Dynamic Modeling of Cell Free Metabolic Networks using Effective Kinetic Models

Joseph A. Wayman, Adithya Sagar and Jeffrey D. Varner*

School of Chemical and Biomolecular Engineering

Cornell University, Ithaca NY 14853

Running Title: Modeling cell free metabolism

To be submitted: *Processes*

*Corresponding author:

Jeffrey D. Varner,

Associate Professor, School of Chemical and Biomolecular Engineering,

244 Olin Hall, Cornell University, Ithaca NY, 14853

Email: jdv27@cornell.edu

Phone: (607) 255 - 4258

Fax: (607) 255 - 9166

Abstract

Keywords: Cell free metabolism, Mathematical modeling

Introduction

Whole-cell bacterial processes are widely used in biotechnology to produce an array of products including high-value protein therapeutics. However, whole-cell processes share the central limitation of requiring cell growth, which redirects resources away from product synthesis, and cell walls, which complicate interrogation and control of intracellular metabolic processes. On the other hand, cell-free metabolic systems offer many advantages over traditional in vivo production methods. For example, cell-free systems can direct scarce metabolic resources exclusively towards a single product of interest. Moreover, with no cell wall, cell free systems can more easily be interrogated, and substrates of the metabolite processes directly controlled. Cell free production offers the unique opportunity to study metabolism without the complication of cell growth and gene expression processes. For modeling, this implies that we need only consider allosteric regulation of enzyme activity when building and testing cell free metabolic models. Of course, modeling allosteric mechanisms is itself a difficult problem when the model is at a whole genome scale. To address this problem, we have developed a an approach based upon the constrained fuzzy logic framework of Morris and Lauffenburger [REFHERE].

In this study, we present an effective cell free metabolic modeling framework, and test this framework using two proof of concept metabolic networks. [FINISH].

Results

Formulation and properties of cell free effective models. We developed two proof of concept metabolic networks to investigate the features of our effective cell free modeling approach (Fig. 1). In both examples, substrate S was converted to the end-products P_1 and P_2 through a series of enzymatically catalyzed reactions, including a branch point at hypothetical metabolite M_2. Several of these reactions involved cofactor dependence (AH or A), and various allosteric regulation mechanisms. Network A included feedback inhibition of the initial pathway enzyme (E_1) by pathway end products P_1 and P_2 (Fig. 1A). On the other hand, network B involved feedback inhibition of E_1 by P_2 and E_6 by P_1 (Fig. 1B). In both networks, branch point enzymes E_3 and E_6 were subject to feed-forward activation by cofactor AH. Lastly, enzyme activity was assumed to decay according to a first-order rate law in both cases. Allosteric regulation of enzyme activity was represented using a novel rule-based strategy, similar in spirit to the Constrained Fuzzy Logic (cFL) approach of Lauffenberger and coworkers [1]. In this formulation, Hill-like transfer functions were used to calculate the influence of metabolite abundance upon target enzyme activity. When an enzyme was potentially sensitive to more than one regulatory influence, logical rules were used to select which transfer function regulated enzyme activity at any given time (Fig. 2). Thus, our test networks involved important features such as cofactor recycling, enzyme activity and metabolite dynamics, as well as multiple overlapping allosteric regulatory mechanisms. As such, developing our effective modeling approach using these simple problems gave us valuable insight into the development of larger network models, without the complication of network size.

The rule based regulatory strategy approximated the behavior of classical allosteric activation and inhibition mechanisms (Fig. 3). We first explored feed-forward substrate activation of enzyme activity (for both positive and negative cooperativity). Consistent with classical data, the rule based strategy predicted a sigmoidal relationship between sub-

strate abundance and reaction rate as a function of the cooperativity parameter (Fig. 3A). For cooperativity parameters less than unity, increased substrate abundance *decreased* the reaction rate. This was consistent with the idea that substrate binding *decreases* at regulatory sites negatively impacts the ability of the enzyme to bind substrate at the active site. On the other hand, as the cooperativity parameter increased past unity, the rate of conversion of substrate S to product P by enzyme E approached a step function. In the presence of an inhibitor, the rule based strategy predicted non-competitive like behavior as a function of the cooperativity parameter (Fig. 3B). When the control gain parameter, κ_{ij} in Eqn. (10), was greater than unity, the inhibitory force was directly proportional to the cooperativity parameter, η in Eqn. (10). Thus, as the cooperativity parameter increased, the maximum reaction rate decreased (Fig. 3B, orange). However, when the gain parameter was less than unity, enzyme inhibition increased with *decreasing* cooperativity, i.e., smaller η yielded increased inhibition (Fig. 3B). Interestingly, our rule based approach was unable to directly simulate competitive inhibition of enzyme activity. For competitive inhibitors, the kinetic component of our rate, $\bar{r}_j$ in (3), could be modified to account for the inhibition (data not shown). Taken together, the rule based strategy captured classical regulatory patterns for both enzyme activation and inhibition. Thus, we are able to model complex kinetic phenomena such as ultrasensitivity, despite an effective description of reaction kinetics.

End product yield was controlled by feedback inhibition, while selectivity was controlled by branch point enzyme inhibition (Fig. 4). A critical test of our modeling approach was to simulate networks with known behavior. If we cannot reproduce the expected behavior of simple networks, then our effect modeling strategy, and particularly the rule-based approximation of allosteric regulation, will not be feasible for large scale problems. We considered two cases, control on/off, for each network configuration. Each of these cases had identical kinetic parameters and initial conditions; the *only* differences between

the cases was the allosteric regulation rules, and the control parameters associated with these rules. As expected, end product accumulation was larger for network A when the control was off (no feedback inhibition of E_1 by P_1 and P_2), as compared to the on case (Fig. 4A). We found this behavior was robust to the choice of underlying kinetic parameters, as we observed that same qualitative response across an ensemble of randomized parameter sets (N = 100). The control on/off response of network B was more subtle. In the off case, the behavior was qualitatively similar to network A. However, for the on case, flux was diverted away from P_2 formation by feedback inhibition of E_6 activity at the M_2 branch point by P_1 (Fig. 4B). Lower E_6 activity at the M_2 branch point allowed more flux toward P_1 formation, hence the yield of P_1 also increased (Fig. 4C). Again, the control on/off behavior was robust to the values of the kinetic parameters, as the same qualitative trend was conserved across an ensemble of possible randomized kinetic parameters (N = 100). Taken together, these simulations suggested that the rule based allosteric control concept could robustly capture expected feedback behavior.

Estimating parameters and effective allosteric regulatory structures. A critical challenge for any dynamic model is the estimation of kinetic parameters. For metabolic processes, there is also the added challenge of identifying the regulation and control structures that manage metabolism. Of course, these issues are not independent; any description of enzyme activity regulation will be a function of system state, which in turn depends upon the kinetic parameters. For cell free systems, regulated gene expression has been removed, however, enzyme activity regulation is still operational. We explored this linkage by estimating model parameters from synthetic data using both network structures. We generated noise-corrupted synthetic measurements of the substrate S, intermediate M_5 and end-product P_1 approximately every 20 min using network A. We then generated an ensemble of model parameter estimates by minimizing the difference between model simulations and the synthetic data using particle swarm optimization, starting from random

initial guesses. The estimation of kinetic parameters was sensitive to the choice of regulatory structure (Fig. 5). PSO identified an ensemble of parameters that bracketed the mean of the synthetic measurements in less than 1000 iterations when the control structure was correct (Fig. 5A and B). However, when there was network mismatch (network B fit against network A synthetic data), PSO unable to identify an ensemble, all else being the same (Fig. 5C and D). Interestingly, the particle swarm generated a *sloppy* parameter ensemble, in the absence of network mismatch (Fig. ZZZ). Taken together, ... [FINISH].

We modified our particle swarm identification strategy to simultaneously search over both kinetic parameters and putative control structures. In addition to our initial networks, we constructed three additional presumptive network models, each with the same enzymatic connectivity but different allosteric regulation of the initial pathway enzyme (Fig. 6). We then initialized a population of particles, each with one of the five potential regulatory programs, and randomized kinetic parameters. Thus, we generated an initial population of particles that had *both* different kinetic parameters as well as different control structures. We biased the distribution of the particle population according to our *a prior* belief of the correct regulatory program. To this end, we considered three different priors, a uniform distribution where each putative regulatory structure represented 20% of the population, and two mixed distributions that were positively or negatively biased towards the correct structure (network A). In both the positively biased, and uniform cases the particle swarm clearly differentiated between the true or closely related structures and those that were materially different (Fig. 7). As expected, the positively biased population (40% of the initial particle population seeded with network A) gave the best results, where the correct structure was preferentially identified (Fig. 7A). On the other hand, the uniform distribution identified a combination of network A (ZZ) and network C (YY) as the most likely control structures (Fig. 7B). Network A and C differ by the regulatory connection between the end-product P_2 and enzyme E_1; in network A P_2 was assumed to inhibit E_1 while in

network C P_2 was assumed to activate E_1. Lastly, when the initial population was biased towards a completely incorrect structure (initial population seeded with 40% network B), the particle swarm misidentified the correct structure (Fig. 7C). [FINISH]

Discussion

In this study, we proposed a effective modeling strategy to dynamically simulate cell free metabolic networks. We tested this strategy using two proof of concept metabolic networks. In both networks, substrate S was converted to the end-products P_1 and P_2 through a series of enzymatically catalyzed reactions, including a branch point at hypothetical metabolite M_2. While both networks had the same enzymatic connectivity, that had differing control structures. [FINISH]

Cybernetic models, other dynamic models of metabolism.

While the results of this study were encouraging, there are several critical next steps that must be accomplished before we can model genome scale cell free metabolic networks. [FINISH]

Materials and Methods

Formulation and solution of the model equations. We used ordinary differential equations (ODEs) to model the time evolution of metabolite (x_i) and scaled enzyme abundance (ϵ_i) in hypothetical cell free metabolic networks:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j \left(\mathbf{x}, \epsilon, \mathbf{k}\right) \qquad i = 1, 2, \ldots, \mathcal{M} \tag{1}$$

$$\frac{d\epsilon_i}{dt} = -\lambda_i \epsilon_i \qquad i = 1, 2, \ldots, \mathcal{E} \tag{2}$$

where $\mathcal{R}$ denotes the number of reactions, $\mathcal{M}$ denotes the number of metabolites and $\mathcal{E}$ denotes the number of enzymes in the model. The quantity $r_j\left(\mathbf{x}, \epsilon, \mathbf{k}\right)$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of metabolite and enzyme abundance, as well as unknown kinetic parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, metabolite i is produced by reaction j. Conversely, if $\sigma_{ij} > 0$, metabolite i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates metabolite i is not connected with reaction j. Lastly, λ_i denotes the scaled enzyme degradation constant. The system material balances are subject to the initial conditions $\mathbf{x}\left(t_o\right) = \mathbf{x}_o$ and $\epsilon\left(t_o\right) = \mathbf{1}$ (initially we have 100% cell-free enzyme abundance).

Each reaction rate was written as the product of two terms, a kinetic term ($\bar{r}_j$) and a regulatory term (v_j):

$$r_j\left(\mathbf{x}, \epsilon, \mathbf{k}\right) = \bar{r}_j v_j \tag{3}$$

We used multiple saturation kinetics to model the reaction term $\bar{r}_j$:

$$\bar{r}_j = k_j^{max} \epsilon_i \left(\prod_{s \in m_j^-} \frac{x_s}{K_{js} + x_s} \right) \tag{4}$$

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the scaled enzyme ac-

tivity which catalyzes reaction j, and K_{js} denotes the saturation constant for species s in reaction j. The product in Eqn. (4) was carried out over the set of *reactants* for reaction j (denoted as m_j^-). The allosteric regulation term v_j depended upon the combination of factors which influenced the activity of enzyme i. For each enzyme, we used a rule based approach to select from competing control factors (Fig. 2). If an enzyme was activated by m metabolites, we modeled this activation as:

$$v_j = \max\left(f_{1j}\left(x\right), \ldots, f_{mj}\left(x\right)\right) \tag{5}$$

Conversely, if enzyme activity was inhibited by a m metabolites, we modeling this inhibition as:

$$v_j = 1 - \max\left(f_{1j}\left(x\right), \ldots, f_{mj}\left(x\right)\right) \tag{6}$$

Lastly, if an enzyme had both m activating and n inhibitory factors, we modeled the regulatory term as:

$$v_j = \min\left(u_j, d_j\right) \tag{7}$$

where:

$$u_j = \max_{j^+}\left(f_{1j}\left(x\right), \ldots, f_{mj}\left(x\right)\right) \tag{8}$$

$$d_j = 1 - \max_{j^-}\left(f_{1j}\left(x\right), \ldots, f_{nj}\left(x\right)\right) \tag{9}$$

where j^+ and j^- denote the sets of activating, and inhibitory factors for enzyme j. If an enzyme had no allosteric factors, we set $v_j = 1$. In this study, each individual factor had the form:

$$f_i\left(\mathbf{x}\right) = \frac{\kappa_{ij}^{\eta} x_j^{\eta}}{1 + \kappa_{ij}^{\eta} x_j^{\eta}} \tag{10}$$

where x_j denotes the abundance of metabolite j, and κ_{ij} and η are control parameters.

The κ_{ij} parameter was species gain parameter, while η was a cooperativity parameter (similar to a Hill coefficient). The model equations were encoded using the Octave programming language, and solved using the LSODE routine in Octave [2].

Estimation of model parameters and structures from synthetic experimental data. Model parameters were estimated by minimizing the difference between simulations and synthetic experimental data (squared residual):

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left(\frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2 \tag{11}$$

where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ, $x_j(\tau, \mathbf{k})$ denotes the simulated value for species j at time τ, and $\omega_j(\tau)$ denotes the experimental measurement variance for species j at time τ. The outer summation is respect to time, while the inner summation is with respect to state. We approximated a realistic model identification scenario, assuming noisy experimental data, limited sampling resolution (approximately 20 minutes per sample) and a limited number of measurable metabolites.

We minimized the model residual using Particle swarm optimization (PSO) [3]. PSO uses a *swarming* metaheuristic to explore parameter spaces. A strength of PSO is its ability to find the global minimum, even in the presence of potentially many local minima, by communicating the local error landscape experienced by each particle collectively to the swarm. Thus, PSO acts both as a local and a global search algorithm. For each iteration, particles in the swarm compute their local error by evaluating the model equations using their specific parameter vector realization. From each of these local points, a globally best error is identified. Both the local and global error are then used to update the parameter

estimates of each particle using the rules:

$$\boldsymbol{\Delta}_i = \theta_1 \boldsymbol{\Delta}_i + \theta_2 \mathbf{r}_1 \left(\mathcal{L}_i - \mathbf{k}_i \right) + \theta_3 \mathbf{r}_2 \left(\mathcal{G} - \mathbf{k}_i \right) \tag{12}$$

$$\mathbf{k}_i = \mathbf{k}_i + \boldsymbol{\Delta}_i \tag{13}$$

where $(\theta_1, \theta_2, \theta_3)$ are adjustable parameters, $\mathcal{L}_i$ denotes local best solution found by particle i, and $\mathcal{G}$ denotes the best solution found over the entire population of particles. The quantities r_1 and r_2 denote uniform random vectors with the same dimension as the number of unknown model parameters ($\mathcal{K} \times 1$). In thus study, we used $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05564, 0.02886)$, which was taken from XXX. The quality of parameter estimates was measured using two criteria, goodness of fit (model residual) and angle between the estimated parameter vector $\mathbf{k}_j$ and the true parameter set $\mathbf{k}^*$:

$$\alpha_j = cos^{-1} \left(\frac{\mathbf{k}_j \cdot \mathbf{k}^*}{\|\mathbf{k}_j\| \, \|\mathbf{k}^*\|} \right) \tag{14}$$

If the candidate parameter set $\mathbf{k}_j$ were perfect, the residual between the model and synthetic data and the angle between $\mathbf{k}_j$ and the true parameter set $\mathbf{k}^*$ would be equal to zero.

We modified our PSO implementation to simultaneously search over kinetic parameters and putative model control structures. In the combined case, each particle potentially carried a different model realization in addition to a different kinetic parameter vector. We kept the update rules the same (along with the update parameters). Thus, each particle competed on the basis of goodness of fit, which allowed different model structures to contribute to the overall behavior of the swarm. We considered five possible model structures (A through E), where network A was the correct formulation (used to generate the synthetic data). We considered a population N = 100 particles, where each particle

in the swarm was assigned a model structure, and a random parameter vector. The PSO algorithm, model equations, and the objective function were encoded and solved in the Octave programming language [2].

Acknowledgements

This study was supported by the National Science Foundation GK12 award (DGE-1045513) and by the National Science Foundation CAREER award (FILLMEIN).

References

1. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. PLoS Comput Biol 7: e1001099.
2. Octave community (2014). GNU Octave 3.8.1. URL www.gnu.org/software/octave/.
3. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the International Conference on Neural Networks. pp. 1942 - 1948.

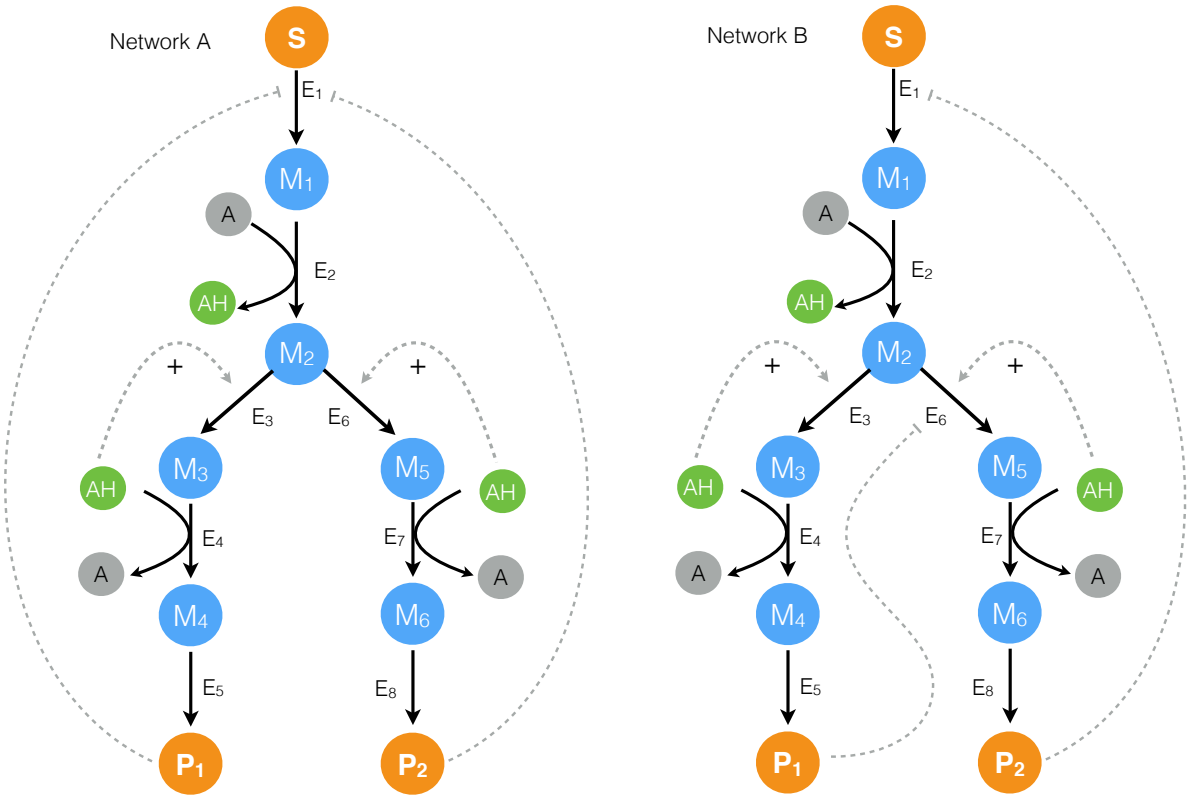

Fig. 1: Proof of concept cell-free metabolic networks considered in this study. Substrate S is converted to products P_1 and P_2 through a series of chemical conversions catalyzed by enzyme(s) E_j. The activity of the pathway enzymes is subject to both positive and negative allosteric regulation.

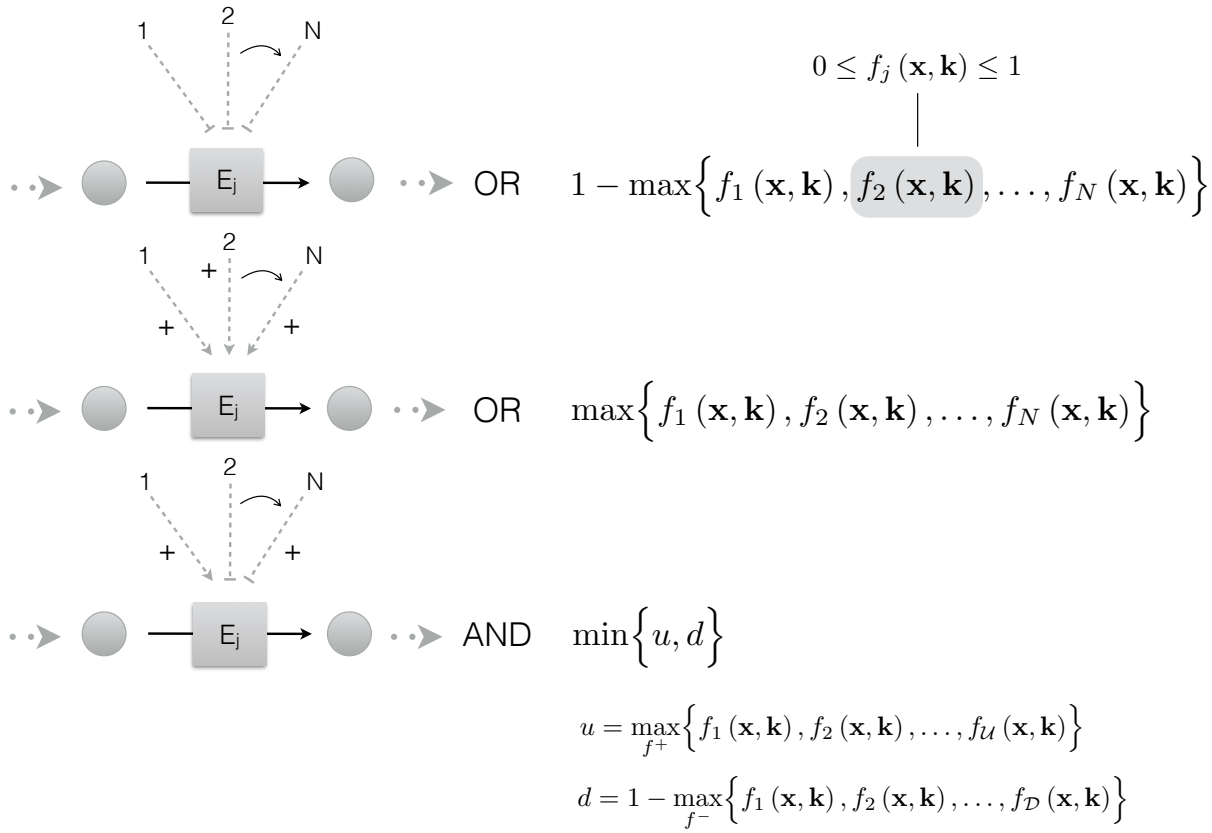

Fig. 2: Schematic of the rule based allosteric enzyme activity control laws.

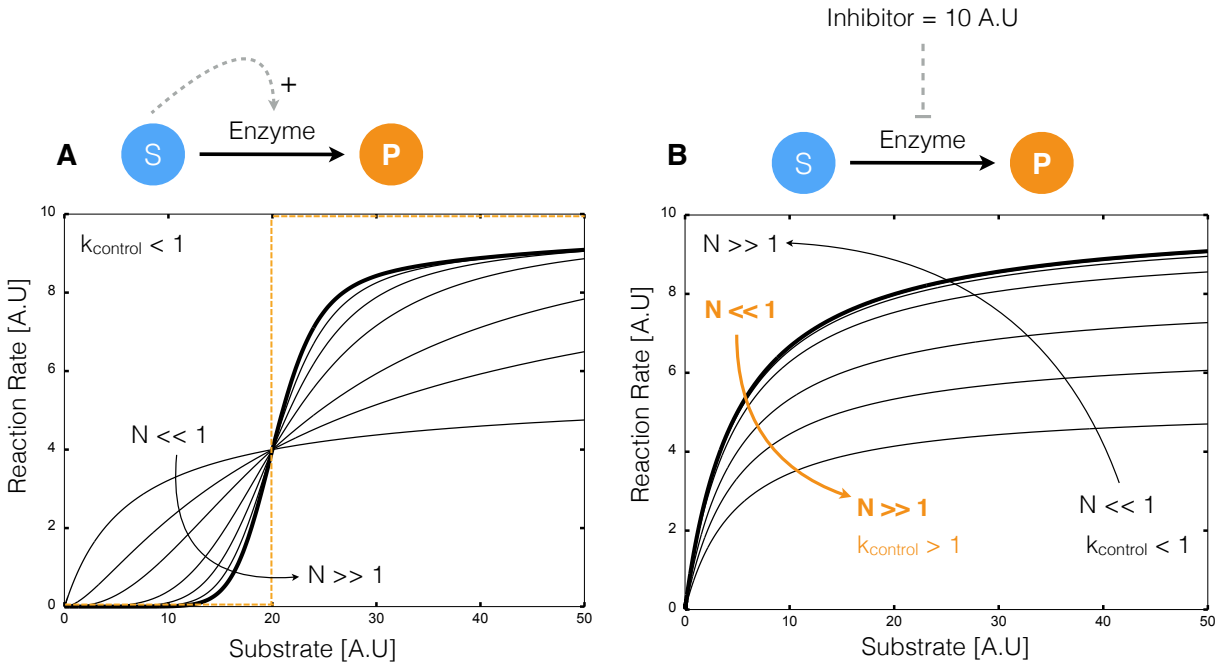

Fig. 3: Kinetics of simple transformations in the presence of activation and inhibition. A:The conversion of substrate S to product P by enzyme E was activated by S. B:The conversion of substrate S to product P by enzyme E was inhibited by inhibitor I.

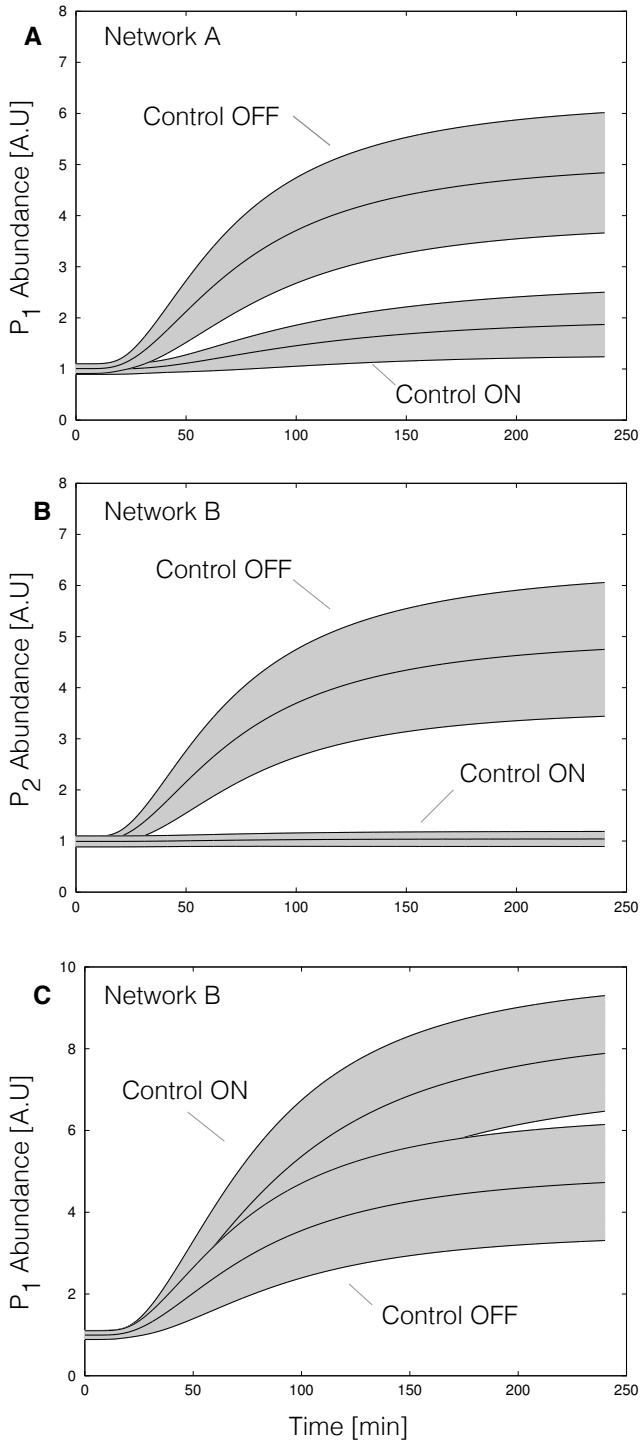

Fig. 4: On/off control simulations for network A and network B for an ensemble of kinetic parameter sets versus time. For each case, N = 100 simulations were conducted using kinetic and initial conditions randomly generated from a hypothetical true parameter set. The gray area represents $\pm$ one standard deviation surrounding the mean. Control parameters were fixed during the ensemble calculations.

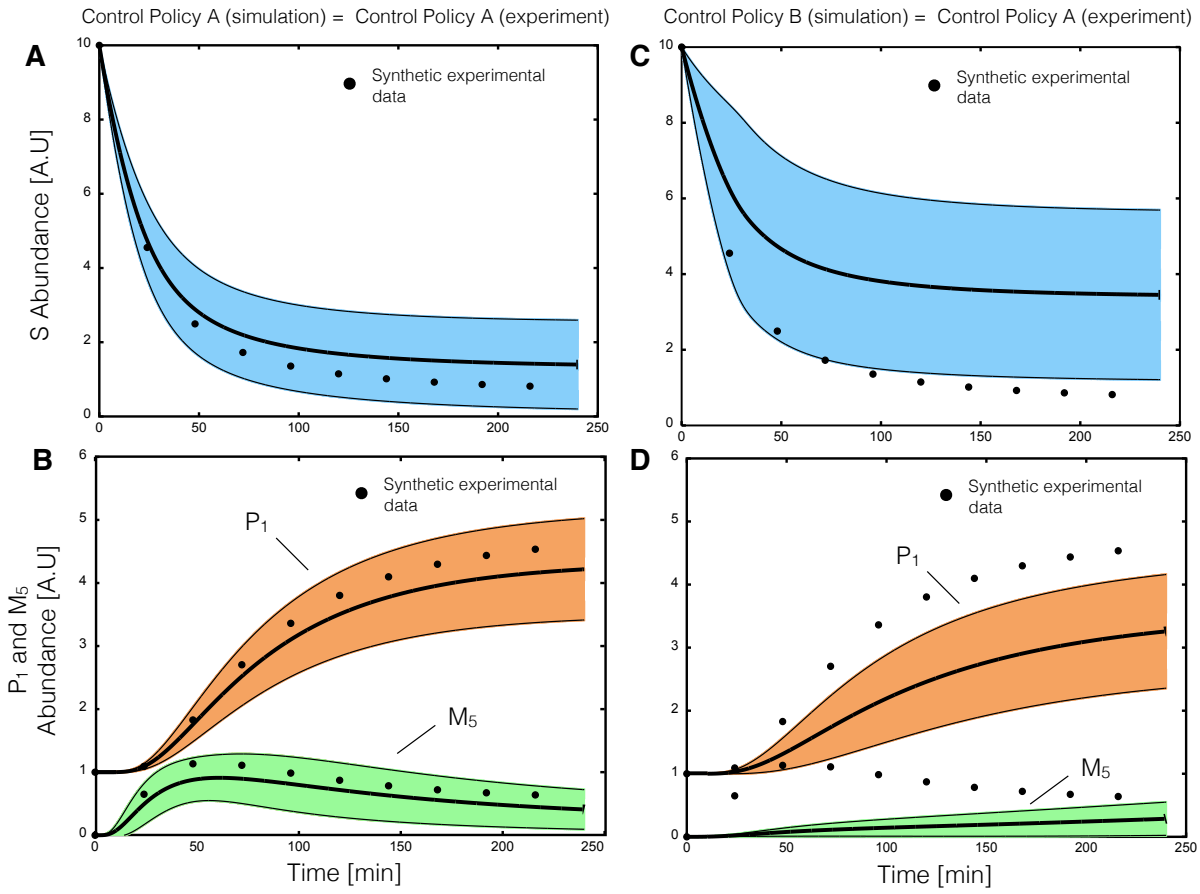

Fig. 5: Parameter estimation from synthetic data for the same and mismatched allosteric control logic.

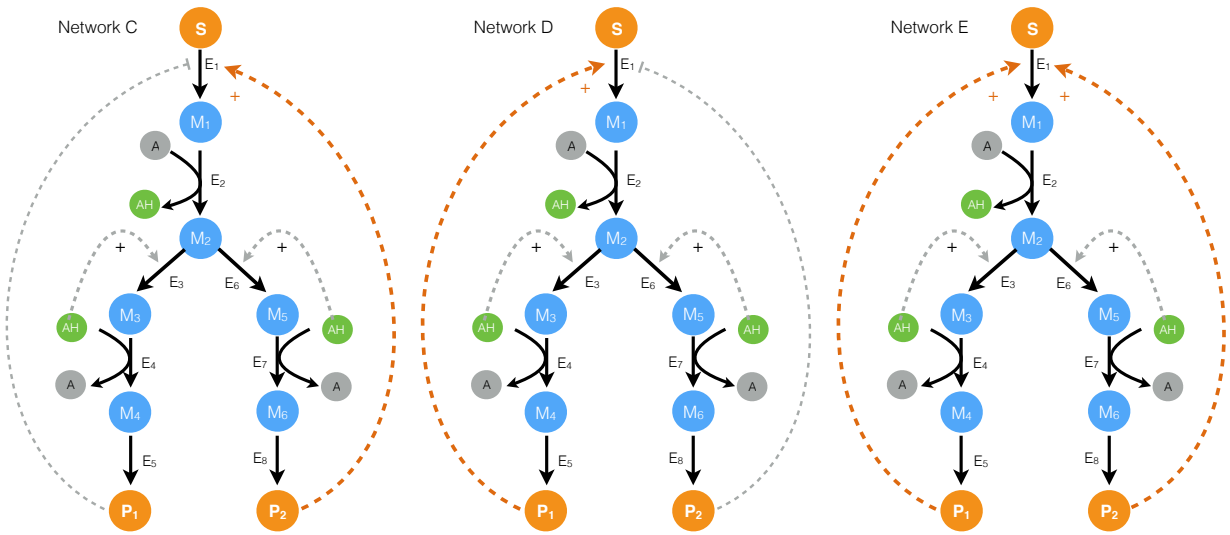

Fig. 6: Schematic of the alternative allosteric control programs used in the structural particle swarm computation. Each network had the same enzymatic connectivity, initial conditions and kinetic parameters, but alternative feedback control structures for the first enzyme in the pathway.

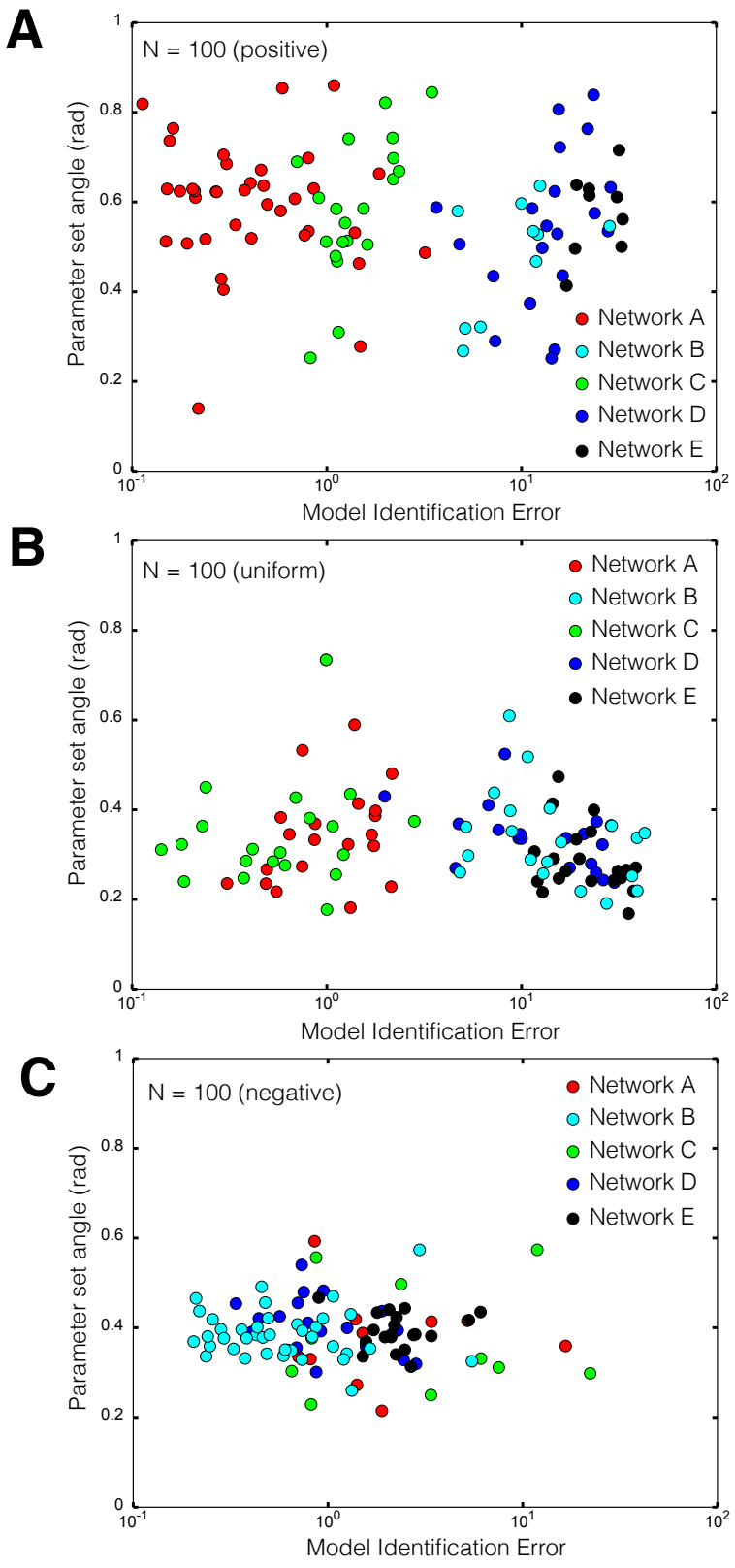

Fig. 7: Combined control structure and kinetic parameter search results. Think about this ...