Sequence Specific Modeling of Cell-Free Protein Synthesis

Michael Vilkhovoy, Nicholas Horvath, and Jeffrey D. Varner*

Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853

E-mail: jdv27@cornell.edu

Phone: +1 (607) 255-4258. Fax: +1 (607) 255-9166

Abstract

In this study, we used sequence specific constraints based modeling to evaluate the performance of synthetic circuits in an *E. coli* cell-free protein synthesis system. A core *E. coli* metabolic model, consisting of glycolysis, pentose phosphate pathway, amino acid biosynthesis and degradation and energy metabolism, was then augmented with sequence specific descriptions of genetic circuits which included mechanistic models of promoter function, transcription and translation. Thus, unlike other synthetic biology modeling efforts, sequence specific constraints based modeling explicitly couples the transcription and translation of circuit components with the availability of metabolic resources. Model parameters for transcription and translation were taken from literature, allowing a first principles prediction of circuit performance. We tested this approach by first simulating T7 induced CAT production and σ_{70}-induced deGFP expression; we then expanded these studies for a range of different proteins. First principles predictions of circuit performance were consistent with measurements for a variety of cases. Further, global sensitivity analysis identified the key metabolic processes that controlled circuit performance in terms of productivity and yield. A

sufficient energy supply with oxidative phosphorylation is instrumental for high carbon yields; in addition, the translation rate is the next step to optimize for higher productivity and yield. Taken together, sequence specific constraints based modeling offers a novel means to *a priori* estimate the performance of cell-free synthetic circuits.

Keywords

Synthetic biology, Constraints based modeling, Biochemical modeling

1 Introduction

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 (*1*). However, CFPS is not new; CFPS in crude *E. coli* extracts has been used since the 1960s to explore fundamentally important biological mechanisms (*2*, *3*). Today, cell-free systems are used in a variety of applications ranging from therapeutic protein production (*4*) to synthetic biology (*5*). Interestingly, many of the challenges confronting in-vivo 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.

Stoichiometric reconstructions of microbial metabolism popularized by constraint based modeling techniques such as flux balance analysis (FBA) have become standard tools to interrogate biological networks (*6*). Since the first genome-scale stoichiometric model of *E. coli*, developed by Edwards and Palsson (*7*), stoichiometric reconstructions of hundreds of organisms, including industrially important prokaryotes such as *E. coli* (*8*) or *B. subtilis* (*9*), are now available (*10*). 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 using linear programming. 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 (*11*), although other metabolic objectives have also been estimated (*12*). Recent advances in constraint-based modeling have overcome the early shortcomings of the platform, including capturing metabolic regulation and control (*13*). Thus, modern constraint-based approaches are extremely useful for the discovery of metabolic engineering strategies and represent the state of the art in metabolic modeling (*14*, *15*).

In this study, we used sequence specific constraints based modeling to evaluate the performance of synthetic circuits in an *E. coli* cell-free protein synthesis (CFPS) system. A core *E. coli* cell free metabolic model, consisting of 132 metabolites and 281 reactions, was developed from literature (*8*). This model, which described glycolysis, pentose phosphate pathway, amino acid biosynthesis and degradation and energy metabolism, was then augmented with sequence specific descriptions of genetic circuits which included mechanistic models of promoter function, transcription and translation. Thus, sequence

specific constraints based modeling explicitly coupled the transcription and translation of circuit components with the availability of metabolic resources. We tested this approach by first simulating the production of CAT and deGFP under two different promoters, and then expanded this for σ_{70}-induced deGFP expression at different plasmid concentrations. First principles predictions of circuit performance were consistent with measurements. We then investigated the productivity and carbon yield for a range of 8 different proteins for a variety of cases. From this, higher carbon number proteins typically had lower productivity rates and carbon yields than that of the lower carbon number proteins. Further, global sensitivity analysis identified the key metabolic processes that controlled circuit performance, showing oxidative phosphorylation is instrumental for maintaining a high carbon yield and the translation rate for productivity. Taken together, sequence specific constraints based modeling offers a novel means to *a priori* estimate the performance of cell free synthetic circuits.

2 Results and discussion

2.1 Model Derivation

The goal of this work was first to construct a modeling framework to describe cell-free protein synthesis systems and to examine its performance in productivity and yield for a protein of interest. One mathematical framework that has found wide use in modeling metabolism is constraint based models such as flux balance analysis (FBA). FBA can predict how cells utilize nutrients to produce products by using the biochemical stoichiometry and thermodynamical feasibility under pseudo steady-state conditions. Traditionally, FBA is used to model *in vivo* processes; however, cell-free systems do not have growth associated reactions or transport through the cell membrane. Thus, to model cell-free metabolism we constructed a cell-free stoichiometric network by removing growth associated reactions from the MG1655 reconstruction (*8*), and incorporating amino acid synthesis and

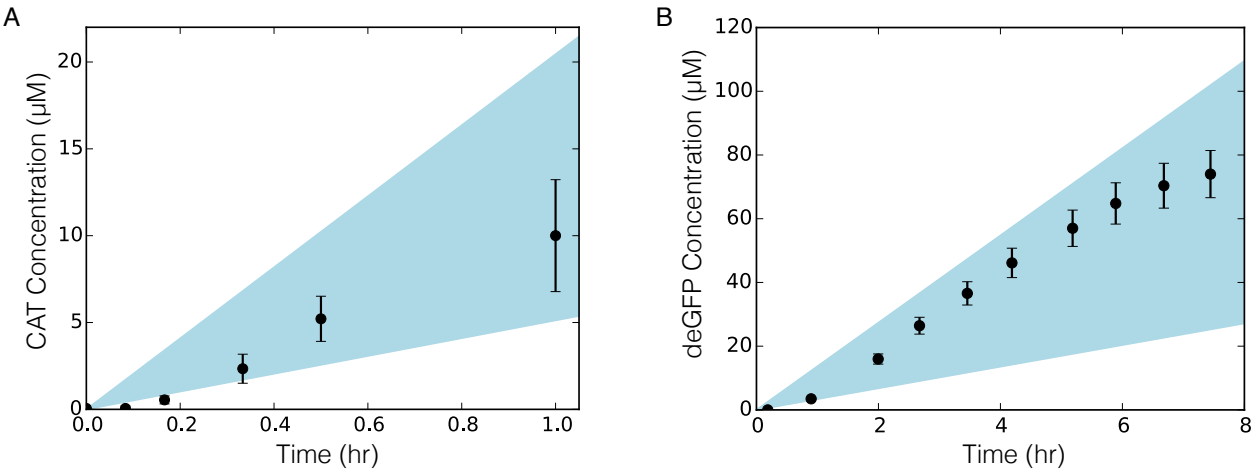

Figure 1: Sequence specific flux balance analysis of protein production vesus time. A. CAT production under a T7 promoter in CFPS *E. coli* extract for 1 hr under glucose consumption. B. deGFP production under a P70 promoter in TXTL 2.0 *E. coli* extract for 8 hr under glucose consumption. 95% CI (blue region) over the ensemble of 100 sets.

transcription/translation associated reactions (*16*) for a protein of interest to be expressed. The network consisted of 281 reactions and 132 species. We developed a cell-free sequence specific flux balance analysis (ssFBA) with a detailed promoter model (*17*) to examine the performance of CFPS. We first validated the ssFBA approach by comparing simulated and measured concentrations of two proteins from two different cell-free *E. coli* extract systems. The first protein, chloramphenicol acetyltransferase (CAT), was produced under a T7 promoter in a glucose/NMP cell-free system (*18*) for a duration of 1 hour under glucose consumption (Fig 1A). The second protein, dual emission green fluorescent protein (deGFP), was produced under a P70 promoter in TXTL 2.0 *E. coli* extract for a duration of 8 hours under maltose consumption (Fig 1B). The ssFBA simulations predicted the production of both proteins for the duration of both CFPS batch reactions. Uncertainty in experimental factors such as RNA polymerase, ribosome concentrations, elongation rates and the upper bounds for oxygen and glucose consumption rates did not alter the qualitative performance of the model. Thus, the metabolic network and molecular description of trancription and translation were consistent with experimental measurements.

Next, ssFBA predicted deGFP production as a function of plasmid concentration (Fig 2). Concentration of deGFP at each plasmid concentration was calculated by multiplying

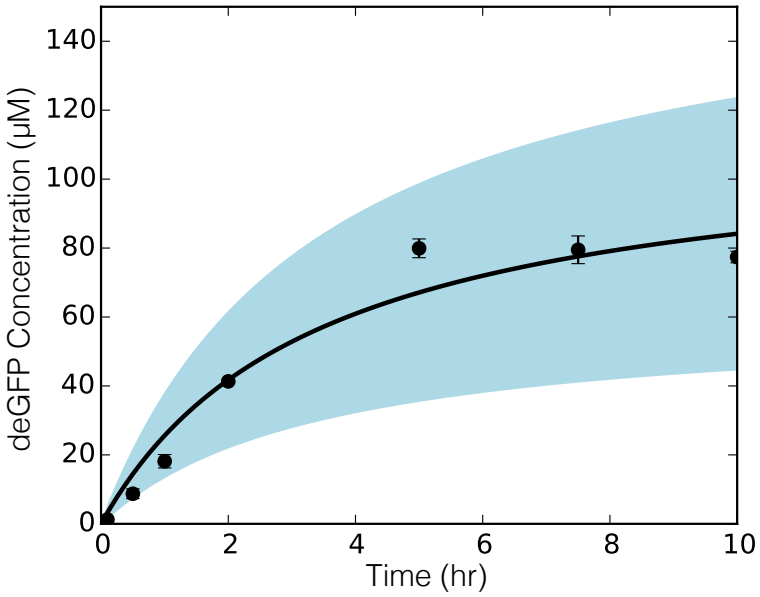

Figure 2: Predicted deGFP concentration (black line) at different plasmid concentrations versus measurements of deGFP (dots) synthesized in TXTL 2.0. 95% CI (blue region) over the ensemble of 100 sets.

the flux of deGFP synthesis by the active time of production, approximately 8 hours in TXTL 2.0 (*19*). The mean of the ensemble shows a good prediction against the measured deGFP levels, even though it under predicted deGFP concentration at the saturating point of 5 nM of plasmid concentration. However, the ensemble and the mean of the ensemble captured the overall saturating dynamics of deGFP production as a function of plasmid concentration.

These results validated our mathematical framework to model CFPS systems and predict the production of two proteins with no adjustable parameters. It also showed that the sequence specific reactions were sufficient to predict the production of two different proteins under two different promoters and cell-free systems. Since the model accurately predicted the behavior of protein production, we wanted to use our mathematical framework to help us understand the performance limits of CFPS and how these shortcomings could be addressed.

2.2 CFPS Productivity

Our next goal was to examine the performance of CFPS productivity for eight different proteins under three different cases. Each of the proteins was produced under a P70 promoter, except for CAT which was produced under a T7 promoter. In all cases, CFPS was supplied with glucose. In the first case, CFPS was supplied with amino acids, and the system was allowed to synthesize amino acids (control). In the second case, CFPS was supplied with amino acids, but the amino acid synthesis was turned off (AA uptake w/o synthesis). We shut off these synthesis reactions since during the cell-free extract preparation the cells are often supplied with amino acids; thus, the enzymes responsible for amino acid synthesis would not be present. In the third case, CFPS was not supplied with amino acids, but the system could synthesize them (AA synthesis w/o uptake). Eight different proteins, ranging in size, were selected to evaluate CFPS performance: bone morphogenetic protein 10 (BMP10), chloramphenicol acetyltransferase (CAT), caspase 9 (CASP9), dual emission green fluorescent protein (deGFP), prothrombin (FII), coagulation factor X (FX), fibroblast growth factor 21 (FGF21), and single chain variable fragment R4 (scFvR4). We used ssFBA to estimate the productivity for each of these proteins for each case (Fig 3A). The first two cases had very similar performance, with the second case (without amino acid synthesis) having a slightly higher productivity for most proteins. This shows the system had sufficient substrates to power the system and synthesize each protein. The third case had the lowest productivity for each protein; thus, the addition of amino acids to CFPS extract is important for maintaining a relatively high productivity. The drop in productivity in the third case is due to the fact that glucose is the only substrate supplied and must be used to provide the necessary energy requirements for transcription and translation as well as synthesize each amino acids required for the production of the protein of interest. The qualitative trend of productivity for the three different cases was the same within each case; however, some proteins had higher productivity than others. For instance, in the second case, FGF21 (fibroblast growth factor 21) had a productivity of

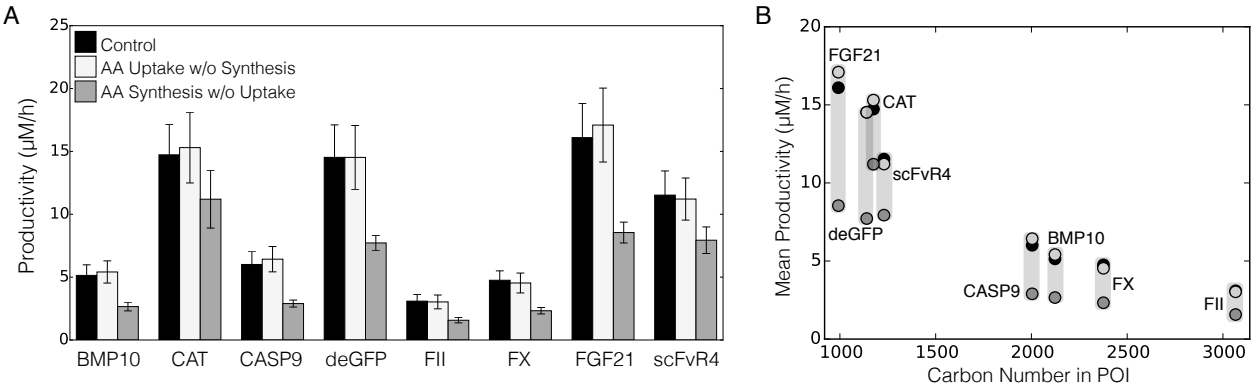

Figure 3: CFPS productivity performace for eight proteins for the control (black), AA uptake w/o synthesis (off white), and AA synthesis w/o uptake (grey). A. Productivity normalized to the specific glucose uptake rate (Error bars represent the 95% CI of the ensemble). B. Mean productivity versus the carbon number for each corresponding protein.

17 (μM/h) whereas FII (prothrombin) had a productivity of 3 (μM/h). To examine this further, the mean productivity was plotted against the carbon number of each protein (Fig 3B). The proteins with the highest productivity had the lowest carbon number, whereas proteins with low productivity had higher carbon numbers. This inverse trend was due to the fact that larger proteins require more amino acids and substrates to assemble them, resulting in lower productivity.

2.3 CFPS Carbon Yield

Following the same outline as in examining the productivity, we calculated the carbon yield for each protein (Fig 4A). The same trends followed, where the case without amino acid synthesis showed the highest carbon yield and the control case had comparable performance. The third case (with no amino acid uptake) had the lowest yields; this is most likely because glucose is utilized to synthesize the necessary amino acids for each protein as well as power the system. To determine the carbon contribution from each substrate (glucose and amino acids) we examined the carbon flux going toward the production of deGFP for all three cases (Table 1). For the first case, the system relied on a mixture of glucose and some amino acids for the production of deGFP with a carbon yield of 20.5%. Once amino acid synthesis was removed from the network (second case), each amino acid

was utilized and the carbon yield increased to 22.2%. Only the necessary amount of amino acids were used for the production of deGFP; thus, it may be hypothesized that all the glucose was used to power CFPS and does not contribute to the carbon yield. If that is the case, the carbon yield without glucose contribution is 100%, whereas for the first case it is 92.5%. Finally for the third case (without amino acids supplemented), the carbon yield was reduced to 14.8% and the system used about twice the amount of glucose as in the first two cases. In this case, glucose is being used to synthesize amino acids and provide the necessary energy requirements to power transcription and translation; this trend was seen across all proteins except for CAT. Interestingly, CAT carbon yield showed similar performance for all three cases; however, it still had a lower yield than the first two cases. Each protein had different energy requirements for transcription and translation depending on its sequence. In the case of CAT, energy requirements for transcription or translation were low enough that its carbon yield was not hindered in the third case as significantly as for the other proteins.

We next investigated the effect of the carbon number of each protein on the carbon yield (Fig 4B). The same inverse qualitative trend is observed as for productivity, but it was less significant. The proteins with the lowest carbon number had the highest yield and the higher carbon number proteins had a lower carbon yield within each case. As the protein length increased, the productivity and carbon yield decreased, suggesting that large proteins may not be feasible for cell-free production. We thus wanted to examine the most sensitive parameters for cell-free productivity and carbon yield in order to optimize CFPS performance.

2.4 CFPS performance sensitivity analysis

To better understand the effect of substrate utilization and the transcription/translation parameters on CFPS performance we performed global sensitivity analysis on the productivity and carbon yield for deGFP, a representative protein (Fig 5). In examining

Table 1: Carbon contribution from glucose and each amino acid toward deGFP production for three cases: control, amino acid uptake without synthesis, and amino acid synthesis without uptake.

Carbon Produced (mM/hr)	Control	AA uptake w/o synthesis	AA synthesis w/o uptake
deGFP	9.8	9.6	9.9
Carbon Consumed (mM/hr)			
GLC	37.3	33.7	66.9
ALA	0.0	0.2	-
ARG	0.3	0.3	-
ASN	0.5	0.4	-
ASP	0.4	0.6	-
CYS	0.2	0.1	-
GLU	2.1	0.6	-
GLN	0.4	0.3	-
GLY	0.3	0.3	-
HIS	0.5	0.5	-
ILE	0.0	0.6	-
LEU	1.0	1.0	-
LYS	1.0	0.9	-
MET	0.0	0.2	-
PHE	1.0	0.9	-
PRO	0.4	0.4	-
SER	0.0	0.2	-
THR	1.0	0.5	-
TRP	0.1	0.1	-
TYR	0.8	0.8	-
VAL	0.6	0.7	-
Sum	47.9	43.3	66.9
Yield	20.5%	22.2%	14.8%
Yield w/o GLC	92.5%	100%	-

productivity performance (Fig 5A), the significance of transcription/translation parameters was fairly constant across all three cases, with the elongation rate by ribosomes being the most significant. This showed that translation rate is instrumental for productivity, and should be the first step investigated in its optimization, prior to examining transcription rates. This is consistent with literature findings. Underwood and coworkers have also shown that an increase in ribosome levels does not significantly increase protein yields or rates; however, adding elongation factors increased yields by 23% at 30 minutes (*20*).

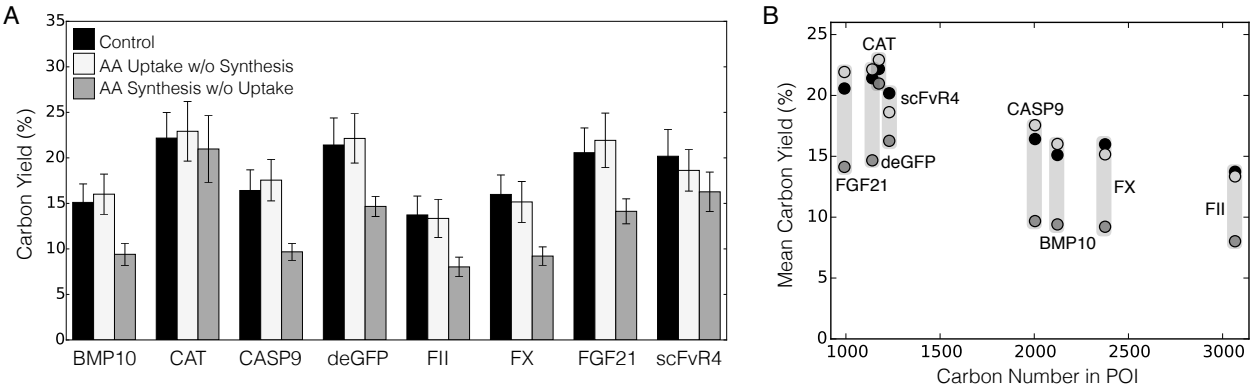

Figure 4: CFPS carbon yield performace for eight proteins for the control (black), AA uptake w/o synthesis (off white), and AA synthesis w/o uptake (grey). A. Carbon yield (Error bars represent the 95% CI of the ensemble). B. Mean carbon yield versus the carbon number for each corresponding protein.

In addition, Li et al. have increased productivity of firefly luciferase by 5-fold in CFPS systems by adding and adjusting factors that affect transcription and translation such as elongation factors, ribosome recyclcing factor, release factors, chaperones, BSA, and tRNAs (*21*). In examining the substrate utilization, glucose uptake was not very important for productivity in the first two cases, but its significance increased when amino acids were removed from CFPS. This makes sense, as amino acid synthesis from glucose became the only way to power protein synthesis. On the other hand, when amino acid synthesis was removed, amino acid uptake became more important, as it was the only source of amino acids.

When considering carbon yield performance (Fig 5B), the substrate and oxygen uptake fluxes became much more important while the sensitivity to transcription/translation parameters decreased slightly. The transcription/translation parameters had the same trend as for productivity, where the ribosome elongation rate was the most sensitive compared to the other transcription/translation parameters and showed significance across all cases. Thus, in investigating CFPS performance in terms of productivity and yield, the ribosome elongation rate is an important parameter to optimize, as has already been shown in literature (*20*, *21*). Glucose and oxygen uptake showed significant importance for carbon yield since they determined the mechanism of energy generation, which is

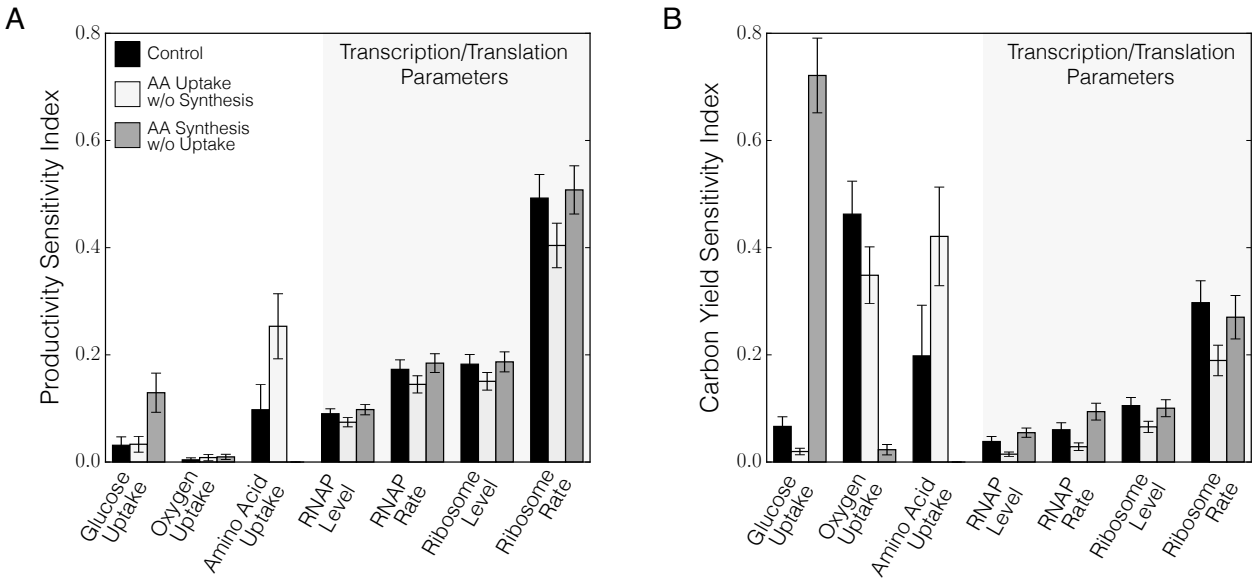

Figure 5: Total order sensitivity of deGFP productivity (A) and carbon yield (B) to specific uptake rates and transcription/translation parameters for three cases: control (black), amino acid uptake without synthesis (off white), and amino acid synthesis without uptake (gray). (Error bars represent 95% CI of the ensemble).

critical for efficient protein production. Meanwhile, productivity was determined primarily by the rate of the most downstream steps, transcription and translation. Glucose is required for both amino acid synthesis and efficient energy generation, both of which were important for a high yield. This resulted in a tradeoff between energy generation to power transcription/translation and amino acid synthesis required to assemble the protein of interest. Across all three cases, substrate utilization (amino acid uptake and glucose) showed to be the next important as these substrates contributed to the carbon yield of deGFP. Interestingly, in the first two cases, oxygen uptake also had a significant effect on carbon yield. This is most likely due to the importance of oxidative phosphorylation for efficient energy generation. With high oxidative phosphorylation, carbon substrates were not wasted to meet the necessary energy requirements of CFPS; instead, carbon was directed toward the production of deGFP. However, in the third case (with no amino acid uptake), oxygen uptake significance decreased, while glucose uptake became the most siginificant. When amino acids were not supplied in CFPS, glucose was the only carbon source for amino acid synthesis and energy generation. The same tradeoff was observed

between utilizing glucose for amino acid synthesis and efficient energy generation; the significance of oxygen was irrelevant compared to that of glucose uptake. In addition, NADH was required to power oxidative phosphorylation; however, the demand for NADPH for amino acid syntheis was higher, since amino acids were required for the production of the protein of interest. NADH was thus drawn away from oxidative phosphorylation to be interconverted to NADPH. This decreased the significance of oxygen uptake, since there was not enough NADH to power oxidative phosphorylation. Jewett and coworkers have reported oxidative phosphorylation still operating in cell-free systems, and that yield decreased from 1.5-fold to 4-fold when oxidative phosphorylation reactions were knocked out(*1*). However, it is unknown how active oxidative phosphorylation is compared to that of *in vivo* systems. Thus, to investigate this further we compared deGFP carbon yield to oxidative phosphorylation flux (Fig 6). Interestingly, oxidative phosphorylation has a somewhat linear correlation with carbon yield for the first two cases (R^2=0.77 for the control, R^2=0.79 for the second case), whereas for the third case, this effect was not as prevalent (R^2=0.46).

When the system can utilize glucose for energy generation, the increase in oxidative phosphorylation has a positive effect on carbon yield. However, when glucose is used to synthesize amino acids and power CFPS, there is a limit on how much oxidative phosphorylation can effect carbon yield due to the limitation of NADH to power oxidative phosphorylation. Thus, it is interesting to see how much of the energetic needs of the system are met by oxidative phosphorylation and where the remaining energy is coming from. To investigate the differences in productivity and carbon yields, we compared the flux distributions for deGFP predicted by ssFBA simulations for the three different cases (Fig.7).

All cases used the first step in the pentose phosphate pathway to generate NADPH; the carbon flux then continued through the Entner–Doudoroff pathway toward pyruvate. The majority of the flux was distributed between lactate and acetate, representing an

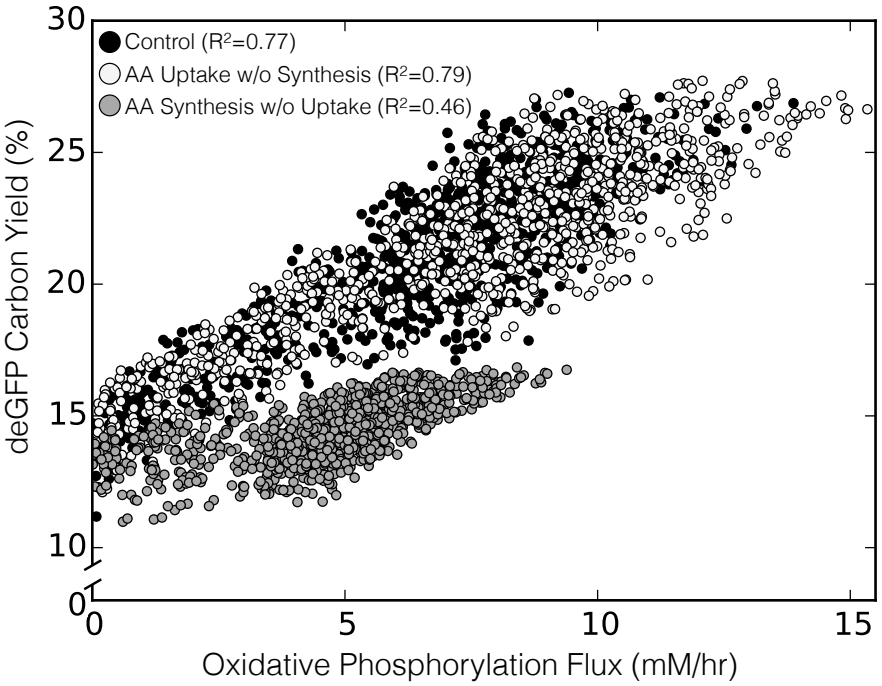

Figure 6: An ensemble of 1000 ssFBA solutions for deGFP carbon yield versus oxidative phosphorylation flux for three cases: control (black), amino acid uptake without synthesis (off white), and amino acid synthesis without uptake (gray).

inefficiency in energy generation as well as the accumulation of wasteful byproducts. For the first two cases, the energy source was primarily oxidative phosphorylation, and to a lesser extent the TCA cycle. The second case with no amino acid synthesis showed to be the most efficient process (highest carbon yield). All the NADPH produced was interconverted to NADH to power oxidative phosphorylation. Oxidative phosphorylation lead to a high redox ratio contributing to the accumulation of acetate. Diverting flux away from acetate accumulation may allow more resources towards the protein of interest. Interestingly, for the first case and especially for the third case, there was simultaneous aerobic and anaerobic energy generation, which is imposible for *in vivo* systems. Cell-free systems do not have the same sense of regulation as *in vivo* systems, so any enzyme present in the extract will catalyze its corresponding reaction. In the third case (third row), there was a low flux through oxidative phosphorylation and a high flux through fumerate dehydrogenase (fumerate to succinate). Fumerate dehydrogenase typically operates in anaerobic conditions and was less efficient than oxidative phosphorylation. One of the

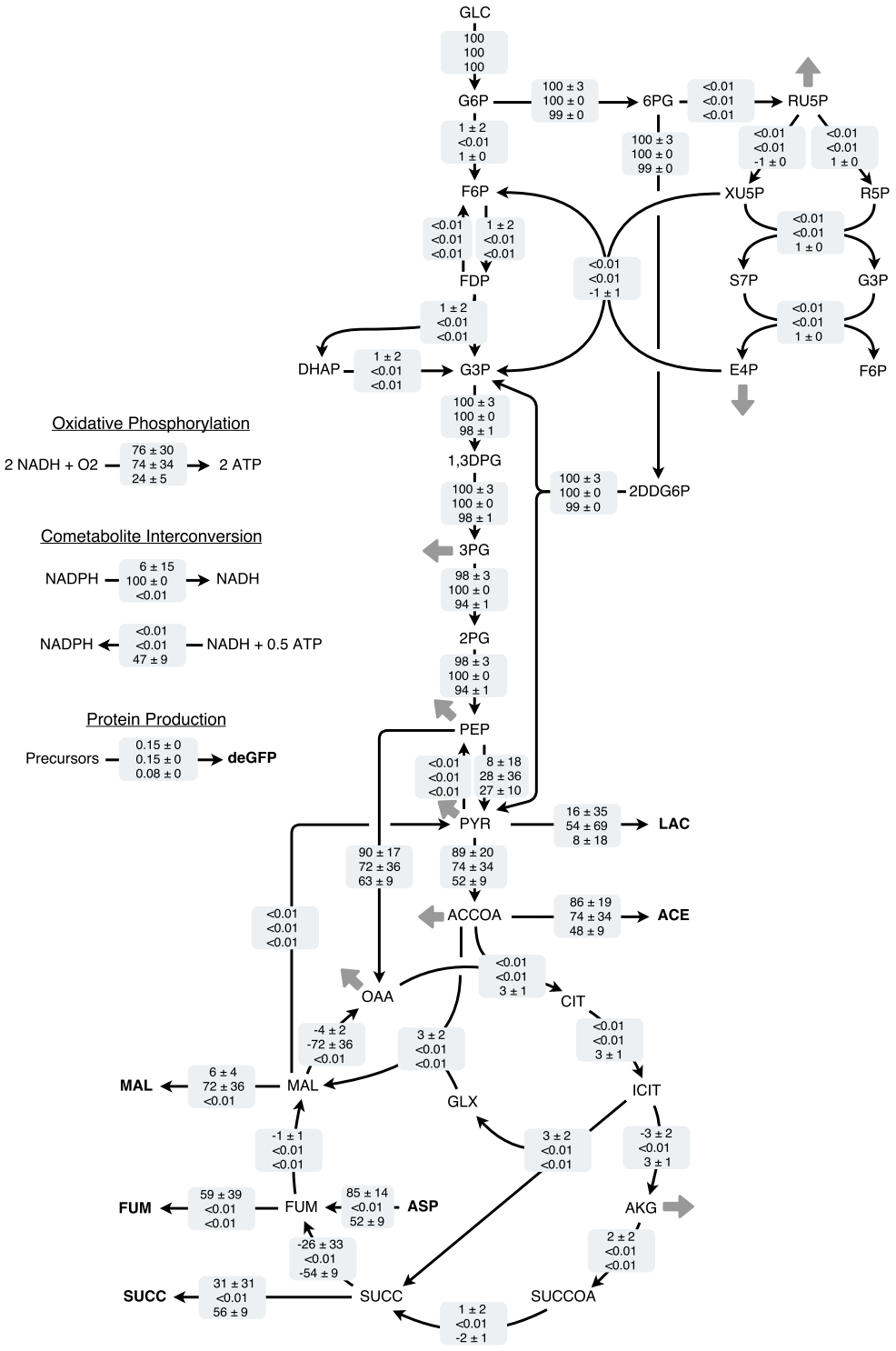

Figure 7: deGFP production flux profile for glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, TCA cycle, NADPH/NADH transfer, and oxidative phosphorylation. Flux across ensemble (mean ± standard deviation) normalized to glucose uptake flux. Flux distribution for three different cases: control (top row), amino acid uptake without synthesis (second row), and amino acid synthesis without uptake (third row). Bold metabolites are allowed to accumulate; grey arrows lead to amino acid synthesis.

reasons for this inefficiency was that some of the NADH was interconverted to NADPH, which is required for amino acid synthesis; thus, NADH was depleted and could not power oxidative phosphorylation as effectively as in the first two cases. It is currently unknown how much oxidative phosphorylation is present in cell-free systems, but our model and the sensitivity analysis suggest that efficient energy generation is essential to higher carbon yields. Thus, removing anaerobic enzymes during the cell-free extract preparation could potentially improve CFPS performance and protein yield.

All proteins with the exception of CAT followed a similar trend, where for the first two cases (when amino acids were supplied) there was a relatively high yield compared to the third case. We also performed a sensitivity analysis on CAT (Fig S1, available in the Supporting Information), which did not show the same trend in yield across the three cases as the other proteins. For CAT, glucose uptake was as important to productivity as the kinetic parameters. The oxygen uptake was the most important factor in yield and maintained this across all three cases. The energy requirements for transcription of CAT were lower than for the other proteins, so it could divert more resources towards efficient energy generation. Thus, the carbon yield was not hindered in the third case, since the system maintained the necessary amino acid synthesis rates and the required energy species. Apart from these differences, there was a general trend of transcription and translation being more significant for productivity and substrate uptake being more significant for yield.

Taken together, this represents sequence specific constraints based modeling to evaluate the performance of synthetic circuits in an *E. coli* CFPS system. We have shown first principle predictions for protein production of deGFP and CAT, under two different promoters and two different cell-free extract systems, with no adjustable parameters and in agreement with experimental measurements. This modeling approach validated trends seen in literature for productivity and yield as a function of carbon number. Furthermore, global sensitivity analysis identified oxidative phosphorylation as being instrumental for

maintaining a high carbon yield and translation rate as being important for productivity. The model also suggested that cell-free systems can simultaneously operate aerobically and anaerobically, which can lead to inefficient production and should be addressed to optimize carbon yield. In conclusion, sequence specific constraints based modeling offers a novel means to *a priori* estimate the performance of cell free synthetic circuits.

Materials and Methods

Formulation and solution of the model equations.

We estimated the theoretical maximum performance of the cell free protein synthesis system using sequence-specific flux balance analysis (ssFBA) (*16*). The sequence-specific flux balance analysis problem was formulated as a linear program:

$$\begin{aligned} \max_{\boldsymbol{w}} \left(w_{TL} = \boldsymbol{\theta}^T \boldsymbol{w} \right) \\ \text{Subject to : } \mathbf{Sw} = \mathbf{0} \\ \alpha_i \leq w_i \leq \beta_i \qquad i = 1, 2, \ldots, \mathcal{R} \end{aligned} \tag{1}$$

where $\mathbf{S}$ denotes the stoichiometric matrix, $\mathbf{w}$ denotes the unknown flux vector, $\boldsymbol{\theta}$ denotes the objective selection vector and α_i and β_i denote the lower and upper bounds on flux w_i, respectively. The stoichiometry of the kinetic model was used for the ssFBA calculations, with the execpetion of the transcription and translation rates. The transcription (TX) and translation (TL) stoichiometry was modeled using the template reactions taken from Allen

and Palsson (*16*):

$$
\begin{aligned}
\mathrm{G_{\mathcal{P}}} + \mathrm{R_1} &\longrightarrow \mathrm{G^*_{\mathcal{P}}} \\
\mathrm{G^*_{\mathcal{P}}} + \sum_{\mathrm{k}\in\{\mathrm{A,C,G,U}\}} \eta_k \cdot \{\mathrm{k}\}\,\mathrm{TP} &\xrightarrow{\mathrm{TX}} \mathrm{mRNA} + \mathrm{G_{\mathcal{P}}} + \mathrm{R_1} + \sum_{\mathrm{k}\in\{\mathrm{A,C,G,U}\}} 2\eta_{\mathrm{k}} \cdot \mathrm{Pi} \\
\mathrm{mRNA} &\longrightarrow \sum_{\mathrm{k}\in\{\mathrm{A,C,G,U}\}} \eta_{\mathrm{k}} \cdot \{\mathrm{k}\}\,\mathrm{MP} \\
\mathrm{mRNA} + \mathrm{R_2} &\longrightarrow \mathrm{R^*_2} \\
\mathrm{ff_j} \cdot \mathrm{AA_j} + \mathrm{ff_j} \cdot \mathrm{tRNA} + \mathrm{ff_j} \cdot \mathrm{ATP} &\longrightarrow \mathrm{ff_j} \cdot \mathrm{AA_j} - \mathrm{tRNA_j} + \\
&\qquad \mathrm{ff_j} \cdot \mathrm{AMP} + 2\mathrm{ff_j} \cdot \mathrm{Pi} \qquad j = 1, 2, \ldots, 20 \\
\mathrm{R^*_2} + \sum_{\mathrm{j}\in\{\mathrm{AA}\}} \mathrm{ff_j} \cdot \Big(\mathrm{AA_j} - \mathrm{tRNA_j} + 2 \cdot \mathrm{GTP}\Big) &\xrightarrow{\mathrm{TL}} \mathcal{P} + \mathrm{R_2} + \mathrm{mRNA} + \\
&\quad + \sum_{\mathrm{j}\in\{\mathrm{AA}\}} \mathrm{ff_j} \cdot \Big(\mathrm{tRNA} + 2 \cdot \mathrm{GDP} + 2 \cdot \mathrm{Pi}\Big)
\end{aligned}
$$

where $G_{\mathcal{P}}$ denotes the gene encoding protein product $\mathcal{P}$, $\mathrm{R_1}$ denotes the concentration of RNA polymerase, $G^*_{\mathcal{P}}$ denotes the gene bounded by the RNA polymerase, η_i and α_j denote the stoichiometric coefficients for nucleotide and amino acid, respectively, Pi denotes inorganic phosphate, $\mathrm{R_2}$ denotes the ribosome concentration, $\mathrm{R^*_2}$ denotes bounded ribosome, and AA_j denotes the j^{th} amino acid.

The transcription rate (w_{TX}) was fixed in the ssFBA calculation at:

$$w_{TX} = V^{max}_{TX}\left(\frac{G}{K_{TX} + G}\right) \tag{2}$$

where G denotes the gene concentration and K_{TX} denotes a transcription saturation coefficient. The maximum rate of transcription V^{max}_{TX} was formulated as:

$$V^{max}_{TX} \equiv \left[R_1\left(\frac{v_{TX}}{l_G}\right)\mathcal{P}\right] \tag{3}$$

The term R_1 denotes the RNA polymerase abundance, v_{TX} denotes the RNA polymerase

elongation rate (nt/hr), l_G denotes the gene length in nucleotides, and the last term $\mathcal{P}$ describes a model of promoter activity. In this study, we considered two promoters: T7 and σ_{70}. The promoter function for the T7 promoter, $\mathcal{P}_{T7}$, was given by:

$$P_{T7} = \frac{K_{T7}}{1 + K_{T7}} \tag{4}$$

where K_{T7} denotes a T7 RNA polymerase binding constant (*17*). The σ_{70} binding promoter, used for all other proteins, was formulated as:

$$P_{\sigma_{70}} = \frac{K_1 + K_2 f_{p70}}{1 + K_1 + K_2 f_{p70}} \tag{5}$$

where K_1 denotes the state of RNA polymerase binding, K_2 is the state of σ_{70} binding along with RNA polymerase, and f_{p70} denotes the fraction of the σ_{70} transcription factor bound to the promoter (modeled as a Hill function).

The translation rate (w_{TL}) was bounded by:

$$0 \leq w_{TL} \leq V_{TL}^{max} \left(\frac{\text{mRNA}_{\text{SS}}}{K_{TL} + \text{mRNA}_{\text{SS}}} \right) \tag{6}$$

where mRNA_{SS} denotes the steady state mRNA abundance and K_{TL} denotes the translation saturation constant. The maximum translation rate V_{TL}^{max} was formulated as:

$$V_{TL}^{max} \equiv \left[K_P R_2 \left(\frac{v_{TL}}{l_P} \right) \right] \tag{7}$$

The term K_P denotes the polysome amplification constant, v_{TL} denotes the ribosome elongation rate (amino acids per hour), l_P denotes the number of amino acids in the protein of interest, and mRNA_{SS} denotes the steady-state mRNA concentration:

$$mRNA_{SS} \simeq \frac{w_{TX}}{\lambda} \tag{8}$$

where λ denotes the rate constant controlling the mRNA degradation rate.

The objective of the sequence-specific flux balance calculation was to maximize the rate of protein translation, w_{TL}. The total glucose uptake rate was bounded by [0,40 mM/h] according to experimental data, while the amino acid uptake rates were bounded by [0,30 mM/h], but did not reach the maximum flux. Gene and protein sequences were taken from literature and are available in the Supporting Information. The sequence-specific flux balance linear program was solved using the GNU Linear Programming Kit (GLPK) v4.52 (22).

Calculation of the carbon yield.

The carbon yield (Y_C^{POI}) was calculated as the ratio of carbon produced as the protein of interest divided by the carbon consumed as reactants (glucose and amino acids):

$$Y_C^{POI} = \frac{\Delta\texttt{POI} \cdot C_{POI}}{\sum_{i=1}^{\mathcal{R}} \max(\Delta m_i, 0) \cdot C_{m_i}} \tag{9}$$

where $\Delta\texttt{POI}$ denotes the abundance of the protein of interest produced, C_{POI} denotes carbon number of the protein of interest, $\mathcal{R}$ denotes the number of reactants, Δm_i denotes the amount of the i^{th} reactant consumed (never allowed to be negative), and C_{m_i} denotes the carbon number of the i^{th} reactant.

Quantification of uncertainty.

An ensemble of 100 sets of flux distributions was calculated for three different cases: control (with amino acid synthesis and uptake), amino acid uptake without synthesis, and amino acid synthesis without uptake. An ensemble of flux distributions was then calculated by randomly sampling the maximum specific glucose uptake rate from within a range of 30 to 40 mM/h, determined from experimental data and randomly sampling oxygen uptake,

RNAP polymerase levels, ribosome levels, and elongation rates in a physiological range determined from literature. RNA polymerase levels were sampled between 60 and 80 nM, ribosome levels between 7 and 16 µM, the RNA polymerase elongation rate between 20 and 30 nt/sec, and the ribosome elongation rate between 1.5 and 3 aa/sec (*19*, *20*).

Global sensitivity analysis.

We conducted a global sensitivity analysis, using the variance-based method of Sobol, to estimate which parameters controlled the performance of synthetic circuits (*23*). We computed the total sensitivity index of each parameter relative to two performance objectives: productivity of the protein of interest and carbon yield. We established the sampling bounds for each parameter from literature. We used the sampling method of Saltelli *et al.* (*24*) to compute a family of $N\,(2d+2)$ parameter sets which obeyed our parameter ranges, where N was a parameter proportional to the desired number of model evaluations and d was the number of parameters in the model. In our case, N = 1000 and d = 7, so the total sensitivity indices were computed from 16,000 model evaluations. The variance-based sensitivity analysis was conducted using the SALib module encoded in the Python programming language (*25*).

Acknowledgement

Please use "The authors thank ..." rather than "The authors would like to thank ...".

The author thanks Mats Dahlgren for version one of achemso, and Donald Arseneau for the code taken from cite to move citations after punctuation. Many users have provided feedback on the class, which is reflected in all of the different demonstrations shown in this document.

Supporting Information Available

The following files are available free of charge.

- Protein Sequences: DNA and protein sequences of each protein of interest.
- CAT Sensitivity: Global sensitivity analysis of CAT productivity and carbon yield.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

1. Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., and Swartz, J. R. (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol 4*, 220.
2. Matthaei, J. H., and Nirenberg, M. W. (1961) Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. *Proc Natl Acad Sci U S A 47*, 1580–8.
3. Nirenberg, M. W., and Matthaei, J. H. (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.
4. Lu, Y., Welsh, J. P., and Swartz, J. R. (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.
5. Hodgman, C. E., and Jewett, M. C. (2012) Cell-free synthetic biology: thinking outside the cell. *Metab Eng 14*, 261–9.

6. Lewis, N. E., Nagarajan, H., and Palsson, B. Ø. (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol 10*, 291–305.

7. Edwards, J. S., and Palsson, B. Ø. (2000) The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A 97*, 5528–33.

8. Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. Ø. (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.

9. Oh, Y.-K., Palsson, B. Ø., Park, S. M., Schilling, C. H., and 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.

10. Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson, B. Ø. (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol 7*, 129–43.

11. Ibarra, R. U., Edwards, J. S., and Palsson, B. Ø. (2002) Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature 420*, 186–9.

12. Schuetz, R., Kuepfer, L., and Sauer, U. (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol Syst Biol 3*, 119.

13. Hyduke, D. R., Lewis, N. E., and Palsson, B. Ø. (2013) Analysis of omics data with genome-scale models of metabolism. *Mol Biosyst 9*, 167–74.

14. McCloskey, D., Palsson, B. Ø., and Feist, A. M. (2013) Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol Syst Biol 9*, 661.

15. Zomorrodi, A. R., Suthers, P. F., Ranganathan, S., and Maranas, C. D. (2012) Mathematical optimization applications in metabolic networks. *Metab Eng 14*, 672–86.

16. Allen, T. E., and Palsson, B. Ø. (2003) Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. *J Theor Biol 220*, 1–18.

17. Moon TS, T. A. S. B. V. C., Lou C (2012) Genetic programs constructed from layered logic gates in single cells. *Nature 491*.

18. Calhoun, K. A., and Swartz, J. R. (2005) An Economical Method for Cell-Free Protein Synthesis using Glucose and Nucleoside Monophosphates. *Biotechnology Progress 21*, 1146–53.

19. Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016) The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synth Biol 5*, 344–55.

20. Underwood, K. A., Swartz, J. R., and Puglisi, J. D. (2005) Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnology and Bioengineering 91*, 425–35.

21. Li, J., Gu, L., Aach, J., and Church, G. M. (2014) Improved Cell-Free RNA and Protein Synthesis System. *PLoS ONE 9*, 1–11.

22. GNU Linear Programming Kit, Version 4.52. 2016; `http://www.gnu.org/software/glpk/glpk.html`.

23. Sobol, I. (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation 55*, 271–80.

24. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications 181*, 259–70.

25. Herman, J. D. http://jdherman.github.io/SALib/.