Toward a Genome Scale Dynamic Model of Cell Free Protein Synthesis in *Escherichia coli*

Nicholas Horvath, Michael Vilkhovoy, Joseph Wayman, Kara Calhoun[1], James Swartz[1] and Jeffrey D. Varner*

School of Chemical and Biomolecular Engineering

Cornell University, Ithaca NY 14853

[1]School of Chemical Engineering

Stanford University, Stanford, California 94305

Running Title: Dynamic model of cell free protein synthesis

To be submitted: *Scientific Reports*

*Corresponding author:

Jeffrey D. Varner,

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

Fill me in.

Keywords: Biochemical engineering, systems biology, cell free protein synthesis

Introduction

Mathematical modeling has long contributed to our understanding of metabolism. Decades before the genomics revolution, mechanistically, structured metabolic models arose from the desire to predict microbial phenotypes resulting from changes in intracellular or extracellular states [1]. The single cell *E. coli* models of Shuler and coworkers pioneered the construction of large-scale, dynamic metabolic models that incorporated multiple, regulated catabolic and anabolic pathways constrained by experimentally determined kinetic parameters [2]. Shuler and coworkers generated many single cell kinetic models, including single cell models of eukaryotes [3, 4], minimal cell architectures [5], as well as DNA sequence based whole-cell models of *E. coli* [6]. Conversely, highly abstracted kinetic frameworks, such as the cybernetic framework, represented a paradigm shift, viewing cells as growth-optimizing strategists [7]. Cybernetic models have been highly successful at predicting metabolic choice behavior, e.g., diauxie behavior [8], steady-state multiplicity [9], as well as the cellular response to metabolic engineering modifications [10]. Unfortunately, traditional, fully structured cybernetic models also suffer from an identifiability challenge, as both the kinetic parameters and an abstracted model of cellular objectives must be estimated simultaneously. However, recent cybernetic formulations from Ramkrishna and colleagues have successfully treated this identifiability challenge through elementary mode reduction [11, 12].

In the post genomics world, large-scale stoichiometric reconstructions of microbial metabolism popularized by static, constraint-based modeling techniques such as flux balance analysis (FBA) have become standard tools [13]. Since the first genome-scale stoichiometric model of *E. coli*, developed by Edwards and Palsson [14], well over 100 organisms, including industrially important prokaryotes such as *E. coli* [15] or *B. subtilis* [16], are now available [17]. Stoichiometric models rely on a pseudo-steady-state assumption to reduce unidentifiable genome-scale kinetic models to an underdetermined linear

algebraic system, which can be solved efficiently even for large systems. Traditionally, stoichiometric models have also neglected explicit descriptions of metabolic regulation and control mechanisms, instead opting to describe the choice of pathways by prescribing an objective function on metabolism. Interestingly, similar to early cybernetic models, the most common metabolic objective function has been the optimization of biomass formation [18], although other metabolic objectives have also been estimated [19]. Recent advances in constraint-based modeling have overcome the early shortcomings of the platform, including capturing metabolic regulation and control [20]. Thus, modern constraint-based approaches have proven extremely useful in the discovery of metabolic engineering strategies and represent the state of the art in metabolic modeling [21, 22]. However, genome-scale kinetic models of industrial important organisms such as *E. coli* have yet to be constructed.

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

obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds several significant advantages for model development, identification and validation. Theoretically, 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 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. While only an initial proof-of-concept, the framework presented here could be an important first step toward genome-scale cell-free kinetic modeling of the biosynthetic capacity of industrially important organisms.

Results

Estimation of an ensmeble of models We constructed an initial model of cell-free *E. coli* metabolism by removing the growth-associated processes from the genome-scale model of Palsson and coworkers [**?**] and adding reactions for the synthesis of chloramphenicol acetyltransferase (CAT), a model protein for which we have a comprehensive training dataset from Swartz and coworkers [**?**]. This initial model captured the core metabolism (glycolysis, pentose phosphate, Enter-Doudoroff, TCA cycle), as well as the synthesis of energy species, amino acids, and CAT. We generated mass balance equations around metabolites and enzymes in the network, modeled as ordinary differential equations with reaction rates equal to the product of a kinetic term and a control term. We used multiple saturation kinetics to model metabolic fluxes and mass action kinetics to model enzyme degradation. We modeled the control term using a rule-based approach in which each control factor had a regulatory transfer function, modeled as a Hill function [**?**], and the control term was calculated as the mean of all transfer functions. The model was trained against the Swartz dataset, which included measurements of glucose, CAT, organic acids (pyruvate, lactate, acetate, succinate, malate), energy species (AXP, GXP, CXP, UXP), and 18 of the 20 proteinogenic amino acids. We generated an ensemble of 18,000 parameter sets by minimizing the error between the Swartz dataset and the metabolite concentrations predicted by the model. Out of these 18,000 sets, we defined the set with the lowest error value as our best-fit set. The ensemble captures the central carbon metabolism, including glucose uptake, CAT production, and the dynamics of the organic acid intermediates (Fig. 2, top). Allosteric control is important to the dynamics of the organic acid intermediates, as without it several of the measurements are not captured by the ensemble or the best-fit set (Fig. 2, bottom). The ensemble also captures the energy species dynamics, particularly the overall energy total (Fig. 2, top) and the totals by base (Fig. 3). The ensemble and the best-fit set also predict some of the amino acid

measurements, while failing to predict others (Fig. 4). This is likely due to a structural deficiency in the model; in some cases, the consumption of an amino acid through CAT synthesis is not enough to explain the decrease shown in the data, and there are no other reactions that consume it. Thus, a more comprehensive biological description is needed to fully explain amino acid dynamics.

Sensitivity analysis We conducted a local sensitivity analysis to determine which of the kinetic and control parameters affected model performance. We calculated performance as area under the CAT curve, which was directly related to CAT synthesis rate, as the culture time was fixed and no CAT degradation was modeled. We randomly chose 180 sets of the 18,000 sets in the ensemble and defined these as our sub-ensemble; for each set in the sub-ensemble, we varied the rate constant and saturation constants of each metabolic flux and measured the resulting change in CAT production to estimate the sensitivity of model performance to that parameter. We did the same for the control parameters, both the order (Hill coefficient) and gain (related to the the dissociation constant). This allowed us to estimate the relative importance of the kinetic and control parameters to CAT production across the ensemble. Of the rate constants, those with the highest positive sensitivities were CAT synthesis, GTP synthesis, GMP synthesis, glutamine synthesis, and aspartate synthesis (Fig. 5, top). This is explained by GTP and the amino acids being reactants for CAT synthesis. Also, GMP synthesis increases the total amount of guanosine, allowing for more GTP production. The rate constants with the largest negative sensitivities were GTP degradation, arginine synthesis, and UMP synthesis. While GTP degradation is obvious, the others can be explained in that they consume ATP as well as several amino acids, all of which are reactants for CAT synthesis. Of the saturation constants, the reverse is seen: the largest positive sensitivities are those associated with arginine synthesis, while the largest negative sensitivities are those associated with CAT synthesis, GTP synthesis, and GMP synthesis (Fig. 5, middle). This is because

an increase in saturation constant causes a decrease in the corresponding reaction rate [**?**]. The control parameters were seen to be the least significant and the most uncertain (Fig. 5, bottom). Only two had a small standard error across the ensemble, relative to the ensemble mean sensitivity: the gain and order for pyruvate acting as an inhibitor on the pdh reaction. This could be because pdh consumes pyruvate and diverts carbon away from the pathways that ultimately contribute to CAT production. Taken together with the lack of change in glucose uptake and CAT production when control is removed (Fig. 2), this suggests that allosteric control is not the limiting factor to CAT production.

We conducted a global sensitivity analysis on the parameters that could be controlled experimentally: the initial conditions of glucose, oxygen, amino acids, and enzymes. We used the variance-based method of Sobol, and the same objective function of area under the CAT curve. Using a parameter set of relatively good fit against data, we defined parameter bounds and generated a Sobol sequence of 111,600 parameter values that fit within those bounds. We then calculated the total-order sensitivity and confidence interval for each of the experimentally controllable initial conditions. As the sensitivities were total-order, they were guaranteed to be non-negative. The largest sensitivities belonged to the initial conditions of the CAT macromolecular synthesis machinery, GTP synthase, GTP degradation, some amino acids such as phenylalanine, proline, and leucine, and some amino acid synthases (Fig. 6). This is all explained by GTP and amino acids being reactants for CAT synthesis. While some of the amino acids and amino acid synthases were among the highest in sensitivity, theirs were also very uncertain relative to the CAT macromolecular synthesis machinery and GTP synthase. The initial conditions of glucose and oxygen were among the least important according to the global sensitivity analysis, suggesting that the model predicts that CAT production can be sustained by consuming initial stores or can be powered by other means.

Calculation of CAT yield We calculated the carbon yield of CAT production for our experimental data and our best-fit parameter set as a function of the initial and final concentrations and the carbon numbers of CAT, glucose, and amino acids. Arginine and glutamate were excluded due to not being present in the Swartz dataset. The experimental data displayed a CAT yield of 0.0865, while the best-fit parameter set displayed a CAT yield of 0.0871. We then used sequence-specific FBA to calculate a theoretical maximum CAT yield of 0.1942. Thus, we showed that our experimental dataset and best-fit parameter set were each producing CAT at 45% of the theoretical maximum. This allowed us to quantify the amount of carbon being diverted to byproducts, and suggests that there is potential for a doubling of CAT production by reducing this diversion of carbon.

Discussion

We constructed a model of cell-free *E. coli* metabolism with CAT production based on saturation kinetics and allosteric control. We trained the model against measurements of glucose, CAT, organic acid intermediates, energy species, and amino acids. We generated an ensemble of 18,000 parameter sets that predict the experimental dataset, with the exception of some amino acids. We calculated the carbon yield of CAT production at 45% of the theoretical maximum obtained by flux balance analysis; this was in line with the CAT yield shown by the dataset. We conducted sensitivity analyses and determined that CAT production was most sensitive to the rate constants, saturation constants, and enzyme initial conditions associated with CAT synthesis, GTP synthesis, GMP synthesis, and amino acid synthesis, and to a lesser extent some amino acid initial conditions. CAT production was much less sensitive to control parameters, suggesting that while allosteric control affects some organic acid intermediates, it does not limit CAT production. The initial conditions of glucose and oxygen were even less important to CAT production, suggesting that it can be driven by other means. However, the low sensitivities to glucose and oxygen only apply in the specific case of the parameter sets that were studied. When other species' initial conditions and associated parameters are different, glucose and oxygen can become more important to CAT production.

In this study, we present the first dynamic, cell-free model of *E. coli* metabolism and biosynthesis at this scope. While the early models of Shuler and coworkers achieved large-scale, dynamic descriptions of single cells [2], and the stoichiometric models associated with the FBA approach are computationally efficient amd widespread [17], none have yet been able to construct genome-scale kinetic models of cell-free *E. coli* metabolism. We have harnessed the advantages of cell-free systems (no cell wall, no cell growth) and integrated traditional kinetics with allosteric control to build an ensemble of kinetic models that fit a comprehensive cell-free dataset, toward genome-scale. This study should

provide a foundation for genome-scale, dynamic modeling of cell-free *E. coli* metabolism, toward industrial-scale biosynthetic production.

One important direction for the future is to expand the biological description, especially with regard to amino acids. Specifically, adding more reactions that consume amino acids would improve the model's ability to predict those that show a decrase in the experimental data. Also, including specific transcription and translation steps for CAT would allow us to more accurately model the complexity and the resource cost of protein synthesis. Another area for future work is to more thoroughly sample parameter space. For the metabolites in the dataset, initial conditions were fixed at the initial data values. All other parameters were varied in a manner so as to best fit the dataset. However, the resulting ensemble may not represent every biological or practical possibility. In a different region of parameter space, the system could behave differently, including the flux distribution through the network, the accuracy and spread of ensemble fits, the relative sensitivities, and the yield as a percentage of the theoretical maximum. Testing the model under a variety of conditions could strengthen or challenge the findings of this study. Further experimentatiion could also be used to gain a deeper understanding of model performance under a variety of conditions. Specifically, CAT production performed in the absence of amino acids could inform the system's ability to manufacture them, while experimentation in the absence of glucose or oxygen could shed light on how important they are to protein synthesis, and under which conditions. Finally, the approach should be extended to other protein products. CAT is only a test protein used for model identification; the modeling framework, and to some extent the parameter values, should be protein agnostic. An important extension of this study would be to apply its insights to other protein applications, where possible.

Materials and Methods

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

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

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

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

The reaction rate was written as the product of a kinetic term ($\bar{r}_j$) and a control term (v_j), $r_j\left(\mathbf{x}, \mathbf{k}\right) = \bar{r}_j v_j$. In this study, we used either saturation or mass action kinetics. The control term $0 \leq v_j \leq 1$ depended upon the combination of factors which influenced rate process j. For each rate, we used a rule-based approach to select from competing control factors. If rate j was influenced by $1, \ldots, m$ factors, we modeled this relationship as $v_j = \mathcal{I}_j\left(f_{1j}\left(\cdot\right), \ldots, f_{mj}\left(\cdot\right)\right)$ where $0 \leq f_{ij}\left(\cdot\right) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on rate j. The function $\mathcal{I}_j\left(\cdot\right)$ is an integration rule which maps the output of regulatory transfer functions into a control variable. Each regulatory

transfer function took the form:

$$f_{ij}\left(\mathcal{Z}_i, k_{ij}, \eta_{ij}\right) = k_{ij}^{\eta_{ij}} \mathcal{Z}_i^{\eta_{ij}} / \left(1 + k_{ij}^{\eta_{ij}} \mathcal{Z}_i^{\eta_{ij}}\right) \tag{3}$$

where $\mathcal{Z}_i$ denotes the abundance factor i, k_{ij} denotes a gain parameter, and η_{ij} denotes a cooperativity parameter. In this study, we used $\mathcal{I}_j \in \{mean\}$ [**?**]. If a process has no modifying factors, $v_j = 1$. We used multiple saturation kinetics to model the reaction term $\bar{r}_j$:

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

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the scaled enzyme activity which catalyzes reaction j, and K_{js} denotes the saturation constant for species s in reaction j. The product in Equation (4) was carried out over the set of *reactants* for reaction j (denoted as m_j^-).

Generation of model ensemble We generated an ensemble of 18,000 parameter sets via a downhill-only random walk Monte Carlo method [**?**]. Beginning with a single parameter set as a starting point, we calculated its cost function, equal to the sum-absolute-error between experimental data and model predictions:

$$cost = \sum_{i=1}^{\mathcal{D}} \left(w_i \sum_{j=1}^{\mathcal{T}} abs\left(x_{ij}^{data} - x_i^{sim}|_{t(j)}\right) \right) \tag{5}$$

where $\mathcal{D}$ denotes the number of datasets, w_i denotes a weight, equal to 5 for the glucose, CAT, pyruvate, lactate, acetate, succinate, and malate datasets, and 1 elsewhere, $\mathcal{T}$ denotes the number of timepoints in the ith dataset, $t(j)$ denotes the jth timepoint, x_{ij}^{data} denotes the value of the ith dataset at the jth timepoint, and $x_i^{sim}|_{t(j)}$ dneotes the simulated value of the metabolite corresponding to the ith dataset, interpolated to the jth

timepoint. We then perturbed model parameters:

$$k_i^{new} = k_i * exp(a\, r_i) \qquad i = 1, 2, \ldots, \mathcal{P} \tag{6}$$

where $\mathcal{P}$ denotes the number of parameters, equal to 652, which includes 163 rate constants, 455 saturation constants, and 34 control parameters, k_i^{new} denotes the new value of the ith parameter, k_i denotes the current value of the ith parameter, a denotes a distribution variance, set to 0.03, and r denotes a random sample from the normal distribution. We stored the parameter set and calculated its cost; if it was less than the previous cost, we used the new parameter set to generate the following set. After generating 180,000 sets we defined the 18,000 sets with the lowest cost values as our ensemble, and the set with the lowest cost value as our best-fit set.

Global and local sensitivity analysis We conducted a global sensitivity analysis, using the variance-based method of Sobol, to estimate which of the experimentally controllable parameters affected the performance of the reduced order model [28]. This included the initial conditions of glucose, oxygen, amino acids, and enzymes. We computed the total sensitivity index of each parameter relative to a performance objective of area under the CAT curve (CAT production). We established the sampling bounds for each parameter from the value of that parameter in the set used to generate the ensemble. We used the sampling method of Saltelli *et al.* [29] to compute a family of $N\,(2d + 2)$ sets which obeyed our parameter ranges, where N was the number of trials, and d was the number of parameters in the model. In our case, N = 300 and d = 185, so the total sensitivity indices were computed from 111,600 model evaluations. The variance-based sensitivity analysis was conducted using the SALib module encoded in the Python programming language [30]. We conducted a local sensitivity analysis to estimate which of the other model parameters affected performance. This included the same parameters that were

varied in the ensemble: rate constants, saturation constants, and control parameters. The local sensitivity for each parameter was calculated across a sub-ensemble of 180 parameter sets, randomly chosen from the ensemble of 18,000 sets:

$$S_{ij} = \frac{p_{ij}}{AUC(p_{ij})} \frac{AUC(p_{ij} + \Delta p_{ij}) - AUC(p_{ij})}{\Delta p_{ij}} \qquad i = 1, 2, \ldots, \mathcal{E} \qquad j = 1, 2, \ldots, \mathcal{P}$$
$$\Delta p_{ij} = 0.001 \; p_{ij} \tag{7}$$

where $\mathcal{E}$ denotes the number of parameter sets in the sub-ensemble, equal to 180, $\mathcal{P}$ denotes the number of parameters, equal to 652, S_{ij} denotes the sensitivity of the jth parameter for the ith parameter set, p_{ij} denotes the value of the jth parameter for the ith parameter set, Δp_{ij} denotes the perturbation of the jth parameter for the ith parameter set, equal to 0.1% of the parameter value, and $AUC()$ denotes the area under the CAT curve. We then calculated the mean and standard error of each local sensitivity across the sub-ensemble of 180 sets.

Calculation of CAT yield The yield on CAT production was calculated for three cases: the experimental data, the best-fit parameter set, and a theoretical maximum yield. In each case the yield was formulated as a ratio of carbon produced as CAT to carbon consumed as reactants (glucose and amino acids):

$$Yield = \frac{\Delta CAT \; C_{CAT}}{\sum_{i=1}^{\mathcal{R}} \max(\Delta m_i, 0) \; C_{m_i}} \tag{8}$$

where ΔCAT denotes the amount of CAT produced, C_{CAT} denotes carbon number of CAT, $\mathcal{R}$ denotes the number of reactants, Δm_i denotes the amount of the ith reactant consumed, never allowed to be negative, and C_{m_i} denotes the carbon number of the ith reactant. Because no data was available for arginine or glutamate, these reactants were left out of all three calculations. In the experimental case and the best-fit set case, yield

was calculated by setting ΔCAT equal to the final minus the initial CAT concentration and setting Δm_i equal to the initial minus the final reactant concentration. The theoretical yield was calculated using flux balance anaylsis (FBA) with a sequence-based analysis on CAT. The sequence specific FBA [31] problem was formulated as:

$$\begin{aligned} &\max_{\boldsymbol{v}} \left(v_{obj} = \boldsymbol{c}^T \boldsymbol{v}\right) \qquad \alpha_i \leq v_i \leq \beta_i \qquad i = 1, 2, \ldots, \mathcal{R} \\ &\text{Subject to}: \ \mathbf{Sv} = \mathbf{0} \end{aligned} \tag{9}$$

where S denotes the stoichiometric matrix, $\mathbf{v}$ denotes the unknown flux vector, $\boldsymbol{c}$ denotes the objective selection vector, and α_i and β_i denote the lower and upper bounds on flux v_i, respectively. The stoichiometric matrix was expanded to include the transcription and translation reactions for producing CAT. The objective v_{obj} was to maximize the specific rate of CAT formation. The specific glucose uptake rate was constrained to allow a maximum flux of 10 mM/hr [32]; the amino acids and oxygen uptake rates were also bound to allow a maximum flux of 10 mM/hr, but did not reach this maximum flux. Glucose, oxygen, and amino acids were modeled as being imported into the system, whereas CAT synthesis was modeled through protein export. The rest of the network followed a psuedo steady-state asusmption where all other metabolites were not allowed to accumulate; thus, the network could be solved by linear programming. The flux balance analysis problem was solved using the GNU Linear Programming Kit (v4.52) [33]. The solution flux vector was used to calculate the theoretical carbon yield of CAT, reformulated in terms of flux:

$$Yield = \frac{v_{CAT}\ C_{CAT}}{\sum_{i=1}^{\mathcal{R}} \max(v_i, 0)\ C_i} \tag{10}$$

where v_{CAT} denotes the CAT export flux and v_i denotes the import flux of the ith substrate.

Acknowledgements

This study was supported by an award from [FILL ME IN].

References

1. Fredrickson AG (1976) Formulation of structured growth models. Biotechnol Bioeng 18: 1481-6.
2. Domach MM, Leung SK, Cahn RE, Cocks GG, Shuler ML (1984) Computer model for glucose-limited growth of a single cell of escherichia coli b/r-a. Biotechnol Bioeng 26: 203-16.
3. Steinmeyer D, Shuler M (1989) Structured model for Saccharomyces cerevisiae. Chem Eng Sci 44: 2017 - 2030.
4. Wu P, Ray NG, Shuler ML (1992) A single-cell model for cho cells. Ann N Y Acad Sci 665: 152-87.
5. Castellanos M, Wilson DB, Shuler ML (2004) A modular minimal cell model: purine and pyrimidine transport and metabolism. Proc Natl Acad Sci U S A 101: 6681-6.
6. Atlas JC, Nikolaev EV, Browning ST, Shuler ML (2008) Incorporating genome-wide dna sequence information into a dynamic whole-cell model of escherichia coli: application to dna replication. IET Syst Biol 2: 369-82.
7. Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT (1985) A cybernetic view of microbial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27: 1-9.
8. Kompala DS, Ramkrishna D, Jansen NB, Tsao GT (1986) Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. Biotechnol Bioeng 28: 1044-55.
9. Kim JI, Song HS, Sunkara SR, Lali A, Ramkrishna D (2012) Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. Biotechnol Prog 28: 1160-6.
10. Varner J, Ramkrishna D (1999) Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. Metab Eng 1: 88-116.
11. 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 BØ (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10: 291-305.
14. Edwards JS, Palsson BØ (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 BØ, 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 BØ (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. Sobol I (2001) Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and Computers in Simulation 55: 271 - 280.
29. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, et al. (2010) Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. Computer Physics Communications 181: 259 - 270.
30. Herman JD. http://jdherman.github.io/salib/.
31. Allen TE, Palsson BØ (2003) Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. Journal of Theoretical Biology 220: 1 - 18.
32. Mahadevan R, Edwards JS, Doyle FJ 3rd (2002) Dynamic flux balance analysis of diauxic growth in escherichia coli. Biophysical Journal 83: 1331 - 1340.

33. (2016). GNU Linear Programming Kit, Version 4.52. URL http://www.gnu.org/software/glpk/glpk.html.

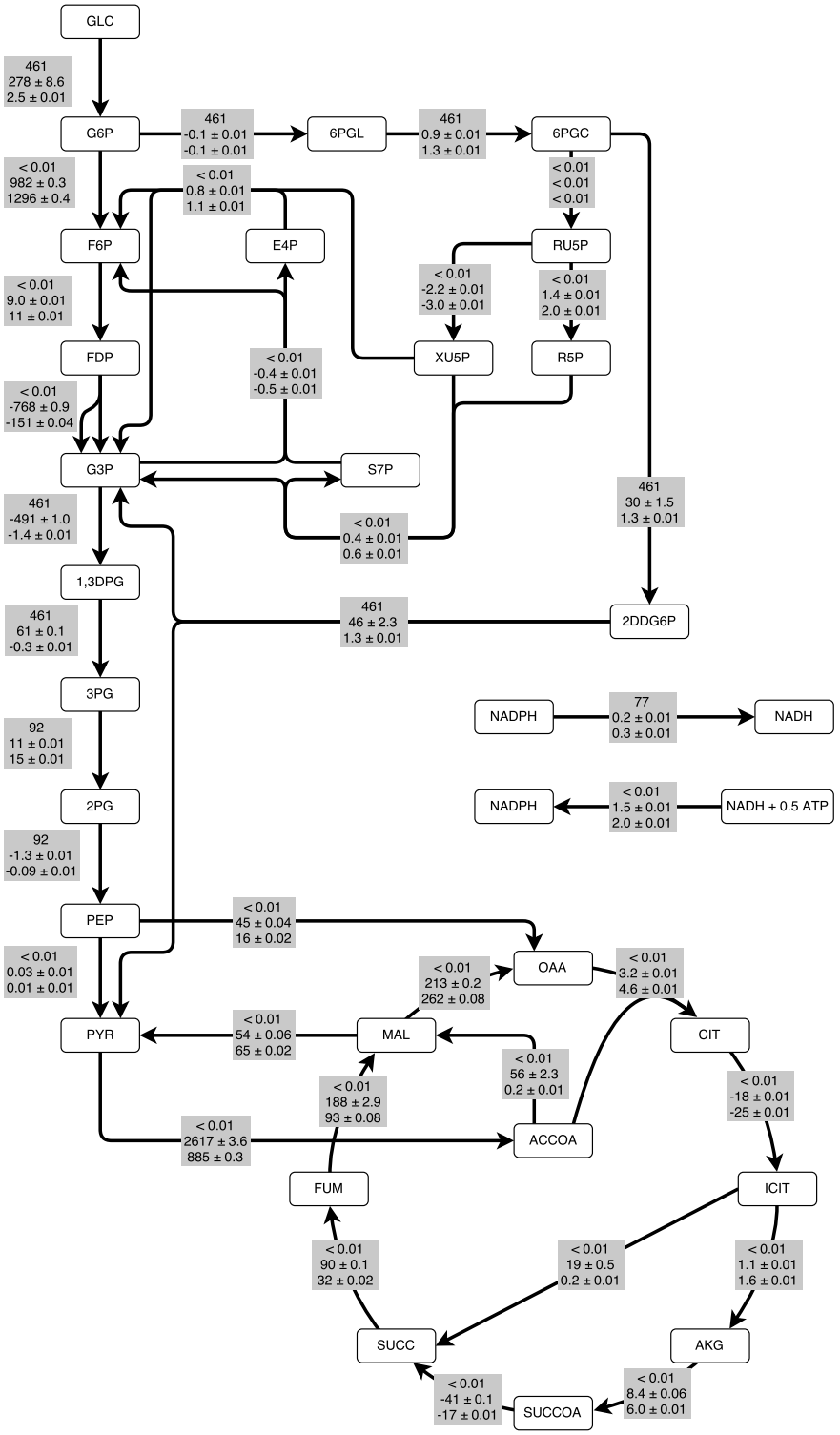

Fig. 1: Flux profile for glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, TCA cycle, and NADPH/NADH transfer. FBA flux value (top), and mean ± standard error across ensemble at 1.5 hrs (middle) and 3 hrs (bottom), normalized to CAT synthesis flux.

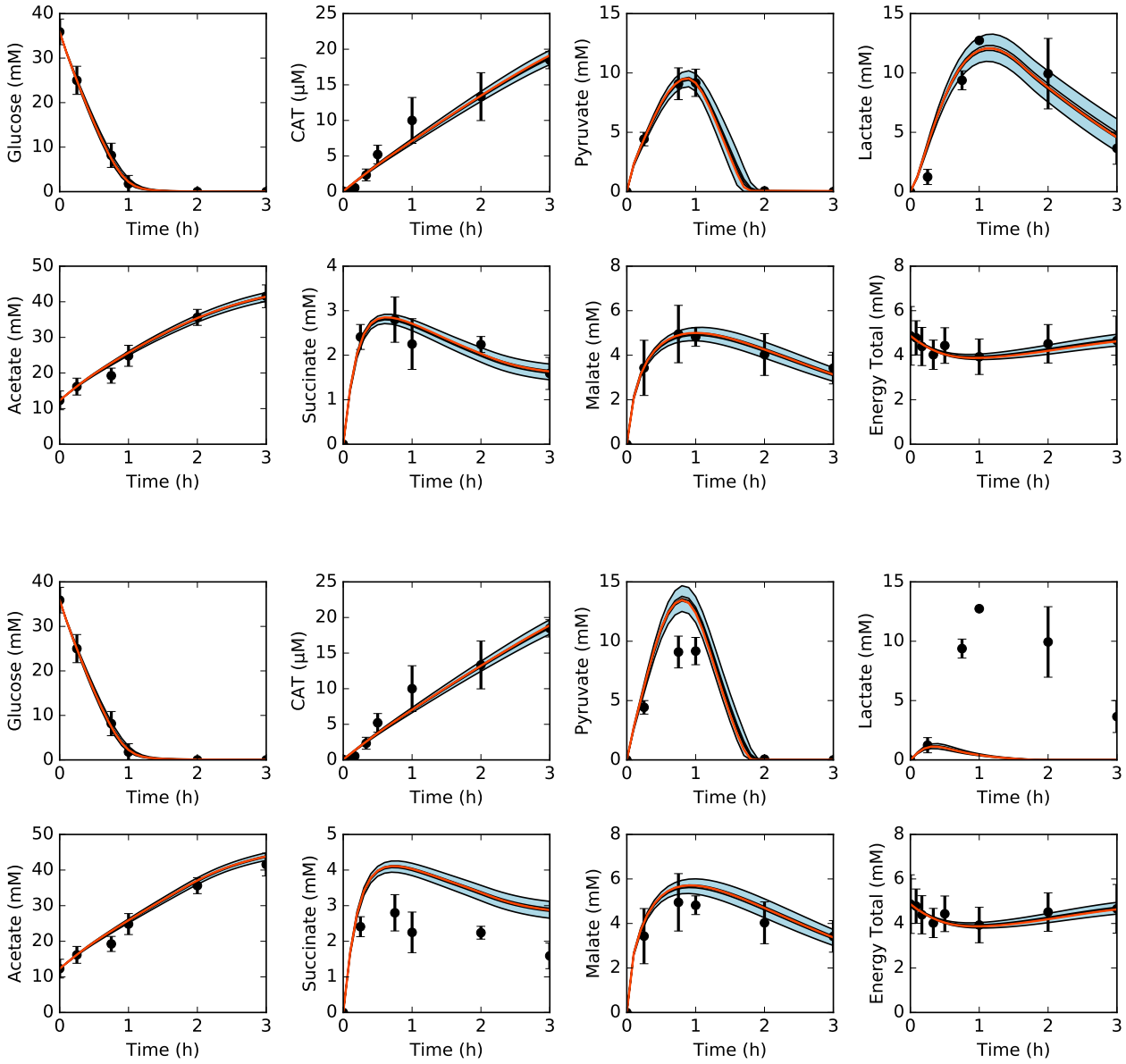

Fig. 2: Central carbon metabolism in the presence (top) and absence (bottom) of allosteric control, including glucose (substrate), CAT (product), and intermediates, as well as total concentration of energy species. Best-fit parameter set (orange line) versus experimental data (points). 95% confidence interval (blue shaded region) and 95% confidence interval of the mean (gray shaded region) over the ensemble of 18,000 sets.

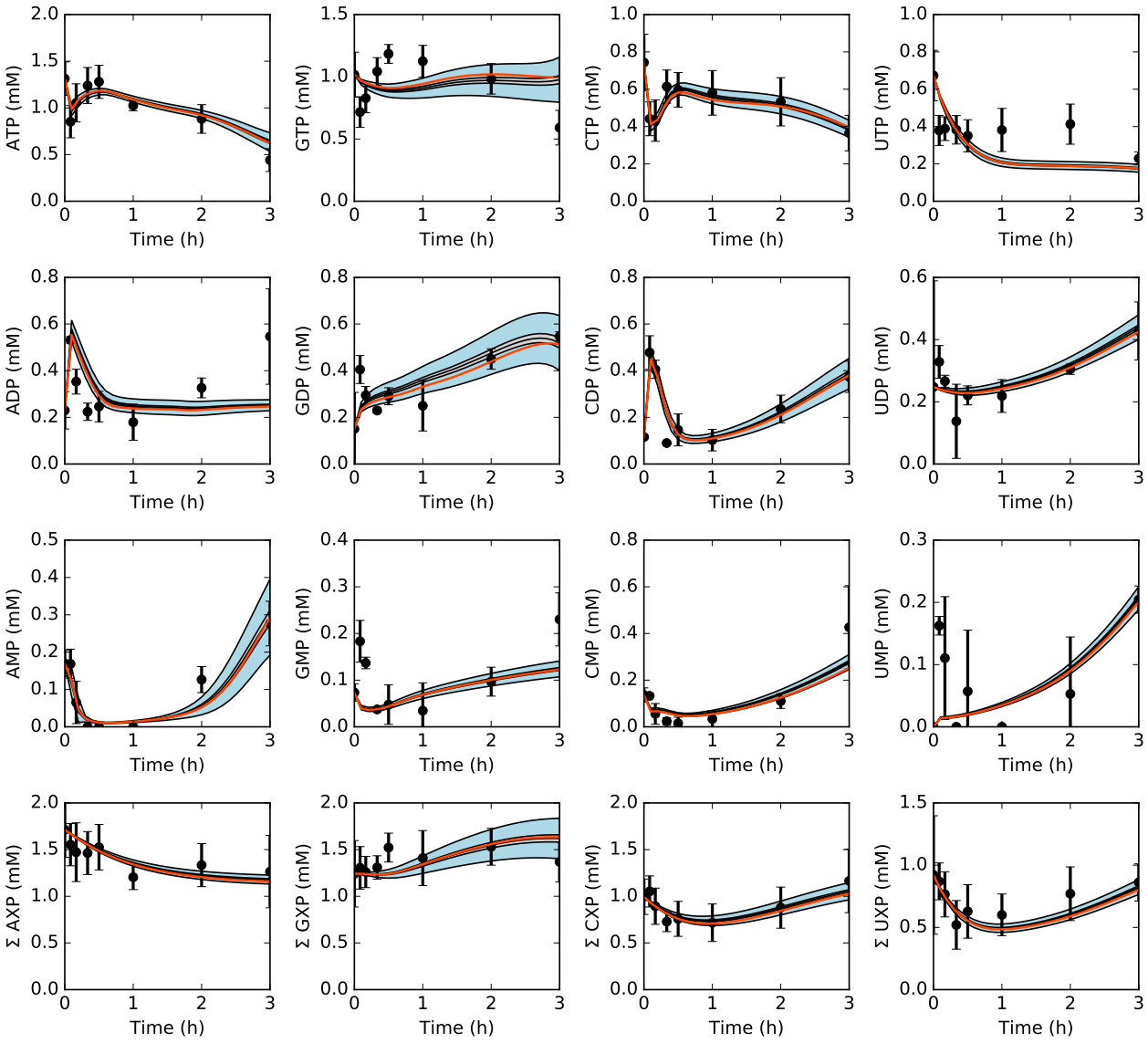

Fig. 3: Energy species and energy totals by base in the presence of allosteric control. Best-fit parameter set (orange line) versus experimental data (points). 95% confidence interval (blue shaded region) and 95% confidence interval of the mean (gray shaded region) over the ensemble of 18,000 sets.

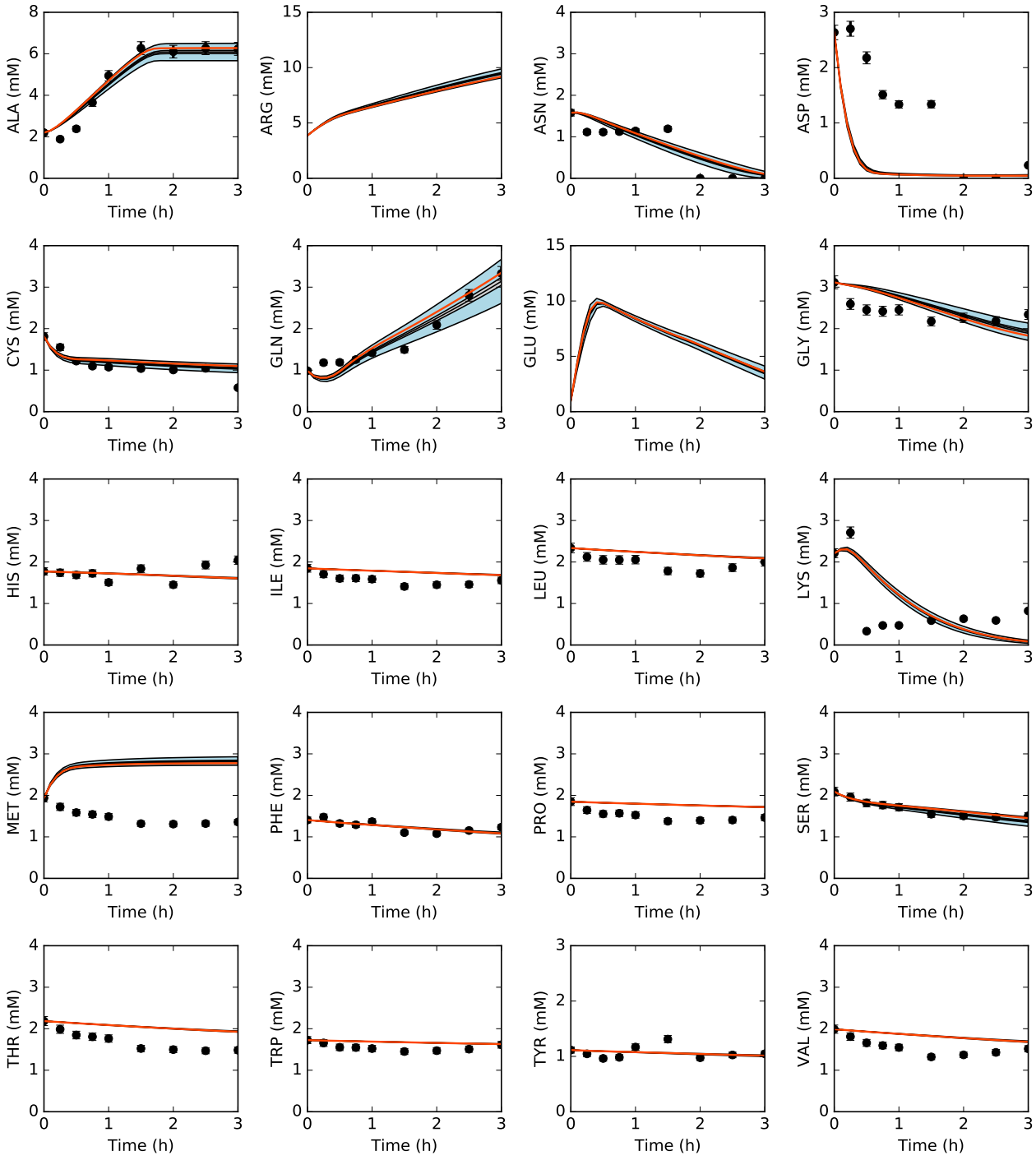

Fig. 4: Amino acids in the presence of allosteric control. Best-fit parameter set (orange line) versus experimental data (points). 95% confidence interval (blue shaded region) and 95% confidence interval of the mean (gray shaded region) over the ensemble of 18,000 sets.

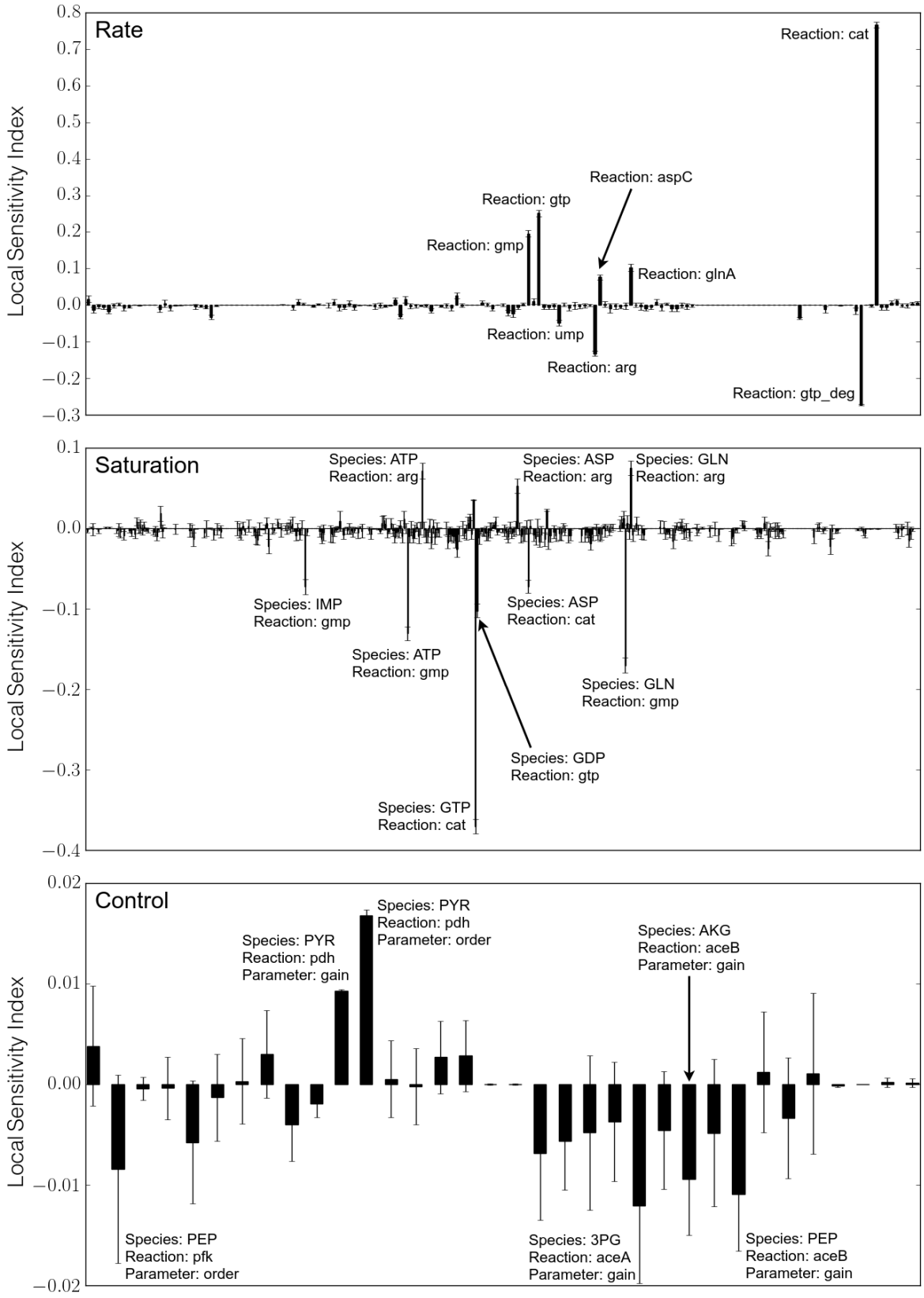

Fig. 5: Mean and standard error of local sensitivities of rate constants (top), saturation constants (middle), and control parameters (bottom).

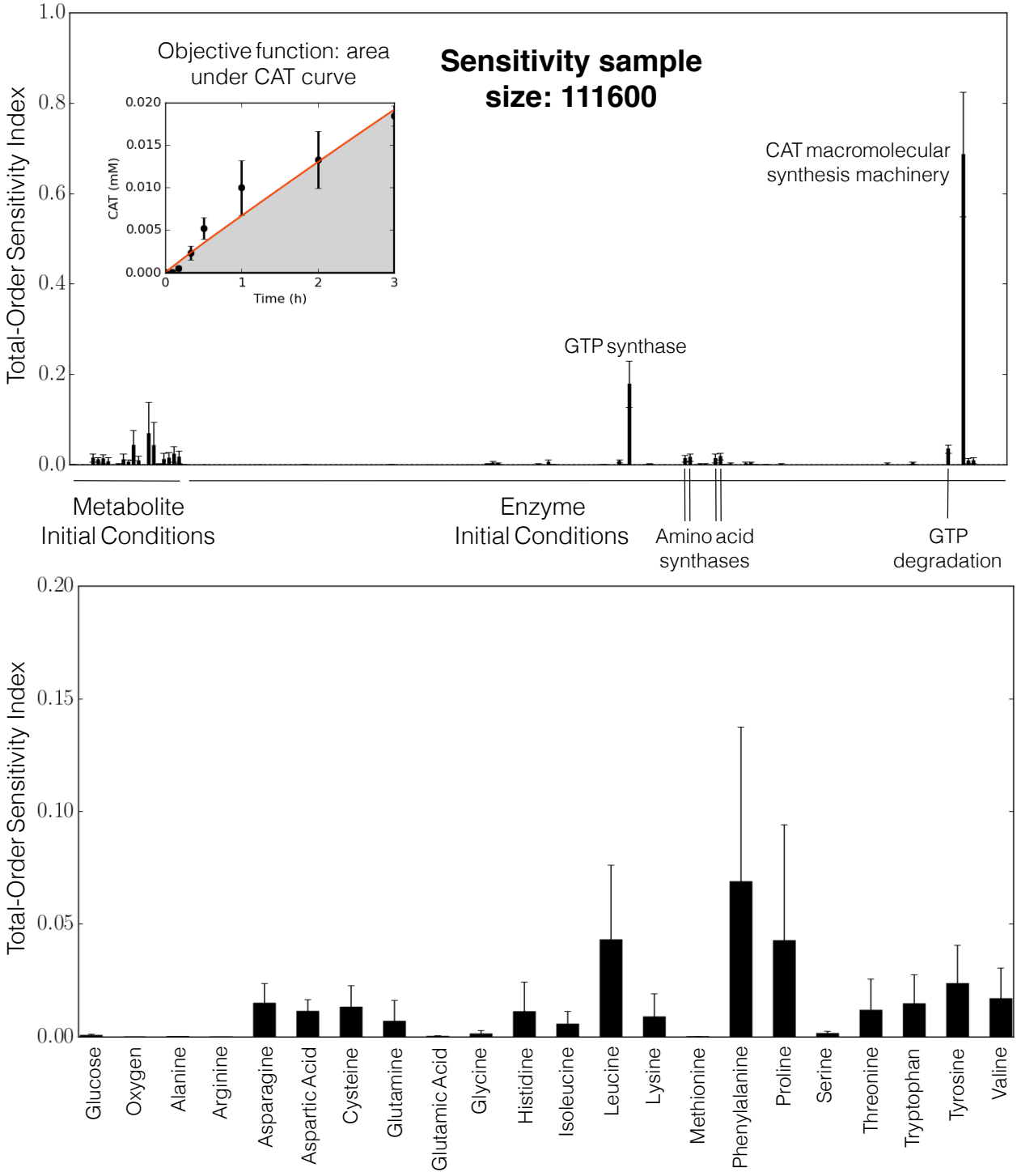

Fig. 6: Total-order global sensitivities for experimentally controllable initial conditions, including glucose, oxygen, amino acids, and enzymes.