# Sequence Specific Modeling of *E. coli* Cell-Free Protein Synthesis

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

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

‡*School of Chemical Engineering, Stanford University, Stanford, CA 94305*

E-mail: jdv27@cornell.edu

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

**Abstract**

Cell-free protein expression has become a widely used research tool in systems and synthetic biology, and a promising technology for personalized medicine. In this study, we used sequence specific constraint based modeling to evaluate the performance of an *E. coli* cell-free protein synthesis system. A core *E. coli* metabolic model, describing glycolysis, the pentose phosphate pathway, amino acid biosynthesis and degradation and energy metabolism, was augmented with sequence specific descriptions of transcription and translation processes, and effective models of promoter function. Thus, sequence specific constraint based modeling explicitly couples transcription and translation processes, and the regulation of gene expression, with the availability of metabolic resources. We tested this approach by simulating the expression of two model proteins chloramphenicol acetyltransferase and dual emission green fluorescent protein for which we have training data sets; we then expanded the simulations

to a range of therapeutically relevant proteins. Protein expression simulations were consistent with measurements for a variety of cases. Further, global sensitivity analysis identified the key metabolic processes that controlled the productivity, energy efficiency, and carbon yield of the process. Taken together, sequence specific constraint based modeling offers a novel means to *a priori* estimate the performance of cell-free synthetic circuits.

## Keywords

Synthetic biology, constraints based modeling, cell-free protein synthesis

## 1 Introduction

Cell-free protein expression has become a widely used research tool in systems and synthetic biology, and a promising technology for personalized protein production. Cell-free systems offer many advantages for the study, manipulation and modeling of metabolism compared to *in vivo* processes. Central amongst these, is direct access to metabolites and the biosynthetic machinery without the interference of a cell wall, or complications associated with cell growth. This allows us to interrogate the chemical environment while the biosynthetic machinery is operating, potentially at a fine time resolution. 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 fundamental 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*). However, if CFPS is to become a mainstream technology for applications such as point of care manufacturing (*6*), we must first understand the performance limits and cost of these systems (*1*). One tool to address these questions is mathematical modeling.

Stoichiometric reconstructions of microbial metabolism, popularized by constraint based approaches such as flux balance analysis (FBA), have become standard tools to interrogate metabolism (*7*). Constraint based methods such as FBA, metabolic flux analysis (MFA) (*8*), and network decomposition approaches such as elementary modes (*9*) or extreme pathways (*10*) model intracellular metabolism using biochemical stoichiometry and other constraints such as thermodynamical feasibility (*11*, *12*) under pseudo steady state conditions. Constraint based approaches have predicted productivity (*13*, *14*), yield (*13*), mutant behavior (*15*), and growth phenotypes (*16*) for biochemical networks of arbitrary complexity, including genome scale networks, using linear programming (*17*). Since the first genome scale stoichiometric model of *E. coli* (*18*), stoichiometric reconstructions of hundreds of organisms, including industrially important prokaryotes such as *E. coli* (*19*) or *B. subtilis* (*20*), are now available (*21*). Stoichiometric reconstructions have been expanded to include the integration of with detailed descriptions of gene expression (ME-Model) (*16*, *22*), and protein structures (GEM-PRO) (*23*, *24*). These expansion have greatly expanded the scope of questions these models can explore. Constraint based methods are powerful tools to estimate the performance of metabolic networks with very few adjustable parameters. However, constraint based methods are typically used to model *in vivo* processes, and have not yet been applied to cell-free metabolism.

In this study, we used sequence specific constraint based modeling to evaluate the performance of *E. coli* cell-free protein synthesis (CFPS). A core *E. coli* cell-free metabolic model was developed from literature (*19*). This model, which described glycolysis, pentose phosphate pathway, amino acid biosynthesis and degradation and energy metabolism, was then augmented with sequence specific descriptions of promoter function, transcription and translation processes. Thus, sequence specific constraints based modeling explicitly coupled transcription and translation with the availability of metabolic resources. We tested this approach by simulating the production of two model proteins, and then investigated the productivity, energy efficiency, and carbon yield for eight different proteins.

From this, higher carbon number proteins typically had lower productivity rates, energy efficiency, 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 as instrumental for maintaining a high energy efficiency and 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 and validation

The cell-free stoichiometric network was constructed by removing growth associated reactions from the *i*AF1260 reconstruction of K-12 MG1655 *E. coli* (*19*), and adding deletions associated with the specific cell free system (see materials and methods). We then added the transcription and translation template reactions of Allen and Palsson for the specific proteins of interest (*22*). A schematic of the metabolic network, which consisted of 264 reactions and 146 species, is shown in Fig. 1A. The network described the major carbon and energy pathways, as well as amino acid biosynthesis and degradation pathways. Using this network, in combination with effective promoter models taken from Moon et al. [TODO:REF], and literature values for cell-free culture parameters (Table 2), we simulated the sequence specific production of two model proteins, chloramphenicol acetyltransferase (CAT) and dual emission green fluorescent protein (deGFP) using different *E. coli* cell-free extracts. The cell free metabolic network, along with all model code and parameters, can be downloaded under an MIT software license from the Varnerlab GitHub site [TODO:REF].

Cell free simulations of the time evolution of CAT and deGFP production were consistent with experimental measurements (Fig. 1B and C). Chloramphenicol acetyltransferase (CAT), was produced under a T7 promoter in a glucose/NMP cell-free system (*25*) for 1 hour using glucose as a carbon and energy source (Fig. 1B). With the exception of the first 10-15 min, the cell-free prediction of CAT abundance was consistent with the measured values (Correlation?). On the other hand, deGFP was produced under a P70a promoter in TXTL 2.0 *E. coli* extract for 8 hours using maltose as a carbon and energy source (Fig. 1C). The cell-free simulation captured the overall trend of deGFP abundance, but failed to capture saturation at the end of the CFPS culture (Correlation?). Uncertainty in experimental factors such as the concentration of RNA polymerase, ribosomes, transcription and translation elongation rates or the upper bounds for oxygen and glucose consumption

rates (modeled as being normally distributed around the parameter values shown in Table 2) did not qualitatively alter the performance of the model (blue region, 95% confidence estimate). Together, these simulations suggested the description of transcription and translation, and its integration with metabolism encoded in the cell free metabolic model, were consistent with experimental measurements. However, these simulations were conducted at a single plasmid concentration. Thus, it was unclear if the model could capture cell free protein synthesis for a range of plasmid concentrations.

Cell free simulations of the deGFP titer for a range of plasmid concentrations were consistent with experimental measurements (Fig. 1D). The cell free deGFP titer at each plasmid concentration was calculated by multiplying the flux of deGFP synthesis by the active time of production, approximately 8 hours in TXTL 2.0 (*26*). The mean of the ensemble (calculated by sampling the uncertainty in the model parameters) captured the saturation of deGFP production as a function of plasmid concentration. However, while the mean and 95% confidence estimate of the ensemble was consistent with measured deGFP levels, the model under predicted the deGFP titer at the saturating point of 5 nM of plasmid concentration. These results validated our mathematical framework to model CFPS systems and predict the production of two proteins with very few adjustable parameters. It also showed that the sequence specific reactions were sufficient to predict the production of two different proteins under different promoters and cell-free systems. Since the model accurately predicted protein production, we used our mathematical framework to understand the performance limits of CFPS.

## 2.2 Analysis of CFPS performance

To better understand the performance of CFPS reactions, we predicted the productivity, energy efficiency and carbon yield for the cell free production of eight model proteins with and without amino acid supplementation (Fig. 2). The expression of each of these proteins was under a P70a promoter, with the exception of CAT which we assumed was

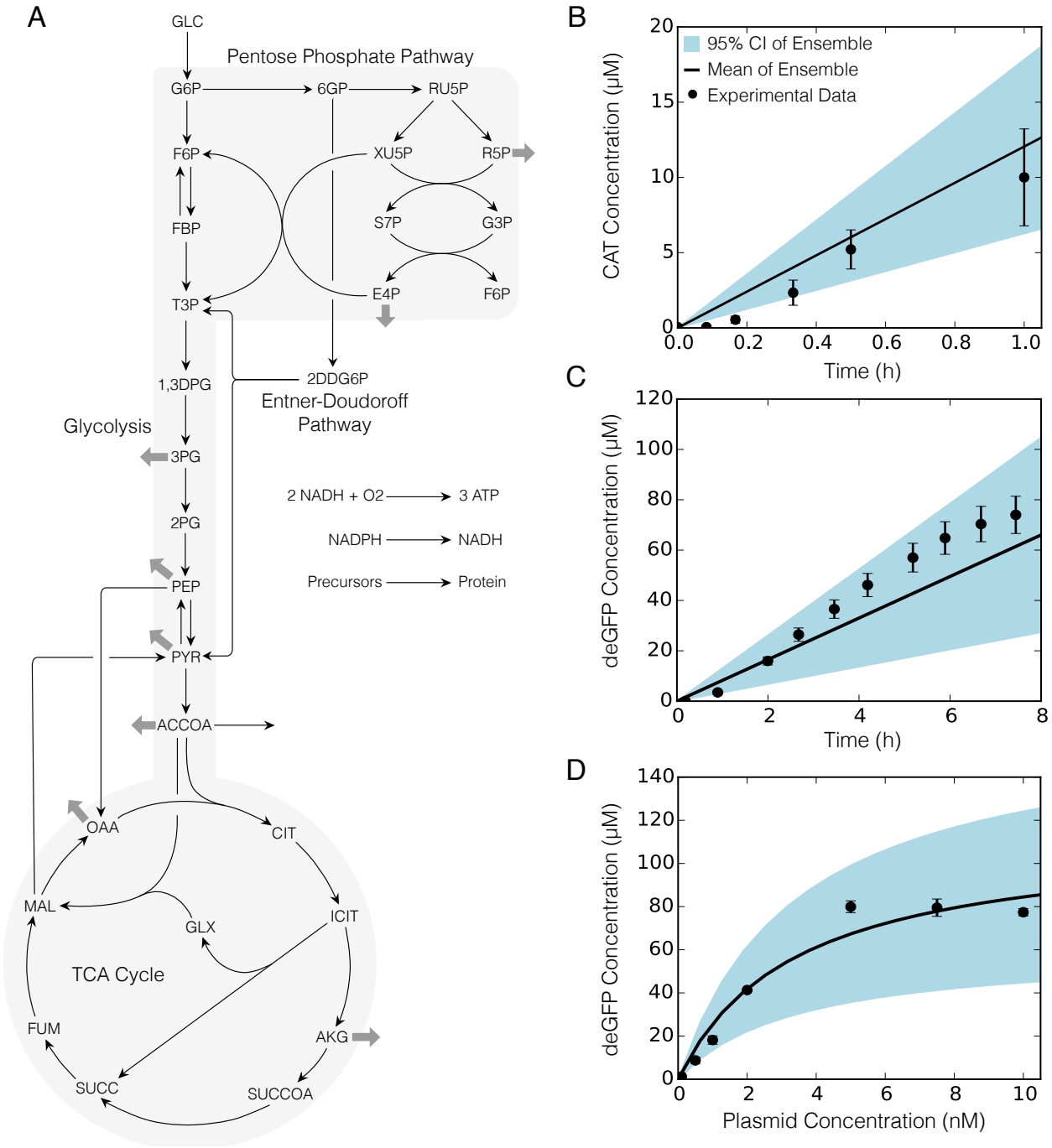


**Figure 1:** Sequence specific flux balance analysis. A. Core metabolic network with glycolysis, pentose phoshate pathway, TCA cycle, Entner-Doudoroff pathway. Thick gray arrows indicate withdrawal of precursors for amino acid synthesis. B. CAT production under a T7 promoter in CFPS *E. coli* extract for 1 h under glucose consumption. Error bars denote the standard deviation of experimental measurements. C. deGFP production under a P70 promoter in TXTL 2.0 *E. coli* extract for 8 h under maltose consumption. Error bars denote a 10% coefficient of variation. D. Predicted deGFP concentration at different plasmid concentrations versus measurements of deGFP synthesized in TXTL 2.0. 95% CI (blue region) over the ensemble of 100 sets, mean of the ensemble (black line), and experimental measurements (dots).

expressed using a T7 promoter. In all cases, CFPS was supplied with glucose; however, we considered different scenarios for amino acid supplementation. First, the CFPS reaction was supplied with amino acids, and simultaneously the system was also allowed to synthesize amino acids from glucose (AA uptake and synthesis). Next, the CFPS reaction was supplied with both glucose and amino acids, but *de novo* amino acid biosynthesis was not allowed (AA uptake w/o synthesis). This scenario was consistent with common cell free extract preparation protocols which often involve amino acid supplementation; thus, we expect the enzymes responsible for amino acid biosynthesis to be largely absent from the CFPS reaction in these cases. Lastly, the cell free reaction was not supplemented with amino acids, and was forced to synthesize them *de novo* from glucose (AA synthesis w/o uptake). Eight 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). An additional case was considered for CAT, where central metabolic fluxes were constrained by experimental metabolic measurements (see supplemental materials).

### 2.2.1 Productivity

The theoretical maximum mean productivity ($\mu$M/h) was inversely proportional to the protein length ($l_P$) and varied between 1 and 12 $\mu$M/h for the proteins sampled (Fig. 2A-B). The theoretical maximum productivity was within a standard deviation with and without amino acid supplementation for each protein, but varied significantly between proteins. For instance, BMP10 (XX aa) had a optimal productivity of approximately 2.5 $\mu$M/h whereas CAT (ZZ aa) had an optimal productivity of approximately 12 $\mu$M/h. To examine the influence of protein length further, the mean optimal productivity was plotted against the carbon number of each protein (Fig. 2B). The optimal productivity and protein length were related by the power-law relationship $\alpha \times (l_P)^{\beta}$, where $\alpha = 6.0 \times 10^6$ [TODO:

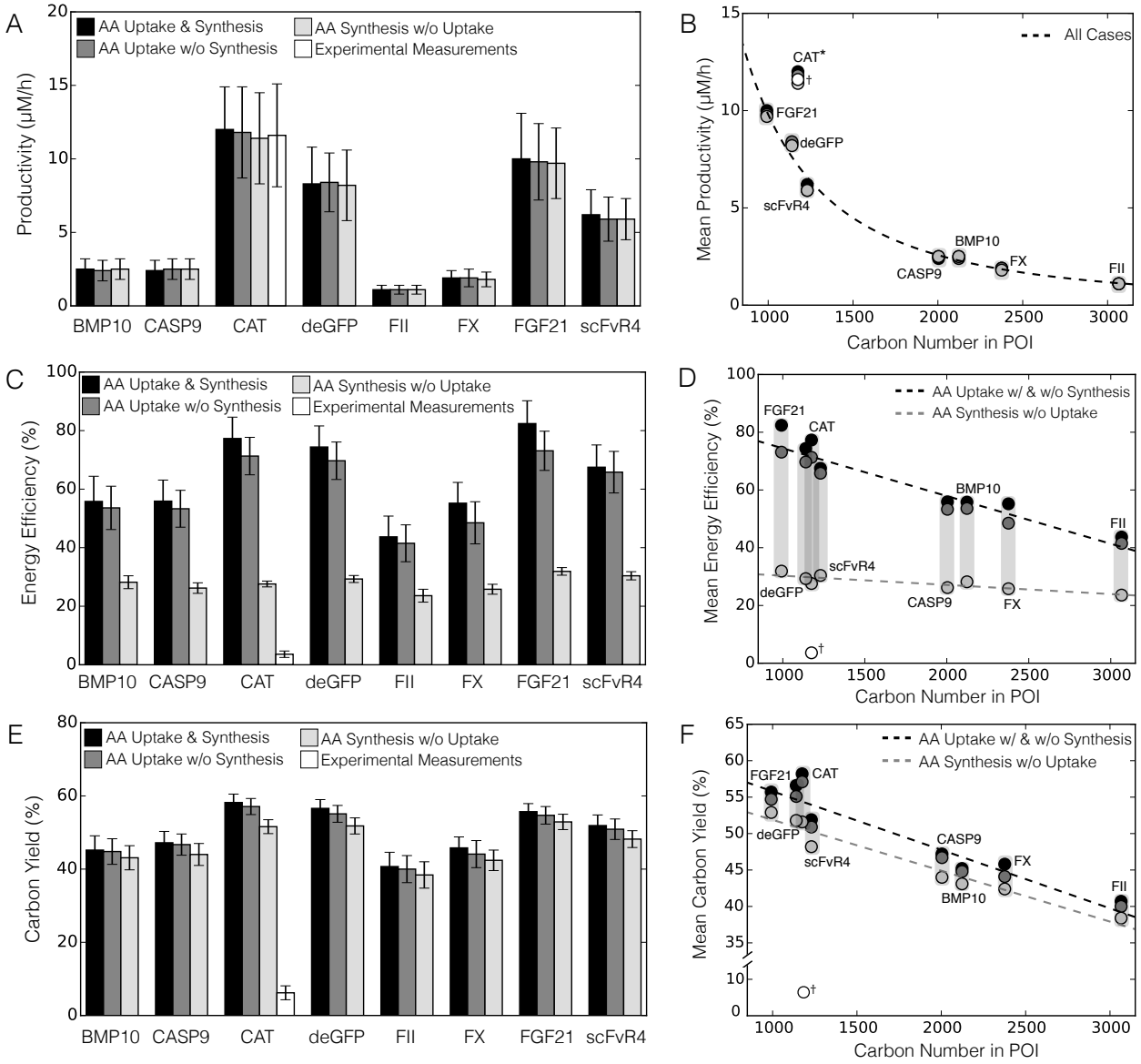


**Figure 2:** CFPS performance of eight proteins for four cases. Amino acid uptake and synthesis (black), AA uptake without synthesis (dark grey), AA synthesis without uptake (light grey), and constrained by experimental measurements, for CAT only (white). A. Productivity. B. Mean productivity versus carbon number. Single trendline (dotted line) calculated across all cases ($y = 6.02 \cdot 10^{6} x^{-1.9}$; $R^2 = 0.99$). C. Energy efficiency. D. Mean energy efficiency versus carbon number. Trendline for cases with amino acids (black dotted line; $y = -1.65 \cdot 10^{-2} x + 90.95$; $R^2 = 0.91$) and trendline for without amino acids (grey dotted line; $y = -3.18 \cdot 10^{-3} x + 33.48$; $R^2 = 0.78$). E. Carbon yield. F. Mean carbon yield versus carbon number. Trendline for cases with amino acids (black dotted line; $y = -8.03 \cdot 10^{-3} x + 63.83$; $R^2 = 0.95$) and trendline for without amino acids (grey dotted line; $y = -6.98 \cdot 10^{-3} x + 58.86$; $R^2 = 0.90$). Error bars: 95% CI; asterisk: protein excluded from trendline; dagger: constrained by experimental measurements and excluded from trendline.

UNITS?] and $\beta = -1.9$. Interestingly, CAT did not obey the power-law relationship; the relatively high productivity of CAT was due to its T7 promoter. However, CAT expressed under a P70a promoter followed the power-law correlation with a productivity

of approximately 8.8 $\mu$M/h (predicted to be XX by the optimal productivity correlation). Taken together, the maximum optimal productivity of a cell free reaction was found to be inversely proportional to protein size, following a power-law relationship for proteins expressed under a P70a promoter.

### 2.2.2 Energy efficiency

The energy efficiency of protein production remained relatively high for when amino acids were supplied in the media, however energy efficiency dropped below 32% when amino acids were not available (Fig. 2C-D). Following the same outline as in examining the productivity, we calculated the energy efficiency of production for each protein. When amino acids were supplied in the media (first two cases), there was a comparable performance in having the highest energy efficiency, despite the network's ability to synthesis amino acids. For the case when amino acids were removed from the media, protein production resulted in a low energy efficiency; below 32% depending on the protein. This was because glucose had to be utilized to synthesize the amino acids necessary for protein synthesis and meet the necessary energy demands of CFPS. We next investigated the effect of protein carbon number on energy efficiency (Fig. 2D). The same inverse trend was observed as for productivity, except that it was linear. Proteins with a high carbon number had a lower energy efficiency since they have a higher transcription and translation cost then smaller proteins. The cases supplemented with amino acids had the same trendline whereas when amino acids were not available there was a significant drop in energy efficiency. Interestingly, without supplemented amino acids (third case) each protein had a similar energy efficiency of about 28% regardless of carbon number. In this case, the energy burden of synthesizing each amino acid required for the assembly of the protein kept the energy efficiency saturated at a relatively low level. However, the experimentally constrained case of CAT production showed even a lower energy efficiency of 3.6 $\pm$ 1.1% compared to the theoretical maximum of 77.3 $\pm$ 7.3%. This shows CFPS systems have a lot of room

for improvement: first, the experimental measurements showed accumulation of certain amino acids; this carbon could potential could be diverted towards a protein of interest. Second, the system had a high accumulation of metabolic byproducts, specifically organic acids, which is a result of inefficient energy utilization.

### 2.2.3 Carbon Yield

The theoretical maximum mean carbon yield was inversely proportional to protein length and varied between 40% and 57% for the proteins sampled (Fig. 2E-F). The relationship between the optimal carbon yield and protein length was linear; $m_Y \times (l_P) + b_Y$ where $m_Y = -8.03 \times 10^{-3}$ [TODO: UNITS] with supplementation, and $m_Y = -6.98 \times 10^{-3}$ [UNITS] without. The y-intercept $b_Y = 63.83$ [UNITS] with supplementation, and $b_Y = 58.86$ [TODO:UNITS] without. The linear carbon yield models showed good predictability for estimating the optimal carbon yield for a range of protein lengths ($R^2 = 0.95$ with amino acids and $R^2 = 0.90$ without supplementation), irrespective of the promoter used to control protein expression. The effective yield models were approximately parallel, with a drop in carbon yield of approximately 7% without supplementation. Only the necessary amount of amino acids were used for the production of the protein of interest; thus, it may be hypothesized that all the glucose was used to power CFPS and did not contribute to the carbon yield. In that case, the carbon yield without the glucose contribution would be 100%. However, in the experimentally constrained case, CAT was produced with a carbon yield of 6.2% compared to the theoretical maximum of 58%. This decrease in carbon yield suggests inefficiencies in CFPS that can potentially be improved. Constraint based calculations assume a metabolic pseudo steady state; thus, intermediate metabolites cannot accumulate within the cell-free extract. In addition, the flux balance analysis calculation is solved by maximizing the flux through the protein production reaction. Therefore, carbon flux will travel through the network to optimize the maximum flux through the protein synthesis reaction. In examining the experimental dataset, there is a high accumulation

of organic acids, especially acetate, pyruvate and lactate. The experimental performance could be improved by diverting this carbon toward the protein of interest by knockouts during the cell-free extract preparation.

## 2.3 Global 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, energy efficiency and carbon yield for the CAT protein (Fig. 3). Surprisingly, RNAP and ribosome abundance had only a modest effect; the translation elongation rate had the largest effect on protein productivity. On the other hand, oxygen followed by glucose consumption had the largest influence on the energy efficiency and carbon yield, respectively. The significance of transcription/translation parameters was robust to amino acid supplementation, with the translation rate being the most sensitive across all cases (Fig. 3A). This suggested that the translation elongation rate, and not transcriptional parameters, controlled productivity. Underwood and coworkers showed that an increase in ribosome levels did not significantly increase protein yields or rates; however, adding elongation factors increased yields by 23% at 30 minutes (*27*). In addition, Li et al. increased the productivity of firefly luciferase production by 5-fold by adjusting factors that affect transcription and translation such as elongation factors, ribosome recycling factor, release factors, chaperones, BSA, and tRNAs (*28*). In examining substrate utilization, glucose consumption was not important for productivity in the presence of amino acid supplementation. However, its importance increased significantly when amino acids were not available. On the other hand, amino acid consumption was only sensitive when amino acids synthesis reactions were blocked, as it was the only source of amino acids for CAT synthesis.

The oxygen consumption rate was the most important factor controlling the energy efficiency of cell free protein synthesis (Fig. 3B). In addition to the oxygen consumption,

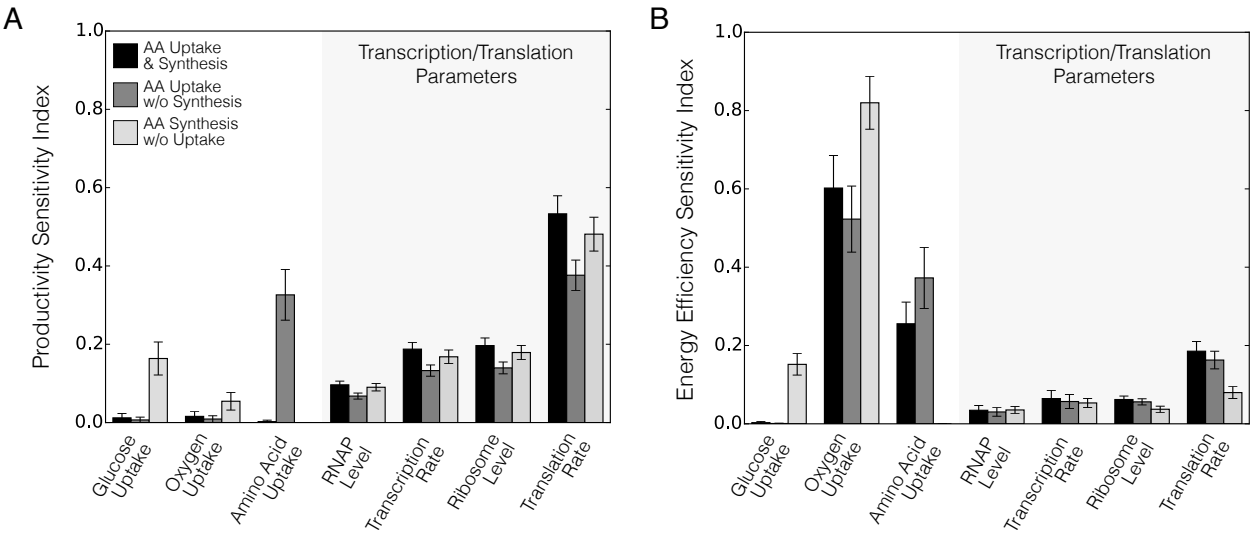


**Figure 3:** Total order sensitivity of deGFP productivity (A) and energy efficiency (B) to specific uptake rates and transcription/translation parameters for three cases: amino acid uptake and synthesis (black), amino acid uptake without synthesis (dark grey), and amino acid synthesis without uptake (light gray). Error bars represent a 95% CI on the sensitivity index.

amino acid utilization was also sensitive. On the other hand, glucose consumption was only sensitive in the absence of amino acid supplementation, since in the absence of amino acids, glucose acted as the sole source of carbon and energy for protein synthesis. In the model, we assumed that ATP could be produced by both substrate level and oxidative phosphorylation. Jewett and coworkers reported that oxidative phosphorylation still operated in cell free systems, and that the protein yield decreased from 1.5-fold to 4-fold when oxidative phosphorylation reactions were inhibited in pyruvate-powered CFPS (*1*). However, it is unknown how active oxidative phosphorylation is for a glucose powered cell free system. Moreover, the connection between oxidative phosphorylation activity and other performance metrics, such as the carbon yield, is also unclear.

To investigate the connection between carbon yield and oxidative phosphorylation further, we calculated the CAT carbon yield as a function of the oxidative phosphorylation flux (Fig. 4). Oxidative phosphorylation had a strong effect on the carbon yield, both with and without amino acid supplementation. In the presence of amino acid supplementation, the carbon yield ranged from 20% to approximately 60%, depending upon the oxidative phosphorylation flux. However, without amino acid supplementation, the carbon yield

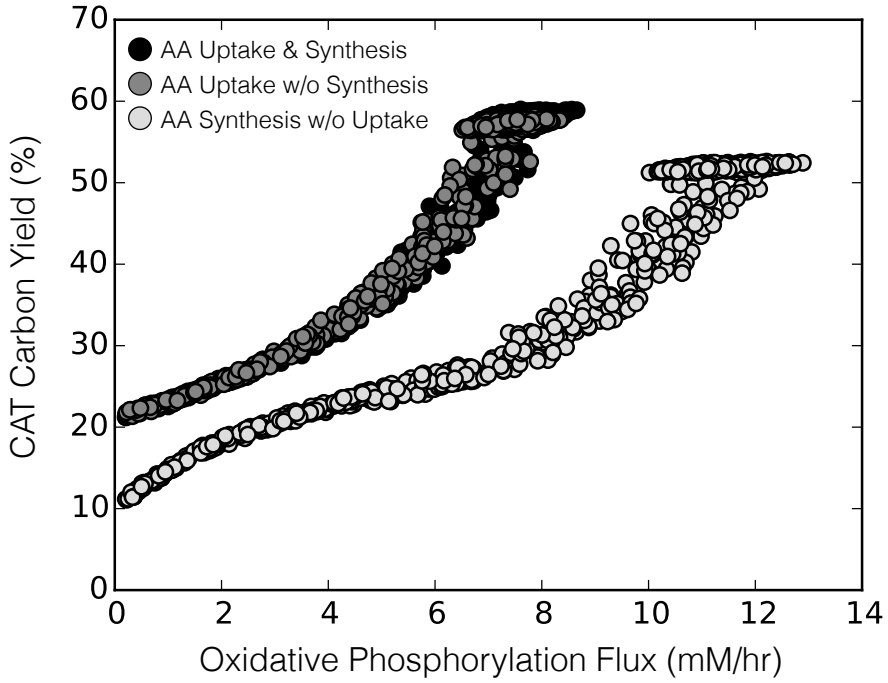


**Figure 4:** The deGFP carbon yield versus oxidative phosphorylation flux, across an ensemble of 1000 flux balance solutions, for amino acid uptake and synthesis (black), amino acid uptake without synthesis (dark grey), and amino acid synthesis without uptake (light gray).

dropped to approximately 10%, and reached a maximum of 50%. In the absence of supplementation, a lower carbon yield was expected for the same oxidative phosphorylation flux, as glucose was utilized for both energy generation and amino acid biosynthesis. In all cases, whenever the carbon yield was below its theoretical maximum, there was an accumulation of both acetate and lactate. The experimental dataset also exhibited a mixture of acetate and lactate accumulation during CAT synthesis, which suggested the CFPS reaction was not operating with optimal oxidative phosphorylation activity. Oxidative phosphorylation is a membrane associated process, however CFPS has no cell membrane. Jewett et al hypothesized that membrane vesicles present in the CFPS reaction carry out oxidative phosphorylation [TODO: REF]; however, the number, size, protein loading and lifetime of these vesicles remains an open area of study. Thus, we expect CFPS will have suboptimal oxidative phosphorylation capacity.

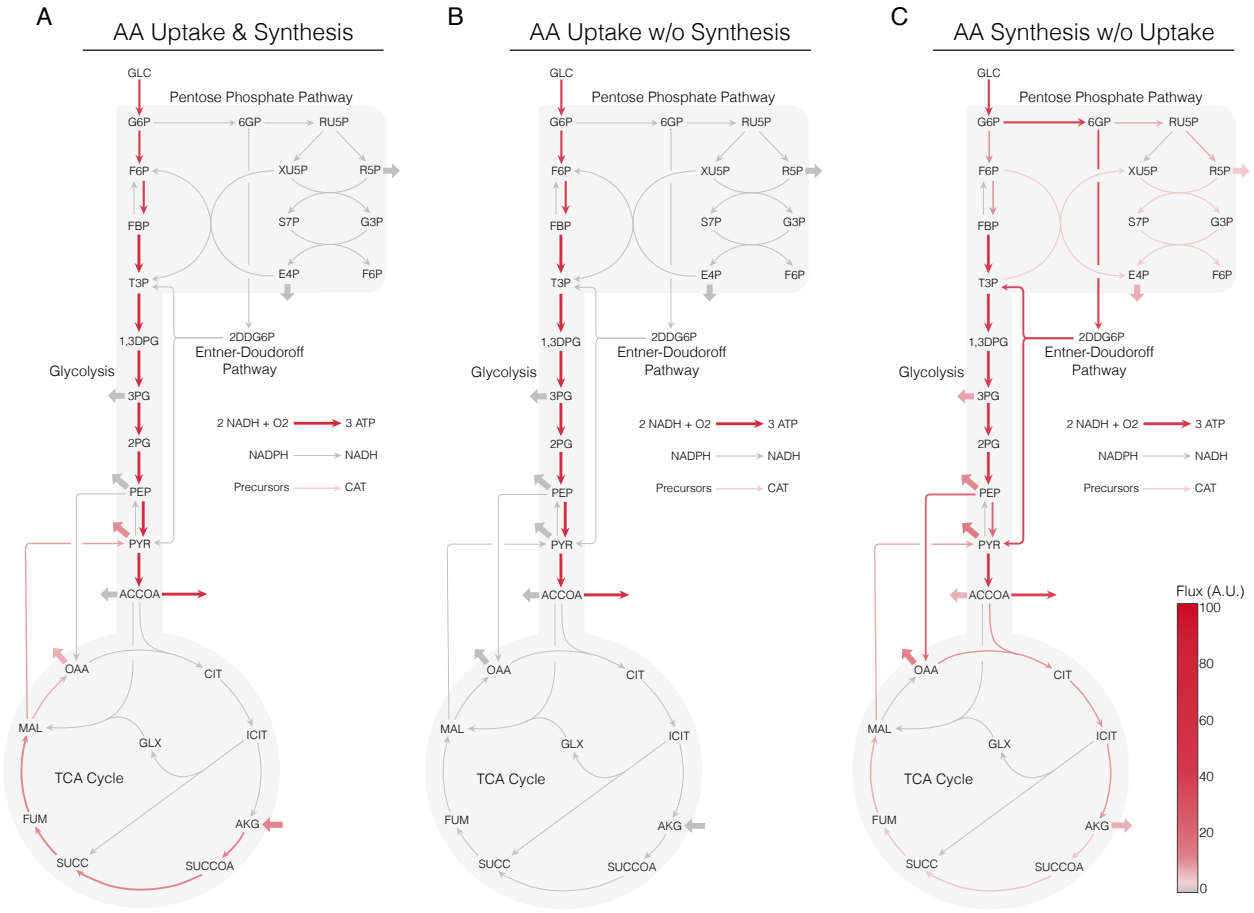


**Figure 5:** Flux profile for glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, TCA cycle, and oxidative phosphorylation, for three different cases: (A) amino acid uptake and synthesis, (B) amino acid uptake without synthesis, and (C) amino acid synthesis without uptake. Mean flux across the ensemble, normalized to glucose uptake flux. Thick arrows indicate flux to or from amino acids.

## 2.4 Optimal Metabolic Flux Distribution

Amino acid supplementation altered the optimal metabolic flux distribution predicted for CAT production (Fig. 5). To investigate the influence of amino acid supplementation, we compared the simulated metabolic flux distributions for CAT production with and without external amino acids. In the presence of amino acid supplementation, and *de novo* amino acid synthesis, there was an incomplete TCA cycle, where a combination of glucose and amino acids powered protein expression (Fig. 5A). Glucose was consumed to produce acetyl-coenzyme A, and associated by-products, while glutamate was converted to alpha-ketoglutarate which traveled to oxaloacetic acid and pyruvate for additional amino acid

biosynthesis. On the other hand, in the presence of amino acid supplementation, but without *de novo* amino acid biosynthesis, there was no TCA cycle flux. In this case, ATP was produced by a combination of substrate level and oxidative phosphorylation, where ubiquinone was regenerated via *nuo* activity, without relying on succinate dehydrogenase in the TCA cycle (Fig. 5B). These first two cases where amino acids were available had similar performance, and their respective metabolic flux distributions had a 99% correlation for all proteins. However, in the absence of amino acid supplementation (where all amino acids were synthesized *de novo* from glucose), the energy efficiency and carbon yield decreased; in this case the TCA cycle was largely complete and there was diversion of metabolic flux into the Entner-Doudoroff pathway to produce NADPH (Fig. 5C). [TODO: We need a link sentence to the experimentally constrained case.]

The case constrained by experimental measurements (Fig. 6), had the highest correlation of 0.66 with the flux distribution from the case supplied with no amino acids and a correlation of 0.52 with the cases supplied with amino acids. Thus there are some differences in the flux distribution compared to the optimum solutions which may provide some insight to improve CFPS performance. Metabolic fluxes were constrained by experimental measurements (available in Supporting Information) where available for the first hour which constrained the feasible solution space. The central carbon organic acids showed good agreement with the data (Fig. 6B). Only certain amino acid synthesis reactions were blocked since during the growth of *E. coli* not all amino acids were supplied (see Materials and Methods). During the cell-free reaction all amino acids were supplied, however glucose still traveled through all the major pathways, and the same metabolic precursors were still utilized for amino acid biosynthesis. In this case, it is unclear which substrate (glucose or amino acids) is used to power CFPS and may in fact be a combination of both. The optimum solutions only produced the required amount of amino acids necessary, however in examining the measurements there is an accumulation of alanine and glutamine which may explain some of the differences in the correlations of the flux

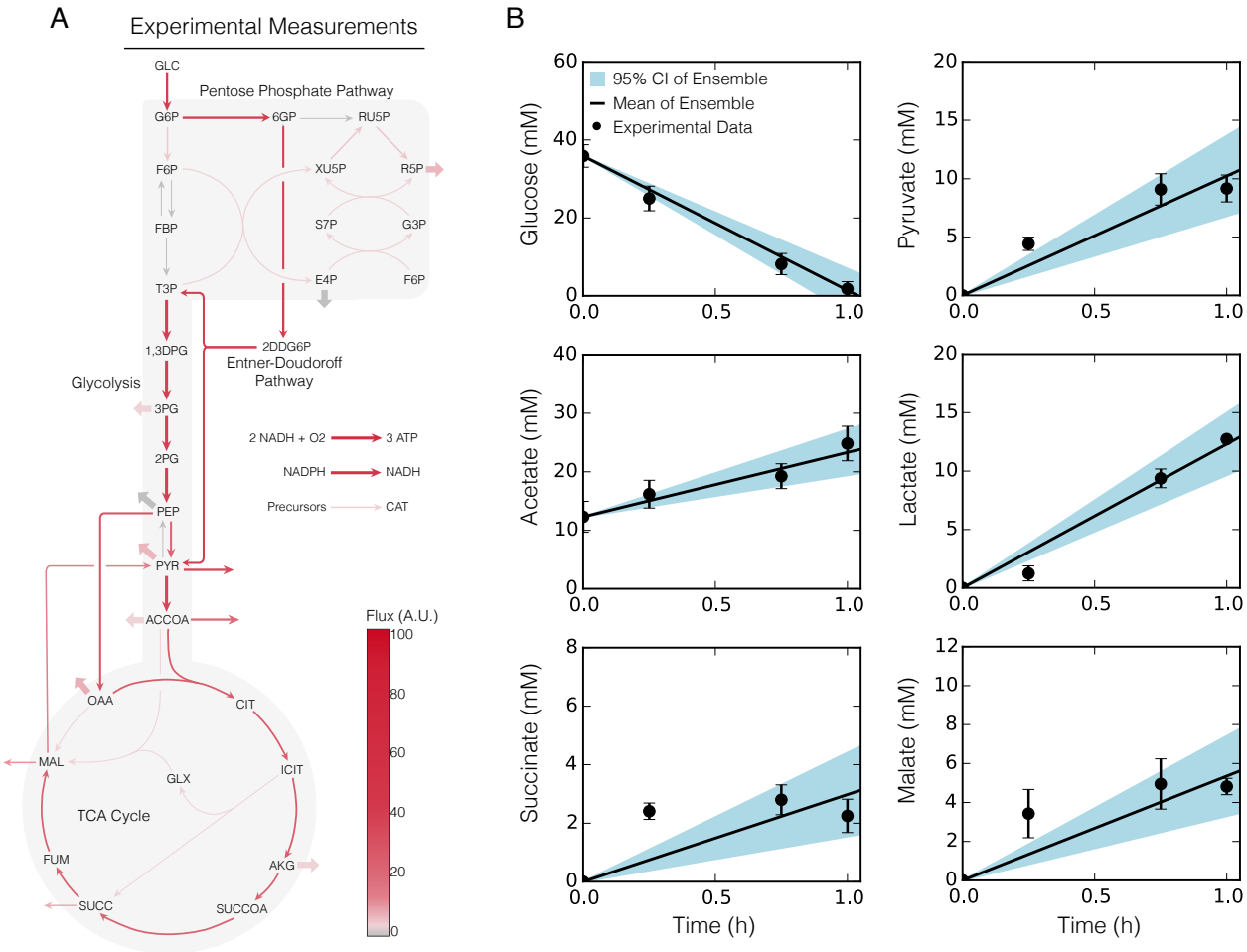


**Figure 6:** Constraint based simulation of CAT production for an experimentally constrained case. (A) Flux profile for glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, TCA cycle, and oxidative phosphorylation. Mean flux across the ensemble, normalized to glucose uptake flux. Thick arrows indicate flux to or from amino acids. (B) Central carbon metabolite measurements versus simulations over a one hour time course.

distributions. Accumulation of pyruvate, lactate, acetate, and other organic acids can be seen (Fig. 6B) and shows an inefficiency of carbon utilization.

### 2.4.1 Potential Alternative Metabolic Optima

One of the interesting features of constraint based approaches is the ability to compute alternative optimal solutions. These solutions have the same objective value, e.g., the productivity but different underlying metabolic flux distributions. Despite the constraints by experimental measurements, it is difficult to calculate the physiological flux distribution

of metabolism (Fig. 7) For example, there is a high flux through the Entner-Douodoroff pathway, but this may be an artifact of the optimal solution. A knockout of the Entner-Doudoroff pathway has no effect on the norm productivity of CAT compared to no knockouts (Fig. 7A). In addition, pairwise knockouts of Entner-Doudoroff and other subgroups (i.e. pentose phosphate pathway, cofactors, folate metabolism, etc.) in the network result in the same optimal solution of CAT productivity. However, there is a difference in the flux distribution with these knockouts (Fig. 7B); the flux balance approach will reroute the flux to optimize the objective function. Interestingly, there are crucial subgroups that are detrimental to CAT productivity. For example, a single group knockout of glycolysis/gluconoeogenesis or oxidative phosphorylation leads to no CAT production. There were also cases of group knockouts that had no effect on productivity or its flux distribution such as isoleucine, leucine, and valine biosynthesis knocked out with histidine biosynthesis. Flux balance analysis has been shown to have multiple alternative optimal solutions with flux variablitiy analysis and mixed-integer linear programming resulting in a poor depiction of the physiological flux distribution (*29*–*31*). In our case, the constraint based simulation reached the same optimal solution of CAT productivity for 40% of the pairwise group knockouts, but had a different flux distribution than compared to no knockouts. To determine which reactions occur in CFPS, adding thermodynamic feasibility constraints to reactions may result in a better depiction of the flux distribution (*11*, *12*). It would also be interesting to track the carbon flux using $C^{13}$ labeling in CFPS to better constrain the flux distribution. However, while $C^{13}$ labeling techniques are well established for *in vivo* processes (*32*), they are an active area of research for cell free systems.

## 2.5 Summary and conclusions

In this study, we developed a sequence specific constraint based modeling approach to predict the performance of cell free protein synthesis reactions for a range of proteins. We showed first principle predictions for protein production of deGFP and CAT were

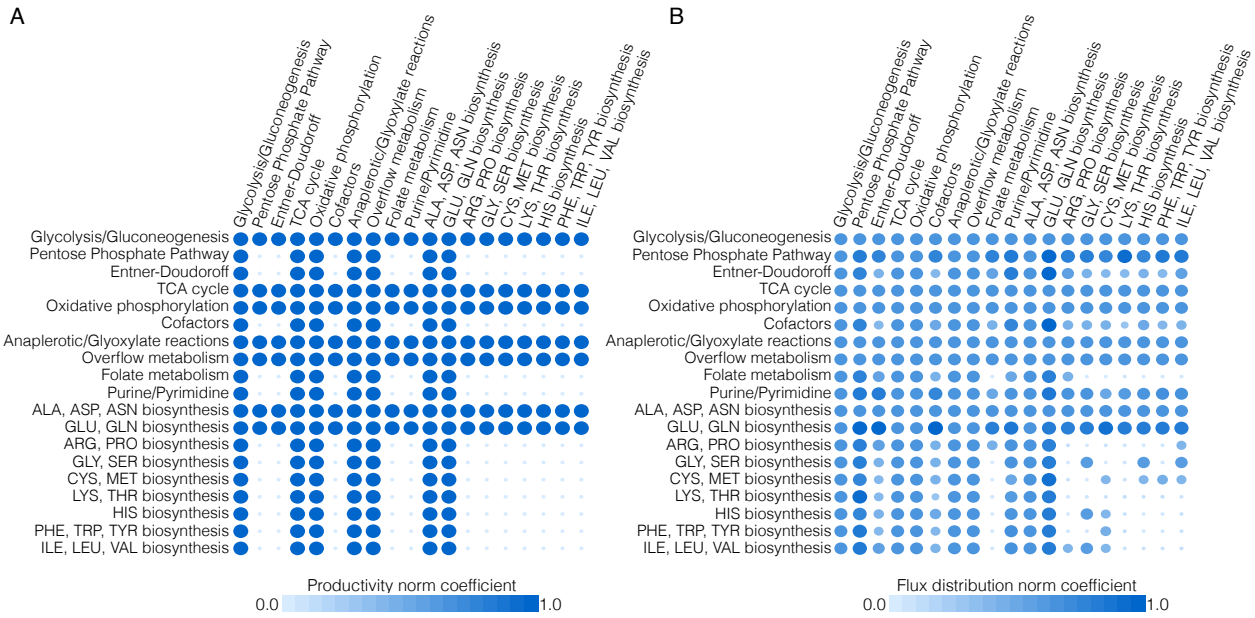


**Figure 7:** The norm effect of pairwise knockouts of subgroups in the cell-free network (A) CAT productivity norm compared to no knockouts in the experimentally constrained case. (B) Flux distribution norm compared to no knockouts in the experimentally constrained case.

in agreement with experimental measurements, under two different promoters and two different cell-free extract systems. This modeling approach suggested correlations for productivity, energy efficiency and carbon yield as a function of the size of the protein. Furthermore, global sensitivity analysis identified oxygen uptake as being instrumental for maintaining a high energy efficiency and carbon yield. The translation rate was identified as the rate limiting step 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 energy efficiency and 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

## Glucose/NMP cell-free protein synthesis.

The glucose/NMP cell-free protein synthesis reaction was performed using the S30 extract in 1.5-mL Eppendorf tubes (working volume of 15 $\mu$L) and incubated in a humidified incubator. The S30 extract was prepared from *E. coli* strain KC6 (A19 ΔtonA ΔtnaA ΔspeA ΔendA ΔsdaA ΔsdaB ΔgshA met+). This K12-derivative has several gene deletions to stabilize amino acid concentrations during the cell-free reaction. The KC6 strain was grown to approximately 3.0 OD595 in a 10-L fermenter (B. Braun, Allentown PA) on defined media with glucose as the carbon source and with the addition of 13 amino acids (alanine, arginine, cysteine, serine, aspartate, glutamate, and glutamine were excluded) (*33*). Crude S30 extract was prepared as described previously (*34*). Plasmid pK7CAT was used as the DNA template for chloramphenical acetyl transferase (CAT) expression by placing the *cat* gene between the T7 promoter and the T7 terminator (*35*). The plasmid was isolated and purified using a Plasmid Maxi Kit (Qiagen, Valencia CA). Cell-free CAT synthesis was performed at 37 °C.

The protein synthesis reaction was conducted using the PANOxSP protocol with slight modifications from that described previously (*36*). Unless otherwise noted, all reagents were purchased from Sigma (St. Louis, MO). The initial mixture included 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 30 mM phosphoenolpyruvate (Roche, Indianapolis IN); 130 mM potassium glutamate; 10 mM ammonium glutamate; 16 mM magnesium glutamate; 50 mM HEPES-KOH buffer (pH 7.5); 1.5 mM spermidine; 1.0 mM putrescine; 34 $\mu$g/mL folinic acid; 170.6 $\mu$g/mL *E. coli* tRNA mixture (Roche, Indianapolis IN); 13.3 $\mu$g/mL pK7CAT plasmid; 100 $\mu$g/mL T7 RNA polymerase; 20 unlabeled amino acids at 2-3 mM each; 5 $\mu$M l-[U-$^{14}$C]-leucine (Amersham Pharmacia, Uppsala Sweden); 0.33 mM nicotinamide adenine dinucleotide (NAD); 0.26 mM coenzyme A (CoA); 2.7 mM sodium oxalate; and 0.24 volumes of *E. coli* S30 extract. This reaction was modified for the energy

source used such that glucose reactions have 30-40 mM glucose in place of PEP. Sodium oxalate was not added since it has a detrimental effect on protein synthesis and ATP concentrations when using glucose or other early glycolytic intermediate energy sources (*37*). The HEPES buffer (pKa $\sim$ 7.5) was replaced with Bis-Tris (pKa $\sim$ 6.5). In addition, the magnesium glutamate concentration was reduced to 8 mM for the glucose reaction since a lower magnesium optimum was found when using a nonphosphorylated energy source (*36*). Finally, 10 mM phosphate was added in the form of potassium phosphate dibasic adjusted to pH 7.2 with acetic acid.

## Protein product and metabolite measurements.

Cell-free reaction samples were quenched at specific timepoints with equal volumes of ice-cold 150 mM sulfuric acid to precipitate proteins. Protein synthesis of CAT was determined from the total amount of $^{14}$C-leucine-labeled product by trichloroacetic acid precipitation followed by scintillation counting as described previously (*25*). Samples were centrifuged for 10 min at 12,000g and 4°C. The supernatant was collected for high performance liquid chromatography (HPLC) analysis. HPLC analysis (Agilent 1100 HPLC, Palo Alto CA) was used to separate nucleotides and organic acids, including glucose. Compounds were identified and quantified by comparison to known standards for retention time and UV absorbance (260 nm for nucleotides and 210 nm for organic acids) as described previously (*25*). The standard compounds quantified with a refractive index detector included inorganic phosphate, glucose, and acetate. Pyruvate, malate, succinate, and lactate were quantified with the UV detector. The stability of the amino acids in the cell extract was determined using a Dionex Amino Acid Analysis (AAA) HPLC System (Sunnyvale, CA) that separates amino acids by gradient anion exchange (AminoPac PA10 column). Compounds were identified with pulsed amperometric electrochemical detection and by comparison to known standards.

## Formulation and solution of the model equations.

The sequence-specific flux balance analysis problem was formulated as a linear program:

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

where **S** denotes the stoichiometric matrix, **w** denotes the unknown flux vector, $\boldsymbol{\theta}$ denotes the objective cost vector and $\mathcal{L}_i$ and $\mathcal{U}_i$ denote the lower and upper bounds on flux $w_i$, respectively. The transcription (T) and translation (X) stoichiometry was modeled based upon the template reactions of Allen and Palsson (22): where $G_{\mathcal{P}}$ denotes the gene encoding

**Table 1:** Transcription and translation template reactions for protein production.

| **Description** | **Template reaction** | | |
|---|---|---|---|
| Transcription initiation | $G_{\mathcal{P}} + R_T$ | $\longrightarrow$ | $G^*_{\mathcal{P}}$ |
| Transcription ($w_T$) | $G^*_{\mathcal{P}} + \sum_{k\in\{A,C,G,U\}} \eta_k \cdot (\{k\}\, TP + H_2O)$ | $\longrightarrow$ | $mRNA + G_{\mathcal{P}} + R_T + \sum_{k\in\{A,C,G,U\}} \eta_k \cdot PPi$ |
| mRNA degradation | $mRNA$ | $\longrightarrow$ | $\sum_{k\in\{A,C,G,U\}} \eta_k \cdot \{k\}\, MP$ |
| Translation initiation | $mRNA + R_X$ | $\longrightarrow$ | $R^*_X$ |
| tRNA charging | $\alpha_j \cdot \left( AA_j + tRNA + ATP + H_2O \right)$ | $\longrightarrow$ | $\alpha_j \cdot \left( AA_j\text{-}tRNA_j + AMP + PPi \right)$ $j = 1, 2, \ldots, 20$ |
| Translation ($w_X$) | $R^*_X + \sum_{j\in\{AA\}} \alpha_j \cdot \left( AA_j\text{-}tRNA_j + 2GTP + 2H_2O \right)$ | $\longrightarrow$ | $\mathcal{P} + R_X + mRNA + \sum_{j\in\{AA\}} \alpha_j \cdot \left( tRNA + 2GDP + 2Pi \right)$ |

protein product $\mathcal{P}$, $R_T$ denotes the concentration of RNA polymerase, $G^*_{\mathcal{P}}$ denotes the gene bounded by the RNA polymerase (open complex), $\eta_i$ and $\alpha_j$ denote the stoichiometric coefficients for nucleotide and amino acid, respectively, $Pi$ denotes inorganic phosphate, $R_X$ denotes the ribosome concentration, $R^*_X$ denotes bound ribosome, and $AA_j$ denotes $j^{th}$ amino acid.

The objective of the sequence specific flux balance calculation was to maximize the rate of protein translation, $w_X$. 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) v4.55 (*38*). For all cases, amino acid degradation reactions were blocked as these enzymes were knocked out during the cell-free extract preparation (*25*, *26*). In the second case, all amino acid synthesis reactions were set to 0 mM/hr since *E. coli* was grown in the presence of amino acids, thus these enzymes would not be present in the cell-free extract media. In the third case, amino acid uptake reactions were set to 0 mM/hr. In the experimental constrained case, *E. coli* was grown in the presence of 13 amino acids (alanine, arginine, cysteine, serine, aspartate, glutamate, and glutamine were excluded) (*33*), thus the synthesis reactions responsible for those 13 amino acid were set to 0 mM/hr.

The bounds on the transcription rate ($\mathcal{L}_T = w_T = \mathcal{U}_T$) were modeled as:

$$w_T = V_T^{max}\left(\frac{G}{K_T + G}\right) \tag{2}$$

where $G$ denotes the gene concentration and $K_T$ denotes a transcription saturation coefficient. The maximum transcription rate $V_T^{max}$ was formulated as:

$$V_T^{max} \equiv \left[R_T\left(\frac{\dot{v}_T}{l_G}\right)u\left(\kappa\right)\right] \tag{3}$$

The term $R_T$ denotes the RNA polymerase concentration (nM), $\dot{v}_T$ denotes the RNA polymerase elongation rate (nt/h), $l_G$ denotes the gene length in nucleotides (nt). The term $u\left(\kappa\right)$ (dimensionless, $0 \leq u\left(\kappa\right) \leq 1$) is an effective model of promoter activity, where $\kappa$ denotes promoter specific parameters. The general form for the promoter models was taken from Moon *et al.* (*39*). In this study, we considered two promoters: T7 and P70a. The

promoter function for the T7 promoter, $u_{T7}$, was given by:

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

where $K_{T7}$ denotes a T7 RNA polymerase binding constant. The P70a promoter function $u_{P70a}$ (which was used for all other proteins) was formulated as:

$$u_{P70a} = \frac{K_1 + K_2 f_{\sigma_{70}}}{1 + K_1 + K_2 f_{\sigma_{70}}} \tag{5}$$

where $K_1$ denotes the weight of RNA polymerase binding alone, $K_2$ denotes the weight of RNAP-$\sigma_{70}$ bound to the promoter, and $f_{p70}$ denotes the fraction of the $\sigma_{70}$ transcription factor bound to RNAP, modeled as a Hill function:

$$f_{\sigma_{70}} = \frac{\sigma_{70}^n}{K_D^n + \sigma_{70}^n} \tag{6}$$

where $\sigma_{70}$ denotes the sigma-factor 70 concentration, $K_D$ denotes the dissociation constant, and $n$ denotes a cooperativity coefficient. The values for all promoter parameters are given in Table 2.

The translation rate ($w_X$) was bounded by:

$$0 \leq w_X \leq V_X^{max} \left( \frac{mRNA^*}{K_X + mRNA^*} \right) \tag{7}$$

where mRNA$^*$ denotes the steady state mRNA abundance and $K_X$ denotes a translation saturation constant. The maximum translation rate $V_X^{max}$ was formulated as:

$$V_X^{max} \equiv \left[ K_P R_X \left( \frac{\dot{v}_X}{l_P} \right) \right] \tag{8}$$

The term $K_P$ denotes the polysome amplification constant, $\dot{v}_X$ denotes the ribosome elongation rate (amino acids per hour), $l_P$ denotes the number of amino acids in the protein of

interest. The steady-state mRNA abundance mRNA$^*$ was estimated as:

$$mRNA^* \simeq \frac{w_T}{\lambda} \tag{9}$$

where $\lambda$ denotes the rate constant controlling the mRNA degradation rate ($hr^{-1}$). All translation parameters are given in Table 2.

**Table 2:** Parameters for sequence specific flux balance analysis

| Description | Parameter | Value | Units | Reference |
|---|---|---|---|---|
| RNA polymerase concentration | $R_T$ | 75 | nM | (*26*) |
| Ribosome concentration | $R_X$ | 1.6 | $\mu$M | (*26*, *27*) |
| | | | | |
| Transcription elongation rate | $\dot{v}_{TX}$ | 25 | nt/sec | (*26*) |
| Translation elongation rate | $\dot{v}_{TL}$ | 2 | aa/sec | (*26*, *27*) |
| Transcription saturation coefficient | $K_{TX}$ | 3.5 | nM | estimated |
| Translation saturation coefficient | $K_{TL}$ | 45.0 | $\mu$M | estimated |
| Polysome amplification constant | $K_P$ | 10 | constant | estimated |
| mRNA degradation rate | $\lambda$ | 5.2 | $hr^{-1}$ | (*26*) |
| | | | | |
| T7 promoter | $K_{T7}$ | 10 | constant | estimated |
| Weight RNA polymerase binding alone P70a | $K_1$ | 0.014 | constant | estimated |
| Weight bound RNAP-$\sigma_{70}$ P70a | $K_2$ | 10 | constant | estimated |
| $\sigma_{70}$ concentration | $\sigma_{70}$ | 35 | nM | (*26*) |
| $\sigma_{70}$ dissociation constant | $K_D$ | 130 | nM | estimated |
| $\sigma_{70}$ hill coefficient | n | 1 | constant | estimated |
| | | | | |
| Gene concentration | $G$ | 5 | nM | (*26*) |
| Gene length of CAT | $l_G$ | 683 | nt | (*35*) |
| Gene length of deGFP | $l_G$ | 660 | nt | (*26*) |
| Protein length of CAT | $l_P$ | 229 | aa | (*35*) |
| Protein length of deGFP | $l_P$ | 219 | aa | (*26*) |

## Calculation of energy efficiency.

Energy efficiency ($\mathcal{E}$) was calculated as the ratio of protein production to glucose consumption, both in terms of equivalent ATP molecules:

$$\mathcal{E} = \frac{\mathtt{q}_{POI} \cdot \left(2\left(\mathrm{ATP_{TX}} + \mathrm{CTP_{TX}} + \mathrm{GTP_{TX}} + \mathrm{UTP_{TX}}\right) + 2 \cdot \mathrm{ATP_{TL}} + \mathrm{GTP_{TL}}\right)}{\mathtt{q}_{GLC} \cdot \mathtt{ATP}_{GLC}} \tag{10}$$

where $\mathtt{q}_{POI} = w_{TX}$ denotes the production rate for the protein of interest, $\mathrm{ATP_{TX}}$, $\mathrm{CTP_{TX}}$, $\mathrm{GTP_{TX}}$, $\mathrm{UTP_{TX}}$ denote the stoichiometric coefficients of each energy species for the transcription of the protein of interest, $\mathrm{ATP_{TL}}$, $\mathrm{GTP_{TL}}$ denote the stoichiometric coefficients of ATP and GTP for the translation of the protein of interest, $\mathtt{q}_{GLC} = w_{GLC}$ denotes the glucose uptake rate, and $\mathrm{ATP_{GLC}}$ denotes the equivalent ATP number for glucose. The energy species stoichiometric coefficients are available in the Supporting Information.

## 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):

$$Y_C^{POI} = \frac{\mathtt{q}_{POI} \cdot C_{POI}}{\sum_{i=1}^{\mathcal{R}} \mathtt{q}_{m_i} \cdot C_{m_i}} \tag{11}$$

where $\mathtt{q}_{POI}$ denotes the flux of the protein of interest produced, $C_{POI}$ denotes carbon number of the protein of interest, $\mathcal{R}$ denotes the number of reactants, $\mathtt{q}_{m_i}$ denotes the uptake flux of the $i^{th}$ reactant, and $C_{m_i}$ denotes the carbon number of the $i^{th}$ reactant.

## Quantification of uncertainty.

Experimental factors taken from literature, for example macromolecular concentrations or elongation rates, have uncertainty associated with their values. To quantify the influence of

this uncertainty on model performance, we randomly sampled the expected physiological ranges for these parameters as determined from literature. An ensemble of N = 100 flux distributions was calculated for the three different cases we considered: control (with amino acid synthesis and uptake), amino acid uptake without synthesis, and amino acid synthesis without uptake. The flux ensemble was calculated by randomly sampling the maximum glucose consumption rate within a range of 0 to 30 mM/h, (determined from experimental data) and randomly sampling RNA 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 12 and 18 μM, the RNA polymerase elongation rate between 20 and 30 nt/sec, and the ribosome elongation rate between 1.5 and 3 aa/s (*26*, *27*).

## Global sensitivity analysis.

We conducted a global sensitivity analysis using the variance-based method of Sobol to estimate which parameters controlled the performance of the cell-free protein synthesis reaction (*40*). We computed the total sensitivity index of each parameter relative to three performance objectives: productivity of the protein of interest, energy efficiency and carbon yield. We established the sampling bounds for each parameter from literature. We used the sampling method of Saltelli *et al.* (*41*) 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 (*42*).

### Pairwise group knockouts.

Pairwise and single group knockouts were simulated by setting the flux bounds for the all the reactions in a group to zero. We grouped reactions in the cell-free network into 19 subgroups (available in Supporting Information). We computed the norm of the productivity of CAT for each pairwise knockout compared to the productivity of CAT with no knockouts. We also computed the norm of the flux distribution for each pairwise knockout compared to the flux distribution with no knockouts.

## 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.
- Supporting Information: Performance trendlines as a function of carbon number, transcription/translation stoichiometric coefficients of energy species, and experimental measurements of CAT production.
- Carbon Yield Sensitivity Analysis: Global sensitivity analysis on deGFP carbon yield.
- Metabolites and reactions of the cell-free stoichiometric network.

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. Pardee, K., Slomovic, S., Nguyen, P., Lee, J., Donghia, N., Burrill, D., Ferrante, T., McSorley, F., Furuta, Y., Vernet, A., Lewandowski, M., Boddy, C., Joshi, N., and Collins, J. (2016) Portable, On-Demand Biomolecular Manufacturing. *Cell 167*, 248 – 259.e12.

7. 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.

8. Wiechert, W. (2001) 13C Metabolic Flux Analysis. *Metabolic Engineering 3*, 195 – 206.

9. Schuster, S., Fell, D. A., and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol 18*, 326–32.

10. Schilling, C. H., Letscher, D., and Palsson, B. O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol 203*, 229–48.

11. Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2006) Thermodynamics-Based Metabolic Flux Analysis. *Biophysical Journal 92*, 192–1805.

12. Hamilton, J. J., Dwivedi, V., and Reed, J. L. (2013) Quantitative Assessment of Thermodynamic Constraints on the Solution Space of Genome-Scale Metabolic Models. *Biophysical Journal 105*, 512–522.

13. Varma, A., and Palsson, B. O. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology 60*, 3724–3731.

14. Sánchez, C., Quintero, J. C., and Ochoa, S. (2015) Flux balance analysis in the production of clavulanic acid by Streptomyces clavuligerus. *Biotechnology Progress 31*, 1226–1236.

15. Edwards, J. S., and Palsson, B. O. (2000) Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics 1*, 1.

16. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., and Palsson, B. Ø. (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology 9*.

17. Covert, M. W., Knight, E. M., Reed, J. L., Herrgård, M. J., and Palsson, B. Ø. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature 429*, 92–6.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. Zhang, Y., Thiele, I., Weekes, D., Li, Z., Jaroszewski, L., Ginalski, K., Deacon, A. M., Wooley, J., Lesley, S. A., Wilson, I. A., Palsson, B., Osterman, A., and Godzik, A. (2009) Three-Dimensional Structural View of the Central Metabolic Network of Thermotoga maritima. *Science 325*, 1544–1549.

24. Chang, R. L., Andrews, K., Kim, D., Li, Z., Godzik, A., and Palsson, B. O. (2013) Structural Systems Biology Evaluation of Metabolic Thermotolerance in Escherichia coli. *Science 340*, 1220–1223.

25. 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.

26. 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.

27. 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.

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

29. Lee, S., Phalakornkule, C., Domach, M. M., and Grossmann, I. E. (2000) Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers & Chemical Engineering 24*, 711 – 716.

30. Mahadevan, R., and Schilling, C. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering 5*, 264 – 276.

31. Schuetz, R., Kuepfer, L., and Sauer, U. (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology 3*.

32. Zamboni, N., Fendt, S.-M., and Sauer, U. (2009) $^{13}$C-based metabolic flux analysis. *Nature Protocols 4*, 878–92.

33. Zawada, J., Richter, B., Huang, E., Lodes, E., Shah, A., and Swartz, J. R. *Fermentation Biotechnology*; Chapter 9, pp 142–156.

34. Jewett, M., Voloshin, A., and Swartz, J. In *Gene cloning and expression technologies*; Weiner, M., and Lu, Q., Eds.; Eaton Publishing: Westborough, MA, 2002; pp 391–411.

35. Kigawa, T., Muto, Y., and Yokoyama, S. (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *Journal of Biomolecular NMR 6*, 129–134.

36. Jewett, M. C., and Swartz, J. R. (2004) Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering 86*, 19–26.

37. Kim, D.-M., and Swartz, J. R. (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnology and Bioengineering 74*, 309–316.

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

39. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells. *Nature 491*, 249–53.

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

41. 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.

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