us better understand lesser known and understood
39 characteristics of metabolism like bistability (Kotte, ET AL. 2014), and their role in human health, but can
40 also improve predictions about the impact of engineering design perturbations on metabolism, and propose
41 alternative designs to achieve metabolite production goals (Khodayari, ET AL. 2016).

42 Kinetic models differ from CBMs in their use of ~~heavily parameterized~~ mechanistic enzyme kinetic rate
43 laws to model enzyme catalyzed fluxes within a metabolic network. ~~These parameters represent various~~
44 ~~aspects of the enzyme kinetic rate laws~~ (Srinivasan, Cluett, AND Mahadevan 2015; Saa AND Nielsen 2017).
45 Hence, the use of kinetic models requires information on ~~the~~ the enzyme kinetic rate laws that will be used
46 to model ~~the~~ the fluxes within a metabolic network, as well as numerical values for ~~the~~ the parameters used in
47 these rate laws. Analyzing the ability of a metabolic network to exhibit dynamic characteristics like multiple
48 steady states and oscillations, irrespective of the structure of the network, is one example where kinetic rate
49 laws and parameter values ~~might~~ play a crucial role (Srinivasan, Cluett, AND Mahadevan 2017).
50 Despite ~~the~~ ^{their} importance, the parameterization of kinetic models is still a problem for which solutions are
51 a subject of debate within the modeling community. Typically, enzyme kinetic rate laws are parameterized
52 based on in vitro observations of enzyme activity, as opposed to observations made under in vivo condi-

53 tions (Heijnen 2005; Smallbone, ET AL. 2007). However, some researchers have questioned their relevance
54 for gleaning information on the dynamics of metabolism under in vivo conditions, as opposed to in vitro
55 conditions (Heijnen 2005; Heijnen AND Verheijen 2013). On the other hand, some reports have shown that
56 despite the large uncertainties associated with parameters estimated based on in vivo experimental data
57 (Link, Christodoulou, AND Sauer 2014), in vitro parameter estimates are a reasonable approximation of val-
58 ues that would be applicable under in vivo conditions (Ron Milo paper comparing in vitro vs invivo enzyme
59 turn over rates in PLoS Computational Biology).

60 Nevertheless, some authors have sought to quantify the uncertainty in in vivo parameter estimates using
61 different techniques (Vanlier, C. Tiemann, ET AL. 2013; Andreozzi, Miskovic, AND Hatzimanikatis 2016),
62 while others have proposed to alleviate as well as constrain the uncertainty in parameter estimates and
63 consequent model predictions by using a Monte Carlo approach to kinetic modeling of metabolism. These
64 approaches ~~do~~ allow for the integration of experimentally observed in vivo data. ORACLE (Wang and
65 Hatzimanikatis, 2004) and Ensemble modeling (Tan and Liao, 2008) are two examples of such an approach.
66 These and other Monte Carlo kinetic modeling methods have been previously reviewed (Srinivasan, Cluett,
67 AND Mahadevan 2015). Bayesian approaches to improve parameter estimation and quantify estimation
68 uncertainty have also been proposed (Saa AND Nielsen 2016).

IS this true about "both"?

Related to
69 In spite of the development of these methods to quantify parameter estimation uncertainty, model pa-
70 rameter identifiability, a necessary, and sometimes sufficient condition to estimate unique kinetic parameter
71 values from experimental data, is often overlooked (Ljung AND Glad 1994; Berthoumieux, ET AL. 2013).
72 Briefly, it concerns with the ability to estimate unique values for all model parameters from observed ex-
73 perimental data. In a model, ~~a~~ parameter is said to be structurally or a priori identifiable if its values
74 can be uniquely estimated independent of all other model parameters *from available experimental data*.
75 However, if parameter ~~a~~ cannot be uniquely estimated independent of ~~each other~~ due to redundant model
76 parameterization, or due to ~~a~~ nonlinear relationship between ~~the~~ model parameters, then the parameter ~~a~~
77 is ~~are~~ said to be structurally non-identifiable. Conversely, if the ability to estimate unique parameter values is
78 compromised due to the inability of the available data to capture the requisite information needed to esti-
79 mate the parameters in the modeled system, and the uncertainty in parameter estimates is unquantifiable,

the importance of i.e. the data needs to be satisfied

80 the parameter is said to be practically non-identifiable (Ljung AND Glad 1994).

81 Authors have proposed to ~~overcome concerns with~~ ^{ways} ~~parameter identifiability~~ ^{1 address} by proposing approximate
82 kinetic models of metabolism that utilize empirical enzyme kinetic rate laws ~~whose~~ ^{with} parameters ~~have no~~ ^{that have no?} phys-
83 ical significance, and are identifiable (Heijnen 2005; Smallbone, ET AL. 2007). Significant work has also
84 been done towards the development of methods for structural identification of parameters in kinetic models
85 of metabolism (Ljung AND Glad 1994; Nikerel, ET AL. 2009; Berthoumieux, ET AL. 2013; Raue, ET AL.
86 2014) (paper from Rudyanto Gunawan on model discrimination and sensitivity analysis).

87 Methods to improve practical identifiability through a priori experimental design have also been de-
88 veloped, with focus on kinetic models of metabolism (Gadkar, Gunawan, AND Doyle 2005; Vanlier, C. a.
89 Tiemann, ET AL. 2014; Raue, ET AL. 2014). Some of these methods are limited by their applicability to
90 approximate kinetic models only (Nikerel, ET AL. 2009; Berthoumieux, ET AL. 2013), while some of them
91 suffer from computational limitations when applied to kinetic models of large metabolic networks (Gadkar,
92 Gunawan, AND Doyle 2005; Raue, ET AL. 2014) (Bangga method using FIM for D-optimal design, ??).

93 In this paper, we propose a scalable methodology that uses available steady state fluxomics, metabolomics
94 and proteomics data to test the practical identifiability of parameters for each individual reaction in kinetic
95 models of metabolism. We demonstrate how the computer algebra-based method that we have developed
96 can also facilitate the design of experiments that are ~~minimal and informative~~ ^{for generating the} to generate data required

97 to estimate unique parameter values for all reaction fluxes in a metabolic network. In doing so, we ~~not~~ ^{can achieve}

98 ~~also~~ ^{as well as} propose the number and types of perturbations that will provide the most useful data for parameter

99 estimation, ~~and~~ ¹ also test the identifiability of different enzyme kinetic rate laws that are typically used

100 to model fluxes in metabolic networks. We illustrate our methodology to ~~identify parameters and design~~

101 ~~experiments to identify parameters~~ ^{using} in a small metabolic network model of gluconeogenesis in *Escherichia coli*

102 (Kotte, ET AL. 2014; Srinivasan, Cluett, AND Mahadevan 2017) under the assumption that all intracellular

103 metabolite concentrations and fluxes can be measured.

104 2 Methods

I think you should add a section here that contains section 2 because it has 6 subsections

105 2.1 Parameter estimation for kinetic models of metabolism

In kinetic models of metabolism, 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.

$$\dot{x} = \mathbf{S}v \quad (1a)$$

should state dimensions of all vectors and matrix S, D, u

$$v = f(x, \theta, u) \quad (1b)$$

What is NP?

- 106 The expression for the nonlinear function (f) used to describe each reaction flux v_i in v , $i = 1, 2, \dots, n$, in particular
- 107 a kinetic model (Equation 1b) is dependent on the enzyme kinetic mechanism that is used to model the reaction (Srinivasan, Cluett, AND Mahadevan 2015). Accordingly, f is typically
- 108 a nonlinear function of the vector of metabolite concentrations (x), the vector of enzyme kinetic parameters (θ) and other input concentrations
- 109 (u).

Parameter estimation methods based on optimization principles are typically used to determine true parameter values based on available experimental data. Under the assumption that all intracellular metabolite concentrations and fluxes can be measured, a parameter estimation problem can be formulated as a nonlinear programming problem (Equation 2) to estimate the values of enzyme kinetic parameters, θ , based on the measured data.

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

m and d need to be defined

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

- 111 Here $y = [x, v]^T$ is the vector of both concentrations (x) and fluxes (v). The minimization of least square error between the measured (y^*) and modeled (y) concentrations and fluxes, weighted by the variance in the

113 experimental data σ_{kl}^* for each concentration and flux, at each time point, is used as an objective function
114 (Equation 2a) for the optimization problem Minimize_{θ} ^{*Least squares*}. The parameter values are determined within fixed
115 upper (θ_u) and lower (θ_l) bounds (Equation 2b).

116 2.2 Structural and practical identifiability of parameters in kinetic models

117 In the Introduction, we briefly mentioned that the ability to estimate unique parameter values from available
118 experimental data is governed by the identifiability of these parameters in the model (Ljung AND Glad 1994;
119 Vanlier, C. A. Tiemann, ET AL. 2012; Berthoumieux, ET AL. 2013; Raue, ET AL. 2014). Below, we provide
120 a formal definition of structural and practical identifiability of parameters.

The parameters in θ in any nonlinear model (Equation 1) are said to be structurally identifiable if, for an input-output mapping defined by $y = [x, v]^T = \Phi(\theta, u)$ for at least one input function u , any two values of parameters θ_1 and θ_2 satisfy the relationship in Equation (3):

$$\Phi(\theta_1, u) = \Phi(\theta_2, u) \iff \theta_1 = \theta_2 \quad (3)$$

121 Accordingly, if parameters in θ have a unique value, a finite number of non-unique values or an infinite
122 number of values for all input functions, they are said to be structurally globally identifiable, locally iden-
123 tifiable or non-identifiable, respectively. So, the structural identifiability of parameters in a dynamic model
124 helps establish the presence or absence of a relationship between the unmeasured and measured concen-
125 trations/fluxes, as well as correlations between different model parameters (Rudiyanto Gunawan paper on
126 model discrimination). Consequently, the effect of model structure and parameterization on the ability to
127 infer true parameter values from experimental data is determined by the structural identifiability of the
128 parameter.

129 Experimental data from many physical systems is usually noisy, and when parameters are estimated on
130 the basis of noisy data, the ability to estimate unique parameter values to satisfy Equation (3) is referred
131 to as practical identifiability. If a single unique parameter satisfying Equation (3) can be found, then θ is
132 said to be globally practically identifiable. Whereas, if parameter estimates with quantifiable uncertainties
133 can be found, then the θ is said to be locally identifiable. The absence of unique parameter estimates for θ

134 leads to practical non-identifiability. The practical identifiability of a parameter is hence contingent upon
135 the nature, quality and quantity of data available to estimate the parameter as opposed to the structure and
136 parameterization of the model.

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

142 2.3 A method to determine practical identifiability of kinetic models of metabolism

143 We provide the mathematical framework for **identification** of parameters in kinetic models of metabolism
144 in this section. A summary of the methodology in the form of a flow diagram is shown in Figure 1. As
145 indicated in Figure 1a, the first step involves the construction of the kinetic model (Equation 1) of the
146 metabolic network with n_r reaction fluxes.

147 For each flux v_i , $i = 1, 2, \dots, n_r$, in the kinetic model, let $\theta \in \mathbb{R}^p$ in Equation (1b). If data from
148 n_E experiments is available for the chosen metabolic network, as stated earlier, for each experiment $j =$
149 $1, 2, \dots, n_E$, we assume that all metabolite concentrations (x) and reaction fluxes (v) are measurable. We
150 discuss the implications of relaxing this assumption in the results section. The pertinent information for
151 each experiment j is available as a vector of concentrations and fluxes, \mathbf{x}_j and \mathbf{v}_j , respectively (Figure 1b).

In order to establish the practical identifiability of kinetic parameters for each flux v_i , $i = 1, 2, \dots, n_r$, 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 first selecting a combination of $p \leq n_E$ experimental data. The fluxes and concentrations from p different experiments are then used to formulate a system of nonlinear algebraic equations in \mathbb{R}^p for each flux v_i , shown in Equation (4).

$$v_{i,j} = f_j(\mathbf{x}_j, \theta, \mathbf{u}_j) \quad \forall j = \{1, 2, \dots, p\} \subset \{1, 2, \dots, n_E\} \quad (4)$$

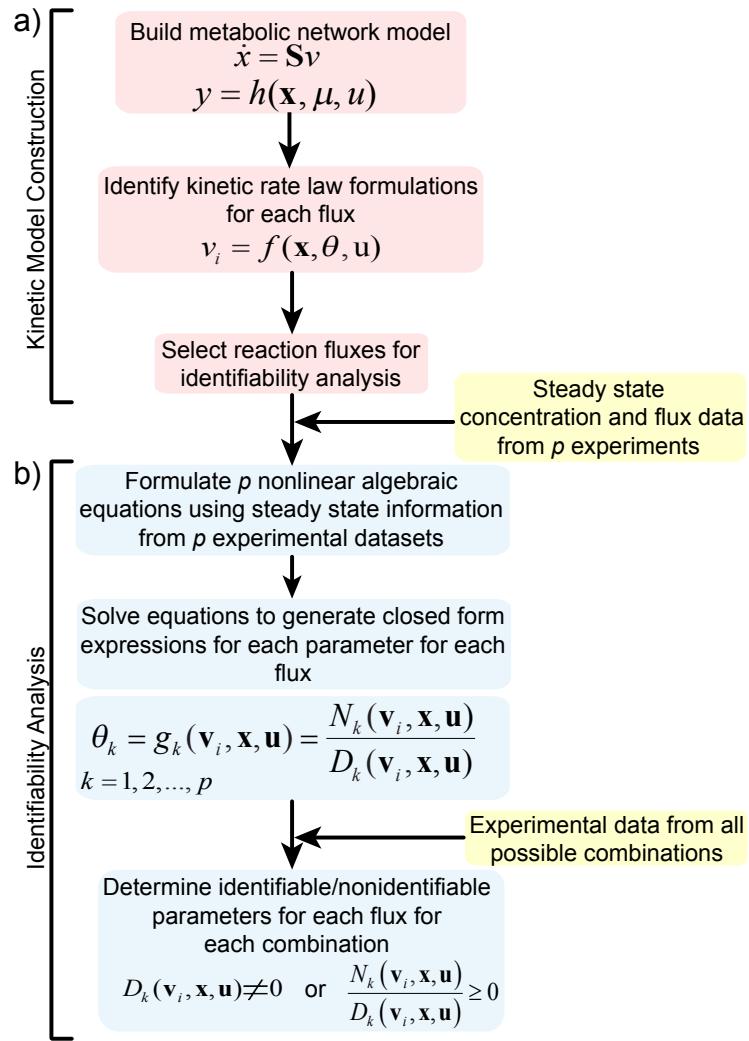


Figure 1. A flow diagram showing the methodology developed to establish practical identifiability of parameters in kinetic models of metabolism. a) The steps for the construction of a kinetic model of a metabolic network. The choice of rate law formulations to describe metabolic fluxes influences the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

152 Here, $v_{i,j}$ refers to the flux v_i obtained from experiment j . \mathbf{x}_j and \mathbf{u}_j are the vector of metabolite and other
 153 input concentrations from each experiment j , and θ is a vector in \mathbb{R}^p , whose elements are denoted by θ_k .

Each equation in (4), indicated by the index j , corresponds to the kinetic rate law expression $f(x, \theta, u)$ for each v_i , $i = 1, 2, \dots, n_r$, described in Equation (1b), written for concentrations $(\mathbf{x}_j, \mathbf{u}_j)$ and fluxes $(v_{i,j})$ obtained from experiment j . Solving the system in Equation (4) results in \mathbb{R}^p nonlinear expressions for each parameter θ_k in $\theta \in \mathbb{R}^p$ (Equation 5), 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). Note that \mathbf{v}_i , \mathbf{x} and \mathbf{u} are used to denote vector of vectors of fluxes for reaction i (\mathbf{v}_i), metabolite (\mathbf{x}) and input (\mathbf{u}) concentrations, respectively, obtained from p experiments denoted by the index $j = 1, 2, \dots, p$.

$$\theta_k = g_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) = \frac{N_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})}{D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})} \quad (5)$$

154 The identifiability of parameter θ_k , $k = 1, 2, \dots, p$, for flux v_i can be established by determining the value
 155 of $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$ (Figure 1b): any parameter θ_k is said to practically identifiable if $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) \neq 0$, and prac-
 156 tically non-identifiable if $D_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u}) = 0$. Furthermore, the physical properties of the kinetic parameters
 157 can be used to distinguish between identifiable and non-identifiable parameter values by designating only
 158 parameters with a positive value of $g_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$ as identifiable (Figure 1b). The solution $g_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$ in Equa-
 159 tion (5) is unique for an identifiable θ_k , and an infinite number of solutions are possible for a non-identifiable
 160 θ_k . However, if there are multiple but finite solutions $g_k(\mathbf{v}_i, \mathbf{x}, \mathbf{u})$, then the corresponding parameter θ_k is
 161 locally identifiable.

162 2.4 Degree of identifiability: A quantitative measure of practical identifiability

163 We express the practical identifiability of kinetic parameters using a simple quantitative term called the
 164 degree of identifiability. We describe the degree of identifiability of any single parameter as the percentage
 165 of all data combinations (used to test for practical identifiability) that can identify that parameter.

166 As an example, if 90% of all the experimental data combinations used for testing can identify a parameter
 167 θ_i , then the degree of identifiability of θ_i is said to be 0.9 or 90%. On the other hand, if only 10% of the
 168 combinations can identify another parameter θ_j , then θ_j has a degree of identifiability of 0.1 or 10%. Further-

169 more, we can create a hierarchy of practically identifiable parameters using their degrees of identifiability.
 170 In the above instance of the two parameters θ_i and θ_j that have degrees of identifiability of 90% and 10%
 171 respectively, θ_i is classified to be more identifiable than θ_j due to its relatively higher degree of identifiability.
 172 Determining this hierarchy of identifiable parameters can help in distinguishing parameters that can be
 173 identified by any type and any combination of experiments from parameters that can be identified by only
 174 a select type and combination of experiments. Such a classification can subsequently be used to design
 175 minimal sets of experiments that can practically identify all kinetic parameters used to model a metabolic
 176 network, going from the least identifiable parameter to the most identifiable parameter.

177 2.5 Kinetic model of gluconeogenesis in *E. coli*

178 A previously proposed kinetic model (Kotte, ET AL. 2014; Srinivasan, Cluett, AND Mahadevan 2017) for
 179 acetate consumption through gluconeogenesis (Figure 2) is used as a case study to illustrate identifiability
 180 analysis for experimental design for parameter estimation in kinetic models of metabolism. The kinetic
 181 model is described below.

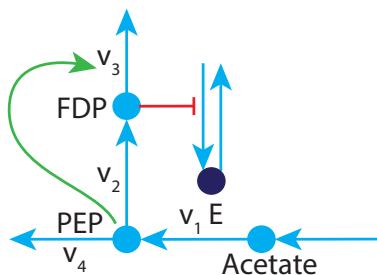


Figure 2. The previously published small metabolic network for gluconeogenesis used to demonstrate our practical identifiability method for kinetic models of metabolism.

$$\frac{d}{dt} pep = v_1 - v_2 - v_4 \quad (6)$$

$$\frac{d}{dt} fdp = v_2 - v_3 \quad (7)$$

$$\frac{d}{dt} E = v_5 - dE \quad (8)$$

184 The kinetic expressions for fluxes v_1 through v_5 are given below. The consumption of acetate through v_1 and
 185 conversion of *pep* through v_2 are expressed in Equations (9) and (11) respectively using Michaelis-Menten
 186 kinetics. The acetate flux through v_1 is also governed by the quantity of available enzyme E.

$$v_1 = k_1^{cat} E \frac{ac}{ac + K_1^{ac}} \quad (9)$$

The model for flux v_1 of the small network (Figure 2), uses the concentration of the enzyme E as a variable (Equation 9). Since we assume that steady state experimental information is only available for metabolite concentrations and fluxes, and not for enzymes (again the details on relaxing this assumption are discussed later), the expression in Equation (9) for v_1 cannot be used for identifying parameters k_1^{cat} and K_1^{ac} . So, we modify the Michaelis-Menten kinetic rate law expression to eliminate the enzyme concentration E as a variable in Equation (10). Consequently k_1^{cat} is replaced by V_1^{max} as a parameter to describe v_1 . The corresponding enzyme binding constant is denoted as $K_1^{ac}(ne)$ to distinguish it from the enzyme binding constant calculated in the presence of measured enzyme concentration data.

$$v_1 = V_1^{max} \frac{ac}{ac + K_1^{ac}(ne)} \quad (10)$$

187 We choose the expression for flux v_1 given in Equation (10) to demonstrate our method for practical identifiability.
 188

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

189

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

190 The allosterically regulated flux v_3 for the consumption of *fdp* is expressed in Equation (12) using the Monod-
 191 Wyman-Changeux (MWC) model for allosterically regulated enzymes, where \tilde{fdp} refers to the ratio of *fdp*
 192 with respect to its allosteric binding constant K_3^{fdp} .

The practically identifiability of parameters of a given flux are determined by solving a system of nonlinear algebraic equations using a computer algebra system (Section 2.3). 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 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 (Liebermeister AND Klipp 2006). The corresponding expression obtained for v_3 is given below (Equation 13).

$$v_3 = V_3^{max} \left(\frac{1}{1 + \frac{K_3^{pep}}{pep}} \right) \left(\frac{\frac{fdp}{K_3^{fdp}}}{1 + \frac{fdp}{K_3^{fdp}}} \right) \quad (13)$$

193 The flux v_4 for the export of *pep* is expressed as a linear equation dependent on *pep* in Equation (14).

$$v_4 = V_4^{max} \cdot pep \quad (14)$$

The production of enzyme E is represented by flux v_5 . The inhibition of this flux by *fdp* is modeled using Hill kinetics, where K_e^{fdp} represents the Hill binding constant for the inhibiting metabolite *fdp*, n_e is the Hill exponent, and V_e^{max} is the maximum reaction rate for v_5 .

$$v_5 = V_e^{max} \left(\frac{1}{1 + \left(\frac{fdp}{K_e^{fdp}} \right)^{n_e}} \right) \quad (15)$$

194 2.6 Experimental design through practical parameter identification

195 Not all metabolite concentrations and fluxes in the model (Equation 1) change for any random experiment.
 196 This makes unambiguous estimation of parameters impossible, either due to the inherent correlation between
 197 changes in different concentrations or fluxes, or due to the homeostasis of the concentrations and fluxes
 198 under the chosen experimental conditions (Heijnen AND Verheijen 2013). In such scenarios, the need to
 199 design experiments to effect a change in, and discriminate between changes in different concentrations/fluxes
 200 becomes necessary.

201 Following the methodology described in Section 2.3, and demonstrated in Section 3.1 for a single flux
 202 using data from a combination of two different experiments, all distinct combinations of data sets obtained
 203 from experiments described in Section S3.1 of the Supplementary Information can be tested for their ability
 204 to practically identify any of the fluxes in the small metabolic network. This step would determine the degree

205 of identifiability (defined in Section 2.4) of each parameter in each flux in the model, and help distinguish
206 experiment combinations that contribute to identifiability from combinations that do not practically identify
207 any parameter in the model (Figure 1b). In doing so, it is possible to obtain a minimal and informative
208 collection of experiments that can be performed to identify as many model parameters as possible (Figure
209 S5). Consequently, the set of experiments can be used to estimate all the identifiable parameters in the
210 model. This is formally explained below.

211 The identifiability of each parameter based on each experiment with index $j = 1, 2, \dots, n_E$ is estab-
212 lished based on the methodology described in Section 2.3 (Figure 1b), and demonstrated in Section 3.1.
213 Subsequently, for any flux v_i , and for any combination of p experimental data sets, if the experimental
214 concentrations and fluxes (\mathbf{x}_j and \mathbf{v}_j , respectively, where $j = 1, 2, \dots, p$) do not satisfy the condition for
215 identifiability for any parameter θ_k in $\theta \in \mathbb{R}^p$ (Figure 1b), then at least one of the p experiments needs to be
216 changed to make parameter θ_k identifiable. Consequently, the corresponding experiment cannot be used for
217 estimating parameter θ_k , and needs to be discarded from the set of all necessary experiments. Furthermore,
218 another experiment from $j = 1, \dots, n_E$ needs to be selected to replace the discarded experiment such that
219 parameter θ_k is identifiable. This process has to be repeated until all parameters in $\theta \in \mathbb{R}^p$ are identifiable
220 for flux v_i . In doing so, we can arrive at a set of p experiments that will always result in practically identi-
221 fiable parameters for flux v_i . Note that if none of the n_E pre-selected experiments satisfy the identifiability
222 condition, then we can design an $(n_E + 1)^{th}$ experiment that can replace one of the experiments that causes
223 practical non-identifiability. This analysis can be performed for each flux in a metabolic network independent
224 of all the other fluxes, making it theoretically scalable even to genome-scale models of metabolism.

225 3 Results

226 First, in Section 3.1, we demonstrate the use of the methodology that we described in Section 2.1 to practically
227 identify parameters in flux v_1 of the small gluconeogenic network (Figure 2) model given in Section 2.5.
228 We discuss the ability of the proposed methodology to determine the structural identifiability of parameters
229 modeling v_1 , v_3 and v_5 in Section 3.2. In Section 3.3 that follows, we show how the demonstrated methodology
230 is capable of practically identifying and estimating parameters for fluxes v_1 , v_2 , v_3 and v_5 using steady state

231 flux values and metabolite concentrations. The various ways in which this information can be used for
 232 designing experiments to generate data that can facilitate estimation of identifiable parameters are discussed
 233 in Section 3.4. The contribution of the uncertainty in the data arising from either the differences between
 234 in vivo and in vitro kinetics, or the noise present in experimentally measured quantities towards identifying
 235 parameters in enzyme kinetic models is discussed finally in Section 3.5.

236 **3.1 Identifying parameters in kinetic models of metabolism: an example**

237 In this section, we illustrate the proposed methodology step by step to identify parameters of flux v_1 in the
 238 small metabolic network (Figure 2 and Section 2.5). We choose the expression for flux v_1 given in Equation
 239 (10) for this demonstration.

Since $\theta = \{V_1^{max}, K_1^{ac}(ne)\} \in \mathbb{R}^2$ for v_1 , as mentioned in Section S3.1, we need steady state concentration
 and flux measurements from at least two different experiments. So, from the $n_E = 21$ different experiments
 described in Section S3.1 and Table S1, we can choose multiple combinations of $p = 2$ experiments to satisfy
 the data requirements for identifying v_1 i.e., in Equation (4) $j = \{1, 2\}$. We label the available concentrations
 and fluxes as $ac^{(j)}$ and $v_1^{(j)}$, respectively. Then, the nonlinear algebraic equations shown in Equation (4)
 can be formulated for v_1 as:

$$v_1^{(j)} = V_1^{max} \frac{ac^{(j)}}{ac^{(j)} + K_1^{ac}(ne)} \quad j = \{1, 2\}$$

Solving this simultaneous system of equations in \mathbb{R}^2 using Mathematica (Wolfram Research, USA), a
 computer algebra system, we get $p = 2$ nonlinear algebraic expressions for parameters V_1^{max} (Equation 16a)
 and $K_1^{ac}(ne)$ (Equation 16b). These expressions have the form shown in Equation (5).

$$\theta_1 = V_1^{max} = \frac{v_1^{(1)} v_1^{(2)} (ac^{(1)} - ac^{(2)})}{v_1^{(2)} ac^{(1)} - v_1^{(1)} ac^{(2)}} \quad (16a)$$

$$\theta_2 = K_1^{ac}(ne) = \frac{ac^{(1)} ac^{(2)} (v_1^{(1)} - v_1^{(2)})}{v_1^{(2)} ac^{(1)} - v_1^{(1)} ac^{(2)}} \quad (16b)$$

240 To test the practical identifiability of the parameters in Equation 16, we substitute any suitable in silico

241 experimental data and determine the value of the denominator of the right hand side expression. Since
 242 the enzyme binding constant ($K_1^{ac}(ne)$) and the maximum reaction rate (V_1^{max}) cannot be negative, we
 243 can further constrain the criteria for identifiability for both these parameters by saying that the evaluated
 244 expressions in Equation (16) should be positive (Figure 1b). The parameter values that are obtained for V_1^{max}
 245 and $K_1^{ac}(ne)$ by substituting in silico steady state experimental data are shown in Supplementary Figure
 246 S1. Due to the numerous possible parameter values seen in Supplementary Figure S1, we can conclude that
 247 both V_1^{max} and K_1^{ac} are practically non-identifiable.

We can also apply the proposed methodology to practically identify parameters in v_1 under the assumption
 that the protein concentration for the enzyme E is also available, in addition to the measured metabolite
 concentrations and fluxes. In doing so, we get two expressions similar to the one shown in Equation (16)
 for k_1^{cat} and K_1^{ac} . Here, the value of V_1^{max} in Equation (10) is substituted with $V_1^{max} = k_1^{cat}E$ instead. The
 corresponding identifiability expressions for k_1^{cat} and K_1^{ac} are given in Equation (17).

$$k_1^{cat} = \frac{v_1^{(1)} v_1^{(2)} (ac^{(1)} - ac^{(2)})}{v_1^{(2)} ac^{(1)} E^{(1)} - v_1^{(1)} ac^{(2)} E^{(2)}} \quad (17a)$$

$$K_1^{ac} = \frac{ac^{(1)} ac^{(2)} (v_1^{(1)} E^{(2)} - v_1^{(2)} E^{(1)})}{v_1^{(2)} ac^{(1)} E^{(1)} - v_1^{(1)} ac^{(2)} E^{(2)}} \quad (17b)$$

248 We show the parameter value for k_1^{cat} and K_1^{ac} that are obtained through the practical identifiability analysis
 249 in Figure 3a when in silico experimental data is substituted in Equation (17). Through Equation (17) and
 250 Figure 3a we are able to show that the uncertainty in the parameter estimates (Supplementary Figure
 251 S1) can be resolved through the incorporation of the available enzyme concentrations. Thus, having more
 252 experimental information can help resolve practical identifiability.

253 In the following section we present results from the identifiability analysis of fluxes v_2 , v_3 and v_5 in the
 254 small metabolic network (Figure 2), using the methodology (Figure 1) that we have demonstrated above for
 255 v_1 .

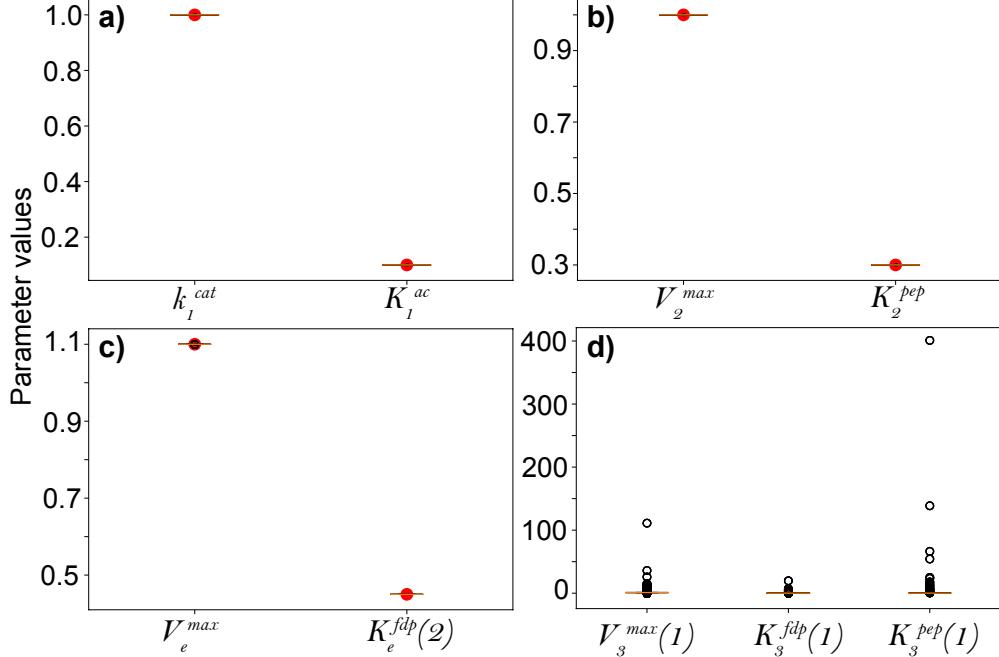


Figure 3. Distribution of predicted parameter values when performing practical identifiability analysis using closed-form solutions for each parameter in flux a) v_1 , b) v_2 , c) v_5 , and d) v_3 . For v_1 , we have assumed that enzyme concentration is available and have accordingly identified and estimated k_1^{cat} , as opposed to V_1^{max} . The parameter values for only the second root of K_e^{fdp} in v_5 ($K_e^{fdp}(2)$) is shown, since $K_e^{fdp}(1)$ is not estimated by any combination of two experiments, and V_e^{max} is estimated by all combinations. Only one of the two roots for v_3 is shown in panel d. The estimated data for the second root has a similar distribution to that of the first root and is shown in the Supplementary Information. Data is generated using the Convenience Kinetic model for allosteric regulation for v_3 .

256 **3.2 Establishing Structural identifiability of parameters based on closed-form
257 solutions**

258 For the proposed methodology (Figure 1) to work, it should be possible to obtain closed form solutions for
259 each parameter in the enzyme kinetic model for each flux as shown in Equation (5). Since the ability to
260 obtain closed-form solutions for each parameter is dependent on the model structure, any parameter that
261 has non-unique closed-form solutions can be called a structurally non-identifiable parameter. However, if
262 the number of solutions that the parameter has are finite, then the parameter is only locally structurally
263 identifiable.

264 We demonstrated the structural identifiability of parameters modeling v_1 in Section 3.1. We have shown
265 that the parameters have only one unique closed-form solution, and accordingly are structurally identifiable.
266 Since v_2 is also expressed using the Michaelis-Menten model, just like v_1 , we find that the parameters (V_2^{max}
267 and K_2^{pep}) are also structurally identifiable. The closed-form expressions for these parameters are similar to
268 the ones shown in Equation 16, with ac replaced by pep , and v_1 replaced by v_2 .

269 However, we find that the parameters used to model v_3 using the Convenience kinetics rate law, and v_5
270 using the Hill kinetic rate law are not structurally identifiable as they are given in Section 2.5.

271 First, for v_3 , we find that the parameters V_3^{max} , K_3^{fdp} and K_3^{pep} have two different closed-form solutions.
272 Thus, based on the presence of non-unique but finite number of possible solutions for these parameters we
273 can classify v_3 as a locally structurally identifiable flux. In order to alleviate local structural identifiability,
274 we reduced the dimension of the parameter space for v_3 . Originally, $\theta \in \mathbb{R}^3$ for v_3 . By reducing the dimension
275 of θ to \mathbb{R}^2 , we were able to obtain a structurally identifiable model for v_3 . To reduce the dimension of the
276 parameter space for v_3 , we fix either K_3^{fdp} or K_3^{pep} as a known quantity, and identify the other unfixed
277 parameter along with V_3^{max} . This results in unique closed-form expressions for both V_3^{max} and the other
278 unfixed parameter (K_3^{pep} or K_3^{fdp}).

279 While v_3 is an allosterically regulated metabolic flux, v_5 describes a transcription/translation reaction
280 using Hill kinetics. We apply our proposed methodology to identify parameters modeling v_5 using only the
281 available experimental data on the metabolite concentrations and the fluxes within the metabolic network.
282 We could not obtain closed form solutions for parameters V_e^{max} , K_e^{fdp} and n_e in v_5 using the computer

283 algebra system. So, instead of changing the model as we did for v_3 (see Section 2.5), we resorted to reducing
284 the dimension of the parameter space by fixing one of the three parameters, the Hill coefficient n_e . We
285 illustrated the consequence of reducing the dimension of the parameter space of the Convenience kinetic
286 model for v_3 earlier. With a fixed and known n_e , K_e^{fdp} has two possible closed-form solutions, which make
287 it a locally structurally identifiable parameter. On the other hand, V_e^{max} has only one unique closed-form
288 solution, and therefore is structurally identifiable.

289 We have now established conditions for structural identifiability of all the major fluxes in the small
290 metabolic network (Figure 2). We have shown how our proposed methodology can be used to establish
291 conditions for structural identifiability using steady state information on the model variables. We next
292 discuss the practical identifiability of the parameters in v_2 , v_3 and v_5 whose parameters are structurally
293 identifiable only under certain conditions.

294 3.3 Relationship between structural and practical parameter identifiability

295 We mention in Section 2.2 that, by definition, unique parameter values based on the model structure are
296 possible for any structurally identifiable parameter. Together with this definition for structural identifiability,
297 we also introduced the concept of practical parameter identifiability. To recall, we mentioned that it should
298 be possible to estimate unique parameter values based on all available experimental data for any practically
299 identifiable parameter.

300 As shown in Figure 1 and illustrated for v_1 in Section 3.1, to determine the practical identifiability
301 of parameters we test for the existence of a non-zero denominator of the closed-form expressions of the
302 parameters. We also reduce the possible space within which a parameter could be practically identifiable
303 by checking for the physiological feasibility of the parameter values that are obtained through this analysis
304 (Figure 1b). If the resulting parameter values obtained from various combinations of experimental data for
305 each closed-form expression are unique, then the parameter is practically identifiable. However, if a non-
306 unique number of parameter values are possible from multiple combinations of experimental steady state
307 data, then the parameter is said to be practically non-identifiable. In conjunction with the conditions for
308 structural identifiability demonstrated earlier in Section 3.2, if the parameter has only one unique closed-form

309 expression, and its value is also unique, then the parameter is both structurally and practically identifiable.
310 If either of these conditions are not satisfied, the parameters can be either locally structurally or practically
311 identifiable or non-identifiable.

312 Accordingly, both v_1 and v_2 are not only structurally identifiable due to the presence of unique closed-
313 form expressions for their parameters, they are also practically identifiable because the parameters in the
314 respective models possess unique values based on distinct combinations of experimental data (Figure 3a and
315 b).

316 Regarding v_5 , we showed earlier in Section 3.2 that the identifiability of v_5 can be analyzed only when the
317 Hill coefficient n_e is held constant. So, in subsequent discussions, the dimension of the v_5 parameter space
318 is kept at \mathbb{R}^2 by fixing the value of n_e . Under these conditions, we find that the structurally identifiable
319 parameter V_e^{max} is also practically identifiable, i.e., it has only one unique value based on all available in silico
320 experimental data (Figure 3c). However, recall that unlike V_e^{max} , K_e^{fdp} is only locally structurally identifiable
321 as it has two possible closed-form expressions. Nonetheless, despite its local structural identifiability, we find
322 that the K_e^{fdp} is also practically identifiable, like V_e^{max} , with only one unique parameter value (Figure 3c).

323 We find that the practical identifiability of v_5 , despite the local structural identifiability of one of its
324 parameters, is due to the enforcement of the physiological relevance criteria on the parameters i.e., only one
325 of the two closed-form expressions for K_e^{fdp} is physiologically relevant. The other solution always acquires
326 a negative value that has no physiological meaning. Thus, by reducing the practically identifiable space of
327 parameters, we have shown that our methodology can establish global practical identifiability even when the
328 parameters are only locally structurally identifiable.

329 Similar to K_e^{fdp} in v_5 , we also explained the local structural identifiability of V_3^{max} , K_3^{fdp} and K_3^{pep}
330 modeling v_3 in Section 3.2. These parameters have two possible closed-form expressions. In Figure 3d we
331 show the numerical values for one of the two possible closed-form expressions for V_3^{max} , K_3^{fdp} and K_3^{pep} . The
332 numerical values for the second closed-form expressions of the three parameters is presented in Supplementary
333 Figure, and they also have a similar distribution. Based on the prior definition for practically identifiable
334 parameters, the numerous possible values that the three parameters can acquire (Figure 3d) leads us to
335 conclude that the parameters in v_3 are practically non-identifiable when they are only structurally locally

336 identifiable. Also, unlike v_5 , which is practically identifiable in the presence of local structural identifiability,
 337 parameters for v_3 are practically non-identifiable even after the reduction in the practically identifiable
 338 parameter space realized using the physiological relevance condition (Figure 1b).

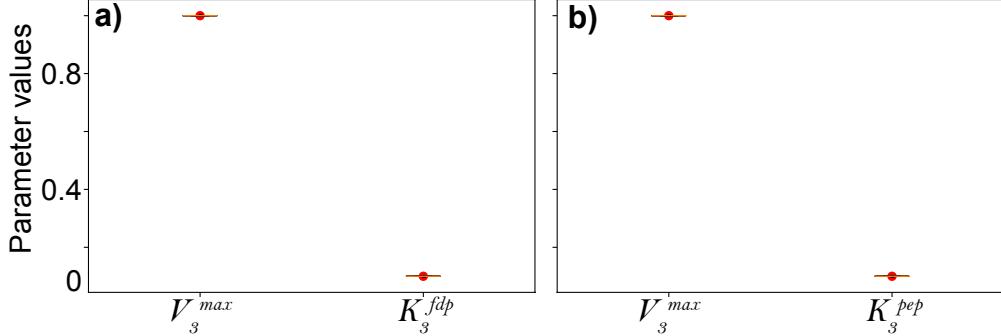


Figure 4. Distribution of predicted parameter values when performing practical identifiability analysis using closed-form solutions for each parameter in flux v_3 . The globally identifiable parameter values of a) V_3^{max} and K_3^{fdp} when K_3^{pep} is held constant, and b) V_3^{max} and K_3^{pep} when K_3^{fdp} is held constant.

339 However, we find that v_3 is practically identifiable when its parameters are also structurally identifiable
 340 (Figure 4). Earlier in Section 3.2 we had mentioned that V_3^{max} and K_3^{fdp} are structurally identifiable only
 341 when K_3^{pep} is fixed, and V_3^{max} and K_3^{pep} are structurally identifiable when K_3^{fdp} is fixed. Under these
 342 scenarios we find the structurally identifiable parameters to also be practically identifiable (Figure 4).

343 In conjunction with the practical identifiability of v_5 established earlier, we see that it is possible to
 344 delineate between structural and practical identifiability of parameters in kinetic models of metabolism only
 345 under certain conditions, and not in others.

346 3.4 A priori experimental design through practical parameter identification

347 The analysis of parameter practical identifiability can be used to gather information on the type of exper-
 348 iments that can provide useful data for parameter estimation. For instance, during practical identifiability
 349 analysis, if either the denominator of the closed-form expression is zero, or if the parameter values that are
 350 obtained are not physiologically feasible (Figure 1b), then the experimental data set concerned is said to
 351 be incapable of practically identifying that said parameter. Consequently, the data from that combination
 352 of experiments is considered non-informative. When this analysis is repeated for multiple combinations

353 of steady state data from the 21 different in silico experiments, we can determine the number of different
 354 experimental data combinations that can practically identify each parameter in each flux (Figure 5).

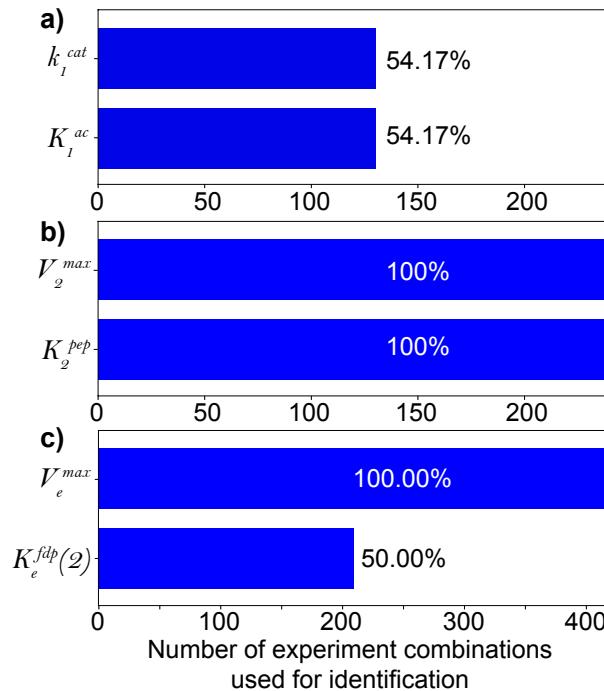


Figure 5. The number of data combination from 21 different in silico experiments that can practically identify each parameter in fluxes a) v_1 , b) v_2 , and c) v_5 when there is no noise in the input experimental data. The percentage of total combinations of experimental data used for analysis (240 for v_1 and v_2 , and 421 for v_5) that can identify each parameter is also specified. v_1 , v_2 and v_5 require data from two experiments for analysis. The contribution of different experiment type towards identifying each parameter is shown in the spider plots.

355 As described in Section 2.4, the information on the number of experimental data sets that can practically
 356 identify each parameter can be used to determine the degree of identifiability of the corresponding parameters.
 357 Subsequently, this information can be used to classify parameters based on their ease of identifiability.

358 In Figure 5 we show the number of experimental data combinations that are capable of identifying each
 359 parameter, and consequently, the degree of identifiability of each parameter (percentage experimental data
 360 combinations that are capable of identifying each parameter) in flux v_1 (Figure 5a), v_2 (Figure 5b) and v_5
 361 (Figure 5c). The degree of identifiability for V_1^{max} and $K_1^{ac}(ne)$ in v_1 is shown in Supplementary Figures

362 S2, and the degree of identifiability for both closed-form expressions of the three parameters of v_3 are shown
363 in Supplementary Figure S3. The degree of identifiability of each parameter is also given in these figures as
364 percentages.

365 It is important to recall that both v_1 (Supplementary Figure S1) and v_3 (Figure 3d) are not practically
366 identifiable (Section 3.3). While v_1 becomes practically identifiable (Figure 3a) if enzyme concentrations are
367 utilized to alleviate uncertainties in the enzyme turn over rates (k_1^{cat}), v_3 becomes practically identifiable only
368 when the dimension of the parameter space is reduced (Figure 4). In these cases the degree of identifiability
369 (Supplementary Figure S2 for v_1 and Supplementary Figure S3 for v_3) refers to the number of experimental
370 data sets that can determine physiologically relevant values for the corresponding parameters.

371 Based on their degrees of identifiability, we see that the maximum reaction rates (V_i^{max}) are more
372 or similarly identifiable in comparison to the corresponding enzyme binding (K_i) constants or the activa-
373 tion/inhibition constants, in the respective reaction rate law models (Figure 5, Supplementary Figures S2
374 and S3). We make this observation despite the fact that the four fluxes are modeled using three different
375 enzyme kinetic rate laws: v_1 and v_2 are modeled using the Michaelis-Menten rate law, v_3 is modeled using
376 the Convenience kinetic rate law and v_5 is modeled as a Hill equation with inhibition. As mentioned earlier,
377 for v_1 (Supplementary Figure S2) and v_3 (Supplementary Figure S3) we have shown that a greater number
378 of experimental data sets can predict physiologically relevant or non-zero positive values for V_1^{max} and V_3^{max}
379 than for $K_1^{ac}(ne)$ and K_3^{fdp} or K_3^{pep} , respectively.

380 The degree of identifiability of v_3 , when its parameters are structurally and practically identifiable (Sec-
381 tions 3.2 and 3.3) are shown in Supplementary Figure S4. Accordingly, we find that with the exception of
382 V_1^{max} (Supplementary Figure S2) and k_1^{cat} in v_1 (Figure 5a), all data sets used to test practical identifiability
383 can determine unique values for parameters when the corresponding parameter is structurally identifiable.

384 We can attribute the difference in the degree of identifiability between v_1 (Figure 5a and Supplemen-
385 tary Figure S2) and the other fluxes (v_2 , v_3 and v_5) to the ability of data from different combinations of
386 experiments to satisfy the conditions for practical identifiability of that parameter, that can be determined
387 a priori. In systems identification terminology, data requirements for parameter identification can be tied to
388 selecting experiments that are persistently excitable for the flux being identified. Any input signal should

389 be rich or informative enough to guarantee full excitement of the dynamics of the system (Ljung AND Glad
390 1994). Only information obtained from such changes in the input can be used to completely identify the
391 system over its entire dynamic range. So, the ability of data from a combination of different experiments
392 to practically identify parameters of a given flux is governed by the ability of the experiment to generate
393 distinct measured concentrations and fluxes that will satisfy the identifiability conditions.

394 In turn, the degree of identifiability of parameters and the informativeness of the corresponding exper-
395 iments used to identify them can be explained by the position of the flux in the metabolic network. The
396 position of any given flux in the metabolic network determines the specific experiment that is persistently
397 excitable enough to identify the parameters of that flux. This dependency can be further elucidated using
398 v_1 and v_2 as examples.

399 We know from Equation (17) and Section 3.1 that for a combination of any two experiments to be capable
400 of identifying v_1 , the experiments must generate data that have distinct acetate concentrations, E and v_1 .
401 We also know, based on our knowledge of the Michaelis-Menten kinetic rate law that changes in the substrate
402 concentration of a reaction can bring about a nonlinear change in the value of the corresponding reaction
403 rate. So, in this instance, since the substrate is an input variable to the model, and v_1 is the correspond-
404 ing uptake flux and E is a system variable, the substrate can be easily perturbed to create persistently excitable
405 experiments to identify parameters in v_1 . We see the consequence of this requirement in the degree of
406 identifiability of k_1^{cat} and K_1^{ac} (Figure 5a). We can generalize this observation for the identification of all
407 uptake fluxes in all metabolic networks, i.e., at a minimum, a change in the input substrate concentration
408 may be necessary for an informative experiment to identify the uptake flux parameters.

409 Similarly, the identification of parameters for v_2 (Figure 5b) requires that persistently excitable experi-
410 ments distinguish between values of both v_2 as well as pep . However, since both of these are system outputs,
411 satisfaction of this condition cannot be guaranteed without an analysis of the dynamics of the metabolic
412 network, and how changes in the input (acetate) bring about changes in the two requisite output quantities.
413 Previous dynamical analysis of the network (Figure 2) has already established the existence of a functional
414 relationship between pep and v_2 , and the input acetate concentration and the levels of expression of the
415 different enzymes within the network (Srinivasan, Cluett, AND Mahadevan 2017). The 100% degree of iden-

416 tifiability seen for v_2 (Figure 5b) confirms the theoretical possibility for any type of perturbation experiment
 417 to be persistently excitable to identify v_2 . Overall, this analysis informs us that the degree of identifiability
 418 and consequently, the type of experiments needed to identify different parameters varies widely depending
 419 on the position of the flux with respect to the inputs and the outputs of the metabolic network, as well as
 420 the various regulatory interactions present within the network (e.g., effect of *pep* on v_3 , or the effect of *fdp*
 421 on v_5 and consequently on v_1 in Figure 2).

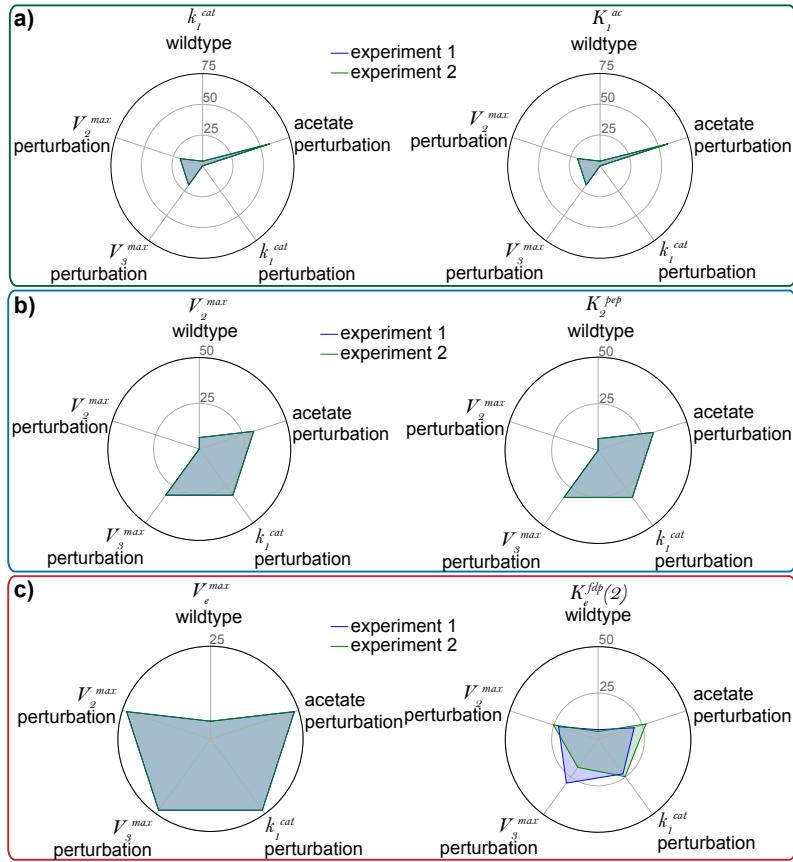


Figure 6. Frequency of each type of perturbation experiment in data sets that can identify parameters in fluxes a) v_1 , b) v_2 and c) v_5 . The frequencies are represented as a percentage of the total number of experiments present within all the data sets that can practically identify the parameters. The frequency of experiments that occur as the first experiment in a combination of two experiments are shown as a blue surface, and the frequency of experiments that occur as the second experiment in the combination are shown as a green surface.

422 From the above example we can summarize that identification of individual fluxes within a metabolic
423 network necessitates a careful consideration of experiments such that the data acquired can satisfy conditions
424 for practical identifiability for all parameters modeling a flux, and subsequently, all fluxes within a network
425 (Heijnen AND Verheijen 2013). To facilitate the design of experiments based on their ability to satisfy
426 requirements for practical identifiability of parameters, we determine the occurrence of each type of steady
427 state perturbation experiment within combinations that can practically identify each parameter (Figure 6,
428 Supplementary Figures S2, S3 and S4). So, with our proposed methodology it is possible to identify the
429 types of perturbation experiments that would be informative for identifying each parameter in each flux with
430 steady state concentration and flux data.

431 As mentioned in Supplementary Section S3.1, we use experimental data from five different types of
432 experiments to test the practical identifiability of parameters in the model (Supplementary Table S1). In
433 Figures 6a, 6b and 6c, the contribution from different experiment types for identifying parameters in v_1 , v_2
434 and v_5 are respectively shown as spider plots.

435 The contribution of experiments that involve changes in the acetate concentrations, which consequently
436 bring about changes in the value of v_1 , contribute to a significant part ($> 50\%$) of the identifiable experimental
437 data combinations for v_1 in comparison to the other types of experiments (Figure 6a and Supplementary
438 Figure S2). This is in agreement with the condition for identifiability that we discussed earlier (Equations 17
439 and 16). Since less than 50% of all data combinations can satisfy these requirements, and can consequently
440 identify v_1 (Figure 5a), we also say that identifiability analysis is crucial to determine the minimum number
441 of experiments, along with the nature of experiments that can help identify parameters for v_1 .

442 With v_2 , we see that the enzyme perturbations as well as the acetate perturbation experiments have
443 similar contributions towards datasets that can identify v_2 (Figure 6b). This also supports our arguments
444 made earlier with regards to the identifiability conditions for v_2 , and the reasons for the difference in the type
445 of experiments that are informative between v_1 and v_2 . Accordingly, we find that in comparison to selecting
446 experiments to identify v_1 , there is very little restriction on the types of experiments that are informative to
447 identify v_2 .

448 We can also extend these observations to justify the observed contribution of experiments towards identi-

449 fying parameters for v_5 (Figure 6c), or determining physiologically relevant parameter values for structurally
450 locally identifiable parameters of v_3 (Supplementary Figures S3 and S4).

451 In all of the above scenarios for v_1 (Figure 6a), v_2 (Figure 6b) and v_3 (Supplementary Figures S3 and
452 S4), the distribution of experiment types between the two (v_1 and v_2) or three (v_3) required experiments is
453 quite similar. Hence, the green/blue/yellow surfaces in the spider plots are superimposed upon each other.
454 This is also seen for experiments identifying V_e^{max} in v_5 (Figure 6c). However, this is not the case for K_e^{fdp}
455 in v_5 (Figure 6c). We see that when two experiments are required to identify K_e^{fdp} , the choice of the first
456 experiment has a bearing on the choice of the second experiment, and vice-versa, so that data with enough
457 information is available for the identification of K_e^{fdp} . Also, since V_e^{max} is globally identifiable using any
458 experiment type (Figures 5c and 6c), the choice and number of experiments required to identify v_5 completely
459 hinges upon the identifiability of K_e^{fdp} from the chosen experiments.

460 So, now we have established a hierarchy of parameters based on their identifiability. Parameters for v_2
461 and V_e^{max} in v_5 that are most identifiable are at the top of the hierarchy, while v_1 and K_e^{fdp} in v_5 fall at the
462 bottom of the list as the experiments needed for their identification require careful consideration. When v_3
463 is structurally identifiable, their parameters also do not require any experimental design considerations.

464 3.5 Parameter non-identifiability due to uncertainty in Experimental Data

465 In all the aforementioned scenarios, the kinetic rate law from which data is derived is known and same as
466 the model for which parameters are estimated. However, in reality, the kinetic rate law based on which
467 metabolic networks function and from which in vivo experimental data is extracted is mostly unknown. The
468 rate laws are primarily inferred through the parameter estimation procedure. This is one of the motivations
469 for the development of approximate kinetic rate law models (Heijnen AND Verheijen 2013; Smallbone, ET AL.
470 2007; Berthoumieux, ET AL. 2013). So, there is a need to see if the methodology that we have developed
471 here is capable of handling the uncertainty that arises due to the mismatch between the model and the data
472 used to identify and estimate the parameters in the model.

473 The scope within which we have defined the model (Section 2.5) makes such an analysis possible by
474 changing the enzyme kinetic rate law used to describe v_3 . Note that the original description (Kotte, ET AL.

475 2014; Srinivasan, Cluett, AND Mahadevan 2017) of the network (Figure 2) uses the Monod-Wyman-Chageaux
476 (MWC) model to describe the flux through v_3 . Whereas, so far we have used a Convenience kinetic rate
477 law description for both data generation as well as identifiability analysis. To determine the ability of our
478 methodology to handle the in vivo-in vitro model uncertainty, we use the MWC model description to generate
479 the in silico experimental data. This data will then be used to identify parameters in all the fluxes, including
480 v_3 that is described by the Convenience kinetic model.

481 First, we find that the spread in the estimated Convenience kinetics parameter values, when v_3 is only
482 locally structurally identifiable, is much larger than when there is mismatch between the model generating
483 the data and the model that is being identified (Supplementary Figure). A more important observation is
484 that even when the parameters are structurally identifiable in v_3 (achieved by assuming either $K_3^{f,dp}$ or K_3^{pep}
485 as a known constant), they can at most only be locally practically identifiable. This is shown by the spread
486 in the estimate values of the structurally identifiable parameters when steady state data based on the MWC
487 model is used in Supplementary Figure.

488 Second, note that the dynamics of the network as represented by an MWC model for v_3 are different
489 from the dynamic characteristics expressed when a Convenience kinetics model is used instead to describe v_3 .
490 Thus, this can bring about a change in the steady state concentrations and fluxes observed for the various
491 in silico experiments listed in Supplementary Table S1. For instance, since the enzyme concentration E is
492 dependent on the dynamics of the network, the uptake flux v_1 can be different between the two models for
493 the same acetate concentration (Equation 9). Consequently, as the enzyme concentration E is not part of
494 the closed-form expression for V_1^{max} and $K_1^{ac}(ne)$ in Equation 16), the difference in the steady state data
495 used for identification can result in a change in the spread (**uncertainty**) observed for estimated values of
496 V_1^{max} and $K_1^{ac}(ne)$ (Supplementary Figure). Thus, while quantifiable, the uncertainty due to mismatch in
497 the in vivo and in vitro information will carry over to the estimated parameters.

498 However, this issue can be resolved if more in vivo information is used for parameter identification. We
499 first observe this scenario when Equation (17), which includes E , is used to identify k_1^{cat} and K_1^{ac} : these
500 parameters are practically identifiable even when in silico steady state data from a mismatched model is
501 used for identification (Supplementary Figure). We also observe this with the identification of v_2 and v_5

502 (Supplementary Figure). For these two fluxes all available and necessary steady state information are part
503 of their identifiability expressions, thereby leaving no room for any uncertainties to propagate from the data
504 through the practical identification process.

505 Apart from the mismatch between the in vivo and the in vitro enzyme kinetic rate laws, uncertainty in
506 experimental data also arises due to the presence of noise in the measured experimental data. This noise
507 could be attributed to the measurement error commonly encountered in process analytics. In order to test the
508 robustness of our methodology to practically identify parameters using steady state data with measurement
509 errors, we used in silico experimental data with 5% additive noise for practical identification, instead of the
510 noise-free data that we have used so far.

511 We found that in every case where parameters are structurally identifiable, the noise did not have any
512 effect on the identifiability of the parameters or their estimated values. We also found that inclusion of all
513 necessary data (e.g., the presence or lack thereof of enzyme concentration E for v_1) can alleviate issues related
514 to using experimental data with errors for identification and estimation: the degrees of identifiability of V_1^{max}
515 and $K_1^{ac}(ne)$ had non-zero standard deviations associated with them, but the degrees of identifiability of k_1^{cat}
516 and K_1^{ac} did not. Using a similar reasoning to the earlier scenario in the presence of mismatches between in
517 vitro and in vivo model, we can say that there is no room for any uncertainties to propagate from the noisy
518 data when all necessary steady state information for identification is available. Thus, both v_2 and v_5 also
519 did not show any differences in either their degree of identifiability or their estimated parameter values.

520 However, for v_3 , whose parameters are only locally structurally identifiable, we found small non-zero stan-
521 dard deviations in the degrees of identifiabilities (Supplementary Figure) when noisy data is used. Although
522 this was seen due to the differences in the number of data combinations that can estimate positive values for
523 each of the three parameters between different noisy experimental data sets, we observe that the standard
524 deviation in the estimated parameter values for each data, between different samples of noisy experimental
525 data, is small (Supplementary Figure).

526 4 Discussions

527 Parameter estimation for kinetic models has always focused on the ability to estimate parameters from
528 existing data without the need for additional experiments, which might not be always possible if parameters
529 are not identifiable from existing experimental data. The presence of noise is typically said to be a significant
530 factor that results in non-identifiability. However, there different reasons for non-identifiability of parameters
531 that we show with our work. First, non-identifiability could be structural to the model used to represent the
532 flux, and cannot be alleviated without reduction in the parameter space. Otherwise, non-identifiability of
533 parameters can be attributed to the lack of information about the dynamics of the system whose parameters
534 are being estimated within the chosen experimental data. The informativeness of experiments can be tied
535 back to their ability to discriminate the dynamics of the system under two or more different input conditions.
536 Thus, the presence of noise only serves to exacerbate the inability of experiments to discriminate the dynamics
537 of the systems.

538 Previously, methods have been developed for practical parameter identification and experimental design
539 for kinetic models of metabolism. These methods for experimental design based on practical identification
540 of parameters rely on solving nonlinear least squares problems using optimization approaches that cannot
541 guarantee global optimal solutions (**Raue2009a**), or calculating the Fischer Information Matrix (FIM) to
542 obtain information on the structural and practical identifiability of parameters in kinetic models. Either
543 of these types of methods become computationally cumbersome for models of large genome-scale, or even
544 central carbon scale metabolic networks. Some authors have eschewed deterministic parameter estimation
545 techniques in favour of Bayesian methods based on probabilistic estimation of parameters and experimental
546 design (**Saa2016a**; Saa AND Nielsen 2016) that has the possibility of overcoming some of the issues with the
547 deterministic techniques.

548 In this document, we have presented a scalable method to practically identify parameters in kinetic mod-
549 els of metabolism, and use it to design experiments that are minimal and informative for estimating the
550 parameters that does not require solutions to non-convex optimization problems. By establishing identifi-
551 ability for each flux within a metabolic network individually, we hope to overcome the scalability obstacle.
552 Furthermore, we believe our method offers an algorithmic alternative to determine persistently excitable

553 experiments that can enable identification of all fluxes within a metabolic network. Using a small metabolic
554 network for gluconeogenesis, we have demonstrated that the identifiability of parameters for a given flux is
555 dependent on the position of the flux within the metabolic network. We have also shown the ability to use
556 our analysis to design the minimal number of experiments that are most informative for identifying all fluxes
557 within a metabolic network.

558 We find that the identifiability of parameters in kinetic models of metabolism using steady state in-
559 formation is dependent on the kinetic rate law used to model the fluxes within metabolism. The impact
560 of the formulation and nonlinearity of a kinetic rate law expression affecting the practical identifiability of
561 parameters in the expression may not be an unique problem isolated to the system that we are investigating.
562 Complicated expressions for describing fluxes have been extensively used to model observed experimental
563 data for different fluxes in a variety of organisms (**Chassagnole2002a; Peskov2012; VanHeerden2014**).
564 However, authors have favored working with approximate kinetic models of metabolism whose parameters
565 are easily identifiable and estimable instead of trying to establish the identifiability of the parameters used
566 in these models ([mention Heijnen papers on resolving identifiability using approximate models here](#)).

567 We have shown that in some instances (e.g., v_5) local practical identifiability could be resolved to obtain
568 global practical identifiability using constraints on the values of the parameters such that they are physically
569 relevant. We have also shown that the structural identifiability of the parameters in any given kinetic rate
570 law model has a bearing on the ability to determine the practical identifiability of parameters using steady
571 state metabolomic, fluxomic and proteomic information. We find that these can sometimes be resolved by
572 reducing the dimension of the parameter space that is being identified: $\theta \in \mathbb{R}^3$ to $\theta \in \mathbb{R}^2$ for both v_3 and
573 v_2 . Additionally, we would also like to point out that discrepancies between in vivo kinetic rate law from
574 which typical experimental data is obtained, and the in vitro rate law used in kinetic models can itself lead
575 to practical parameter non-identifiability or local identifiability. This can lead to uncertainty in parameter
576 estimates made from in vivo experimental data.

577 Our work adds to this existing body of work wherein we develop a method for practical identifiability
578 tailored for use with nonlinear enzyme kinetic rate laws that are typically used to model fluxes in metabolic
579 networks. With our work we hope to change the status quo in the application of systems identification

580 techniques for kinetic models of metabolic networks. Our methodology fills the niche gap of experimental
581 design for parameter estimation by providing a way to design informative experiments to obtain data required
582 for parameter estimation by spending the least amount of resources. In the future, we believe our work can
583 be extended and formulated as a mixed integer linear programming problem that can be solved to determine
584 the type and total minimum number of experiments necessary to estimate all parameters in kinetic models
585 of genome-scale metabolic networks.

586 References

- 587 Andreozzi, S., A. Chakrabarti, ET AL. (2016) Identification of metabolic engineering targets for the enhance-
588 ment of 1,4-butanediol production in recombinant *E. coli* using large-scale kinetic models, *Metab. Eng.*
589 35, 148–159.
- 590 Andreozzi, S., L. Miskovic, AND V. Hatzimanikatis (2016) iSCHRUNK – In Silico Approach to Characteri-
591 zation and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks, *Metab.*
592 *Eng.* 33, 158–168.
- 593 Apaolaza, I., ET AL. (2017) An in-silico approach to predict and exploit synthetic lethality in cancer
594 metabolism, *Nat. Commun.* 8.1, 459.
- 595 Berthoumieux, S., ET AL. (2013) On the identifiability of metabolic network models, *J. Math. Biol.* 67.6-7,
596 1795–1832.
- 597 Bordbar, A., D. McCloskey, ET AL. (2015) Personalized Whole-Cell Kinetic Models of Metabolism for
598 Discovery in Genomics and Pharmacodynamics, *Cell Syst.* 1.4, 283–292.
- 599 Bordbar, A., J. M. Monk, ET AL. (2014) Constraint-based models predict metabolic and associated cellular
600 functions, *Nat. Rev. Genet.* 15.2, 107–120.
- 601 Chandrasekaran, S., ET AL. (2017) Comprehensive Mapping of Pluripotent Stem Cell Metabolism Using
602 Dynamic Genome-Scale Network Modeling, *Cell Rep.* 21.10, 2965–2977.
- 603 Di Filippo, M., ET AL. (2016) Zooming-in on cancer metabolic rewiring with tissue specific constraint-based
604 models, *Comput. Biol. Chem.* 62, 60–69.

- 605 Gadkar, K. G., R. Gunawan, AND F. J. Doyle (2005) Iterative approach to model identification of biological
606 networks, *BMC Bioinformatics* 6.1, 155.
- 607 Heijnen, J. J. (2005) Approximative kinetic formats used in metabolic network modeling, *Biotechnol. Bioeng.*
608 91.5, 534–545.
- 609 Heijnen, J. J. AND P. J. T. Verheijen (2013) Parameter identification of in vivo kinetic models: Limitations
610 and challenges, *Biotechnol. J.* 8.7, 768–775.
- 611 Khodayari, A., ET AL. (2016) A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying
612 flux data for multiple mutant strains, *Nat. Commun.* 7, 13806.
- 613 Kotte, O., ET AL. (2014) Phenotypic bistability in Escherichia coli's central carbon metabolism. en, *Mol.*
614 *Syst. Biol.* 10.7, 736.
- 615 Liebermeister, W. AND E. Klipp (2006) Bringing metabolic networks to life: convenience rate law and
616 thermodynamic constraints. *Theor. Biol. Med. Model.* 3, 41.
- 617 Link, H., D. Christodoulou, AND U. Sauer (2014) Advancing metabolic models with kinetic information,
618 *Curr. Opin. Biotechnol.* 29.1, 8–14.
- 619 Ljung, L. AND T. Glad (1994) On global identifiability for arbitrary model parametrizations, *Automatica*
620 30.2, 265–276.
- 621 Maia, P., M. Rocha, AND I. Rocha (2016) In Silico Constraint-Based Strain Optimization Methods: the
622 Quest for Optimal Cell Factories. *Microbiol. Mol. Biol. Rev.* 80.1, 45–67.
- 623 Nikerel, I. E., ET AL. (2009) Model reduction and a priori kinetic parameter identifiability analysis using
624 metabolome time series for metabolic reaction networks with linlog kinetics, *Metab. Eng.* 11.1, 20–30.
- 625 Rau, A., ET AL. (2014) Comparison of approaches for parameter identifiability analysis of biological systems,
626 *Bioinformatics* 30.10, 1440–1448.
- 627 Saa, P. A. AND L. K. Nielsen (2016) Construction of feasible and accurate kinetic models of metabolism: A
628 Bayesian approach. *Sci. Rep.* 6, 29635.
- 629 Saa, P. A. AND L. K. Nielsen (2017) Formulation, construction and analysis of kinetic models of metabolism:
630 A review of modelling frameworks, *Biotechnol. Adv.* 35.8, 981–1003.

- 631 Smallbone, K., ET AL. (2007) Something from nothing - Bridging the gap between constraint-based and
632 kinetic modelling, *FEBS J.* 274.21, 5576–5585.
- 633 Srinivasan, S., W. R. Cluett, AND R. Mahadevan (2015) Constructing kinetic models of metabolism at
634 genome-scales: A review. *Biotechnol. J.* 10.9, 1345–59.
- 635 — (2017) Model-based design of bistable cell factories for metabolic engineering, *Bioinformatics*.
- 636 Vanlier, J., C. A. Tiemann, ET AL. (2012) A Bayesian approach to targeted experiment design, *Bioinfor-*
637 *matics* 28.8, 1136–1142.
- 638 Vanlier, J., C. Tiemann, ET AL. (2013) Parameter uncertainty in biochemical models described by ordinary
639 differential equations, *Math. Biosci.* 246.2, 305–314.
- 640 Vanlier, J., C. a. Tiemann, ET AL. (2014) Optimal experiment design for model selection in biochemical
641 networks Optimal experiment design for model selection in biochemical networks, 1–22.
- 642 Zerfaß, C., J. Chen, AND O. S. Soyer (2018) Engineering microbial communities using thermodynamic
643 principles and electrical interfaces, *Curr. Opin. Biotechnol.* 50, 121–127.