

# 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](mailto: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

## 1 Introduction

2 Mathematical modeling has long contributed to our understanding of metabolism. Decades  
3 before the genomics revolution, mechanistically, structured metabolic models arose from  
4 the desire to predict microbial phenotypes resulting from changes in intracellular or extra-  
5 cellular states [1]. The single cell *E. coli* models of Shuler and coworkers pioneered the  
6 construction of large-scale, dynamic metabolic models that incorporated multiple, regu-  
7 lated catabolic and anabolic pathways constrained by experimentally determined kinetic  
8 parameters [2]. Shuler and coworkers generated many single cell kinetic models, includ-  
9 ing single cell models of eukaryotes [3, 4], minimal cell architectures [5], as well as DNA  
10 sequence based whole-cell models of *E. coli* [6]. Conversely, highly abstracted kinetic  
11 frameworks, such as the cybernetic framework, represented a paradigm shift, viewing  
12 cells as growth-optimizing strategists [7]. Cybernetic models have been highly successful  
13 at predicting metabolic choice behavior, e.g., diauxie behavior [8], steady-state multiplicity  
14 [9], as well as the cellular response to metabolic engineering modifications [10]. Unfortu-  
15 nately, cybernetic models also suffer from an identifiability challenge, as both the kinetic  
16 parameters and an abstracted model of cellular objectives must be estimated simultane-  
17 ously.

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

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

37 Cell-free systems offer many advantages for the study, manipulation and modeling of  
38 metabolism compared to *in vivo* processes. Central amongst these advantages is direct  
39 access to metabolites and the microbial biosynthetic machinery without the interference of  
40 a cell wall. This allows us to control as well as interrogate the chemical environment while  
41 the biosynthetic machinery is operating, potentially at a fine time resolution. Second,  
42 cell-free systems also allow us to study biological processes without the complications  
43 associated with cell growth. Cell-free protein synthesis (CFPS) systems are arguably the  
44 most prominent examples of cell-free systems used today [21]. However, CFPS is not  
45 new; CFPS in crude *E. coli* extracts has been used since the 1960s to explore funda-  
46 mentally important biological mechanisms [22, 23]. Today, cell-free systems are used  
47 in a variety of applications ranging from therapeutic protein production [24] to synthetic  
48 biology [25]. Interestingly, many of the challenges confronting genome-scale kinetic mod-  
49 eling can potentially be overcome in a cell-free system. For example, there is no complex  
50 transcriptional regulation to consider, transient metabolic measurements are easier to  
51 obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds  
52 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.

## 79 Results

### 80 Formulation and properties of effective cell-free metabolic models.

81 We developed two proof-of-concept metabolic networks to investigate the features of our effective bio-  
82 chemical network modeling approach (Fig. 1). In both examples, substrate S was con-  
83 verted to the end products  $P_1$  and  $P_2$  through a series of enzymatically catalyzed reac-  
84 tions, including a branch point at hypothetical metabolite  $M_2$ . Several of these reactions  
85 involved cofactor dependence (AH or A), and various allosteric regulatory mechanisms  
86 modified the activity of pathway enzymes. Network A included feedback inhibition of the  
87 initial pathway enzyme ( $E_1$ ) by pathway end products  $P_1$  and  $P_2$  (Fig. 1A). On the other  
88 hand, network B involved feedback inhibition of  $E_1$  by  $P_2$  and  $E_6$  by  $P_1$  (Fig. 1B). In both  
89 networks, branch point enzymes  $E_3$  and  $E_6$  were subject to feed-forward activation by  
90 reduced cofactor AH. Lastly, it is known experimentally that cell-free systems have a finite  
91 operational lifespan. Loss of biosynthetic capability could be a function of many factors,  
92 e.g., cofactor or metabolite limitations. We modeled the loss of biosynthetic capability as  
93 a non-specific first-order decay of enzyme activity.

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

105 unmodeled factors. When an enzyme was potentially sensitive to more than one regula-  
106 tory input, logical integration rules were used to select which regulatory transfer function  
107 influenced enzyme activity at any given time. Thus, our test networks involved important  
108 features such as cofactor recycling, enzyme activity and metabolite dynamics, as well as  
109 multiple overlapping allosteric regulatory mechanisms.

110 The rule-based regulatory strategy approximated the behavior of classical allosteric  
111 activation and inhibition mechanisms (Fig. 3). We considered the enzyme catalyzed con-  
112 version of substrate S to a product P, where the overall reaction rate was modeled as the  
113 product of a Michaelis-Menten term and an effective allosteric control variable reflecting  
114 the particular regulatory interaction. We first explored feed-forward substrate activation  
115 of enzyme activity (for both positive and negative cooperativity). Consistent with clas-  
116 sical data, the rule-based strategy predicted a sigmoidal relationship between substrate  
117 abundance and reaction rate as a function of the cooperativity parameter (Fig. 3A). For  
118 cooperativity parameters less than unity, increased substrate abundance *decreased* the  
119 reaction rate. This was consistent with the idea that substrate binding *decreased* at reg-  
120 ulatory sites, which negatively impacted substrate binding at the active site. On the other  
121 hand, as the cooperativity parameter increased past unity, the rate of conversion of sub-  
122 strate S to product P by enzyme  $E$  approached a step function. In the presence of an  
123 inhibitor, the rule-based strategy predicted non-competitive like behavior as a function of  
124 the cooperativity parameter (Fig. 3B). When the control gain parameter,  $\kappa_{ij}$  in Eqn. (10),  
125 was greater than unity, the inhibitory force was directly proportional to the cooperativity  
126 parameter,  $\eta$  in Eqn. (10). Thus, as the cooperativity parameter increased, the maximum  
127 reaction rate decreased (Fig. 3B). Interestingly, our rule-based approach was unable to  
128 directly simulate competitive inhibition of enzyme activity. Taken together, the rule-based  
129 strategy captured classical regulatory patterns for both enzyme activation and inhibition.  
130 Thus, we are able to model complex kinetic phenomena such as ultrasensitivity, despite

131 an effective description of reaction kinetics.

132 End product yield was controlled by feedback inhibition, while product selectivity was  
133 controlled by branch point enzyme inhibition (Fig. 4). A critical test of our modeling  
134 approach was to simulate networks with known behavior. If we cannot reproduce the ex-  
135 pected behavior of simple networks, then our effective modeling strategy, and particularly  
136 the rule-based approximation of allosteric regulation, will not be feasible for genome-scale  
137 cell-free problems. We considered two cases, control ON/OFF, for each network config-  
138 uration. Each of these cases had identical kinetic parameters and initial conditions; the  
139 *only* differences between the cases were the allosteric regulation rules and the control  
140 parameters associated with these rules. As expected, end product accumulation was  
141 larger for network A when the control was OFF (no feedback inhibition of  $E_1$  by  $P_1$  and  
142  $P_2$ ), as compared to the ON case (Fig. 4A). We found this behavior was robust to the  
143 choice of underlying kinetic parameters, as we observed that same qualitative response  
144 across an ensemble of randomized parameter sets ( $N = 100$ ), for fixed control param-  
145 eters. The control ON/OFF response of network B was more subtle. In the OFF case,  
146 the behavior was qualitatively similar to network A. However, for the ON case, flux was  
147 diverted away from  $P_2$  formation by feedback inhibition of  $E_6$  activity at the  $M_2$  branch  
148 point by  $P_1$  (Fig. 4B). Lower  $E_6$  activity at the  $M_2$  branch point allowed more flux toward  $P_1$   
149 formation, hence the yield of  $P_1$  also increased (Fig. 4C). Again, the control ON/OFF be-  
150 havior of network B was robust to changes in kinetic parameters, as the same qualitative  
151 trend was conserved across an ensemble of randomized parameters ( $N = 100$ ), for fixed  
152 control parameters. Taken together, these simulations suggested that the rule-based al-  
153 losteric control concept could robustly capture expected feedback behavior for networks  
154 with uncertain kinetic parameters.

155 **Estimating parameters and effective allosteric regulatory structures.** A critical chal-  
156 lenge for any dynamic model is the estimation of kinetic parameters. For metabolic pro-

157 cesses, there is also the added challenge of identifying the regulation and control struc-  
158 tures that manage metabolism. Of course, these issues are not independent; any descrip-  
159 tion of enzyme activity regulation will be a function of system state, which in turn depends  
160 upon the kinetic parameters. For cell-free systems, regulated gene expression has been  
161 removed, however, enzyme activity regulation is still operational. We explored this link-  
162 age by estimating model parameters from synthetic data using both network structures.  
163 We generated noise corrupted synthetic measurements of the substrate S, intermediate  
164  $M_5$  and end product  $P_1$  approximately every 20 min using network A. We then generated  
165 an ensemble of model parameter estimates by minimizing the difference between model  
166 simulations and the synthetic data using particle swarm optimization (PSO), starting from  
167 random initial parameter guesses. The estimation of kinetic parameters was sensitive to  
168 the choice of regulatory structure (Fig. 5). PSO identified an ensemble of parameters that  
169 bracketed the mean of the synthetic measurements in less than 1000 iterations when the  
170 control structure was correct (Fig. 5A and B). However, with control mismatch (network  
171 B simulated with network A parameters), model simulations were not consistent with the  
172 synthetic data (Fig. 5C and D). Taken together, these results suggested that we could  
173 perhaps simultaneously estimate both parameters and network control architectures, as  
174 incorrect control structures would be manifest as poor model fits.

175 We modified our particle swarm identification strategy to simultaneously search over  
176 both kinetic parameters and putative control structures. In addition to our initial networks,  
177 we constructed three additional presumptive network models, each with the same enzy-  
178 matic connectivity but different allosteric regulation of the pathway enzymes (Fig. 6). We  
179 then initialized a population of particles, each with one of the five potential regulatory pro-  
180 grams and randomized kinetic parameters. Thus, we generated an initial population of  
181 particles that had *both* different kinetic parameters as well as different control structures.  
182 We biased the distribution of the particle population according to our *a priori* belief of the

183 correct regulatory program. To this end, we considered three different priors, a uniform  
184 distribution where each putative regulatory structure represented 20% of the population  
185 and two mixed distributions that were either positively or negatively biased towards the  
186 correct structure (network A). In both the positively biased and uniform cases the PSO  
187 clearly differentiated between the true or closely related structures and those that were  
188 materially different (Fig. 7). As expected, the positively biased population (40% of the  
189 initial particle population seeded with network A) gave the best results, where the correct  
190 structure was preferentially identified (Fig. 7A). On the other hand, when given a uniform  
191 distribution, the PSO approach identified a combination of network A and network C as  
192 the most likely control structures (Fig. 7B). Network A and C differ by the regulatory con-  
193 nection between the end product  $P_2$  and enzyme  $E_1$ ; in network A, end product  $P_2$  was  
194 assumed to inhibit  $E_1$ , while in network C, end product  $P_2$  activated  $E_1$ . Lastly, when the  
195 initial population was biased towards incorrect structures (initial population seeded with  
196 90% incorrect structures), the particle swarm *misidentified* the correct allosteric structure  
197 (Fig. 7C). Interestingly, while each particle swarm identified parameter sets that minimized  
198 the simulation error, the estimated parameter values were not necessarily similar to the  
199 true parameters. The angle between the estimated and true parameters was not consis-  
200 tently small across the swarms (identical parameters would give an angle of zero). This  
201 suggested that our particle swarm approach identified a *sloppy* ensemble, i.e., parame-  
202 ter estimates that were individually incorrect but collectively exhibited the correct model  
203 behavior.

204 We calculated control program output and scaled metabolic flux for the positively, uni-  
205 formly and negatively biased particle swarms (Fig. 8). Network A and network C models  
206 from the positively (Fig. 8A) and uniformly (Fig. 8B) biased particle swarms showed sim-  
207 ilar operational patterns, despite differences in kinetic parameters and control structures.  
208 While models from the negatively biased population had error values similar to the correct

209 structures in the previous swarms, they have different flux and control profiles (Fig. 8C).  
210 In all cases, regardless of network configuration or parameter values, the rate of enzyme  
211 decay was small compared to the other fluxes, and all networks had qualitatively similar  
212 trends for  $E_3$  and  $E_6$  control. Moreover, consistent with the correct model structure, pro-  
213 duction of end product  $P_1$  was the preferred branch for all model configurations. However,  
214 there was variability in  $P_2$  production flux across the population of models, especially for  
215 the uniform swarm when compared with the other cases. High  $P_1$  branch flux resulted  
216 in end product inhibition of  $E_1$  in both network A and network C, however in network D  
217 and E, high  $P_1$  flux induced  $E_1$  activation. These trends were manifested in different flux  
218 profiles, where the negatively biased population appeared more uniform across the pop-  
219 ulation compared with the other swarms, and had higher  $E_1$  specific activity. Interestingly,  
220 the behavior of network A and network C highlighted an artifact of our integration rule;  
221 both a positive or negative feedback connection from  $P_2$  to  $E_1$  were ignored because the  
222  $P_1$  inhibition of  $E_1$  was dominate. Thus, while theoretically distinct, network A and net-  
223 work C appeared operationally to the PSO algorithm to be that same network. On the  
224 other hand, networks B, D and E showed distinct behavior that was not consistent with  
225 the true network. These architectures exhibited either limited inhibition (network B) or  
226 activation (network D and E) of  $E_1$  activity, resulting in significantly different metabolic  
227 flux profiles. However, the PSO was able to find low error parameter solutions, despite  
228 the mismatch in the control structures (error values similar, but not better than the best  
229 network A and network C estimates). Taken together, these results suggested that a  
230 uniform sampling approach could potentially yield an unbiased estimate of both kinetic  
231 parameters and control structures. However, the negatively biased particle swarm results  
232 illustrated a potential shortcoming of the approach, namely convergence to a local error  
233 minimum despite a significantly incorrect control structure. This suggested that estimated  
234 model structures will need to be further evaluated, for example by generating falsifiable

235 experimental designs which could distinguish between low error solutions.

## 236 Discussion

237 In this study, we presented an effective kinetic modeling strategy to dynamically simu-  
238 late cell-free biochemical networks. Our proposed strategy integrated traditional kinetic  
239 modeling with an effective rules based approach to dynamically describe metabolic reg-  
240 ulation and control. We tested this approach by developing kinetic models of hypotheti-  
241 cal cell-free metabolic networks. In particular, we tested whether our effective modeling  
242 approach could describe classically expected behavior, and second whether we could si-  
243 multaneously estimate kinetic parameters and regulatory connectivity, in the absence of  
244 specific mechanistic knowledge, from synthetic experimental data. Toward these ques-  
245 tions, we explored five hypothetical cell-free networks. In each network, a substrate S  
246 was converted to the end products  $P_1$  and  $P_2$  through a series of enzymatically catalyzed  
247 reactions, including a branch point at a hypothetical metabolite  $M_2$ . Each network also  
248 included the same cofactors and cofactor recycle architecture. However, while all five  
249 networks shared the same enzymatic connectivity, each had different allosteric regulatory  
250 connectivity. We found that simple effective rules, when integrated with traditional enzyme  
251 kinetic expressions, could capture complex allosteric patterns such as ultrasensitivity, or  
252 non-competitive inhibition in the absence of specific mechanistic information. Moreover,  
253 when integrated into network models, these rules captured classical regulatory patterns  
254 such as product-induced feedback inhibition. Lastly, we simultaneously estimated kinetic  
255 parameters and discriminated between competing regulatory structures, using synthetic  
256 data in combination with a modified particle swarm approach. If we considered all putative  
257 regulatory architectures to be equally likely, we were able to estimate a *sloppy* ensemble  
258 of models with the correct architecture and kinetic parameters.

259 The proposed modeling strategy shares features with other popular techniques, but  
260 has also has several key differences. At its core, our effective modeling approach is sim-  
261 ilar to regulatory constraints based methods, and to the cybernetic modeling paradigm

262 developed by Ramkrishna and colleagues. Covert, Palsson and coworkers drastically im-  
263 proved the predictability of constraints based approaches by integrating boolean rules into  
264 the calculation of metabolic fluxes [27]. If the regulated intracellular flux problem is cou-  
265 pled with time-dependent extracellular balances, these models can predict complex be-  
266 havior such as diauxie growth or the switch between aerobic and anaerobic metabolism.  
267 Another important features of this approach is that it scales with biological complexity.  
268 For example, Covert *et al.*, showed that a genome scale model of *E. coli* augmented  
269 with a boolean rule layer, correctly predicted approximately 80% of the outcomes of a  
270 high-throughput growth phenotyping experiment in *E. coli*. Further, they showed that they  
271 could learn new biology by iteratively refining the model and its associated rules [28].  
272 However, while regulated flux balance analysis is a powerful technique, it does not easily  
273 allow the calculation of time-resolved metabolite abundance. Additionally, the boolean  
274 rules which populate the regulatory layer are limited to ON/OFF decisions; for qualitative  
275 predictions of gene expression this is a reasonable limitation. However, boolean rules will  
276 likely be less effective at capturing dynamic allosteric regulation in a cell free metabolic  
277 system. On the other hand, the strength of cybernetic models is the integration of optimal  
278 metabolic control heuristics with traditional kinetic pathway modeling. Cybernetic models  
279 are highly predictive; they have successfully predicted mutant behavior from limited wild-  
280 type data [10, 29, 30], steady-state multiplicity [9], strain specific metabolic function [31]  
281 and have been used in bioprocess control applications [32]. However, cybernetic control  
282 heuristics are not mechanistic, instead they are the output of an optimal decision with re-  
283 spect to a set of hypothetical physiological objectives. Thus, they are abstractions which  
284 are difficult to translate into a specific biological mechanism. Our approach addresses the  
285 shortcomings of both regulatory constraints based models and cybernetic models. First,  
286 similar to cybernetic models, the core of our approach is a kinetic model. Thus, we are  
287 able to directly calculate the time evolution of metabolism, for example the dynamic abun-

288 dance of network metabolites. Second, similar to regulatory flux balance analysis, our  
289 control laws describe specific mechanistic motifs such a activation or inhibition of enzyme  
290 activity. However, our rules are continuous, thus they potentially allow a finer grained  
291 description metabolic regulation and control mechanisms. Lastly, we can naturally incor-  
292 porate non modeled factors and categorical factors or factor combinations into our control  
293 law formulations.

294 There are several critical questions that should be explored following this proof of  
295 concept study. It is unclear how parameter identification will scale to genome scale net-  
296 works, and second it is unclear how we will identify allosteric connectivity at a genome  
297 scale. The enzymatic connectivity for genome scale cell free networks can easily be es-  
298 tablished by stripping away the growth and cell wall machinery from whole cell genome  
299 reconstructions. Then metabolic fluxes can transformed into kinetic expressions using  
300 heuristics such multiple saturation kinetics, which are then modified by our rule based  
301 control variables. This leaves a large number of unknown kinetic constants that must be  
302 estimated from time-resolved metabolite measurements. We showed that particle swarm  
303 optimization quickly identified an ensemble of model parameters, at least for proof of  
304 concept metabolic networks using synthetic data. This suggested that we can expect  
305 reasonable model predictions, despite only partial parameter knowledge, as network size  
306 grows if we have properly designed experiments. Brown and Sethna showed in a model  
307 of signal transduction that good predictions were possible despite only order of magni-  
308 tude estimates of parameter values [33]. Sethna and coworkers later showed that model  
309 performance is often controlled by only a few parameter combinations, a characteristic  
310 seemingly universal to multi-parameter models referred to as *sloppiness* [34]. We have  
311 also demonstrated *sloppy* behavior in a wide variety of signal transduction processes  
312 [35–40]. Thus, given our previous experience with models with hundreds of unknown pa-  
313 rameters, we expect parameter estimation to be a manageable challenge. On the other

314 hand, a critical challenge will be the estimation of allosteric connectivity at a genome  
315 scale. The regulation of glycolytic enzymes, such as phosphofructokinase I, has been  
316 studied for many years [41, 42]. The allosteric regulation of metabolic enzymes can also  
317 be established from organism specific databases, such as EcoCyc [43] or more general  
318 allosteric databases such as the AlloSteric Database [44]. However, for those enzymes  
319 that have not been well studied, we will need to infer allosteric interactions from exper-  
320 imental data. In general, the reverse engineering of regulatory network structure from  
321 data is a very difficult problem. There are many different approaches from the reverse  
322 engineering of gene regulatory networks that perhaps could be adopted to this problem,  
323 however this remains an open question.

## 324 **Materials and Methods**

325 **Formulation and solution of the model equations.** We used ordinary differential equa-  
 326 tions (ODEs) to model the time evolution of metabolite ( $x_i$ ) and scaled enzyme abundance  
 327 ( $\epsilon_i$ ) in hypothetical cell-free metabolic networks:

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

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

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

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

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

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

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

340 where  $k_j^{max}$  denotes the maximum rate for reaction  $j$ ,  $\epsilon_i$  denotes the scaled enzyme ac-

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

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

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

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

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

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

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

353 where:

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

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

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

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

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

358 where  $\mathcal{Z}_j$  denotes the abundance of the  $j$  factor (e.g., metabolite abundance), and  $\kappa_{ij}$  and  
 359  $\eta$  are control parameters. The  $\kappa_{ij}$  parameter was species gain parameter, while  $\eta$  was a  
 360 cooperativity parameter (similar to a Hill coefficient). The model equations were encoded  
 361 using the Octave programming language and solved using the LSODE routine in Octave  
 362 [45].

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

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

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

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

372 We minimized the model residual using particle swarm optimization (PSO) [46]. PSO  
 373 uses a *swarming* metaheuristic to explore parameter spaces. A strength of PSO is its abil-  
 374 ity to find the global minimum, even in the presence of potentially many local minima, by  
 375 communicating the local error landscape experienced by each particle collectively to the  
 376 swarm. Thus, PSO acts both as a local and a global search algorithm. For each iteration,

377 particles in the swarm compute their local error by evaluating the model equations using  
 378 their specific parameter vector realization. From each of these local points, a globally best  
 379 error is identified. Both the local and global error are then used to update the parameter  
 380 estimates of each particle using the rules:

$$\Delta_i = \theta_1 \Delta_i + \theta_2 r_1 (\mathcal{L}_i - \mathbf{k}_i) + \theta_3 r_2 (\mathcal{G} - \mathbf{k}_i) \quad (12)$$

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

381 where  $(\theta_1, \theta_2, \theta_3)$  are adjustable parameters,  $\mathcal{L}_i$  denotes local best solution found by par-  
 382 ticle  $i$ , and  $\mathcal{G}$  denotes the best solution found over the entire population of particles.  
 383 The quantities  $r_1$  and  $r_2$  denote uniform random vectors with the same dimension as  
 384 the number of unknown model parameters ( $\mathcal{K} \times 1$ ). In thus study, we used  $(\theta_1, \theta_2, \theta_3) =$   
 385  $(1.0, 0.05564, 0.02886)$ . The quality of parameter estimates was measured using two crite-  
 386 ria, goodness of fit (model residual) and angle between the estimated parameter vector  
 387  $\mathbf{k}_j$  and the true parameter set  $\mathbf{k}^*$ :

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

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

391 We modified our PSO implementation to simultaneously search over kinetic parame-  
 392 ters and putative model control structures. In the combined case, each particle potentially  
 393 carried a different model realization in addition to a different kinetic parameter vector. We  
 394 kept the update rules the same (along with the update parameters). Thus, each parti-  
 395 cle competed on the basis of goodness of fit, which allowed different model structures

396 to contribute to the overall behavior of the swarm. We considered five possible model  
397 structures (A through E), where network A was the correct formulation (used to generate  
398 the synthetic data). We considered a population  $N = 100$  particles, where each particle  
399 in the swarm was assigned a model structure, and a random parameter vector. The PSO  
400 algorithm, model equations, and the objective function were encoded and solved in the  
401 Octave programming language [45].

## 402 **Acknowledgements**

403 This study was supported by an award from the National Science Foundation (MCB  
404 #1411715) and the Army Research Office (ARO #59155-LS).

## References

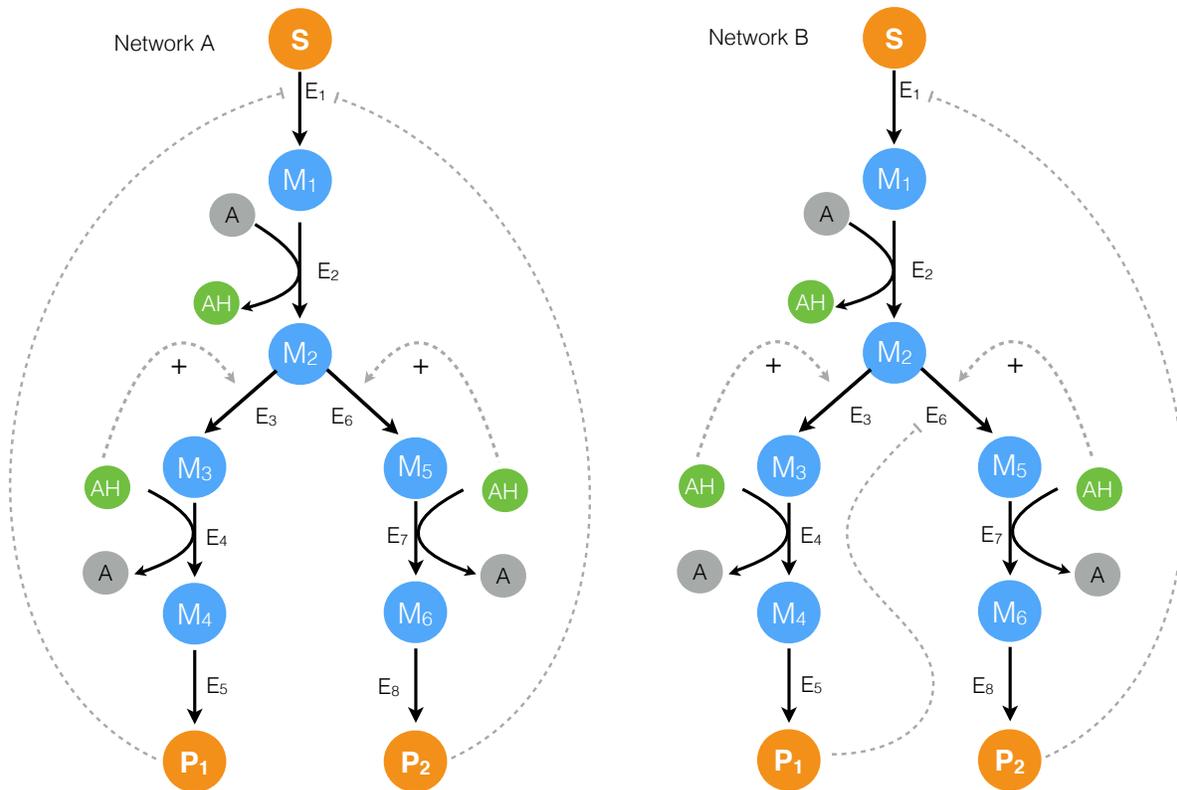
1. Fredrickson AG (1976) Formulation of structured growth models. *Biotechnol Bioeng* 18: 1481-6.
2. Domach MM, Leung SK, Cahn RE, Cocks GG, Shuler ML (1984) Computer model for glucose-limited growth of a single cell of *Escherichia coli* b/r-a. *Biotechnol Bioeng* 26: 203-16.
3. Steinmeyer D, Shuler 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) Exact 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-

- 431 phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10:  
432 291-305.
- 433 12. Edwards JS, Palsson BO (2000) The escherichia coli mg1655 in silico metabolic  
434 genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S*  
435 *A* 97: 5528-33.
- 436 13. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-  
437 scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for  
438 1260 orfs and thermodynamic information. *Mol Syst Biol* 3: 121.
- 439 14. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale re-  
440 construction of metabolic network in bacillus subtilis based on high-throughput phe-  
441 notyping and gene essentiality data. *J Biol Chem* 282: 28791-9.
- 442 15. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of bio-  
443 chemical networks in microorganisms. *Nat Rev Microbiol* 7: 129-43.
- 444 16. Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli k-12 undergoes adaptive  
445 evolution to achieve in silico predicted optimal growth. *Nature* 420: 186-9.
- 446 17. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions  
447 for predicting intracellular fluxes in escherichia coli. *Mol Syst Biol* 3: 119.
- 448 18. Hyduke DR, Lewis NE, Palsson BØ (2013) Analysis of omics data with genome-scale  
449 models of metabolism. *Mol Biosyst* 9: 167-74.
- 450 19. McCloskey D, Palsson BØ, Feist AM (2013) Basic and applied uses of genome-scale  
451 metabolic network reconstructions of escherichia coli. *Mol Syst Biol* 9: 661.
- 452 20. Zomorodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical opti-  
453 mization applications in metabolic networks. *Metab Eng* 14: 672-86.
- 454 21. Jewett MC, Calhoun KA, Voloshin A, Wu JJ, Swartz JR (2008) An integrated cell-  
455 free metabolic platform for protein production and synthetic biology. *Mol Syst Biol* 4:  
456 220.

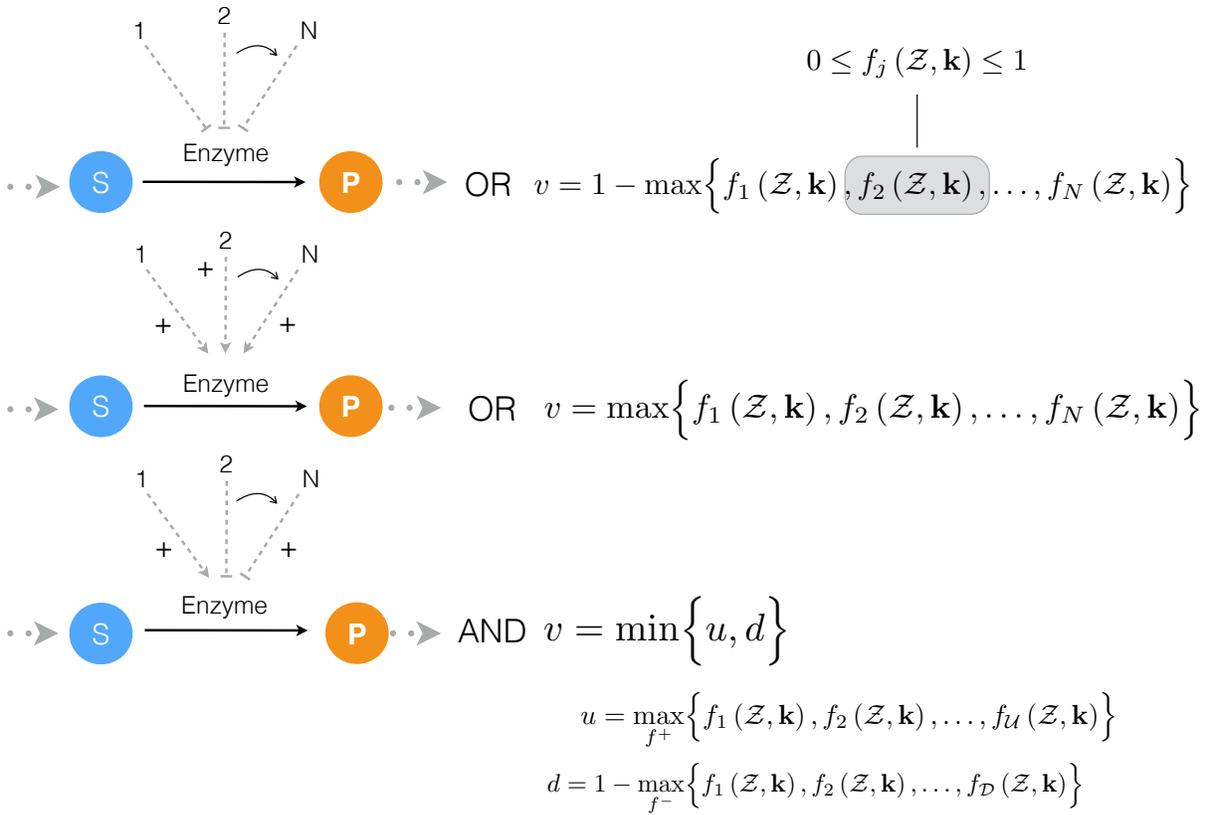
- 457 22. MATTHAEI JH, NIRENBERG MW (1961) Characteristics and stabilization of dnaase-  
458 sensitive protein synthesis in e. coli extracts. Proc Natl Acad Sci U S A 47: 1580-8.
- 459 23. NIRENBERG MW, MATTHAEI JH (1961) The dependence of cell-free protein syn-  
460 thesis in e. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl  
461 Acad Sci U S A 47: 1588-602.
- 462 24. Lu Y, Welsh JP, Swartz JR (2014) Production and stabilization of the trimeric influenza  
463 hemagglutinin stem domain for potentially broadly protective influenza vaccines. Proc  
464 Natl Acad Sci U S A 111: 125-30.
- 465 25. Hodgman CE, Jewett MC (2012) Cell-free synthetic biology: thinking outside the cell.  
466 Metab Eng 14: 261-9.
- 467 26. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Train-  
468 ing signaling pathway maps to biochemical data with constrained fuzzy logic: quan-  
469 titative analysis of liver cell responses to inflammatory stimuli. PLoS Comput Biol 7:  
470 e1001099.
- 471 27. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux  
472 balance models of metabolism. J Theor Biol 213: 73-88.
- 473 28. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-  
474 throughput and computational data elucidates bacterial networks. Nature 429: 92-6.
- 475 29. Varner JD (2000) Large-scale prediction of phenotype: concept. Biotechnol Bioeng  
476 69: 664-78.
- 477 30. Song HS, Ramkrishna D (2012) Prediction of dynamic behavior of mutant strains from  
478 limited wild-type data. Metab Eng 14: 69-80.
- 479 31. Song HS, Ramkrishna D (2011) Cybernetic models based on lumped elementary  
480 modes accurately predict strain-specific metabolic function. Biotechnol Bioeng 108:  
481 127-40.
- 482 32. Gadkar KG, Doyle FJ 3rd, Crowley TJ, Varner JD (2003) Cybernetic model predictive

- 483 control of a continuous bioreactor with cell recycle. *Biotechnol Prog* 19: 1487-97.
- 484 33. Brown KS, Sethna JP (2003) Statistical mechanical approaches to models with many  
485 poorly known parameters. *Phys Rev E Stat Nonlin Soft Matter Phys* 68: 021904.
- 486 34. Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compres-  
487 sion underlies emergent theories and predictive models. *Science* 342: 604-7.
- 488 35. Luan D, Zai M, Varner JD (2007) Computationally derived points of fragility of a human  
489 cascade are consistent with current therapeutic strategies. *PLoS Comput Biol* 3:  
490 e142.
- 491 36. Song SO, Varner J (2009) Modeling and analysis of the molecular basis of pain in  
492 sensory neurons. *PLoS One* 4: e6758.
- 493 37. Tasseff R, Nayak S, Salim S, Kaushik P, Rizvi N, et al. (2010) Analysis of the molecular  
494 networks in androgen dependent and independent prostate cancer revealed fragile  
495 and robust subsystems. *PLoS One* 5: e8864.
- 496 38. Tasseff R, Nayak S, Song SO, Yen A, Varner JD (2011) Modeling and analysis  
497 of retinoic acid induced differentiation of uncommitted precursor cells. *Integr Biol*  
498 (Camb) 3: 578-91.
- 499 39. Nayak S, Siddiqui JK, Varner JD (2011) Modelling and analysis of an ensemble of  
500 eukaryotic translation initiation models. *IET Syst Biol* 5: 2.
- 501 40. Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and  
502 analysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol* 7:  
503 e1002263.
- 504 41. Berg JM, Tymoczko JL, Stryer L (2002) *Biochemistry*. W.H. Freeman.
- 505 42. Peskov K, Goryanin I, Demin O (2008) Kinetic model of phosphofructokinase-1 from  
506 *escherichia coli*. *J Bioinform Comput Biol* 6: 843-67.
- 507 43. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, et al. (2013)  
508 *Ecocyc: fusing model organism databases with systems biology*. *Nucleic Acids Res*

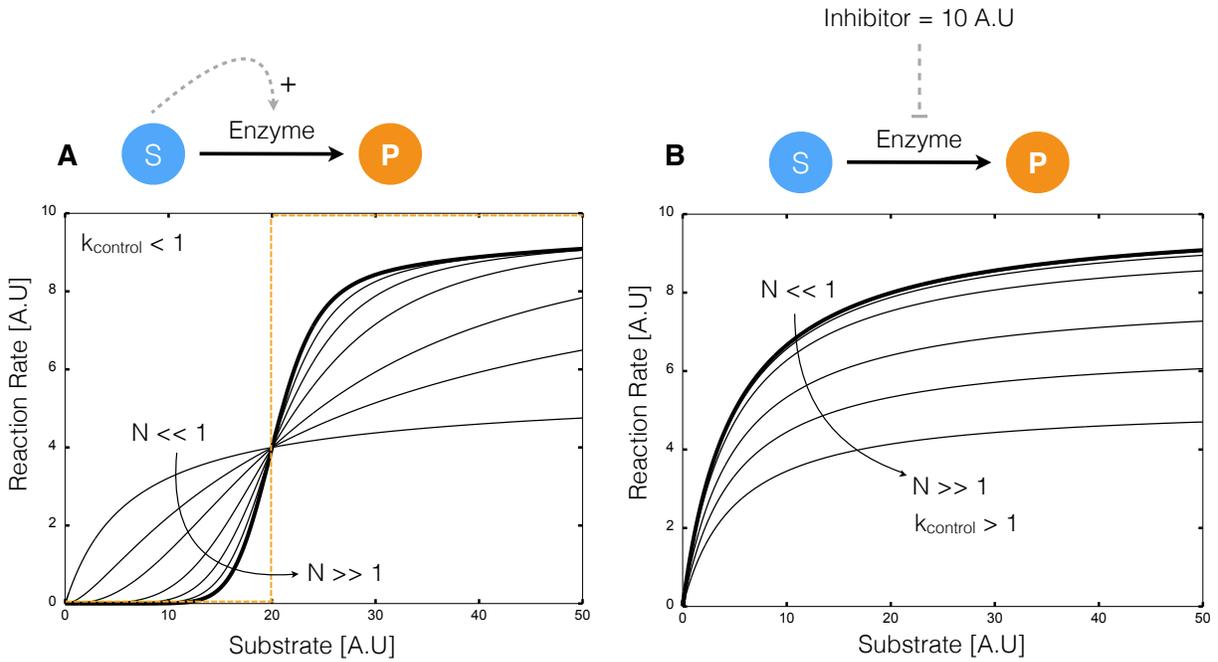
- 509 41: D605-12.
- 510 44. Huang Z, Mou L, Shen Q, Lu S, Li C, et al. (2014) Asd v2.0: updated content and  
511 novel features focusing on allosteric regulation. *Nucleic Acids Res* 42: D510-6.
- 512 45. Octave community (2014). GNU Octave 3.8.1. URL [www.gnu.org/software/octave/](http://www.gnu.org/software/octave/).
- 513 46. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: *Proceedings of the*  
514 *International Conference on Neural Networks*. pp. 1942 - 1948.



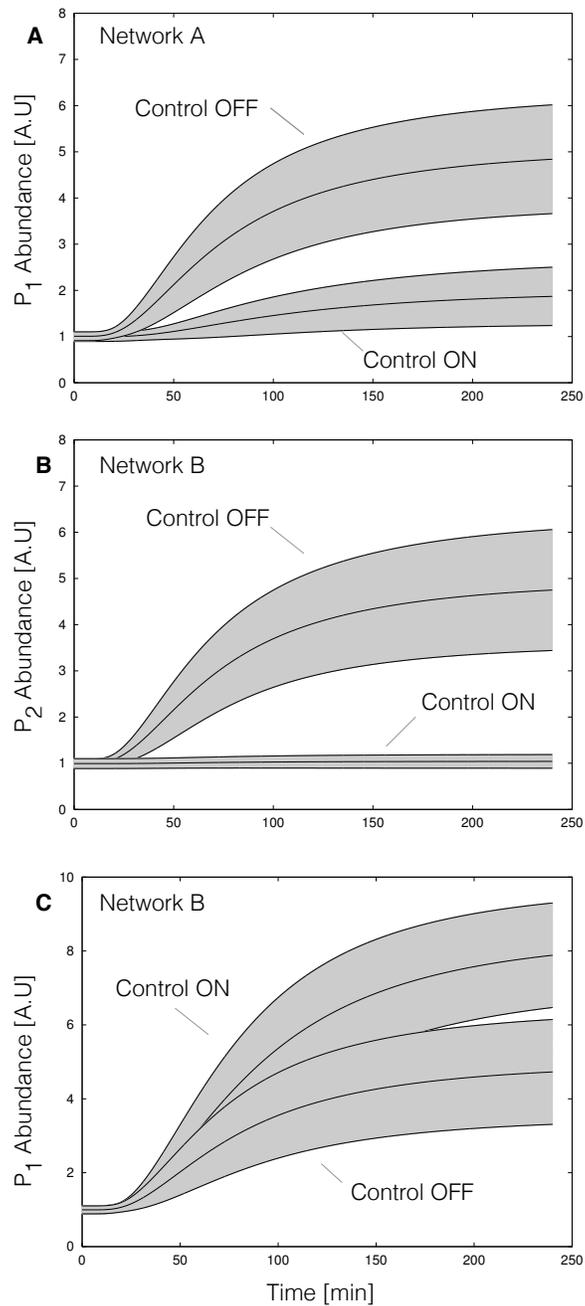
**Fig. 1:** Proof of concept cell-free metabolic networks considered in this study. Substrate  $S$  is converted to products  $P_1$  and  $P_2$  through a series of chemical conversions catalyzed by enzyme(s)  $E_j$ . The activity of the pathway enzymes is subject to both positive and negative allosteric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



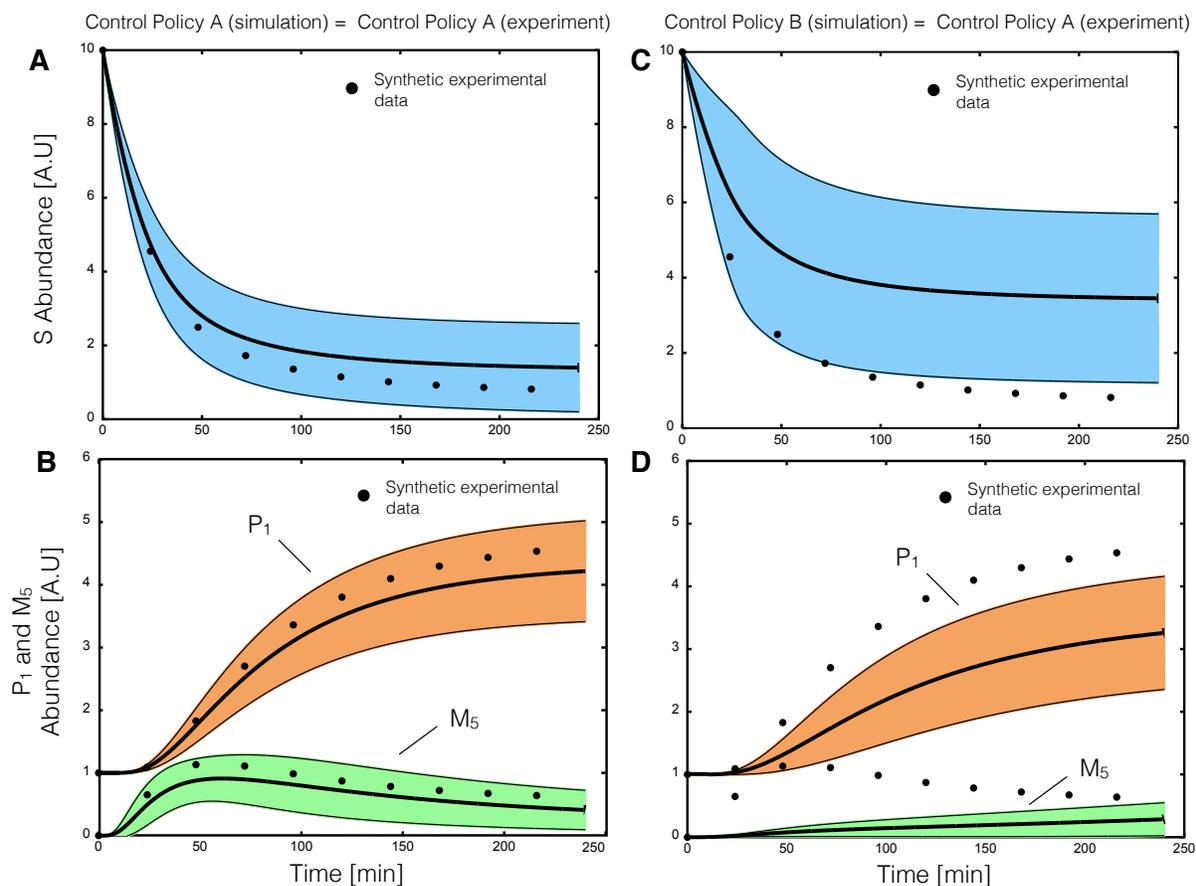
**Fig. 2:** Schematic of rule-based allosteric enzyme activity control laws. Traditional enzyme kinetic expressions, e.g., Michaelis–Menten or multiple saturation kinetics, are multiplied by an enzyme activity control variable  $0 \leq v_j \leq 1$ . Control variables are functions of many possible regulatory 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  $f_1, \dots$  influencing the activity of enzyme  $E_j$ .



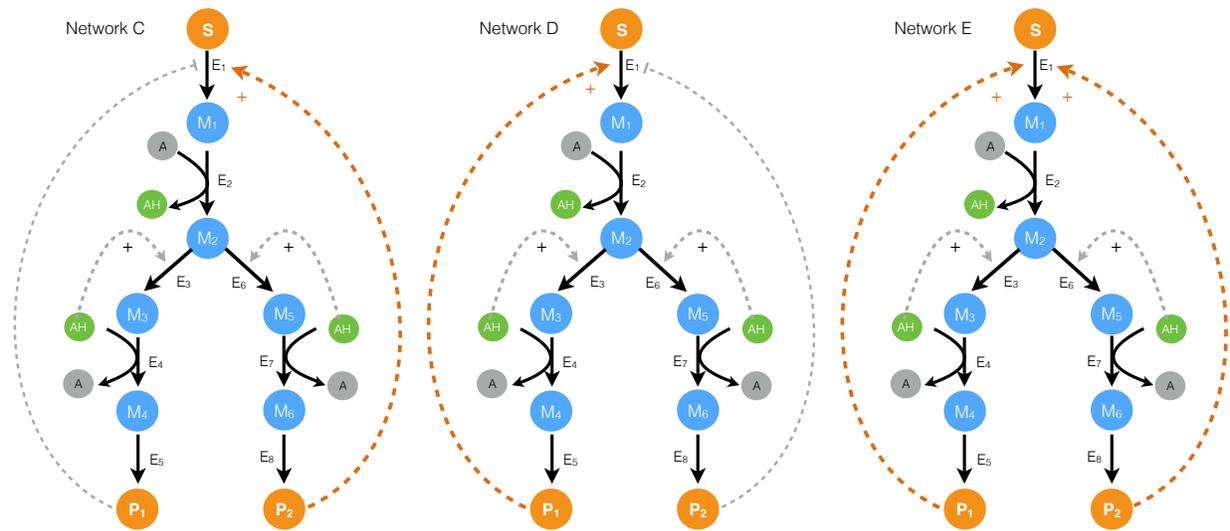
**Fig. 3:** Kinetics of simple transformations in the presence of activation and inhibition. **A:**The conversion of substrate  $S$  to product  $P$  by enzyme  $E$  was activated by  $S$ . For a fixed control gain parameter  $k_{control}$ , the reaction rate approached a step for increasing control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



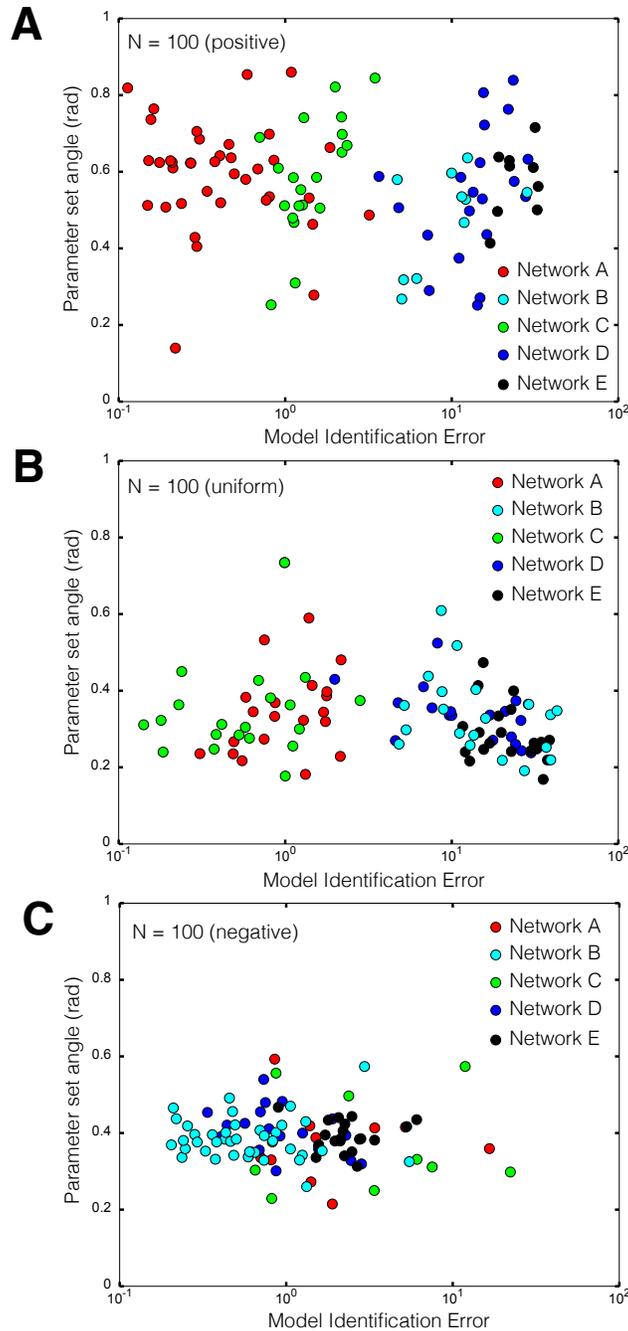
**Fig. 4:** ON/OFF control simulations for network A and network B for an ensemble of kinetic parameter sets versus time ( $N = 100$ ). For each case,  $N = 100$  simulations were conducted using kinetic and initial conditions generated randomly from a hypothetical true parameter set. The gray area represents  $\pm$  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).



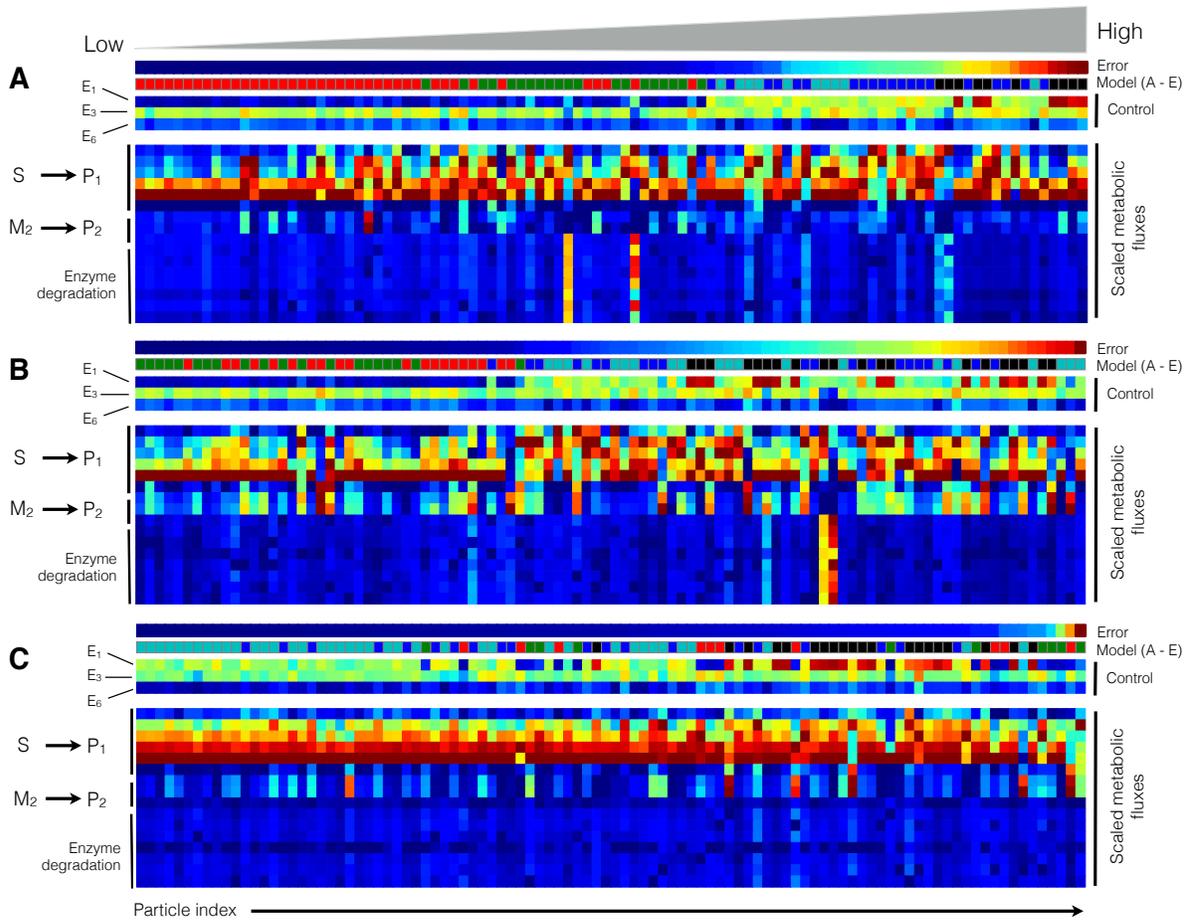
**Fig. 5:** Parameter estimation from synthetic data for the same and mismatched allosteric control logic using particle swarm optimization (PSO). Synthetic experimental data was generated from a hypothetical parameter set using Network A, where substrate  $S$ , end product  $P_1$  and intermediate  $M_5$  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.