

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

1 Introduction

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

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

19 **Results**

20 **Formulation and properties of cell free effective models.** We developed two proof of
21 concept metabolic networks to investigate the features of our effective cell free modeling
22 approach (Fig. 1). In both examples, substrate S was converted to the end-products P_1
23 and P_2 through a series of enzymatically catalyzed reactions, including a branch point
24 at hypothetical metabolite M_2 . Several of these reactions involved cofactor dependence
25 (AH or A), and various allosteric regulation mechanisms. Network A included feedback
26 inhibition of the initial pathway enzyme (E_1) by pathway end products P_1 and P_2 (Fig. 1A).
27 On the other hand, network B involved feedback inhibition of E_1 by P_2 and E_6 by P_1 (Fig.
28 1B). In both networks, branch point enzymes E_3 and E_6 were subject to feed-forward ac-
29 tivation by cofactor AH . Lastly, enzyme activity was assumed to decay according to a
30 first-order rate law in both cases. Allosteric regulation of enzyme activity was represented
31 using a novel rule-based strategy, similar in spirit to the Constrained Fuzzy Logic (cFL)
32 approach of Lauffenberger and coworkers [1]. In this formulation, Hill-like transfer func-
33 tions were used to calculate the influence of metabolite abundance upon target enzyme
34 activity. When an enzyme was potentially sensitive to more than one regulatory influence,
35 logical rules were used to select which transfer function regulated enzyme activity at any
36 given time (Fig. 2). Thus, our test networks involved important features such as cofac-
37 tor recycling, enzyme activity and metabolite dynamics, as well as multiple overlapping
38 allosteric regulatory mechanisms. As such, developing our effective modeling approach
39 using these simple problems gave us valuable insight into the development of larger net-
40 work models, without the complication of network size.

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

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

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

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

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

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

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

₁₂₃ network C P_2 was assumed to activate E_1 . Lastly, when the initial population was biased
₁₂₄ towards a completely incorrect structure (initial population seeded with 40% network B),
₁₂₅ the particle swarm misidentified the correct structure (Fig. 7C). [FINISH]

126 **Discussion**

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

133 Cybernetic models, other dynamic models of metabolism.

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

137 **Materials and Methods**

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

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \quad i = 1, 2, \dots, \mathcal{M} \quad (1)$$

$$\frac{d\epsilon_i}{dt} = -\lambda_i \epsilon_i \quad i = 1, 2, \dots, \mathcal{E} \quad (2)$$

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

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

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

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

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

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

154 tivity which catalyzes reaction j , and K_{js} denotes the saturation constant for species s in
 155 reaction j . The product in Eqn. (4) was carried out over the set of *reactants* for reaction j
 156 (denoted as m_j^-).

157 The allosteric regulation term v_j depended upon the combination of factors which in-
 158 fluenced the activity of enzyme i . For each enzyme, we used a rule based approach to
 159 select from competing control factors (Fig. 2). If an enzyme was activated by m metabo-
 160 lites, we modeled this activation as:

$$v_j = \max(f_{1j}(x), \dots, f_{mj}(x)) \quad (5)$$

161 where $0 \leq f_{ij}(x) \leq 1$ was a regulatory transfer function that calculated the influence of
 162 metabolite i on the activity of enzyme j . Conversely, if enzyme activity was inhibited by a
 163 m metabolites, we modeling this inhibition as:

$$v_j = 1 - \max(f_{1j}(x), \dots, f_{mj}(x)) \quad (6)$$

164 Lastly, if an enzyme had both m activating and n inhibitory factors, we modeled the regu-
 165 latory term as:

$$v_j = \min(u_j, d_j) \quad (7)$$

166 where:

$$u_j = \max_{j^+}(f_{1j}(x), \dots, f_{mj}(x)) \quad (8)$$

$$d_j = 1 - \max_{j^-}(f_{1j}(x), \dots, f_{nj}(x)) \quad (9)$$

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

¹⁶⁹ functional forms for $0 \leq f_{ij}(x) \leq 1$. However, in this study, each individual transfer function
¹⁷⁰ took the form:

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

¹⁷¹ where x_j denotes the abundance of metabolite j , and κ_{ij} and η are control parameters.
¹⁷² The κ_{ij} parameter was species gain parameter, while η was a cooperativity parameter
¹⁷³ (similar to a Hill coefficient). The model equations were encoded using the Octave pro-
¹⁷⁴ gramming language, and solved using the LSODE routine in Octave [2].

¹⁷⁵ **Estimation of model parameters and structures from synthetic experimental data.**

¹⁷⁶ Model parameters were estimated by minimizing the difference between simulations and
¹⁷⁷ synthetic experimental data (squared residual):

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

¹⁷⁸ where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ , $x_j(\tau, \mathbf{k})$ denotes the sim-
¹⁷⁹ ulated value for species j at time τ , and $\omega_j(\tau)$ denotes the experimental measurement
¹⁸⁰ variance for species j at time τ . The outer summation is respect to time, while the inner
¹⁸¹ summation is with respect to state. We approximated a realistic model identification sce-
¹⁸² nario, assuming noisy experimental data, limited sampling resolution (approximately 20
¹⁸³ minutes per sample) and a limited number of measurable metabolites.

¹⁸⁴ We minimized the model residual using Particle swarm optimization (PSO) [3]. PSO
¹⁸⁵ uses a *swarming* metaheuristic to explore parameter spaces. A strength of PSO is its abil-
¹⁸⁶ ity to find the global minimum, even in the presence of potentially many local minima, by
¹⁸⁷ communicating the local error landscape experienced by each particle collectively to the
¹⁸⁸ swarm. Thus, PSO acts both as a local and a global search algorithm. For each iteration,
¹⁸⁹ particles in the swarm compute their local error by evaluating the model equations using

190 their specific parameter vector realization. From each of these local points, a globally best
 191 error is identified. Both the local and global error are then used to update the parameter
 192 estimates of each particle using the rules:

$$\Delta_i = \theta_1 \Delta_i + \theta_2 \mathbf{r}_1 (\mathcal{L}_i - \mathbf{k}_i) + \theta_3 \mathbf{r}_2 (\mathcal{G} - \mathbf{k}_i) \quad (12)$$

$$\mathbf{k}_i = \mathbf{k}_i + \Delta_i \quad (13)$$

193 where $(\theta_1, \theta_2, \theta_3)$ are adjustable parameters, \mathcal{L}_i denotes local best solution found by par-
 194 ticle i , and \mathcal{G} denotes the best solution found over the entire population of particles.
 195 The quantities r_1 and r_2 denote uniform random vectors with the same dimension as
 196 the number of unknown model parameters ($K \times 1$). In this study, we used $(\theta_1, \theta_2, \theta_3) =$
 197 $(1.0, 0.05564, 0.02886)$, which was taken from XXX. The quality of parameter estimates
 198 was measured using two criteria, goodness of fit (model residual) and angle between the
 199 estimated parameter vector \mathbf{k}_j and the true parameter set \mathbf{k}^* :

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

200 If the candidate parameter set \mathbf{k}_j were perfect, the residual between the model and syn-
 201 thetic data and the angle between \mathbf{k}_j and the true parameter set \mathbf{k}^* would be equal to
 202 zero.

203 We modified our PSO implementation to simultaneously search over kinetic parame-
 204 ters and putative model control structures. In the combined case, each particle potentially
 205 carried a different model realization in addition to a different kinetic parameter vector. We
 206 kept the update rules the same (along with the update parameters). Thus, each parti-
 207 cle competed on the basis of goodness of fit, which allowed different model structures
 208 to contribute to the overall behavior of the swarm. We considered five possible model

209 structures (A through E), where network A was the correct formulation (used to generate
210 the synthetic data). We considered a population $N = 100$ particles, where each particle
211 in the swarm was assigned a model structure, and a random parameter vector. The PSO
212 algorithm, model equations, and the objective function were encoded and solved in the
213 Octave programming language [2].

214 **Acknowledgements**

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

²¹⁷ **References**

- ²¹⁸ 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.
- ²²² 2. Octave community (2014). GNU Octave 3.8.1. URL www.gnu.org/software/octave/.
- ²²³ 3. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the International Conference on Neural Networks. pp. 1942 - 1948.
- ²²⁴

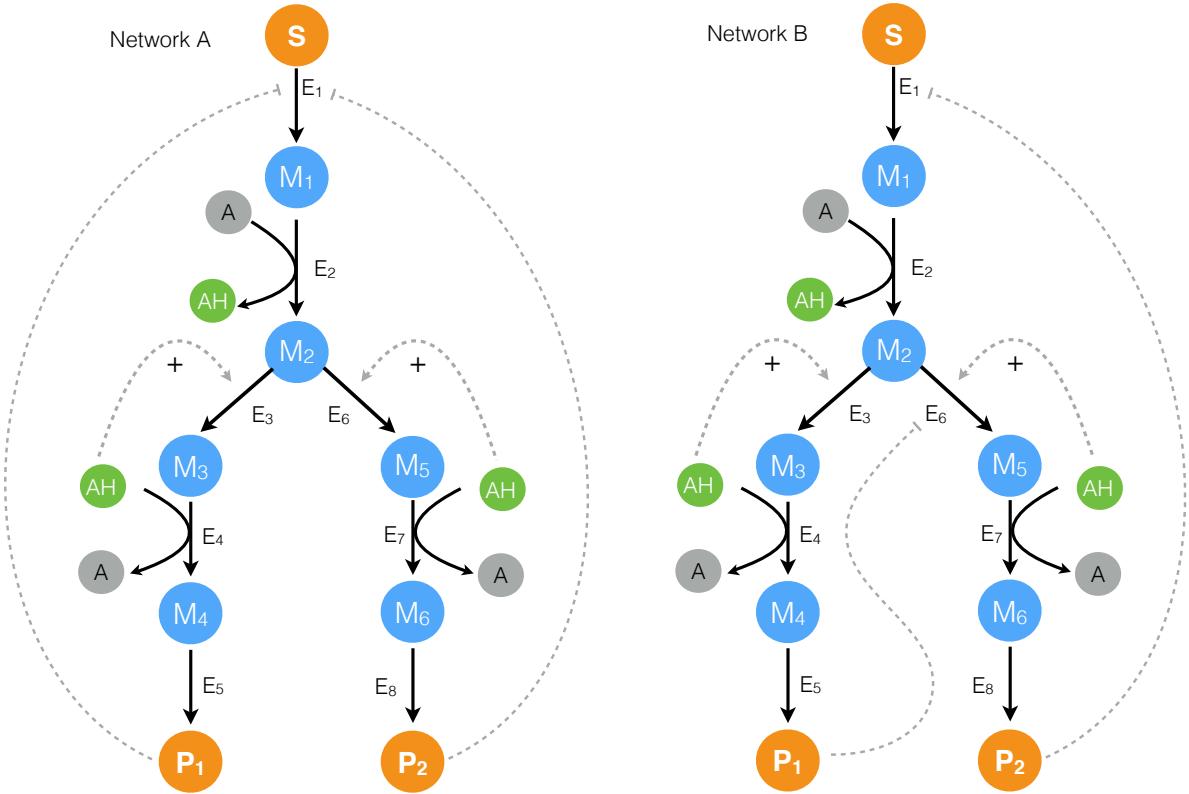


Fig. 1: Proof of concept cell-free metabolic networks considered in this study. Substrate S is converted to products P_1 and P_2 through a series of chemical conversions catalyzed by enzyme(s) E_j . The activity of the pathway enzymes is subject to both positive and negative allosteric regulation.

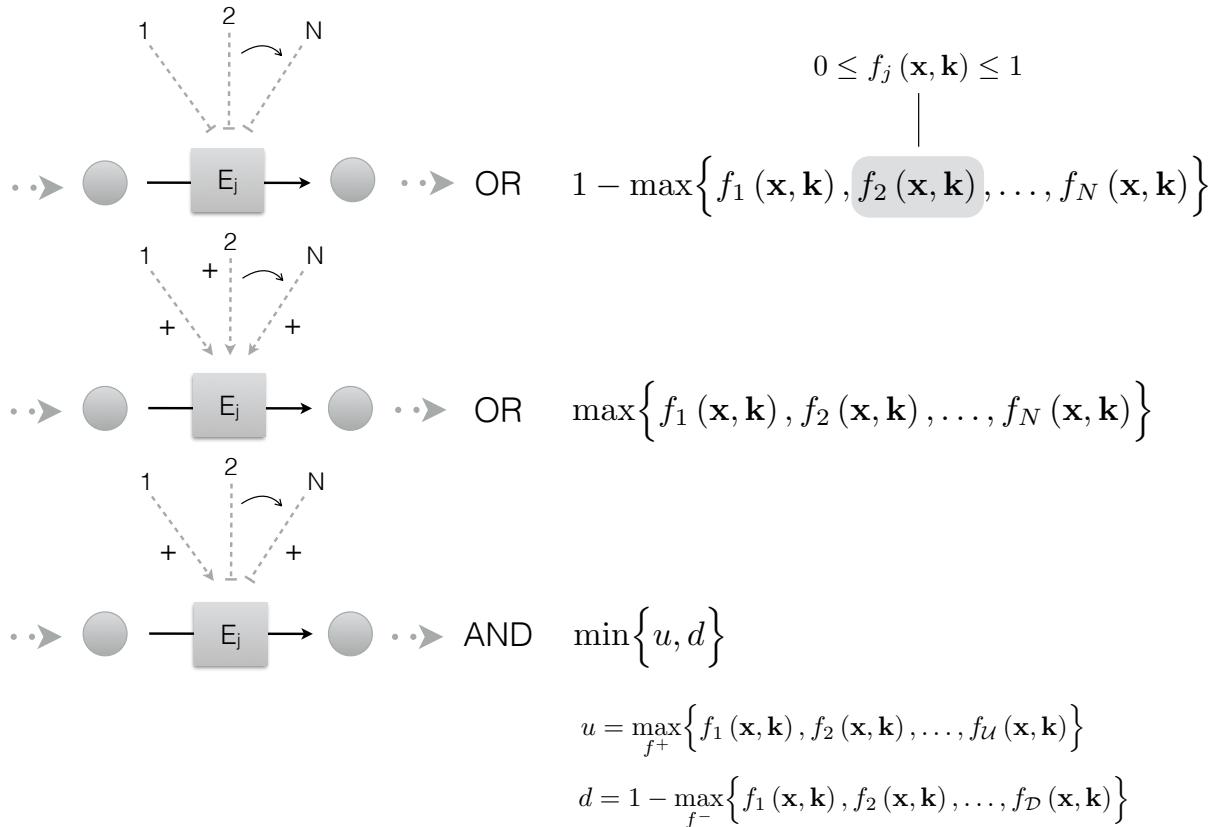


Fig. 2: Schematic of the rule based allosteric enzyme activity control laws.

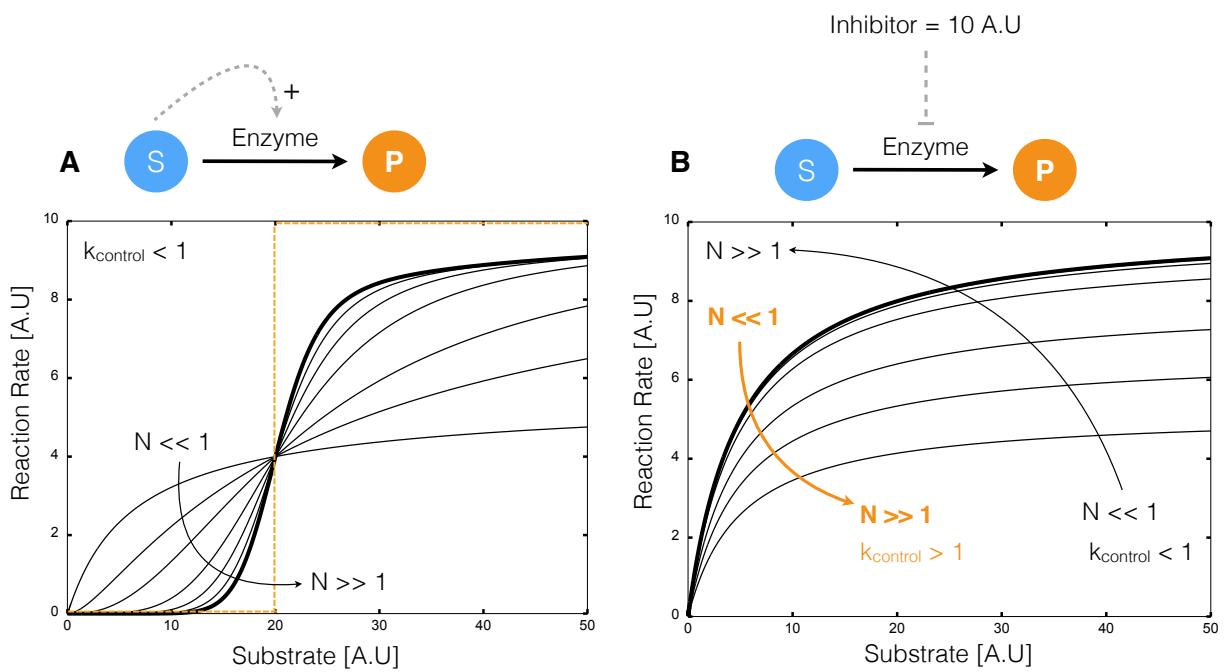


Fig. 3: Kinetics of simple transformations in the presence of activation and inhibition. A: The conversion of substrate S to product P by enzyme E was activated by S . B: The conversion of substrate S to product P by enzyme E was inhibited by inhibitor I .

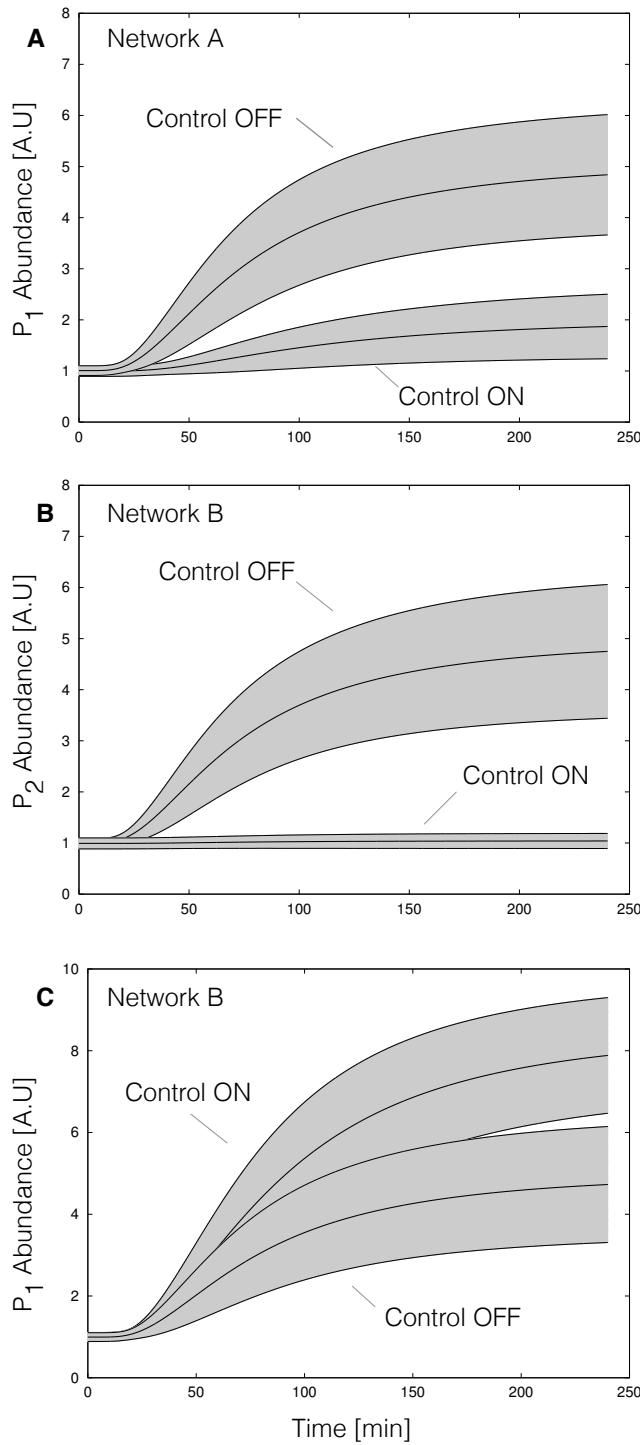


Fig. 4: On/off control simulations for network A and network B for an ensemble of kinetic parameter sets versus time. For each case, $N = 100$ simulations were conducted using kinetic and initial conditions randomly generated from a hypothetical true parameter set. The gray area represents \pm one standard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

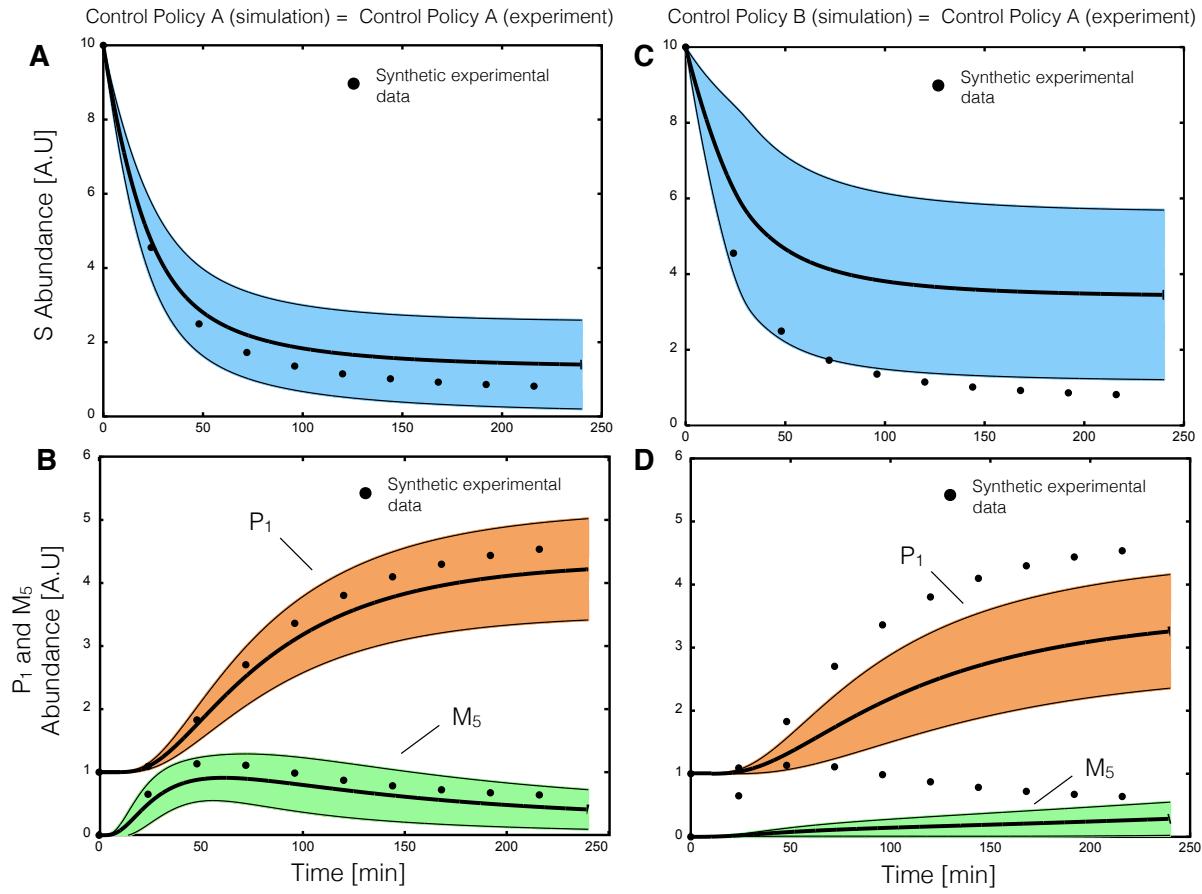


Fig. 5: Parameter estimation from synthetic data for the same and mismatched allosteric control logic.

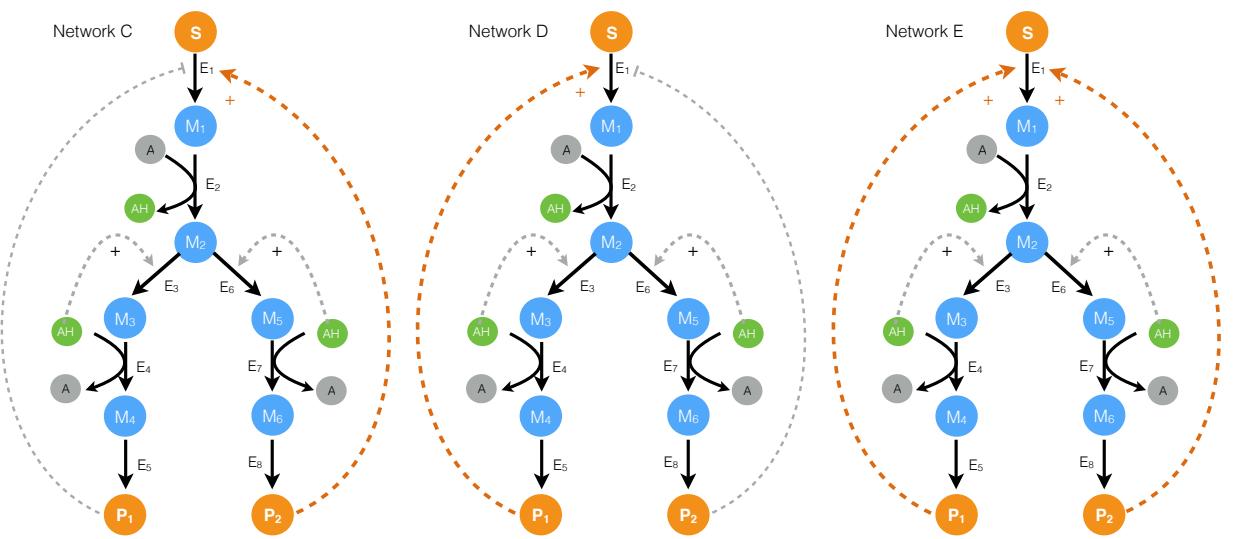


Fig. 6: Schematic of the alternative allosteric control programs used in the structural particle swarm computation. Each network had the same enzymatic connectivity, initial conditions and kinetic parameters, but alternative feedback control structures for the first enzyme in the pathway.

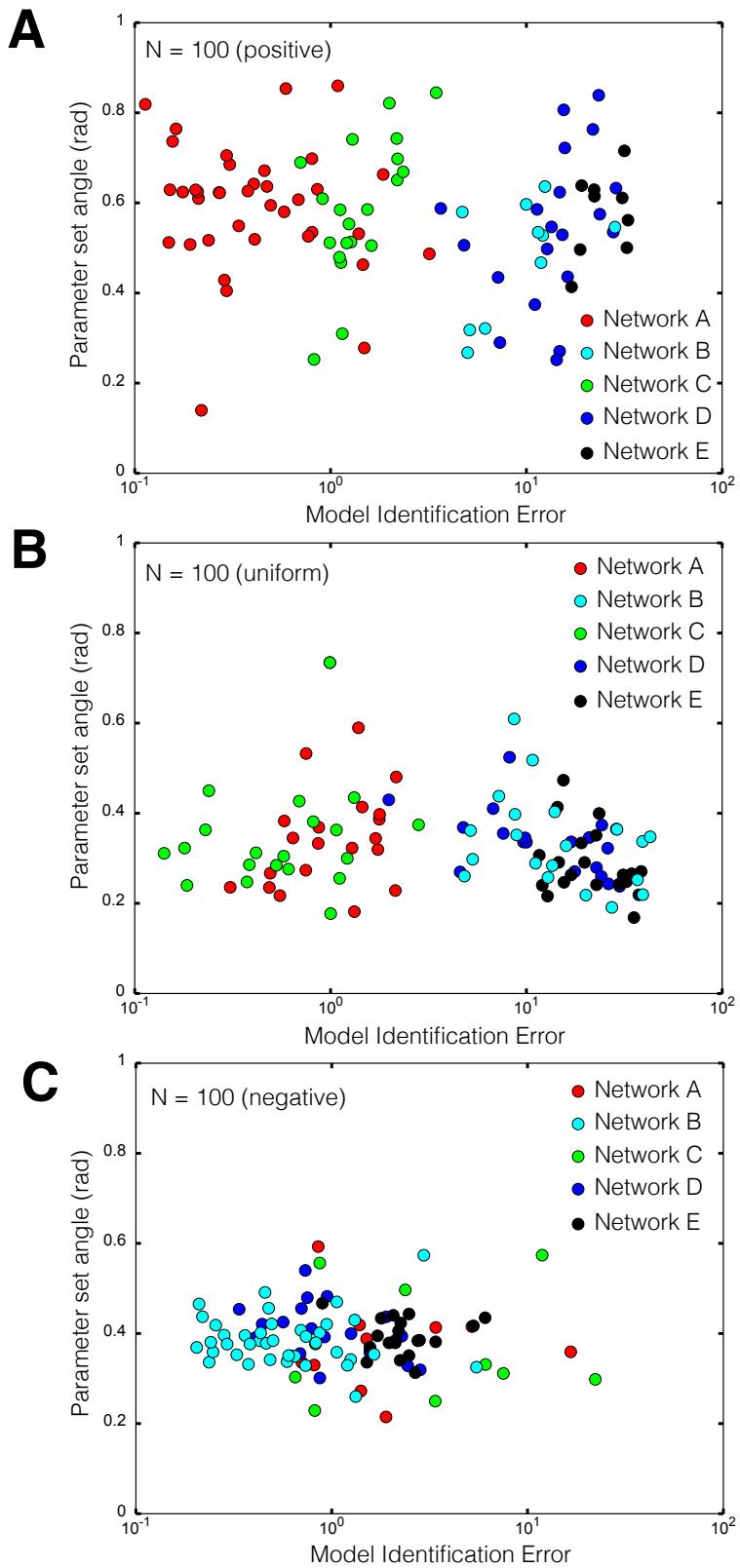


Fig. 7: Combined control structure and kinetic parameter search results. Think about this ...