Optimization problem for estimation of kinetic parameters from noisy experimental data. The noisy data is generated through the use of additive noise in flux equations that use a known kinetic model to predict the responses of a metabolic network to different perturbations.

$$v^* = g(x, p) + \mathcal{N}(0, 1) \tag{1}$$

where $\mathcal{N}(0, 1)$ is Gaussian noise with zero mean and a standard deviation of 1.

$$\frac{d}{dt} pep = v_1 - v_2 - v_4 \tag{2}$$

$$\frac{d}{dt} fdp = v_2 - v_3 \tag{3}$$

$$\frac{d}{dt} E = v_{e,max} \left(\frac{1}{1 + \left(\frac{fdp}{K_e^{fdp}} \right)^{n_e}} \right) - dE \tag{4}$$

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 (5) and (6) respectively using Michaelis-Menten kinetics. The acetate flux through v_1 is also governed by the quantity of available enzyme E.

$$v_1 = k_1^{cat} E \frac{acetate}{acetate + K_1^{acetate}} \tag{5}$$

$$v_2 = V_2^{max} \frac{pep}{pep + K_2^{pep}} \tag{6}$$

$$v_3 = V_3^{max} \frac{f\tilde{dp} \left(1 + f\tilde{dp}\right)^3}{\left(1 + f\tilde{dp}\right)^4 + L_3 \left(1 + \frac{pep}{K_3^{pep}}\right)^{-4}} \tag{7}$$

The allosterically regulated flux v_3 for the consumption of *fdp* is expressed in Equation (7) 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 (8).

$$v_4 = k_4^{cat}.pep \tag{8}$$

We denote the known noisy flux and concentration information with the superscript *. Accordingly, fluxes v^* and concentrations x^* are the noisy information that will be used for estimation. Fluxes (v) and concentrations (x) that predicted by the model (without noise) are denoted without the superscript.

The optimization formulation to estimate fluxes given below is based on the minimization of the tolerance for the L2-norm of the difference between the model predicted flux and the noisy flux.

$$\min_{x,p,e} \; e \tag{9a}$$

$$\text{st } |v - v^*| \leq \varepsilon \tag{9b}$$

$$|x - x^*| \leq \varepsilon \tag{9c}$$

$$Sv = 0 \tag{9d}$$

$$v = f(x,p) + e \tag{9e}$$

$$x_{min} \leq x \leq x_{max} \tag{9f}$$

$$p_{min} \leq p \leq p_{max} \tag{9g}$$

$$e_{min} \leq e \leq e_{max} \tag{9h}$$

In Equation (9) above, ε is the absolute tolerance between the predicted and the measured quantities. The relationship between the predicted and the measured quantities are given by the nonlinear constraints in Equations (9b) and (9c). The stoichiometric constraints for steady state are given by Equation (9d), and Equation (9e) specifies the kinetic rate laws for the predicted fluxes as a function of the concentrations (x) and the parameters (p). Equations (9f) and (9g) are the bounds for the concentrations and parameters that are to be estimated by the optimization problem. The above optimization problem is to be solved for fixed values of $\varepsilon \in \{.01, .05, .1\}$.

Changes in problem formulation due to feasibility issues: Since the above problem (Equation 9) does not result in feasible global optimal solutions (using SCIP), the constraints are relaxed and the feasible problem is presented below.

$$\min_{x,p_i,e_i} \; e_i \tag{10a}$$

$$\text{st } |v_i - v^*| \leq 0.2 \tag{10b}$$

$$|x - x^*| \leq 0.3 \tag{10c}$$

$$|Sv| \leq 1\text{x}10^{-8} \tag{10d}$$

$$v_i = f(x,p_i) + e \tag{10e}$$

$$v_{j \neq i} = f(x,p_{j \neq i}) \tag{10f}$$

$$x_{min} \leq x \leq x_{max} \tag{10g}$$

$$p_{i,min} \leq p_i \leq p_{i,max} \tag{10h}$$

$$e_{i,min} \leq e_i \leq e_{i,max} \tag{10i}$$

In Equation (10) above, the index i represents the flux whose parameters are being optimized and index j represents all other fluxes whose parameters are held constant.

The steady state constraint has been relaxed (from a equality constraint to an inequality constraint) along with permitting the concentrations and the estimated fluxes to deviate significantly from their experimentally observed values (20 - 30%).

The bounds for the flux noise were set at (0,20) in the above problem.

Suggested modifications to address infeasibility (July 4) After discussion with Prof. M, the following changes were made to the problem formulation:

- Experimental fluxes may/may not be at steady state. Accordingly, using the steady state constraint in the above formulation can render the optimization problem infeasible. Suggestion is to remove this constraint(s).
- Include the norm minimization constraint as an objective.
- Add the noise in fluxes and concentrations as part of the bound constraint.

$$\min_{x,p_j} \ \|v_i - v^*\| \qquad i \in \mathcal{R}^n,\ j \in \mathcal{R}^l \tag{11a}$$

$$\text{st } v_i = f(x, p_j) \tag{11b}$$

$$x_{min}(1-\epsilon) \leq x \leq x_{max}(1+\epsilon) \tag{11c}$$

$$p_{j,min} \leq p_j \leq p_{j,max} \tag{11d}$$

- The noise ϵ is fixed at 5% or 10% to signify experimental noise.

More changes to address nonlinearity due to division(July 13) Prof. M suggested changing the nonlinear formulation of the flux equations into just a bilinear form by removing the denominator and formulating it as nonlinear algebraic equation.

$$\min_{x,p_j,v_i} \ \|v_i - v^*\| \qquad i \in \mathcal{R}^n,\ j \in \mathcal{R}^l \tag{12a}$$

$$\text{st } N(x, p_j) - v_i D(x, p_j) = 0 \tag{12b}$$

$$x_{min}(1-\epsilon) \leq x \leq x_{max}(1+\epsilon) \tag{12c}$$

$$p_{j,min} \leq p_j \leq p_{j,max} \tag{12d}$$

$$v_{i,min}(1-\epsilon) \leq v_i \leq v_{i,max}(1+\epsilon) \tag{12e}$$

- Also, the initial noise added to the steady state solutions was reduced from 50% to 5% or 10%.
- ϵ is fixed at 50%.
- Solution is obtained using multisolve with IPOPT instead of aiming to obtain a global solution with SCIP.
- Using approximately 14 million points for the second phase of multistart in OPTI

Changes in formulation This formulation assumes that the concentrations and the fluxes corresponding to each perturbation are variables in the optimization problem formulated to estimate kinetic parameters.

Let $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{p} \in \mathbb{R}^l$ and $\mathbf{v} \in \mathbb{R}$ be the vector of concentrations, parameters and fluxes respectively. If the number of perturbations used for parameter estimation is in $\mathbb{R}^p$, then the vector of concentrations, fluxes and parameters that need to be estimated in the optimization problem changes to $\mathbf{x} \in \mathbb{R}^{mp}, \mathbf{v} \in \mathbb{R}^p$ and $\mathbf{p} \in \mathbb{R}^l$ respectively.

Let $i \in \mathbb{R}^{mp}$, $j \in \mathbb{R}^p$ and $k \in \mathbb{R}^l$ represent the indices for concentrations, fluxes and parameters respectively.

$$\min_{\mathbf{x},\mathbf{p},\mathbf{v}} \ \|\mathbf{v} - \mathbf{v}^*\| \tag{13a}$$

$$\text{st } N(\mathbf{x},\mathbf{p}) - v_j D(\mathbf{x},\mathbf{p}) = 0 \qquad \forall\ j \in \mathbb{R}^p \tag{13b}$$

$$x_i^*(1-\epsilon) \le x_i \le x_i^*(1+\epsilon) \qquad \forall\ i \in \mathbb{R}^{mp} \tag{13c}$$

$$v_j^*(1-\epsilon) \le v_j \le v_j^*(1+\epsilon) \qquad \forall\ j \in \mathbb{R}^p \tag{13d}$$

$$\mathbf{p}_{min} \le \mathbf{p} \le \mathbf{p}_{max} \tag{13e}$$

Identifiability Analysis: Parameter uncertainties accompanying parameter inference and estimation in traditional deterministic models can also be reduced. Parameter uncertainties in deterministic models can be reduced by making sure that parameters in a given model are structurally (identifiability) and practically (estimability) identifiable (Nikerel et al., 2006). It has been shown that reducing the number of both structurally and practically non-identifiable parameters can result in reduction of model uncertainty propagated through uncertain parameter estimates (Hengl et al., 2007). 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. Assume a dynamic model for a model state variable vector x is represented by Equation 2.5 where u is the vector of input trajectories to the model and p is the vector of model parameters, and the observables y obtained from experimental data are represented in Equation 2.6 where $\hat{\text{I}}\text{t}_\text{ç}$ is the intrinsic noise associated with the data. If elements of p are estimated based on the noisy data y, then a parameter pi in p is structurally identifiable if and only if there is a unique solution to the system in Equations 2.5-2.6 (Hengl et al., 2007). If many solutions exist for pi, then the parameter is said to be locally identifiable. Consequently, absence of any solution would result in the parameter being termed as non-identifiable in the context of Equations 2.5 and 2.6.

Structural identifiability can also be interpreted in an alternative way based on maximum likelihood estimates (MLE) with respect to a given parameter (Raue, et al., 2009). A parameter is non-identifiable if a unique minimum, to the maximum likelihood estimate with respect to that parameter does not exist. Hence, parameters cannot be truly inferred if they are structurally

non-identifiable. Several techniques used to differentiate structurally identifiable parameters from non-identifiable parameters were reviewed by McLean & McAuley (2012). The advantages and disadvantages of using these techniques for highly nonlinear models that are frequently encountered in biology were discussed therein as well. In contrast to structural parameter identifiability, practical identifiability of a parameter helps in determining the ability and effects of the available experimental data on estimating model parameter values uniquely. 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. It can be gathered that both the model and the data play an important role in the ability to estimate parameters of a model. However, structural identifiability of any parameter in a given model depends on the fundamental biological process being modeled. Modifications to model structure to alleviate structural identifiability should not upend the purpose of the model, i.e. to represent the biological process and the underlying dynamics. Hence improvements in this regard are few and far between as more biological knowledge becomes available. Conversely, the ability to improve confidence in parameter estimates by decreasing practical non-identifiability of model parameters is a convenient solution. This can be done by improving the quality and quantity of data obtained for parameter estimation. Experimental design (ED) makes it possible to predetermine the quality and quantity of measurements obtained from experiments for the purpose of parameter estimation. The Fisher information matrix (FIM)-based ED strategies (McLean & McAuley 2012) are one of the many ED methodologies out there. Most if not all are based on senstivity-based measures of the model predicted state variable to each parameter in the model. However, biological models are highly nonlinear and are subject to high-dimensional parameterization. As a result, increase in network size for large models (genome-scale models for example) increases the computational complexity of calculating such sensitivity matrices required for ED. Techniques based on graphical analysis of the sensitivity curves are also rendered moot due to the high-dimensional parameter space (McLean & McAuley 2012). Profile likelihood (PL) based ED for parameter estimation aims to overcome this shortcoming in order to reduce practical as well as structural parameter identifiability. A complete description of PL based ED is provided in Appendix C. Although PL has been applied to small (4-12 parameters) and medium scale (15-40 parameters) networks (Raue, et al., 2009; Steiert et al., 2012), an application to a large scale network has not been accomplished. In addition, currently there is an absence of an ED-based strategy for estimating parameters for large scale transcriptional regulatory and metabolic network models.

Results: These are the results from the parameter estimation algorithm that we have been running. I have tried estimating parameters for the Kotte model by using noisy perturbation data.

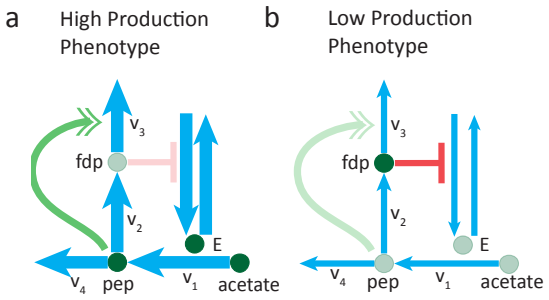

Figure 1: The two stable phenotypes for acetate consumption through the gluconeogenic model. a) The high production phenotype which has a high flux throughout the network as a result of a high *pep* concentration that reduces the inhibition on acetate uptake and b) The low production phenotype that is characterized by a relatively low *pep* concentration and consequently increased inhibition on the acetate uptake as a result of a smaller *fdp*.

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

Estimation for v_1 : The steady state data from the above perturbations in Table 1 were used to estimate the value of the parameters k_1^{cat} and $K_1^{acetate}$ for flux v_1.

The comparison of the estimated and the experimental (noisy model generated) data are shown below. Notice that the model estimates do not match well with the noisy data for perturbation P1 (Table 1). This is the observed case for all other parameter estimations. This could be due to the effect of perturbation to the acetate uptake flux (v_1) on the steady states of the system. (Note that a higher acetate uptake flux lets the system reach the higher steady state).

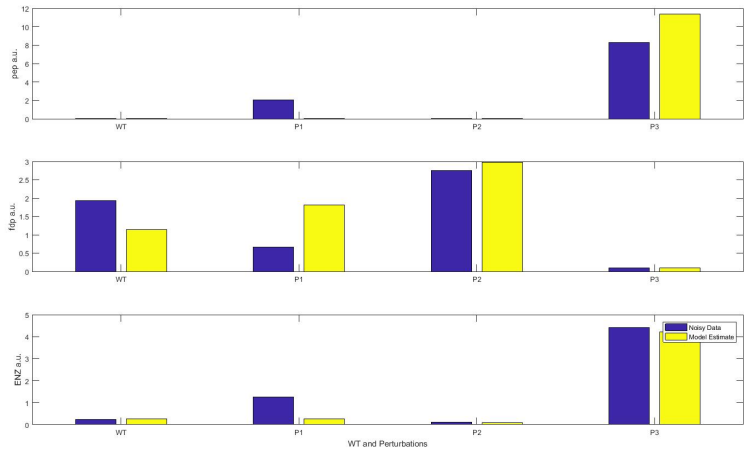

Figure 2: Plot comparing noisy model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_1. pep, fdp and enzyme E (Top to bottom)

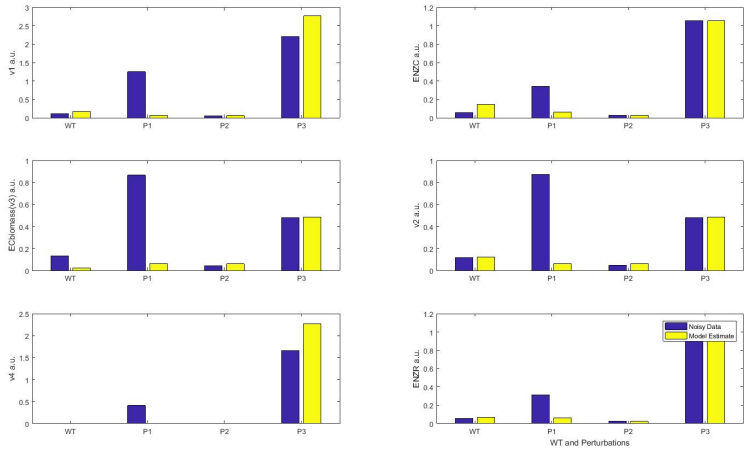

Figure 3: plot comparing noisy model generated steady state fluxes with model estimates obtained using an optimized model for v_1.

Estimation for v_2 : The steady state data from the above perturbations in Table 1 were used to estimate the value of the parameters V_2^{max} and K_2^{pep} for flux v_2.

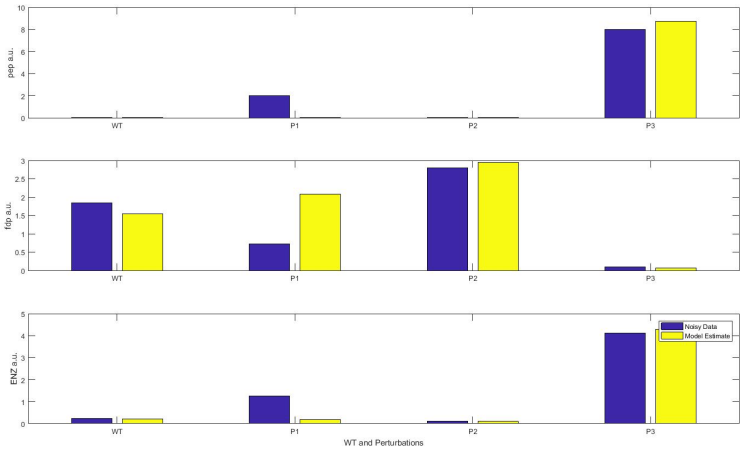

Figure 4: Plot comparing noisy model generated steady state metabolite concentrations with model estimates obtained using an optimized model for v_2. pep, fdp and enzyme E (Top to bottom)

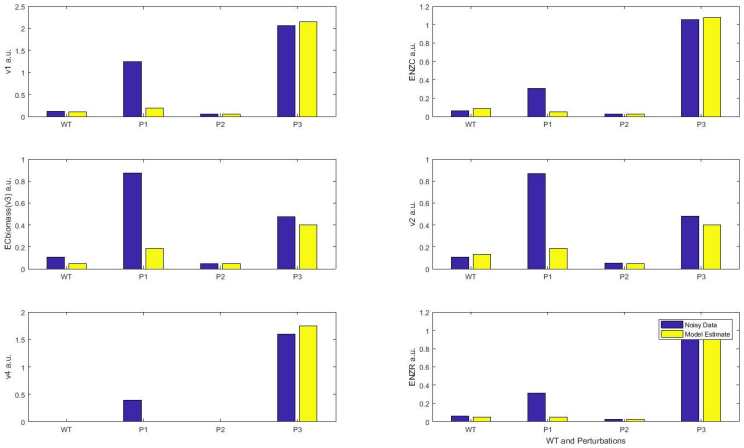

Figure 5: plot comparing noisy model generated steady state fluxes with model estimates obtained using an optimized model for v_2.

The estimation of parameter for flux v_2 also runs into the same issues mentioned above for flux v_1.

Estimation for v_3 : The steady state data from the above perturbations in Table 1 were used to estimate the value of the parameters V_3^{max}, K_3^{fdp} and K_3^{pep} for flux v_2.

The optimization cannot find an optimal solution for this problem. The optimization problem using IPOPT with the OPTI Toolbox ends with a cannot find an optimal solution error despite increasing the number of multistart points considerably.

- Noise to original steady state was 5%.
- Bounds for both fluxes and concentrations in the aforementioned formulation were fixed at 50%.