

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 starting from an unbiased collection of possible allosteric structures using particle swarm optimization. However, when starting with an initial population that was heavily enriched with incorrect structures, our particle swarm approach could converge to an incorrect structure. 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.

tant 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, traditional, fully structured cybernetic models also suffer from an identifiability chal-
¹⁶ lenge, as both the kinetic parameters and an abstracted model of cellular objectives must
¹⁷ be estimated simultaneously. However, recent cybernetic formulations from Ramkrishna
¹⁸ and colleagues have successfully treated this identifiability challenge through elementary
¹⁹ mode reduction, though the techniques replace detailed biological mechanism with an
²⁰ optimization heuristic [11, 12].

²¹ In the post genomics world, large-scale stoichiometric reconstructions of microbial
²² metabolism popularized by static, constraint-based modeling techniques such as flux bal-
²³ ance analysis (FBA) have become standard tools [13]. Since the first genome-scale sto-
²⁴ chiometric model of *E. coli*, developed by Edwards and Palsson [14], well over 100 organ-
²⁵ isms, including industrially important prokaryotes such as *E. coli* [15] or *B. subtilis* [16],
²⁶ are now available [17]. Stoichiometric models rely on a pseudo-steady-state assump-

27 tion to reduce unidentifiable genome-scale kinetic models to an underdetermined linear
28 algebraic system, which can be solved efficiently even for large systems. Traditionally,
29 stoichiometric models have also neglected explicit descriptions of metabolic regulation
30 and control mechanisms, instead opting to describe the choice of pathways by prescrib-
31 ing an objective function on metabolism. Interestingly, similar to early cybernetic mod-
32 els, the most common metabolic objective function has been the optimization of biomass
33 formation [18], although other metabolic objectives have also been estimated [19]. Re-
34 cent advances in constraint-based modeling have overcome the early shortcomings of
35 the platform, including capturing metabolic regulation and control [20]. Thus, modern
36 constraint-based approaches have proven extremely useful in the discovery of metabolic
37 engineering strategies and represent the state of the art in metabolic modeling [21, 22].
38 However, genome-scale kinetic models of industrial important organisms such as *E. coli*
39 have yet to be constructed.

40 Cell-free systems offer many advantages for the study, manipulation and modeling of
41 metabolism compared to *in vivo* processes. Central amongst these advantages is direct
42 access to metabolites and the microbial biosynthetic machinery without the interference of
43 a cell wall. This allows us to control as well as interrogate the chemical environment while
44 the biosynthetic machinery is operating, potentially at a fine time resolution. Second,
45 cell-free systems also allow us to study biological processes without the complications
46 associated with cell growth. Cell-free protein synthesis (CFPS) systems are arguably the
47 most prominent examples of cell-free systems used today [23]. However, CFPS is not
48 new; CFPS in crude *E. coli* extracts has been used since the 1960s to explore funda-
49 mentally important biological mechanisms [24, 25]. Today, cell-free systems are used
50 in a variety of applications ranging from therapeutic protein production [26] to synthetic
51 biology [27]. Interestingly, many of the challenges confronting genome-scale kinetic mod-
52 eling can potentially be overcome in a cell-free system. For example, there is no complex

53 transcriptional regulation to consider, transient metabolic measurements are easier to
54 obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds
55 several significant advantages for model development, identification and validation. The-
56oretically, genome-scale cell-free kinetic models may be possible for industrially important
57 organisms, such as *E. coli* or *B. subtilis*, if a simple, tractable framework for integrating
58 allosteric regulation with enzyme kinetics can be formulated.

59 In this study, we present an effective biochemical network modeling framework for
60 building dynamic cell-free metabolic models. The key innovation of our approach is the
61 seamless integration of simple effective rules encoding complex regulation with traditional
62 kinetic pathway modeling. This integration allows the description of complex regulatory
63 interactions, such as time-dependent allosteric regulation of enzyme activity, in the ab-
64 sence of specific mechanistic information. The regulatory rules are easy to understand,
65 easy to formulate and do not rely on overarching theoretical abstractions or restrictive as-
66 sumptions. We tested our approach by modeling the time evolution of several hypothetical
67 cell-free metabolic networks. In particular, we tested whether our effective modeling ap-
68 proach could describe classically expected enzyme kinetic behavior, and second whether
69 we could simultaneously estimate kinetic parameters and regulatory connectivity, in the
70 absence of specific mechanistic knowledge, from synthetic experimental data. Toward
71 these questions, we explored five hypothetical cell-free networks. Each network shared
72 the same enzymatic connectivity, but had different allosteric regulatory connectivity. We
73 found that simple effective rules, when integrated with traditional enzyme kinetic expres-
74 sions, captured complex allosteric patterns such as ultrasensitivity or non-competitive in-
75 hibition in the absence of mechanistic information. Second, when integrated into network
76 models, these rules captured classical regulatory patterns such as product-induced feed-
77 back inhibition. Lastly, we showed, at least for the network architectures considered here,
78 that we could simultaneously estimate kinetic parameters and allosteric connectivity from

79 synthetic data starting from an unbiased collection of possible allosteric structures using
80 particle swarm optimization. However, when starting with an initial population that was
81 heavily enriched with incorrect structures, our particle swarm approach could converge
82 to an incorrect structure. While only an initial proof-of-concept, the framework presented
83 here could be an important first step toward genome-scale cell-free kinetic modeling of
84 the biosynthetic capacity of industrially important organisms.

85 **Results**

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

100 Allosteric regulation of enzyme activity was modeled by combining individual regula-
101 tory contributions to the activity of pathway enzymes into a control coefficient using an
102 integration rule (Fig. 2). This strategy is similar in spirit to the Constrained Fuzzy Logic
103 (cFL) approach of Lauffenburger and coworkers which has been used to effectively model
104 signal transduction pathways important in human health [28]. In our formulation, Hill-like
105 transfer functions $0 \leq f(\mathcal{Z}) \leq 1$ were used to calculate the influence of factor abundance
106 upon target enzyme activity. In this context, factors can be individual metabolite levels
107 or some function, e.g., the product of metabolite levels. However, more generally, factors
108 can also correspond to non-modeled influences, categorial variables or other abstract
109 quantities. In the current study, we simply let \mathcal{Z} correspond to the abundance of individ-
110 ual metabolites, however in general this can be a complex function of both modeled and

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

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

137 an effective description of reaction kinetics.

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

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

cesses, there is also the added challenge of identifying the regulation and control structures that manage metabolism. Of course, these issues are not independent; any description of enzyme activity regulation will be a function of system state, which in turn depends upon the kinetic parameters. For cell free systems, regulated gene expression has been removed, however, enzyme activity regulation is still operational. We explored this linkage by estimating model parameters from synthetic data using both network structures. We generated synthetic measurements of the substrate S, intermediate M_5 and end product P_1 approximately every 20 min using network A. This data set is similar to published cell free studies both in terms of network coverage, and sampling frequency [23]. 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.

189 We biased the distribution of the particle population according to our *a prior* belief of the
190 correct regulatory program. To this end, we considered three different priors, a uniform
191 distribution where each putative regulatory structure represented 20% of the population
192 and two mixed distributions that were either positively or negatively biased towards the
193 correct structure (network A). In both the positively biased and uniform cases the PSO
194 clearly differentiated between the true or closely related structures and those that were
195 materially different (Fig. 7). As expected, the positively biased population (40% of the
196 initial particle population seeded with network A) gave the best results, where the correct
197 structure was preferentially identified (Fig. 7A). On the other hand, when given a uni-
198 form distribution, the PSO approach identified a combination of network A and network
199 C as the most likely control structures (Fig. 7B). Network A and C differ by the regula-
200 tory connection between the end product P_2 and enzyme E_1 ; in network A, end product
201 P_2 was assumed to inhibit E_1 , while in network C, end product P_2 activated E_1 . Lastly,
202 when the initial population was heavily biased towards incorrect structures (initial popu-
203 lation seeded with 90% incorrect structures), the particle swarm *misidentified* the correct
204 allosteric structure (Fig. 7C). Interestingly, while each particle swarm identified param-
205 eter sets that minimized the simulation error, the estimated parameter values were not
206 necessarily similar to the true parameters. The angle between the estimated and true pa-
207 rameters was not consistently small across the swarms (identical parameters would give
208 an angle of zero). This suggested that our particle swarm approach identified a *sloppy* en-
209 semble, i.e., parameter estimates that were individually incorrect but collectively exhibited
210 the correct model behavior.

211 We calculated control program output and scaled metabolic flux for the positively, uni-
212 formly and negatively biased particle swarms (Fig. 8). Network A and network C models
213 from the positively (Fig. 8A) and uniformly (Fig. 8B) biased particle swarms showed sim-
214 ilar operational patterns, despite differences in kinetic parameters and control structures.

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

- ²⁴¹ model structures will need to be further evaluated, for example by generating falsifiable
- ²⁴² experimental designs which could distinguish between low error solutions.

243 **Discussion**

244 In this study, we presented an effective kinetic modeling strategy to dynamically simu-
245 late cell-free biochemical networks. Our proposed strategy integrated traditional kinetic
246 modeling with an effective rules based approach to dynamically describe metabolic reg-
247 ulation and control. We tested this approach by developing kinetic models of hypotheti-
248 cal cell-free metabolic networks. In particular, we tested whether our effective modeling
249 approach could describe classically expected behavior, and second whether we could si-
250 multaneously estimate kinetic parameters and regulatory connectivity, in the absence of
251 specific mechanistic knowledge, from synthetic experimental data. Toward these ques-
252 tions, we explored five hypothetical cell-free networks. In each network, a substrate S
253 was converted to the end products P_1 and P_2 through a series of enzymatically catalyzed
254 reactions, including a branch point at a hypothetical metabolite M_2 . Each network also
255 included the same cofactors and cofactor recycle architecture. However, while all five
256 networks shared the same enzymatic connectivity, each had different allosteric regulatory
257 connectivity. We found that simple effective rules, when integrated with traditional enzyme
258 kinetic expressions, could capture complex allosteric patterns such as ultrasensitivity, or
259 non-competitive inhibition in the absence of specific mechanistic information. Moreover,
260 when integrated into network models, these rules captured classical regulatory patterns
261 such as product-induced feedback inhibition. Lastly, we simultaneously estimated kinetic
262 parameters and discriminated between competing regulatory structures, using synthetic
263 data in combination with a modified particle swarm approach. If we considered all putative
264 regulatory architectures to be equally likely, we were able to estimate a *sloppy* ensem-
265 ble of models with the correct architecture and kinetic parameters. Thus, we identified
266 parameter values that were different from their true values, but nonetheless produced
267 reasonable model performance (low error). This suggested that we captured important
268 parameter combinations (stiff combinations), while simultaneously missing other param-

269 eter combinations (sloppy combinations). This was similar to the earlier study of Brown
270 and Sethna [29], which showed that reasonable model predictions were possible, despite
271 sometimes only order of magnitude parameter estimates, if the stiff parameter combina-
272 tions were well constrained.

273 The proposed modeling strategy shares features with other popular techniques, but
274 also has several key differences. At its core, our effective modeling approach is similar
275 to regulatory constraint-based methods, and to the cybernetic modeling paradigm devel-
276 oped by Ramkrishna and colleagues. Covert, Palsson and coworkers drastically improved
277 the predictability of constraint-based approaches by integrating Boolean rules into the cal-
278 culation of metabolic fluxes [30]. If the regulated intracellular flux problem is coupled with
279 time-dependent extracellular balances, these models can predict complex behavior such
280 as diauxie growth or the switch between aerobic and anaerobic metabolism. Another im-
281 portant feature of this approach is that it scales with biological complexity. For example,
282 Covert *et al.* showed that a genome-scale model of *E. coli* augmented with a Boolean
283 rule layer, correctly predicted approximately 80% of the outcomes of a high-throughput
284 growth phenotyping experiment in *E. coli*. Further, they showed that they could learn new
285 biology by iteratively refining the model and its associated rules [31]. However, while regu-
286 lated flux balance analysis is a powerful technique, it does not easily allow the calculation
287 of time-resolved metabolite abundance. Additionally, the Boolean rules which populate
288 the regulatory layer are limited to ON/OFF decisions; for qualitative predictions of gene
289 expression this is a reasonable limitation. However, Boolean rules will likely be less ef-
290 fective at capturing dynamic allosteric regulation in a cell-free metabolic system. On the
291 other hand, the strength of cybernetic models is the integration of optimal metabolic con-
292 trol heuristics with traditional kinetic pathway modeling. Cybernetic models are highly
293 predictive; they have successfully predicted mutant behavior from limited wild-type data
294 [10, 32, 33], steady-state multiplicity [9], strain specific metabolic function [12] and have

295 been used in bioprocess control applications [34]. However, cybernetic control heuristics
296 are not mechanistic, instead they are the output of an optimal decision with respect to a
297 set of hypothetical physiological objectives. Thus, they are abstractions which are difficult
298 to translate into a specific biological mechanism. Our approach addresses the shortcom-
299 ings of both regulatory constraint-based models and cybernetic models. First, similar to
300 cybernetic models, the core of our approach is a kinetic model. Thus, we are able to di-
301 rectly calculate the time evolution of metabolism, for example the dynamic abundance of
302 network metabolites. Second, similar to regulatory flux balance analysis, our control laws
303 describe specific mechanistic motifs, such as activation or inhibition of enzyme activity.
304 However, our rules are continuous, thus they potentially allow a finer grained descrip-
305 tion of metabolic regulation and control mechanisms. Lastly, we can naturally incorporate
306 unmodeled factors and categorical factors or combinations thereof into our control law
307 formulations. Though requiring a more complex description of cellular metabolism, our
308 approach may even be extended to simulate cell-based systems by incorporating the
309 same control laws into transcription factor activation and gene expression regulation.

310 There are several critical questions that should be explored following this proof-of-
311 concept study. It is unclear how parameter identification will scale to genome-scale net-
312 works, and second it is unclear how we will identify allosteric connectivity at a genome-
313 scale. The enzymatic connectivity for genome-scale cell-free networks can easily be es-
314 tablished by stripping away the growth and cell wall machinery from whole cell genome
315 reconstructions. Then metabolic fluxes can be transformed into kinetic expressions using
316 heuristics such multiple saturation kinetics, which are then modified by our rule-based
317 control variables. This leaves a large number of unknown kinetic constants that must
318 be estimated from time-resolved metabolite measurements. Ensemble modeling is a
319 well-established approach for parameter identification in large-scale deterministic mod-
320 els. Liao and coworkers developed a method that generates an ensemble of kinetic mod-

321 els that all approach the same steady-state, one determined by fluxomics measurements
322 [35]. The best subpopulation of candidate models were selected based on their agree-
323 ment with further measurements of genetically perturbed systems. Our work relies on
324 heuristic search optimization to identify kinetic models consistent with steady-state and
325 dynamic time-series measurements of cellular species [36–41]. Instead of estimating a
326 single yet highly uncertain parameter set, both approaches estimate an ensemble of pa-
327 rameter sets whose model behavior recapitulates experimental measurements. Here, we
328 showed that particle swarm optimization quickly identified an ensemble of model param-
329 eters, at least for proof-of-concept metabolic networks using synthetic data. This sug-
330 gested that we can expect reasonable model predictions, despite only partial parameter
331 knowledge, as network size grows if we have properly designed experiments. Brown and
332 Sethna showed in a model of signal transduction that good predictions were possible de-
333 spite only order of magnitude estimates of parameter values [29]. Sethna and coworkers
334 later showed that model performance is often controlled by only a few parameter com-
335 binations, a characteristic seemingly universal to multi-parameter models referred to as
336 *sloppiness* [42]. We have also demonstrated *sloppy* behavior in a wide variety of sig-
337 nal transduction processes [36–41]. Thus, given our previous experience with models
338 containing hundreds of unknown parameters, we expect parameter estimation to be a
339 manageable challenge.

340 A critical challenge will be the estimation of allosteric connectivity at a genome scale.
341 The regulation of glycolytic enzymes, such as phosphofructokinase I, has been studied
342 for many years [43, 44]. The allosteric regulation of metabolic enzymes can also be estab-
343 lished from organism specific databases, such as EcoCyc [45], or more general allosteric
344 databases, such as the AlloSteric Database [46]. However, for those enzymes that have
345 not been well studied, we will need to infer allosteric interactions from experimental data.
346 In general, the reverse engineering of regulatory network structure from data is a very

347 difficult problem. Recently, Sauer and colleagues have developed a systematic, model-
348 based approach for the identification of allosteric regulation *in vivo* [47]. They tested the
349 effects of many putative allosteric protein-metabolite interactions on the performance of
350 a kinetic model of glycolysis against dynamic metabolomic and fluxomic measurements.
351 A method similar to this may be easily applied to cell-free systems in order to identify
352 relevant *in vitro* allosteric interactions. Because omics measurements of cell-free environ-
353 ments are easy to obtain, identification of large-scale allosteric control structures may be
354 possible. Also, there are many different approaches from the reverse engineering of gene
355 regulatory networks that perhaps could be adopted to this problem, however this remains
356 an open question. For example, one could imagine designing pulse chase experiments
357 which maximally distinguish between competing allosteric models, similar to the earlier
358 work of Kremling et al [48], or iteratively estimate model structures similar to Doyle and
359 coworkers [49]. Lastly, the choice of max/min integration rules or the particular form of the
360 transfer functions could be generalized to include other rule types and functions. Theo-
361 retically, an integration rule is a function whose domain is a set of transfer function inputs,
362 and whose range is $v \in [0, 1]$. Thus, integration rules other than max/min could be used,
363 such as the mean or the product, assuming the range of the transfer functions is always
364 $f \in [0, 1]$. Alternative integration rules such as the mean might have different properties
365 which could influence model identification or performance. For example, a mean integra-
366 tion rule would be differentiable, which allows derivative-based optimization approaches
367 to be used. The particular form of the transfer function could also be explored. We choose
368 a Hill-like function because of its prominence in the systems and synthetic biology com-
369 munity. However, the only mathematical requirement for a transfer function is that it map
370 a non-negative continuous or categorical variable into the range $f \in [0, 1]$. Thus, many
371 types of transfer functions are possible.

372 **Materials and Methods**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

401 where:

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

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

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

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

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

406 where \mathcal{Z}_j denotes the abundance of the j factor (e.g., metabolite abundance), and κ_{ij}
 407 and η are control parameters. The κ_{ij} parameter represents a species gain parameter,
 408 while η is a cooperativity parameter (similar to a Hill coefficient). In the case $\eta > 1$, the
 409 allosteric interaction displays positive cooperativity. For $\eta < 1$, the interaction is negatively
 410 cooperative. Finally, if $\eta = 1$, the interaction displays no cooperativity. The effect of
 411 different values of η on reaction rate can be seen in Figure 3. The model equations were
 412 encoded using the Octave programming language and solved using the LSODE routine
 413 in Octave (v 3.8.1; www.octave.org). In some cases, metabolic fluxes (or other quantities)
 414 were scaled according to:

$$\hat{r}_j(t = \tau) = \left(\frac{r_j - \min \mathbf{r}}{\max \mathbf{r} - \min \mathbf{r}} \right) \Big|_{t=\tau} \quad (11)$$

415 where $0 \leq \hat{r}_j(t = \tau) \leq 1$ denotes the scaled value for flux j evaluated at time τ . We have
 416 used this scaling in a variety of other contexts [41, 50].

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

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^S \left(\frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2 \quad (12)$$

420 where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ , $x_j(\tau, \mathbf{k})$ denotes the sim-
 421 ulated value for species j at time τ , and $\omega_j(\tau)$ denotes the experimental measurement
 422 variance for species j at time τ . The outer summation is respect to time, while the in-

ner 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 assumed a constant coefficient of variation of 10% for the synthetic data set.

We minimized the model residual using particle swarm optimization (PSO) [51]. 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:

$$\Delta_i = \theta_1 \Delta_i + \theta_2 \mathbf{r}_1 (\mathcal{L}_i - \mathbf{k}_i) + \theta_3 \mathbf{r}_2 (\mathcal{G} - \mathbf{k}_i) \quad (13)$$

$$\mathbf{k}_i = \mathbf{k}_i + \Delta_i \quad (14)$$

where Δ_i denotes the perturbation to the vector of parameters \mathbf{k}_i for particle i . $(\theta_1, \theta_2, \theta_3)$ are adjustable parameters, \mathcal{L}_i denotes the best local 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 ($K \times 1$). In this 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}^* :

$$\alpha_j = \cos^{-1} \left(\frac{\mathbf{k}_j \cdot \mathbf{k}^*}{\|\mathbf{k}_j\| \|\mathbf{k}^*\|} \right) \quad (15)$$

⁴⁴⁴ If the candidate parameter set k_j were perfect, the residual between the model and syn-
⁴⁴⁵ thetic data and the angle between k_j and the true parameter set k^* would be equal to
⁴⁴⁶ zero.

⁴⁴⁷ We modified our PSO implementation to simultaneously search over kinetic parame-
⁴⁴⁸ ters and putative model control structures. In the combined case, each particle potentially
⁴⁴⁹ carried a different model realization in addition to a different kinetic parameter vector. We
⁴⁵⁰ kept the update rules the same (along with the update parameters). Thus, each parti-
⁴⁵¹ cle competed on the basis of goodness of fit, which allowed different model structures
⁴⁵² to contribute to the overall behavior of the swarm. We considered five possible model
⁴⁵³ structures (A through E), where network A was the correct formulation (used to generate
⁴⁵⁴ the synthetic data). We considered a population of 100 particles, where each particle in
⁴⁵⁵ the swarm was assigned a model structure, and a random parameter vector. The PSO
⁴⁵⁶ algorithm, model equations, and the objective function were encoded and solved in the
⁴⁵⁷ Octave programming language (v 3.8.1; www.octave.org).

⁴⁵⁸ Acknowledgements

⁴⁵⁹ This study was supported by an award from the National Science Foundation (MCB
⁴⁶⁰ #1411715) and the Army Research Office (ARO #59155-LS).

461 **References**

- 462 1. Fredrickson AG (1976) Formulation of structured growth models. Biotechnol Bioeng
463 18: 1481-6.
- 464 2. Domach MM, Leung SK, Cahn RE, Cocks GG, Shuler ML (1984) Computer model
465 for glucose-limited growth of a single cell of escherichia coli b/r-a. Biotechnol Bioeng
466 26: 203-16.
- 467 3. Steinmeyer D, Shuler M (1989) Structured model for *Saccharomyces cerevisiae*.
468 Chem Eng Sci 44: 2017 - 2030.
- 469 4. Wu P, Ray NG, Shuler ML (1992) A single-cell model for cho cells. Ann N Y Acad Sci
470 665: 152-87.
- 471 5. Castellanos M, Wilson DB, Shuler ML (2004) A modular minimal cell model: purine
472 and pyrimidine transport and metabolism. Proc Natl Acad Sci U S A 101: 6681-6.
- 473 6. Atlas JC, Nikolaev EV, Browning ST, Shuler ML (2008) Incorporating genome-wide
474 dna sequence information into a dynamic whole-cell model of escherichia coli: appli-
475 cation to dna replication. IET Syst Biol 2: 369-82.
- 476 7. Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT (1985) A cybernetic view of micro-
477 bial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27: 1-9.
- 478 8. Kompala DS, Ramkrishna D, Jansen NB, Tsao GT (1986) Investigation of bacterial
479 growth on mixed substrates: experimental evaluation of cybernetic models. Biotech-
480 nol Bioeng 28: 1044-55.
- 481 9. Kim JI, Song HS, Sunkara SR, Lali A, Ramkrishna D (2012) Exacting predictions by
482 cybernetic model confirmed experimentally: steady state multiplicity in the chemostat.
483 Biotechnol Prog 28: 1160-6.
- 484 10. Varner J, Ramkrishna D (1999) Metabolic engineering from a cybernetic perspective:
485 aspartate family of amino acids. Metab Eng 1: 88-116.
- 486 11. Song HS, Morgan JA, Ramkrishna D (2009) Systematic development of hybrid cy-

- bernetic models: application to recombinant yeast co-consuming glucose and xylose.
Biotechnol Bioeng 103: 984-1002.
12. Song HS, Ramkrishna D (2011) Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. Biotechnol Bioeng 108: 127-40.
13. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10: 291-305.
14. Edwards JS, Palsson BO (2000) The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97: 5528-33.
15. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. Mol Syst Biol 3: 121.
16. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282: 28791-9.
17. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7: 129-43.
18. Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420: 186-9.
19. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. Mol Syst Biol 3: 119.
20. Hyduke DR, Lewis NE, Palsson BØ (2013) Analysis of omics data with genome-scale models of metabolism. Mol Biosyst 9: 167-74.
21. McCloskey D, Palsson BØ, Feist AM (2013) Basic and applied uses of genome-scale

- metabolic network reconstructions of escherichia coli. Mol Syst Biol 9: 661.
22. Zomorrodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical optimization applications in metabolic networks. Metab Eng 14: 672-86.
23. Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4: 220.
24. 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.
25. 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.
26. 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.
27. Hodgman CE, Jewett MC (2012) Cell-free synthetic biology: thinking outside the cell. Metab Eng 14: 261-9.
28. 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.
29. Brown KS, Sethna JP (2003) Statistical mechanical approaches to models with many poorly known parameters. Phys Rev E Stat Nonlin Soft Matter Phys 68: 021904.
30. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213: 73-88.
31. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429: 92-6.

- 539 32. Varner JD (2000) Large-scale prediction of phenotype: concept. *Biotechnol Bioeng*
540 69: 664-78.
- 541 33. Song HS, Ramkrishna D (2012) Prediction of dynamic behavior of mutant strains from
542 limited wild-type data. *Metab Eng* 14: 69-80.
- 543 34. Gadkar KG, Doyle FJ 3rd, Crowley TJ, Varner JD (2003) Cybernetic model predictive
544 control of a continuous bioreactor with cell recycle. *Biotechnol Prog* 19: 1487-97.
- 545 35. Tran LM, Rizk ML, Liao JC (2008) Ensemble modeling of metabolic networks. *Biophys*
546 *J* 95: 5606-17.
- 547 36. Luan D, Zai M, Varner JD (2007) Computationally derived points of fragility of a human
548 cascade are consistent with current therapeutic strategies. *PLoS Comput Biol* 3:
549 e142.
- 550 37. Song SO, Varner J (2009) Modeling and analysis of the molecular basis of pain in
551 sensory neurons. *PLoS One* 4: e6758.
- 552 38. Tasseff R, Nayak S, Salim S, Kaushik P, Rizvi N, et al. (2010) Analysis of the molecular
553 networks in androgen dependent and independent prostate cancer revealed fragile
554 and robust subsystems. *PLoS One* 5: e8864.
- 555 39. Tasseff R, Nayak S, Song SO, Yen A, Varner JD (2011) Modeling and analysis
556 of retinoic acid induced differentiation of uncommitted precursor cells. *Integr Biol*
557 (*Camb*) 3: 578-91.
- 558 40. Nayak S, Siddiqui JK, Varner JD (2011) Modelling and analysis of an ensemble of
559 eukaryotic translation initiation models. *IET Syst Biol* 5: 2.
- 560 41. Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and
561 analysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol* 7:
562 e1002263.
- 563 42. Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compres-
564 sion underlies emergent theories and predictive models. *Science* 342: 604-7.

- 565 43. Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry. W.H. Freeman.
- 566 44. Peskov K, Goryanin I, Demin O (2008) Kinetic model of phosphofructokinase-1 from
567 escherichia coli. J Bioinform Comput Biol 6: 843-67.
- 568 45. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, et al. (2013)
569 Ecocyc: fusing model organism databases with systems biology. Nucleic Acids Res
570 41: D605-12.
- 571 46. Huang Z, Mou L, Shen Q, Lu S, Li C, et al. (2014) Asd v2.0: updated content and
572 novel features focusing on allosteric regulation. Nucleic Acids Res 42: D510-6.
- 573 47. Link H, Kochanowski K, Sauer U (2013) Systematic identification of allosteric protein-
574 metabolite interactions that control enzyme activity in vivo. Nat Biotechnol 31: 357-
575 61.
- 576 48. Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, et al. (2004) A benchmark for
577 methods in reverse engineering and model discrimination: problem formulation and
578 solutions. Genome Res 14: 1773-85.
- 579 49. Gadkar KG, Gunawan R, Doyle FJ 3rd (2005) Iterative approach to model identifica-
580 tion of biological networks. BMC Bioinformatics 6: 155.
- 581 50. Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models
582 using pareto optimal ensemble techniques (poets). Biotechnol J 5: 768-80.
- 583 51. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the
584 International Conference on Neural Networks. pp. 1942 - 1948.

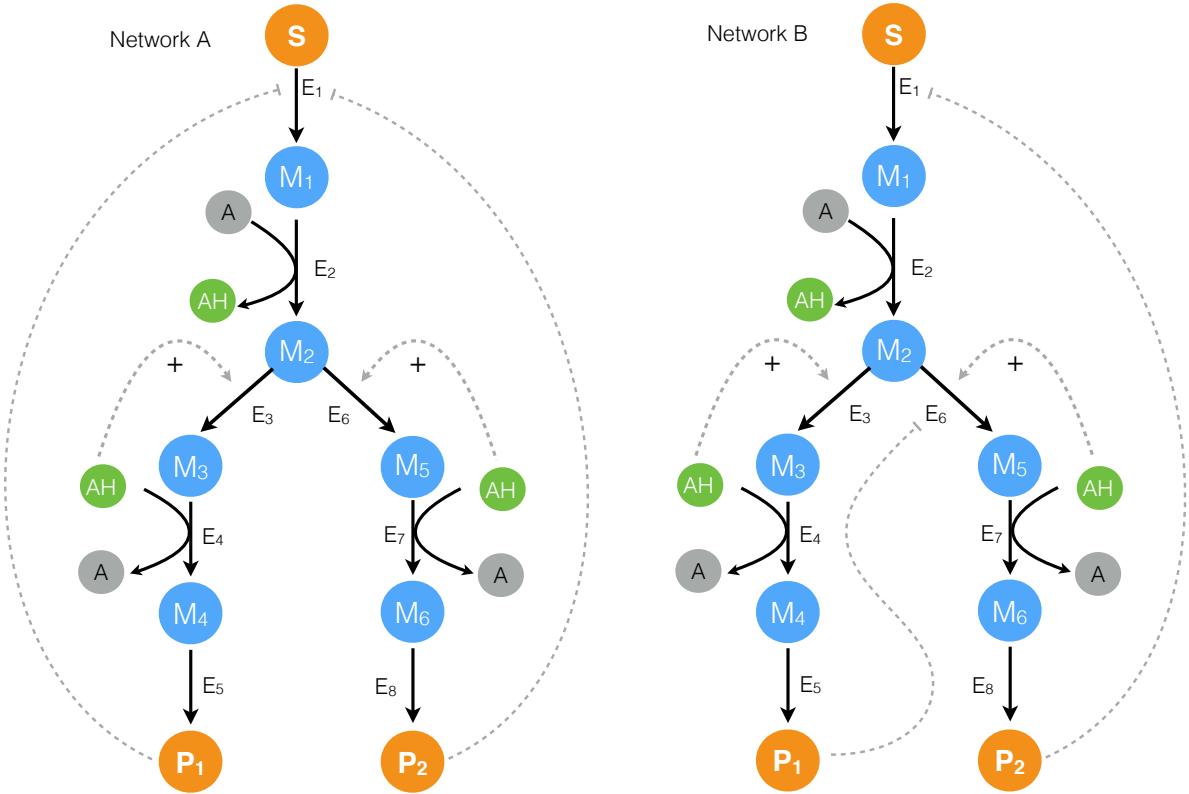


Fig. 1: Proof-of-concept cell-free metabolic networks considered in this study. Substrate S is converted to products P_1 and P_2 through a series of chemical conversions catalyzed by enzyme(s) E_j . The activity of the pathway enzymes is subject to both positive and negative allosteric regulation.

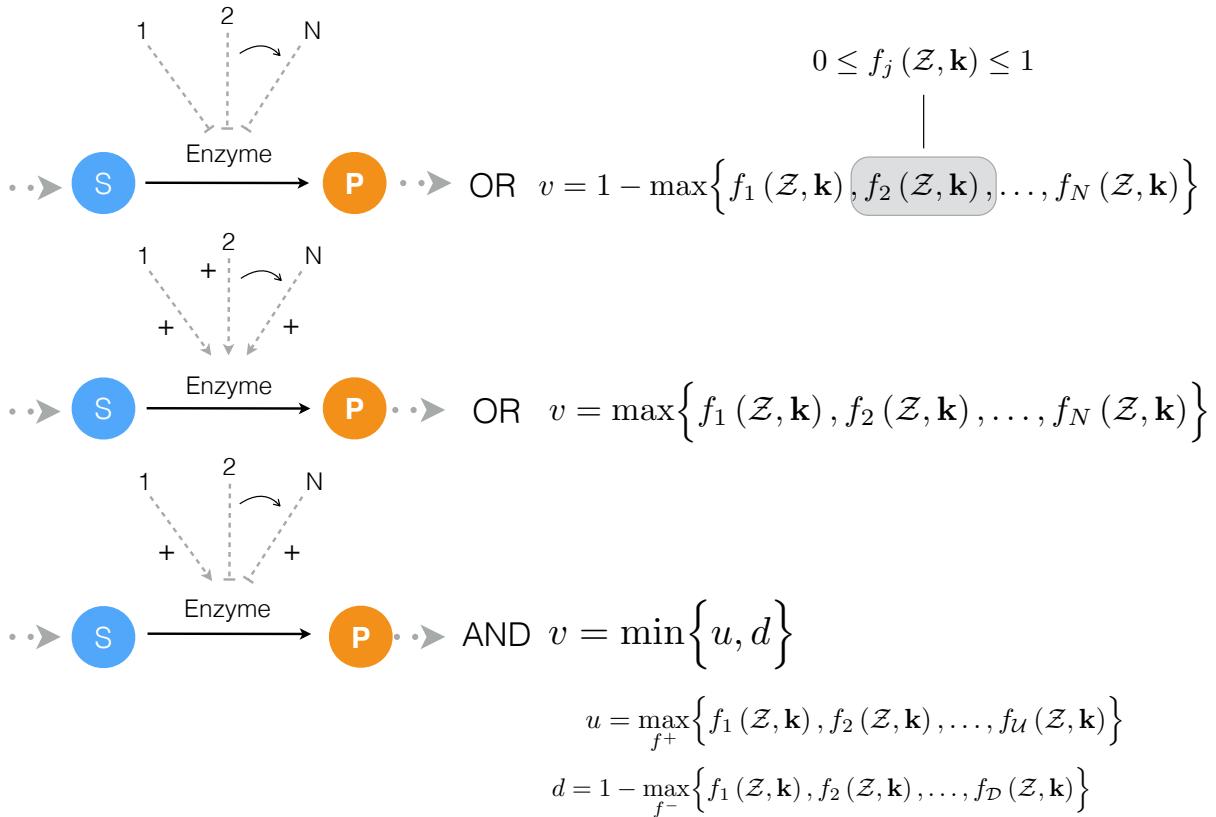


Fig. 2: Schematic of rule-based allosteric enzyme activity control laws. Traditional enzyme kinetic expressions, e.g., Michaelis–Menten or multiple saturation kinetics, are multiplied by an enzyme activity control variable $0 \leq v_j \leq 1$. Control variables are functions of many possible regulatory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

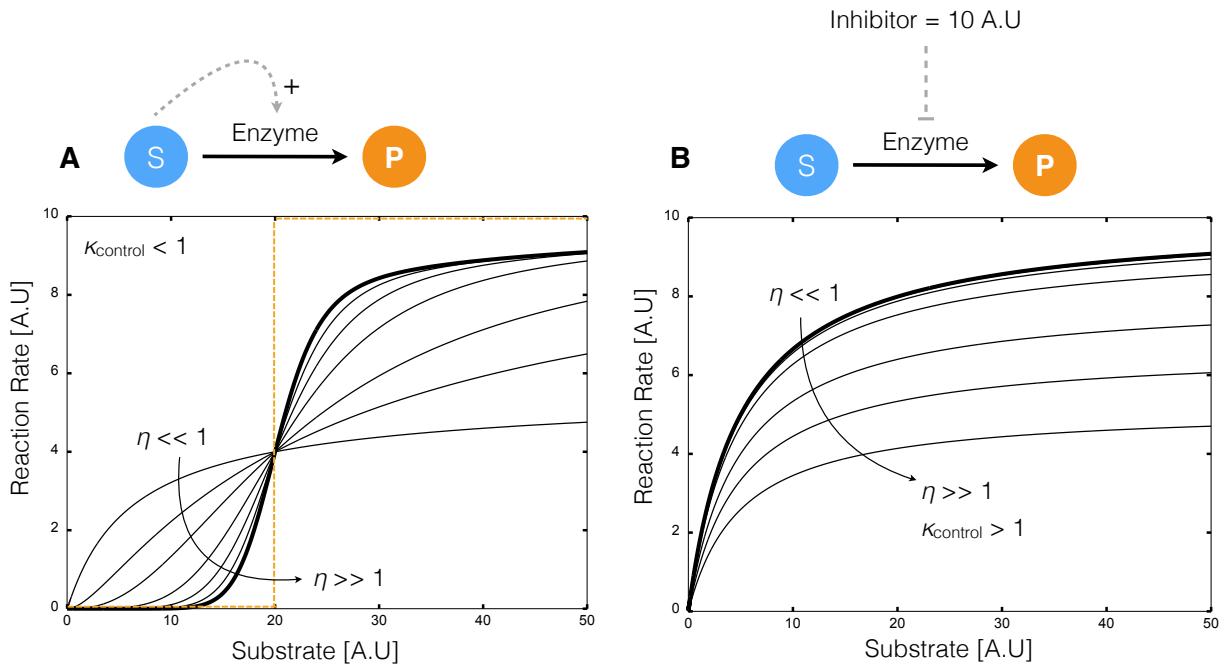


Fig. 3: Kinetics of simple transformations in the presence of activation and inhibition. **A:** The conversion of substrate S to product P by enzyme E was activated by S . For a fixed control gain parameter $\kappa_{control}$, the reaction rate approached a step for increasing cooperativity control parameter η . For activation simulations $\kappa_{control} = 0.05$ and $\eta = \{0.01, 0.1, 1, 2, 4, 6, 8, 10\}$. **B:** The conversion of substrate S to product P by enzyme E with inhibitor I . For a fixed control gain parameter $\kappa_{control}$, the reaction rate approximated non-competitive inhibition for increasing cooperativity control parameter η . For the inhibition simulations $\kappa_{control} = 1.5$ and $\eta = \{0.01, 0.1, 1, 2, 4, 6, 8, 10\}$.

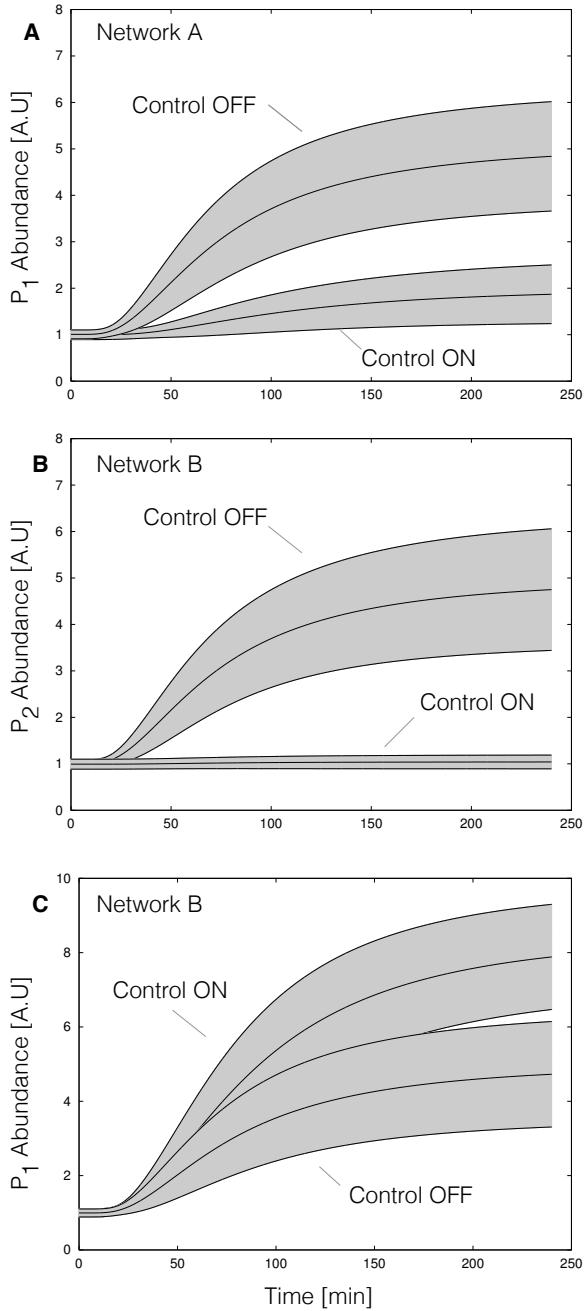


Fig. 4: ON/OFF control simulations for Network A and Network B for an ensemble of 100 kinetic parameter sets versus time. For each case, simulations were conducted using kinetic and initial conditions generated randomly from a hypothetical true parameter set. The gray area represents \pm one standard deviation surrounding the mean. Control parameters were fixed during the ensemble calculations. **A:** End product P₁ abundance versus time for Network A. The abundance of P₁ decreased with end product inhibition of E₁ activity (Control-ON) versus the no inhibition case (Control-OFF). **B:** End product P₂ abundance versus time for Network B. Inhibition of branch point E₆ by end product P₁ decreased P₂ abundance (Control-ON) versus the no inhibition case (Control-OFF). **C:** End product P₁ abundance versus time for Network A. Inhibition of branch point E₆ by end product P₁ decreased P₁ abundance (Control-ON) versus the no inhibition case (Control-OFF).

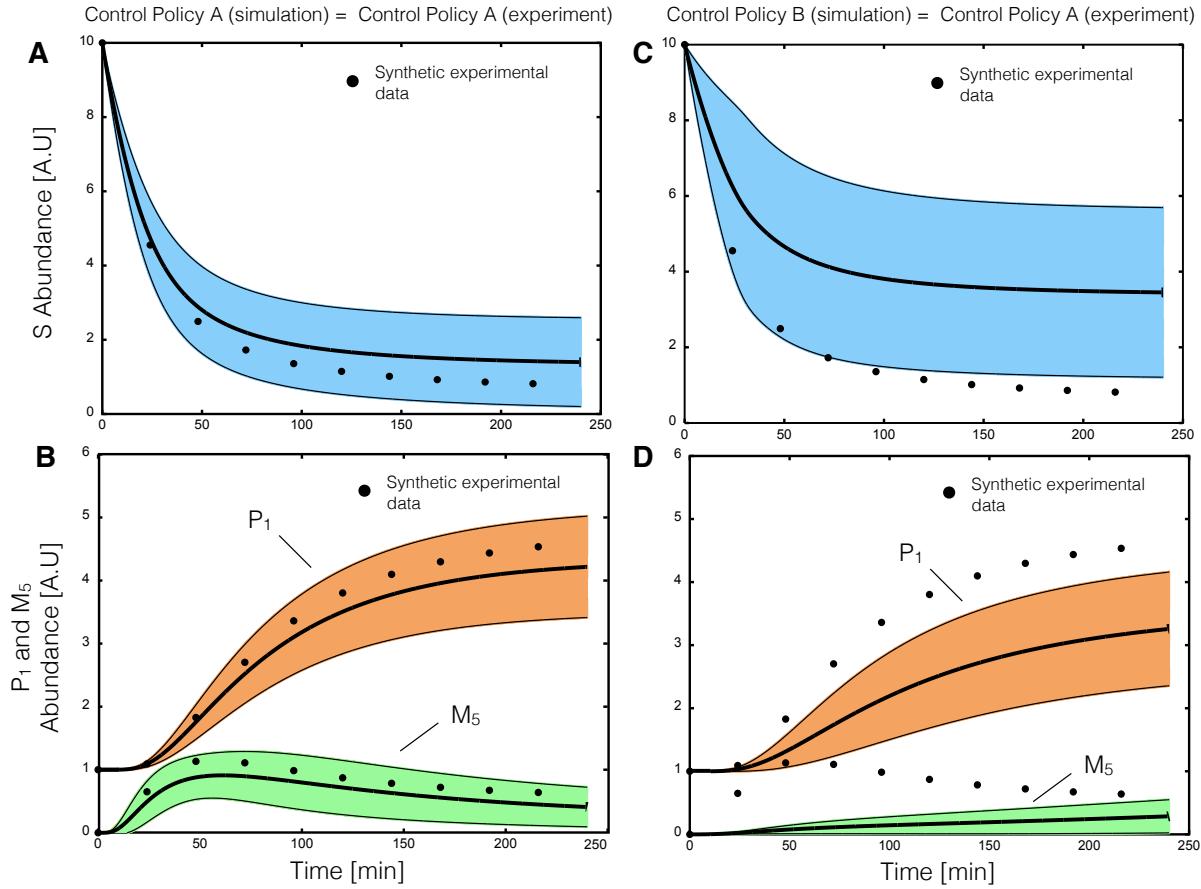


Fig. 5: Parameter estimation from synthetic data for the same and mismatched allosteric control logic using particle swarm optimization (PSO). Synthetic experimental data was generated from a hypothetical parameter set using Network A, where substrate S , end product P_1 and intermediate M_5 were sampled 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 20 parameter sets 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. The synthetic data plotted here was unperturbed by noise. However, we assumed a coefficient of variation of 10% for the synthetic data during parameter estimation.

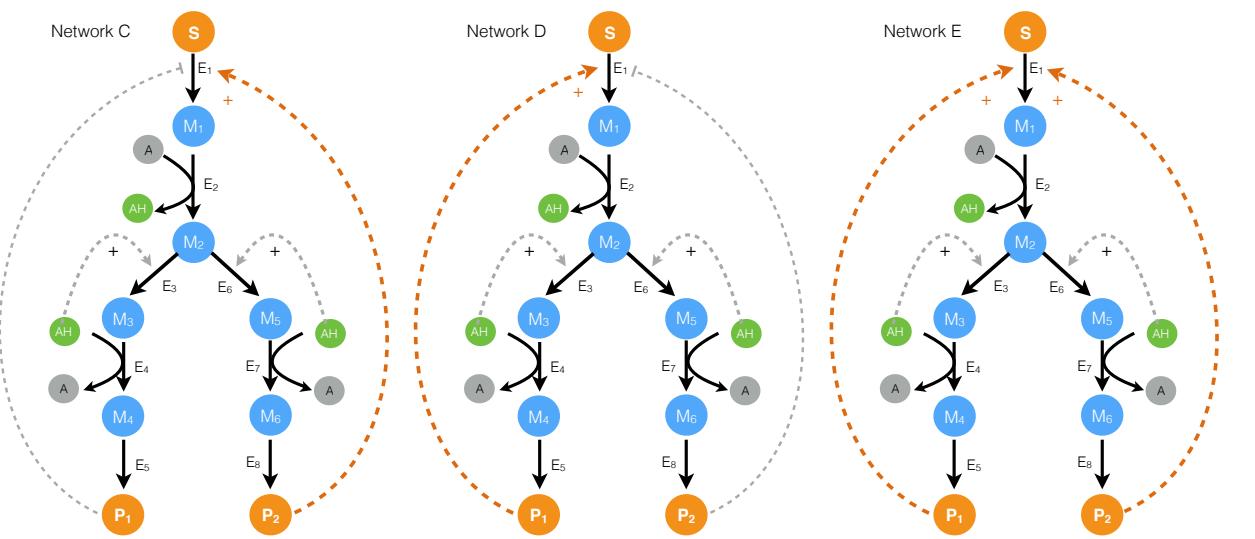


Fig. 6: Schematic of the alternative allosteric control programs used in the structural particle swarm computation. Each network had the same enzymatic connectivity, initial conditions and kinetic parameters, but alternative feedback control structures for the first enzyme in the pathway.

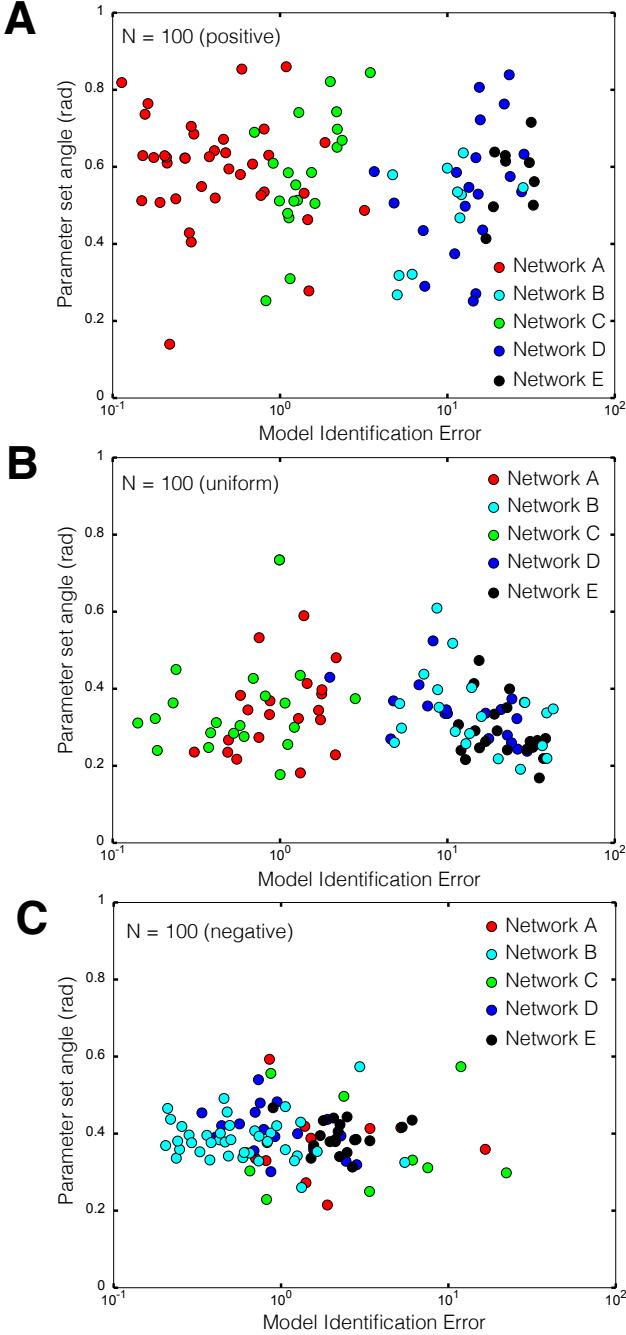


Fig. 7: Combined control and kinetic parameter search using modified particle swarm optimization (PSO). A population of 100 particles was initialized with randomized kinetic parameters and one of five possible control configurations (Network A - E). Simulation error was minimized for a synthetic data 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 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 100 uniformly distributed particles (A,B,C,D,E) = (20%, 20%, 20%, 20% and 20%). **C:** Simulation error versus parameter set angle for 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.

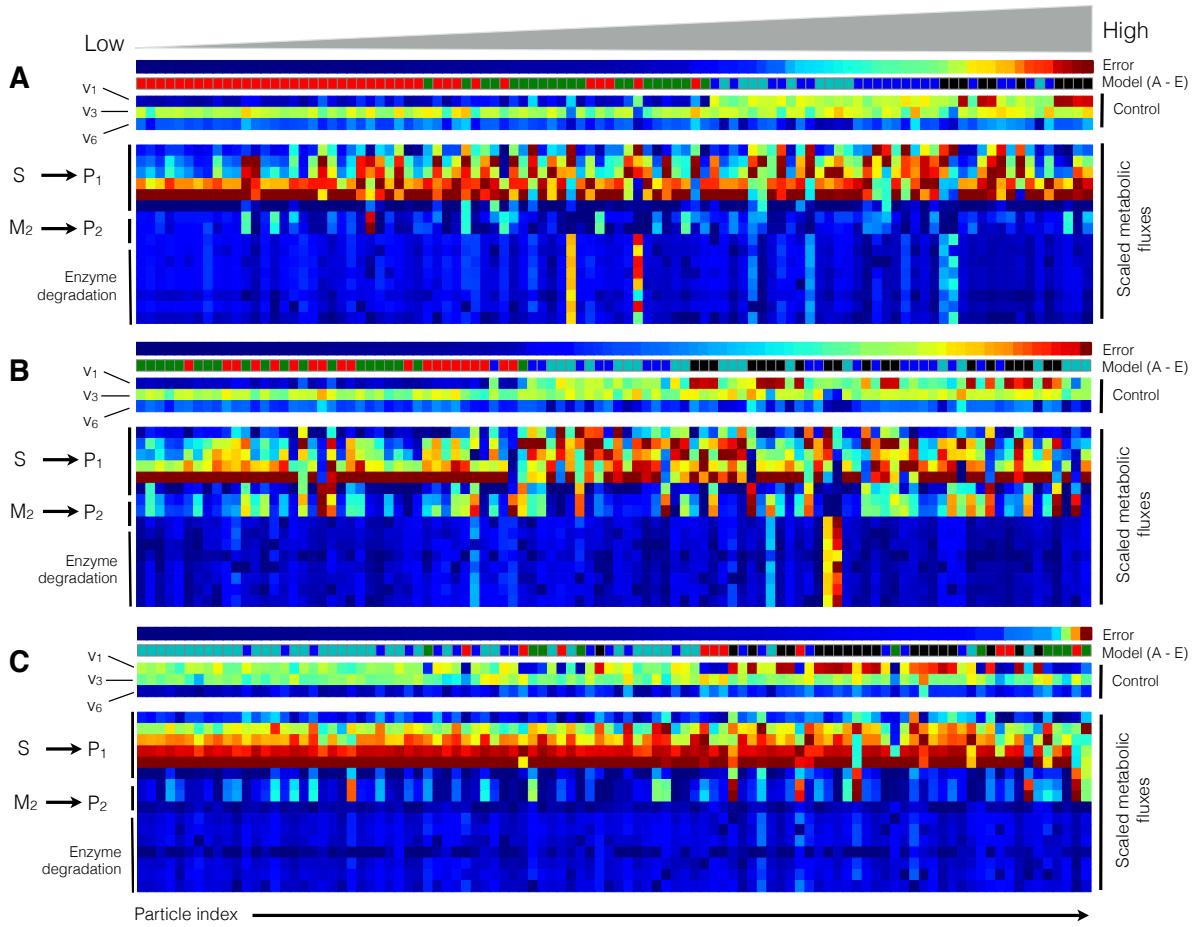


Fig. 8: Metabolic flux and control variables as a function of network type and particle index at $t = 100$ min. The particle error, the control variables governing E_1 , E_3 and E_6 activity (v_1 , v_3 and v_6) and the scaled metabolic flux were calculated for the positively (top), uniformly (middle) and negatively (bottom) biased particle swarms ($N = 100$). Blue denotes a low value, while red denotes a high value for the respective quantity being plotted. 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.