Dynamic Modeling of Cell-Free Biochemical 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: Effective models of 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

Cell-free systems offer many advantages for the study, manipulation and modeling of metabolism compared to *in vivo* processes. Many of the challenges confronting genome-scale kinetic modeling can potentially be overcome in a cell-free system. For example, there is no complex transcriptional regulation to consider, transient metabolic measurements are easier to obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds several significant advantages for model development, identification and validation. Theoretically, genome-scale cell-free kinetic models may be possible for industrially important organisms, such as *E. coli*, if a simple, tractable framework for integrating allosteric regulation with enzyme kinetics can be formulated. Toward this unmet need, we present an effective biochemical network modeling framework for building dynamic cell-free metabolic models. The key innovation of our approach is the integration of simple effective rules encoding complex allosteric regulation with traditional kinetic pathway modeling. We tested our approach by modeling the time evolution of several hypothetical cell-free metabolic networks. We found that simple effective rules, when integrated with traditional enzyme kinetic expressions, captured complex allosteric patterns such as ultrasensitivity or non-competitive inhibition in the absence of mechanistic information. Second, when integrated into network models, these rules captured classic regulatory patterns such as product-induced feedback inhibition. Lastly, we showed, at least for the network architectures considered here, that we could simultaneously estimate kinetic parameters and allosteric connectivity from synthetic data. While only an initial proof-of-concept, the framework presented here could be an important first step toward genome-scale cell-free kinetic modeling of the biosynthetic capacity of industrially important organisms.

Keywords: Cell-free metabolism, Mathematical modeling

Introduction

Mathematical modeling has long contributed to our understanding of metabolism. Decades before the genomics revolution, mechanistically, structured metabolic models arose from the desire to predict microbial phenotypes resulting from changes in intracellular or extracellular states [1]. The single cell *E. coli* models of Shuler and coworkers pioneered the construction of large-scale, dynamic metabolic models that incorporated multiple, regulated catabolic and anabolic pathways constrained by experimentally determined kinetic parameters [2]. Shuler and coworkers generated many single cell kinetic models, including single cell models of eukaryotes [3, 4], minimal cell architectures [5], as well as DNA sequence based whole-cell models of *E. coli* [6]. Conversely, highly abstracted kinetic frameworks, such as the cybernetic framework, represented a paradigm shift, viewing cells as growth-optimizing strategists [7]. Cybernetic models have been highly successful at predicting metabolic choice behavior, e.g., diauxie behavior [8], steady-state multiplicity [9], as well as the cellular response to metabolic engineering modifications [10]. Unfortunately, cybernetic models also suffer from an identifiability challenge, as both the kinetic parameters and an abstracted model of cellular objectives must be estimated simultaneously.

In the post genomics world, large-scale stoichiometric reconstructions of microbial metabolism popularized by static, constraint-based modeling techniques such as flux balance analysis (FBA) have become standard tools [11]. Since the first genome-scale stoichiometric model of *E. coli*, developed by Edwards and Palsson [12], well over 100 organisms, including industrially important prokaryotes such as *E. coli* [13] or *B. subtilis* [14], are now available [15]. Stoichiometric models rely on a pseudo-steady-state assumption to reduce unidentifiable genome-scale kinetic models to an underdetermined linear algebraic system, which can be solved efficiently even for large systems. Traditionally, stoichiometric models have also neglected explicit descriptions of metabolic regulation

and control mechanisms, instead opting to describe the choice of pathways by prescribing an objective function on metabolism. Interestingly, similar to early cybernetic models, the most common metabolic objective function has been the optimization of biomass formation [16], although other metabolic objectives have also been estimated [17]. Recent advances in constraint-based modeling have overcome the early shortcomings of the platform, including capturing metabolic regulation and control [18]. Thus, modern constraint-based approaches have proven extremely useful in the discovery of metabolic engineering strategies and represent the state of the art in metabolic modeling [19, 20]. However, genome-scale kinetic models of industrial important organisms such as *E. coli* have yet to be constructed.

Cell-free systems offer many advantages for the study, manipulation and modeling of metabolism compared to *in vivo* processes. Central amongst these advantages is direct access to metabolites and the microbial biosynthetic machinery without the interference of a cell wall. This allows us to control as well as interrogate the chemical environment while the biosynthetic machinery is operating, potentially at a fine time resolution. Second, cell-free systems also allow us to study biological processes without the complications associated with cell growth. Cell-free protein synthesis (CFPS) systems are arguably the most prominent examples of cell-free systems used today [21]. However, CFPS is not new; CFPS in crude *E. coli* extracts has been used since the 1960s to explore fundamentally important biological mechanisms [22, 23]. Today, cell-free systems are used in a variety of applications ranging from therapeutic protein production [24] to synthetic biology [25]. Interestingly, many of the challenges confronting genome-scale kinetic modeling can potentially be overcome in a cell-free system. For example, there is no complex transcriptional regulation to consider, transient metabolic measurements are easier to obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds several significant advantages for model development, identification and validation. The-

oretically, genome-scale cell-free kinetic models may be possible for industrially important organisms, such as *E. coli* or *B. subtilis*, if a simple, tractable framework for integrating allosteric regulation with enzyme kinetics can be formulated.

In this study, we present an effective biochemical network modeling framework for building dynamic cell-free metabolic models. The key innovation of our approach is the seamless integration of simple effective rules encoding complex regulation with traditional kinetic pathway modeling. This integration allows the description of complex regulatory interactions, such as time-dependent allosteric regulation of enzyme activity, in the absence of specific mechanistic information. The regulatory rules are easy to understand, easy to formulate and do not rely on overarching theoretical abstractions or restrictive assumptions. We tested our approach by modeling the time evolution of several hypothetical cell-free metabolic networks. In particular, we tested whether our effective modeling approach could describe classically expected enzyme kinetic behavior, and second whether we could simultaneously estimate kinetic parameters and regulatory connectivity, in the absence of specific mechanistic knowledge, from synthetic experimental data. Toward these questions, we explored five hypothetical cell-free networks. Each network shared the same enzymatic connectivity, but had different allosteric regulatory connectivity. We found that simple effective rules, when integrated with traditional enzyme kinetic expressions, captured complex allosteric patterns such as ultrasensitivity or non-competitive inhibition in the absence of mechanistic information. Second, when integrated into network models, these rules captured classical regulatory patterns such as product-induced feedback inhibition. Lastly, we showed, at least for the network architectures considered here, that we could simultaneously estimate kinetic parameters and allosteric connectivity from synthetic data. While only an initial proof-of-concept, the framework presented here could be an important first step toward genome-scale cell-free kinetic modeling of the biosynthetic capacity of industrially important organisms.

Results

Formulation and properties of effective cell-free metabolic models. We developed two proof-of-concept metabolic networks to investigate the features of our effective biochemical network 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 regulatory mechanisms modified the activity of pathway enzymes. 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 reduced cofactor AH. Lastly, it is known experimentally that cell-free systems have a finite operational lifespan. Loss of biosynthetic capability could be a function of many factors, e.g., cofactor or metabolite limitations. We modeled the loss of biosynthetic capability as a non-specific first-order decay of enzyme activity.

Allosteric regulation of enzyme activity was modeled by combining individual regulatory contributions to the activity of pathway enzymes into a control coefficient using an integration rule (Fig. 2). This strategy is similar in spirit to the Constrained Fuzzy Logic (cFL) approach of Lauffenburger and coworkers which has been used to effectively model signal transduction pathways important in human health [26]. In our formulation, Hill-like transfer functions $0 \leq f(\mathcal{Z}) \leq 1$ were used to calculate the influence of factor abundance upon target enzyme activity. In this context, factors can be individual metabolite levels or some function, e.g., the product of metabolite levels. However, more generally, factors can also correspond to non-modeled influences, categorial variables or other abstract quantities. In the current study, we simply let $\mathcal{Z}$ correspond to the abundance of individual metabolites. When an enzyme was potentially sensitive to more than one regulatory

input, logical integration rules were used to select which regulatory transfer function influenced enzyme activity at any given time. Thus, our test networks involved important features such as cofactor recycling, enzyme activity and metabolite dynamics, as well as multiple overlapping allosteric regulatory mechanisms.

The rule-based regulatory strategy approximated the behavior of classical allosteric activation and inhibition mechanisms (Fig. 3). We considered the enzyme catalyzed conversion of substrate S to a product P, where the overall reaction rate was modeled as the product of a Michaelis-Menten term and an effective allosteric control variable reflecting the particular regulatory interaction. 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 substrate 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 *decreased* at regulatory sites, which negatively impacted substrate binding 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). Interestingly, our rule-based approach was unable to directly simulate competitive inhibition of enzyme activity. 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 product 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 effective modeling strategy, and particularly the rule-based approximation of allosteric regulation, will not be feasible for genome-scale cell-free 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 were 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), for fixed control parameters. 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 of network B was robust to changes in kinetic parameters, as the same qualitative trend was conserved across an ensemble of randomized parameters (N = 100), for fixed control parameters. Taken together, these simulations suggested that the rule-based allosteric control concept could robustly capture expected feedback behavior for networks with uncertain kinetic parameters.

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 struc-

tures 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 (PSO), starting from random initial parameter 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, with control mismatch (network B simulated with network A parameters), model simulations were not consistent with the synthetic data (Fig. 5C and D). Taken together, these results suggested that we could perhaps simultaneously estimate both parameters and network control architectures, as incorrect control structures would be manifest as poor model fits.

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 pathway enzymes (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 either positively or negatively biased towards the correct structure (network A). In both the positively biased and uniform cases the PSO 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, when given a uniform distribution, the PSO approach identified a combination of network A and network C 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, end product P_2 was assumed to inhibit E_1, while in network C, end product P_2 activated E_1. Lastly, when the initial population was biased towards incorrect structures (initial population seeded with 90% incorrect structures), the particle swarm *misidentified* the correct allosteric structure (Fig. 7C). Interestingly, while each particle swarm identified parameter sets that minimized the simulation error, the estimated parameter values were not necessarily similar to the true parameters. The angle between the estimated and true parameters was not consistently small across the swarms (identical parameters would give an angle of zero). This suggested that our particle swarm approach identified a *sloppy* ensemble, i.e., parameter estimates that were individually incorrect but collectively exhibited the correct model behavior.

We calculated control program output and scaled metabolic flux for the positively, uniformly and negatively biased particle swarms (Fig. 8). Network A and network C models from the positively (Fig. 8A) and uniformly (Fig. 8B) biased particle swarms showed similar operational patterns, despite differences in kinetic parameters and control structures. While models from the negatively biased population had error values similar to the correct structures in the previous swarms, they have different flux and control profiles (Fig. 8C).

In all cases, regardless of network configuration or parameter values, the rate of enzyme decay was small compared to the other fluxes, and all networks had qualitatively similar trends for E_3 and E_6 control. Moreover, consistent with the correct model structure, production of end product P_1 was the preferred branch for all model configurations. However, there was variability in P_2 production flux across the population of models, especially for the uniform swarm when compared with the other cases. High P_1 branch flux resulted in end product inhibition of E_1 in both network A and network C, however in network D and E, high P_1 flux induced E_1 activation. These trends were manifested in different flux profiles, where the negatively biased population appeared more uniform across the population compared with the other swarms, and had higher E_1 specific activity. Interestingly, the behavior of network A and network C highlighted an artifact of our integration rule; both a positive or negative feedback connection from P_2 to E_1 were ignored because the P_1 inhibition of E_1 was dominate. Thus, while theoretically distinct, network A and network C appeared operationally to the PSO algorithm to be that same network. On the other hand, networks B, D and E showed distinct behavior that was not consistent with the true network. These architectures exhibited either limited inhibition (network B) or activation (network D and E) of E_1 activity, resulting in significantly different metabolic flux profiles. However, the PSO was able to find low error parameter solutions, despite the mismatch in the control structures (error values similar, but not better than the best network A and network C estimates). Taken together, these results suggested that a uniform sampling approach could potentially yield an unbiassed estimate of both kinetic parameters and control structures. However, the negatively biased particle swarm results illustrated a potential shortcoming of the approach, namely convergence to a local error minimum despite a significantly incorrect control structure. This suggested that estimated model structures will need to be further evaluated, for example by generating falsifiable experimental designs which could distinguish between low error solutions.

Discussion

In this study, we presented an effective kinetic modeling strategy to dynamically simulate cell-free biochemical networks. Our proposed strategy integrated traditional kinetic modeling with an effective rules based approach to dynamically describe metabolic regulation and control. We tested this approach by developing kinetic models of hypothetical cell-free metabolic networks. In particular, we tested whether our effective modeling approach could describe classically expected behavior, and second whether we could simultaneously estimate kinetic parameters and regulatory connectivity, in the absence of specific mechanistic knowledge, from synthetic experimental data. Toward these questions, we explored five hypothetical cell-free networks. In each network, a 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 a hypothetical metabolite M_2. Each network also included the same cofactors and cofactor recycle architecture. However, while all five networks shared the same enzymatic connectivity, each had different allosteric regulatory connectivity. We found that simple effective rules, when integrated with traditional enzyme kinetic expressions, could capture complex allosteric patterns such as ultrasensitivity, or non-competitive inhibition in the absence of specific mechanistic information. Moreover, when integrated into network models, these rules captured classical regulatory patterns such as product-induced feedback inhibition. Lastly, we simultaneously estimated kinetic parameters and discriminated between competing regulatory structures, using synthetic data in combination with a modified particle swarm approach.

[CONTRAST WITH CYBERNETIC MODELS].

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 ME]

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 were 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:

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

where $0 \leq f_{ij}\left(x\right) \leq 1$ was a regulatory transfer function that calculated the influence of metabolite i on the activity of enzyme j. 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:

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

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

The quantities j^+ and j^- denoted the sets of activating and inhibitory factors for enzyme j. If an enzyme had no allosteric factors, we set $v_j = 1$. There are many possible functional

forms for $0 \leq f_{ij}(x) \leq 1$. However, in this study, each individual transfer function took the form:

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

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 [27].

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) [28]. 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 $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05564, 0.02886)$. 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 [27].

Acknowledgements

This study was supported by an award from the National Science Foundation (MCB #1411715) and the Army Research Office (ARO #59155-LS).

References

1. Fredrickson AG (1976) Formulation of structured growth models. Biotechnol Bioeng 18: 1481-6.
2. Domach MM, Leung SK, Cahn RE, Cocks GG, Shuler ML (1984) Computer model for glucose-limited growth of a single cell of escherichia coli b/r-a. Biotechnol Bioeng 26: 203-16.
3. Steinmeyer D, Shuler M (1989) Structured model for Saccharomyces cerevisiae. Chem Eng Sci 44: 2017 - 2030.
4. Wu P, Ray NG, Shuler ML (1992) A single-cell model for cho cells. Ann N Y Acad Sci 665: 152-87.
5. Castellanos M, Wilson DB, Shuler ML (2004) A modular minimal cell model: purine and pyrimidine transport and metabolism. Proc Natl Acad Sci U S A 101: 6681-6.
6. Atlas JC, Nikolaev EV, Browning ST, Shuler ML (2008) Incorporating genome-wide dna sequence information into a dynamic whole-cell model of escherichia coli: application to dna replication. IET Syst Biol 2: 369-82.
7. Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT (1985) A cybernetic view of microbial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27: 1-9.
8. Kompala DS, Ramkrishna D, Jansen NB, Tsao GT (1986) Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. Biotechnol Bioeng 28: 1044-55.
9. Kim JI, Song HS, Sunkara SR, Lali A, Ramkrishna D (2012) Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. Biotechnol Prog 28: 1160-6.
10. Varner J, Ramkrishna D (1999) Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. Metab Eng 1: 88-116.
11. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-

366 phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10:
367 291-305.
368 12. Edwards JS, Palsson BO (2000) The escherichia coli mg1655 in silico metabolic
369 genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S
370 A 97: 5528-33.
371 13. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-
372 scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for
373 1260 orfs and thermodynamic information. Mol Syst Biol 3: 121.
374 14. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale re-
375 construction of metabolic network in bacillus subtilis based on high-throughput phe-
376 notyping and gene essentiality data. J Biol Chem 282: 28791-9.
377 15. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of bio-
378 chemical networks in microorganisms. Nat Rev Microbiol 7: 129-43.
379 16. Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli k-12 undergoes adaptive
380 evolution to achieve in silico predicted optimal growth. Nature 420: 186-9.
381 17. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions
382 for predicting intracellular fluxes in escherichia coli. Mol Syst Biol 3: 119.
383 18. Hyduke DR, Lewis NE, Palsson BØ (2013) Analysis of omics data with genome-scale
384 models of metabolism. Mol Biosyst 9: 167-74.
385 19. McCloskey D, Palsson BØ, Feist AM (2013) Basic and applied uses of genome-scale
386 metabolic network reconstructions of escherichia coli. Mol Syst Biol 9: 661.
387 20. Zomorrodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical opti-
388 mization applications in metabolic networks. Metab Eng 14: 672-86.
389 21. Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell-
390 free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4:
391 220.

22. MATTHAEI JH, NIRENBERG MW (1961) Characteristics and stabilization of dnaase-sensitive protein synthesis in e. coli extracts. Proc Natl Acad Sci U S A 47: 1580-8.
23. NIRENBERG MW, MATTHAEI JH (1961) The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 47: 1588-602.
24. Lu Y, Welsh JP, Swartz JR (2014) Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. Proc Natl Acad Sci U S A 111: 125-30.
25. Hodgman CE, Jewett MC (2012) Cell-free synthetic biology: thinking outside the cell. Metab Eng 14: 261-9.
26. 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.
27. Octave community (2014). GNU Octave 3.8.1. URL www.gnu.org/software/octave/.
28. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the International Conference on Neural Networks. pp. 1942 - 1948.

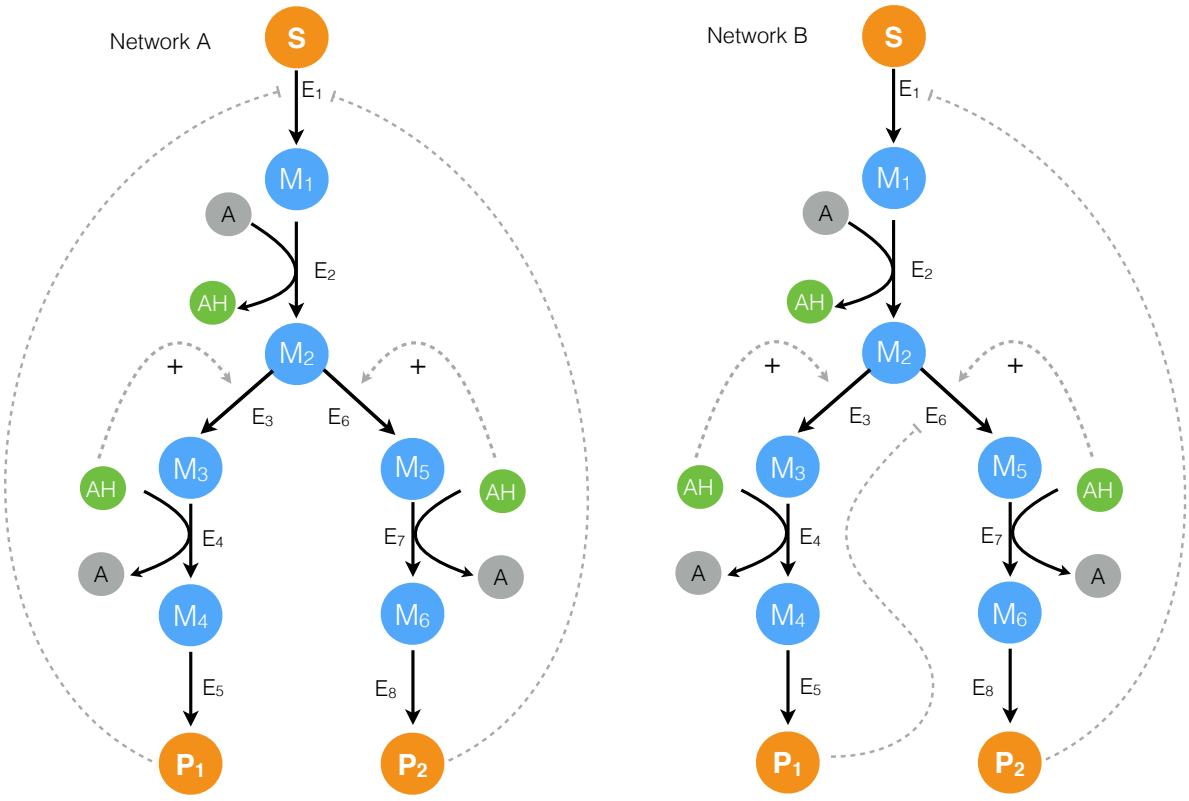

Fig. 1: Proof of concept cell-free metabolic networks considered in this study. Substrate S is converted to products P_1 and P_2 through a series of chemical conversions catalyzed by enzyme(s) E_j. The activity of the pathway enzymes is subject to both positive and negative allosteric regulation.

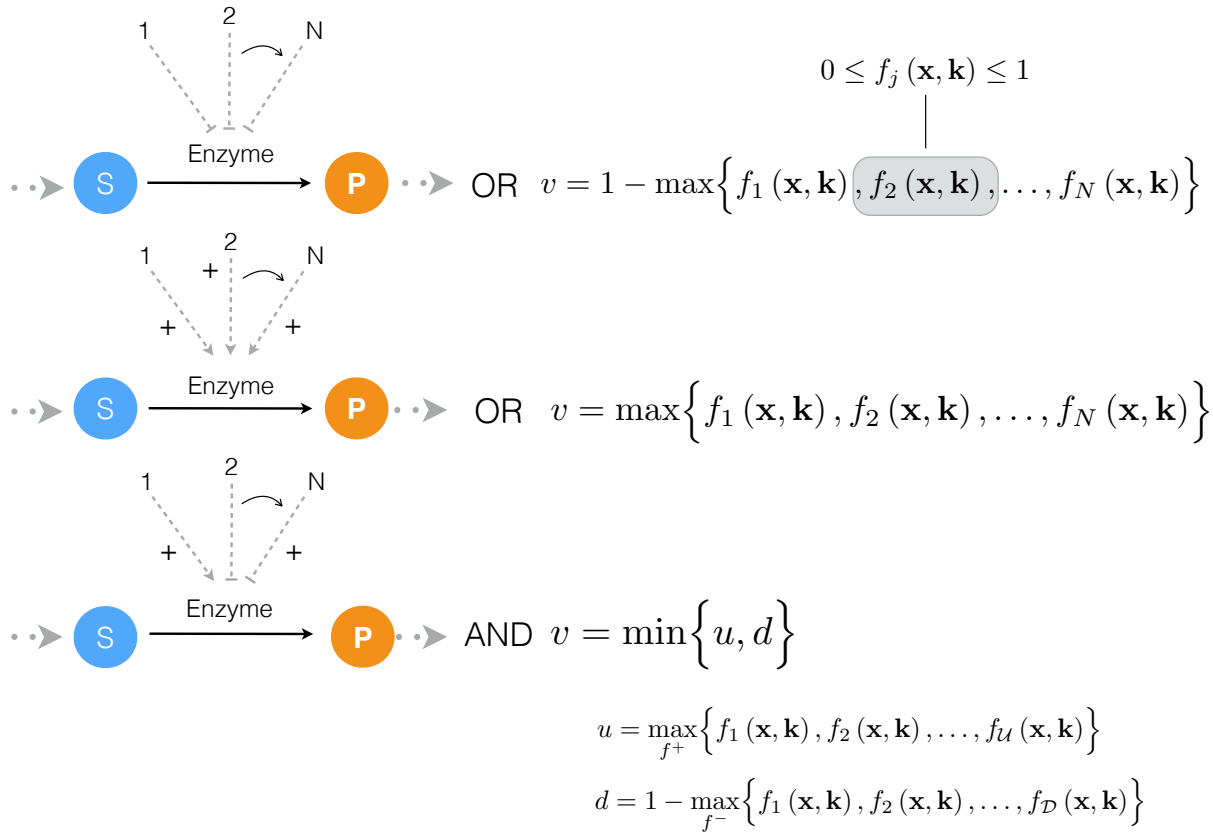

Fig. 2: Schematic of rule-based allosteric enzyme activity control laws. Traditional enzyme kinetic expressions, e.g., Michaelis–Menten or multiple saturation kinetics, are multiplied by an enzyme activity control variable $0 \leq v_j \leq 1$. Control variables are functions of many possible regulatory factors encoded by arbitrary functions of the form $0 \leq f_j(\mathcal{Z}) \leq 1$. At each simulation time step, the v_j variables are calculated by evaluating integration rules such as the max or min of the set of factors $f_1, \ldots$ influencing the activity of enzyme E_j.

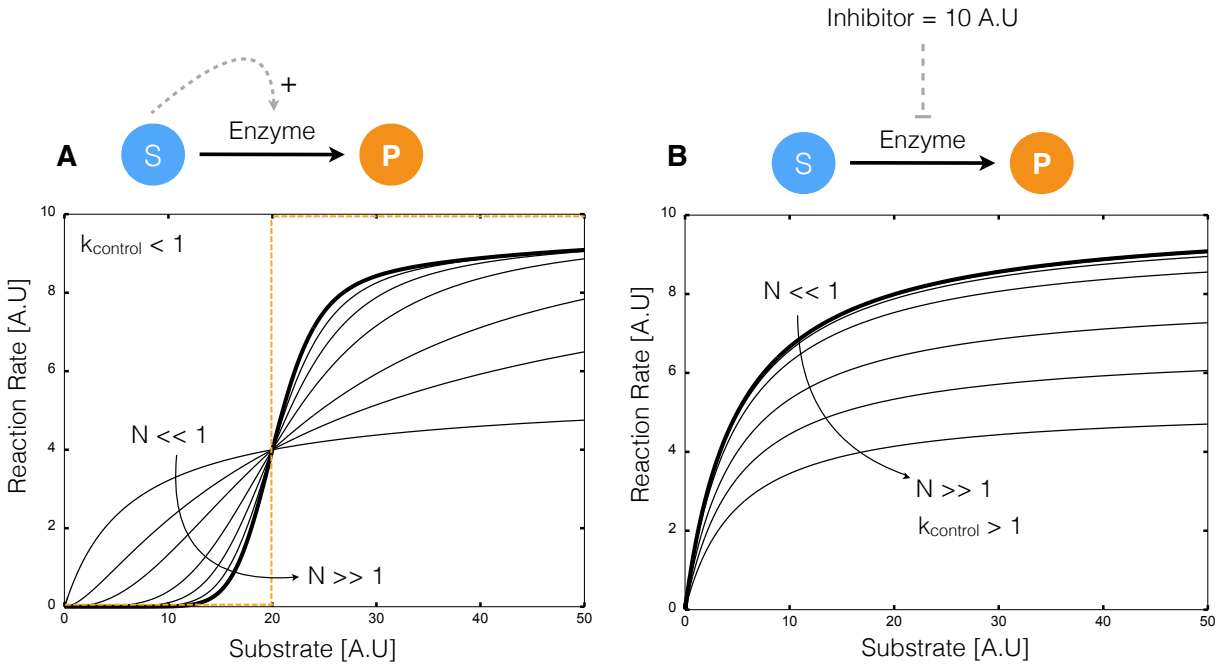

Fig. 3: Kinetics of simple transformations in the presence of activation and inhibition. **A**:The conversion of substrate S to product P by enzyme E was activated by S. For a fixed control gain parameter $k_{control}$, the reaction rate approached a step for increasing control order N. **B**:The conversion of substrate S to product P by enzyme E with inhibitor I. For a fixed control gain parameter $k_{control}$, the reaction rate approximated non-competitive inhibition for increasing control order N.

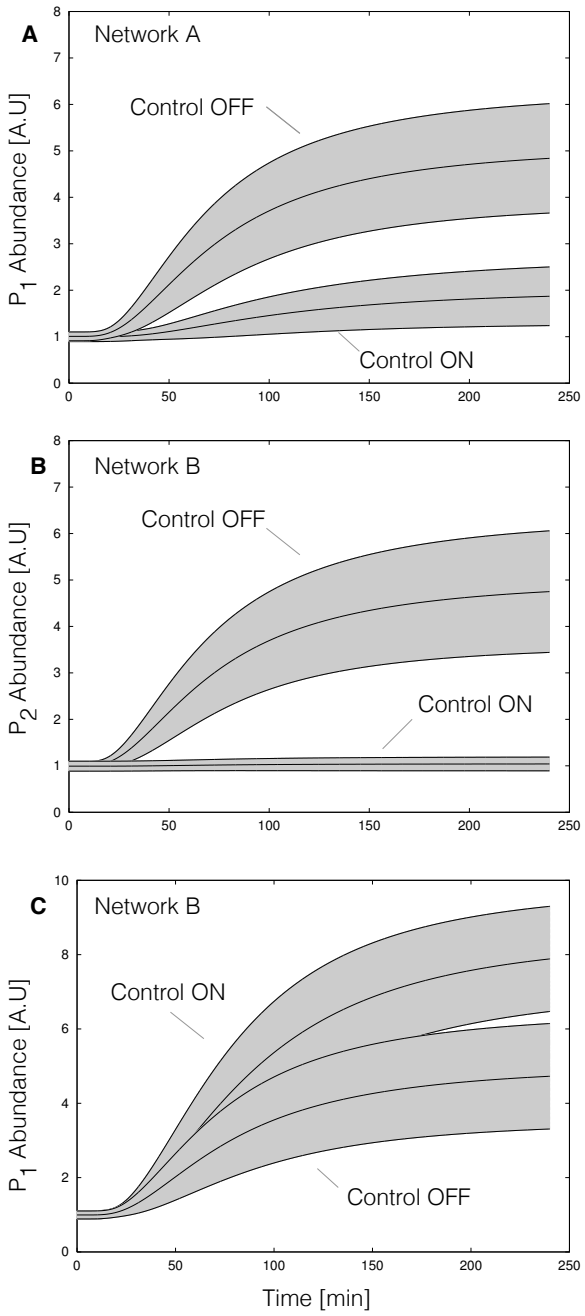

Fig. 4: ON/OFF control simulations for network A and network B for an ensemble of kinetic parameter sets versus time (N = 100). For each case, N = 100 simulations were conducted using kinetic and initial conditions generated randomly from a hypothetical true parameter set. The gray area represents $\pm$ one standard deviation surrounding the mean. Control parameters were fixed during the ensemble calculations. **A**: End product P_1 abundance versus time for Network A. The abundance of P_1 decreased with end product inhibition of E_1 activity (Control-ON) versus the no inhibition case (Control-OFF). **B**: End product P_2 abundance versus time for Network B. Inhibition of branch point E_6 by end product P_1 decreased P_2 abundance (Control-ON) versus the no inhibition case (Control-OFF). **C**: End product P_1 abundance versus time for Network A. Inhibition of branch point E_6 by end product P_1 decreased P_1 abundance (Control-ON) versus the no inhibition case (Control-OFF).

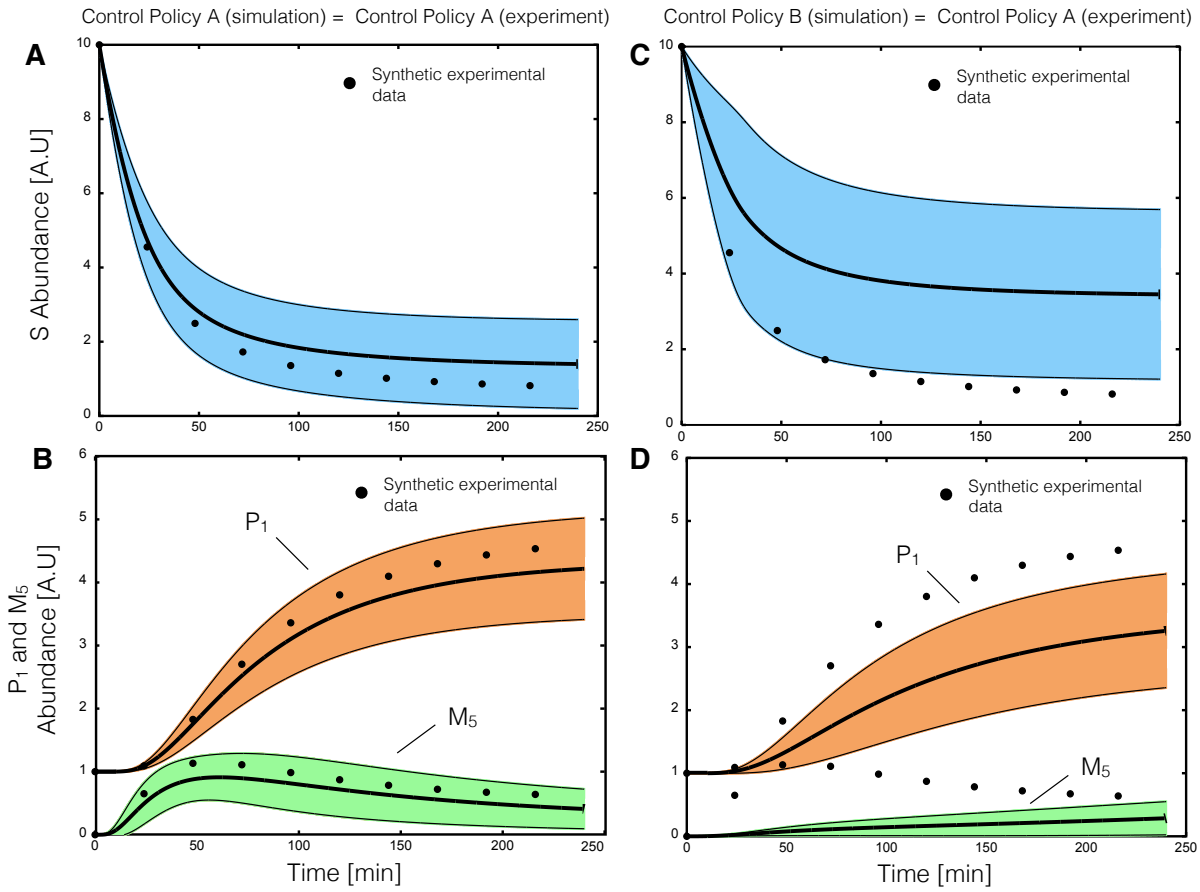

Fig. 5: Parameter estimation from synthetic data for the same and mismatched allosteric control logic using particle swarm optimization (PSO). Synthetic experimental data was generated from a hypothetical parameter set using Network A, where substrate S, end product P_1 and intermediate M_5 were sampled approximately every 20 minutes. For cases **A,B** 20 particles were initialized with randomized parameters and allowed to search for 300 iterations. **A,B**: PSO estimated an ensemble of parameters sets (N = 20) consistent with the synthetic experimental data assuming the correct enzymatic and control connectivity starting from randomized initial parameters. **C,D**: In the presence of control mismatch (Network B control policy simulated with Network A kinetic parameters) the ensemble of models did not describe the synthetic data.

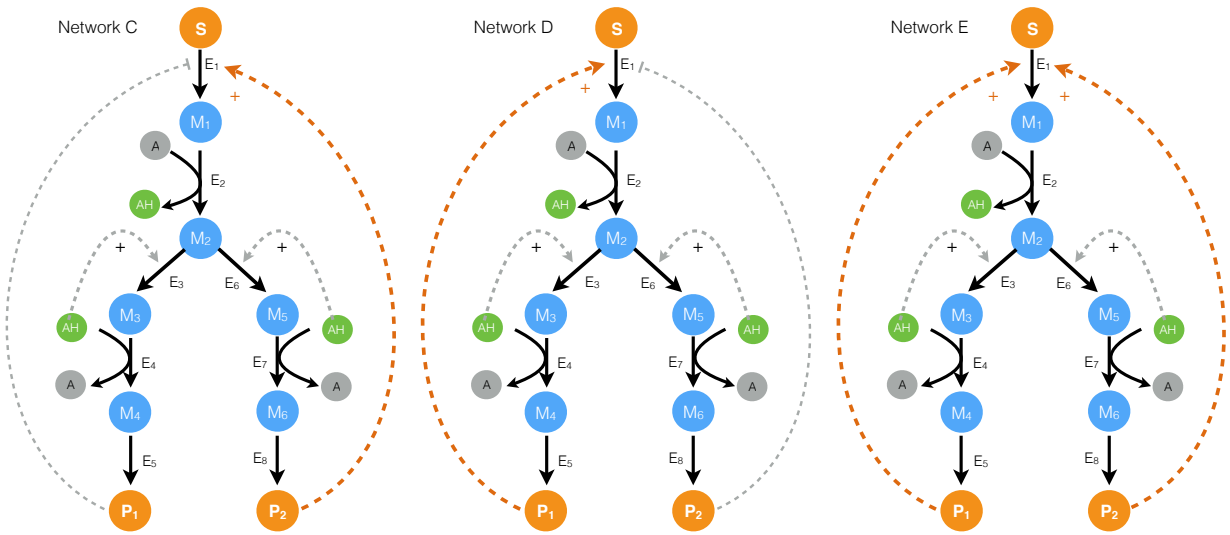

Fig. 6: Schematic of the alternative allosteric control programs used in the structural particle swarm computation. Each network had the same enzymatic connectivity, initial conditions and kinetic parameters, but alternative feedback control structures for the first enzyme in the pathway.

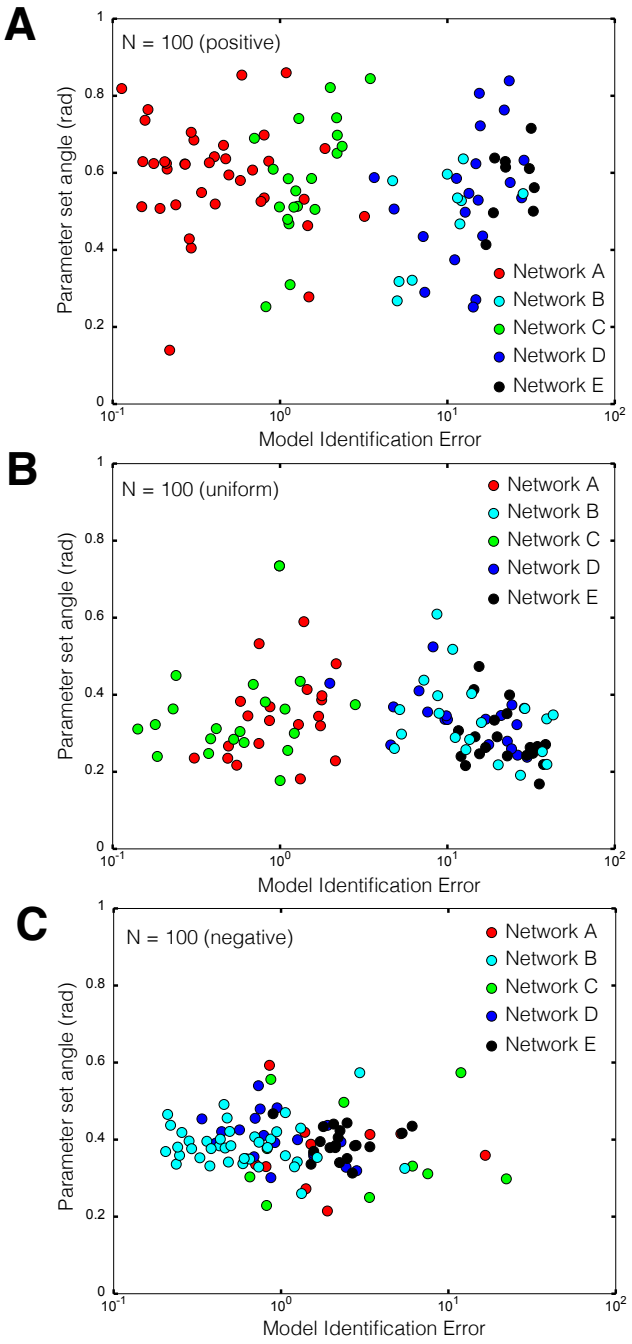

Fig. 7: Combined control and kinetic parameter search using modified particle swarm optimization (PSO). A population of N = 100 particles was initialized with randomized kinetic parameters and one of five possible control configurations (Network A - E). Simulation error was minimized for a synthetic data set (S, end product P_1 and intermediate M_5 sampled approximately every 20 min) generated using Network A. **A**: Simulation error versus parameter set angle for N = 100 particles biased toward the correct regulatory program (A,B,C,D,E) = (40%, 10%, 20%, 20% and 10%). **B**: Simulation error versus parameter set angle for N = 100 uniformly distributed particles (A,B,C,D,E) = (20%, 20%, 20%, 20% and 20%). **C**: Simulation error versus parameter set angle for N = 100 negatively biased particles (A,B,C,D,E) = (10%, 40%, 10%, 20% and 20%). Network A (the correct structure) was preferentially identified for positively and uniform biased particle distributions, but misidentified in the presence of a large incorrect bias.

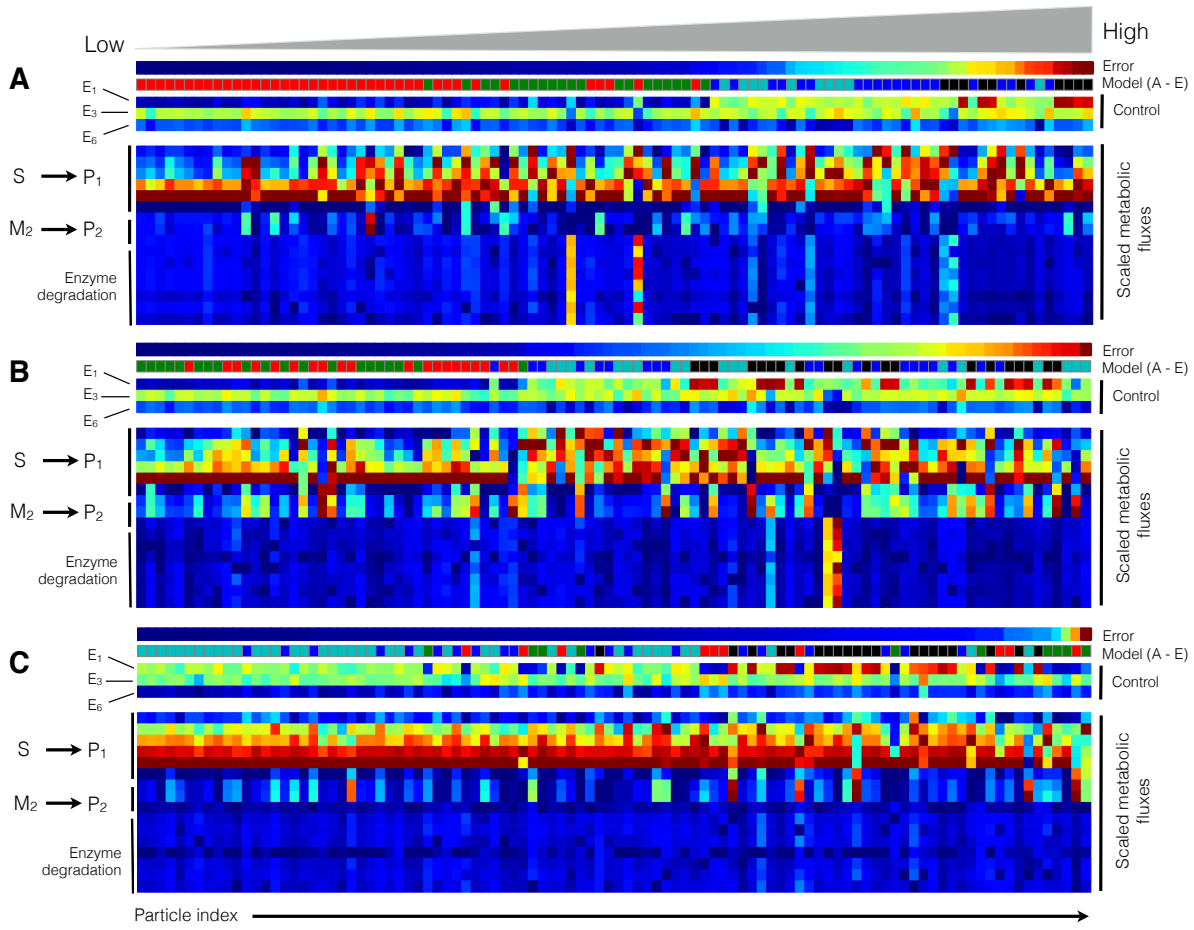

Fig. 8: Metabolic flux and control variables as a function of network type and particle index at t = 100 min. The control variables governing E_1, E_3 and E_6 activity and the scaled metabolic flux and were calculated for the positively, uniformly and negatively biased particle swarms (N = 100). The particles from each swarm were sorted based upon simulation error (low to high error). **A**: Model performance for the positively biased particle swarm as a function of particle index. **B**: Model performance for the uniformly biased particle swarm as a function of particle index. **C**: Model performance for the negatively biased particle swarm as a function of particle index. Models with significant control mismatch showed distinct control and flux patterns versus those models with the correct or closely related control policies. In particular, models with the correct control policy showed stronger inhibition of E_1 activity, leading to decreased flux from S$\rightarrow$P$_1$. Conversely, models with significant mismatch had increased E_1 activity, leading to an altered flux distribution. This is especially apparent in the negatively biased particle swarm.