Systematic evaluation of parameterization for genome-scale metabolic models of cultured mammalian cells

Song-Min Schinn[1], Carly Morrison[2], Wei Wei[2], Lin Zhang[2], Nathan E. Lewis[1,3,4]

1. Department of Pediatrics, University of California, San Diego
2. Pfizer, Biotherapeutics Pharmaceutical Sciences, Andover, MA
3. Department of Bioengineering, University of California, San Diego
4. Novo Nordisk Foundation Center for Biosustainability at UC San Diego

Abstract

Genome-scale metabolic models describe cellular metabolism with mechanistic detail. Given their high complexity, such models need to be parameterized correctly to yield accurate predictions and avoid overfitting. Effective parameterization has been well-studied for microbial models, but it remains unclear for higher eukaryotes, including mammalian cells. To address this, we enumerated model parameters that describe key features of cultured mammalian cells – including cellular composition, bioprocess performance metrics, mammalian-specific pathways, and biological assumptions behind model formulation approaches. We tested these parameters by building thousands of metabolic models and evaluating their ability to predict the growth rates of a panel of phenotypically diverse Chinese Hamster Ovary cell clones. We found the following considerations to be most critical for accurate parameterization: (1) cells limit metabolic activity to maintain homeostasis, (2) cell morphology and viability change dynamically during a growth curve, and (3) cellular biomass has a particular macromolecular composition. Depending on parameterization, models predicted different metabolic phenotypes, including contrasting mechanisms of nutrient utilization and energy generation, leading to varying accuracies of growth rate predictions. Notably, accurate parameter values broadly agreed with experimental measurements. These insights will guide future investigations of mammalian metabolism.

Introduction

Genome-scale metabolic network models contextualize large-scale biological data with curated biochemical knowledge. Thus, such models are valuable platforms for integrating and analyzing disparate omics data[1], to probe the molecular basis of metabolism[2,3], disease[4–6], and human-microbiome interactions[7,8]. A general challenge of metabolic network models is their under-constrained nature due to high degrees of freedom. To address this challenge, these models can be constrained by the following established steps: (1) create a context-specific sub-model consisting of the most relevant metabolic reactions for a given context by interpreting gene expression data[9–11], (2) input experimental data as boundary conditions, and (3) hypothesize likely metabolic flux values by applying optimality principles to the model reaction network[12–14]. These steps parameterize the model with underlying assumptions of cellular makeup[15,16], metabolic pathway organization[17,18] and 'cellular objectives'[13,14,19–21]. In the last two decades, parameterization has been extensively explored and refined for model microbes such as *Escherichia coli*[22,23], *Saccharomyces cerevisae*[24,25] and others[26,27]. However, parameters for higher eukaryotes remain under-characterized, despite notable recent advances[28–30].

Recently, there has been increased interest in applying metabolic network models to investigate cultured mammalian cells, such as Chinese hamster ovary (CHO) cells[31–36,11,37], which are the preeminent expression system for large-scale biotherapeutics manufacturing. Metabolism is particularly relevant to cell line performance, given its role in meeting the anabolic and bioenergetic needs of cell growth and protein production[38,39]; therefore it has been a key target in cell line development[40–44]. Metabolic network models can be used with a wide array of *in silico* methods[45,46] to facilitate industrial bioprocess applications[47–49] – e.g. predict metabolic phenotype[50,33,51], identify metabolic bottlenecks[36], and optimize media formulation[35,52].

Mammalian cells are distinct from microbes in their metabolic pathways, auxotrophic requirements, sub-compartmentalization and regulation[53], and therefore require parameterization that capture such differences. A few studies have investigated mammalian-specific parameters describing cell biomass[54], novel metabolic pathways[32] and cellular objectives[55,34]. These parameters, however, have not been rigorously evaluated against wide-ranging conditions;

meanwhile, other features of mammalian metabolism and growth remain unexplored altogether. To address this, we systematically investigate a panel of parameters for describing diverse cultured mammalian cell metabolic phenotypes, with a focus on CHO cells. Our work identifies key parameters for predicting metabolic phenotypes, challenges conventional parameterization approaches, and guides future modeling efforts for mammalian metabolism.

Results

A cell line-specific model was constructed for phenotypically diverse CHO clones

Cultured mammalian cells, such as CHO cells, consume media nutrients – e.g. glucose, amino acids and cofactors – to produce biomass, recombinant proteins and byproducts. Depending on genotype and environmental conditions, CHO cells vary in their metabolic rates and pathway usage, resulting in high metabolic heterogeneity. Here, we have examined a variety of CHO metabolic phenotypes to identify general modeling principles and avoid overfitting. Specifically, we investigated 10 CHO clones that differed in recombinant antibody, bioreactor conditions and gene knockout treatment (see Methods). Consequently, the cells exhibited varying patterns of nutrient consumption and byproduct secretion (Fig. 1A), ultimately leading to divergent growth and productivity phenotypes (Fig. 1B; Fig. S1). In addition, we also examined temporal variation in metabolism[56,57]. We observed the 10 clones between culture days 4 to 11, during which the cells transitioned from exponential proliferation towards low growth and increased protein secretion (Fig. 1C). In short, we surveyed a wide range of cross-clonal and cross-temporal diversity in metabolic phenotypes for our investigation.

We hypothesized that these diverse CHO metabolisms could be described by a well-parameterized metabolic network model. To test this, we constructed a cell line-specific metabolic network model by modifying a previously published genome-scale model of CHO metabolism[31]. Briefly, this was done by the following steps: (1) transcriptomics data were analyzed to quantify the relative activity of 210 metabolic tasks to hypothesize a list of active metabolic genes[10,58,59], (2) these active genes and their associated model reactions were refined into a fully functional model via the mCADRE algorithm[60], and (3) the subsequent draft model was manually curated.

Mammalian-specific parameters embedded biological assumptions onto model

To enable the curated model of describing diverse CHO metabolisms, we identified salient features of mammalian cell cultures and formulated them as model parameters (Table 1; see Supplementary Document for detailed and technical discussions on their computational implementation, including annotated code). First, we considered cellular biomass and its macromolecular makeup (Table 1, parameters 1-2). These parameters are important since metabolic network models describe proliferation indirectly as a production of biomass constituents – e.g. proteins, lipids, nucleic acids and carbohydrates – by a predefined ratio[12].

Second, we considered that cellular mass and viability change considerably during a fed-batch culture, indicating internal shifts in metabolism and productivity[61]. For example, in the presented experiments, cellular mass increased by as much as 40-70% throughout the culture period (Fig. S2). Several clones also suffered from moderate amounts of cell death due to declining viability in the later stages of culture (Fig. S3). Up to now, modeling efforts have treated these features as static, even for multi-week time-course studies. Here we present parameters that reflect the time-course nature of cellular mass and viability (Table 1, parameters 3-5) to refine data input and interpretation of model results.

Third, we improved descriptions of mammalian metabolic pathways which are not found in microbes (Table 1, parameters 6-8). For example, we included in the model an intricate multi-organelle secretory pathway[62] to improved estimations of metabolic expenditures towards producing complex recombinant proteins. Similarly, we also considered metabolic pathways for producing various byproducts from amino acid catabolism[63], and cellular death and turnover[64].

Fourth, we reevaluated the biological premises behind applying the optimality principle to the metabolic network model (Table 1, parameter 9). This mathematical operation assumes that cellular metabolism operates at some optimal condition and is crucial to constraining the metabolic network model. As an example, proliferative bacteria have often been described to *maximize biomass production* for rapid growth, which is computationally formulated as a biomass objective function[12]. This objective function presumes that biomass production is principally limited by nutrient availability. Meanwhile, a variety of alternative objective

functions have also been explored[13,14] – e.g. optimal energy generation per substrate[65], minimized redox potential[55,66,67], efficient use of enzyme capacity[68] and streamlined amino acid transport[34]. Compared to microbes, mammalian cells are larger, more complex and wired for multicellularity, all of which increase the costs of homeostatic maintenance and limit proliferation. Therefore, we compare the conventional *maximize biomass production* function against alternative objective functions describing mammalian metabolism as chiefly limited by various homeostasis requirements – e.g. thermal[69], proteomic[68,70,71], reactive oxygen species or redox homeostasis[72].

Experimental measurements help produce robust model parameters

The above-described model parameters were systematically varied to evaluate their impact on accurately predicting CHO metabolism. The nine parameters were randomly permuted in value to within-literature-reported ranges, producing a total of 4000 different parameter settings. These parameter settings were tested against experimental data for 10 clones between culture days 4 to 11, resulting in a total of 320,000 model predictions (Fig. 1a). Each of these predictions yielded a model-predicted growth rate that could be compared to experimental measured growth rates (Fig. 1b). For ease of interpretation, prediction errors were preprocessed to yield an 'accuracy' metric (Fig. 2c). Specifically, the prediction errors were transformed by a negative-log function and normalized so that the average error metric for all clones and timepoints would lie between 0 and 1. For reference, when growth rates were predicted to within 5% of experimental measurements for all 80 points, the accuracy metric equaled 0.8.

Overall, the various parameter settings resulted in a wide range of prediction accuracies (Fig. S4). About one-eighth of the parameter settings robustly described at least eight of the ten clones with high accuracy (accuracy > 0.8). Half of the parameter settings predicted only two or fewer clones with high accuracy, indicating that the models were highly overfit or poorly parameterized. The remaining one-third of the parameter settings predicted a moderate range of clones with high accuracy, suggesting various degrees of overfitting. Overfit parameters of one clone tended to also describe biologically similar clones better, recapitulating known differences in cell line lineage and bioprocess conditions (Fig. S5).

Experimental data helped produce robust models. A model parameterized according to experimental measurements (Table 1, parenthesized) yielded an average accuracy of 0.82 across all clones and timepoints (Fig. 2b; 2c, blue line). Likewise, parameter settings with values close to experimental measurements (within ±15%) resulted in comparable accuracies (0.83 ± 0.11). Only a minority of parameter settings (2%) significantly deviated from experimental values yet managed to perform comparably. The value of experimental measurements was especially apparent for the *cell death rate* parameter describing late-stage cell death observed only in some clones. The *cell death rate* parameter helped the model recapitulate observed cross-clonal variation (Fig. S6).

Selection of objective function has greatest influence on model prediction

We estimated the importance of the parameters for prediction accuracy using a linear regression model (Fig. 3). Of the nine parameters, *objective function formulation* had an outsized influence on prediction accuracy. Specifically, the *maximize biomass production* function was correlated strongly with poor predictions. In contrast, the *minimize cytosolic NADPH regeneration* function was well correlated with accurate predictions. The following parameters also improved model predictions: *cell death rate*, *biomass composition* and *consideration of dynamic dry weights to calculate growth rate*. Interestingly, the time-course of biomass weight was more important than the baseline weight value itself. The remaining parameters had negligible impact on model accuracy (Fig. S7).

The *objective function formulations* provided detailed and mechanistic hypotheses of the intracellular metabolic activities underlying the growth rate predictions (Fig. 4). The *maximize biomass production* objective function predicted an efficient and highly active metabolism, which utilized almost all consumed substrates towards generating energy and biomass. As a result, the model consistently overestimated growth rates across clones and timepoints (Fig. S8). In contrast, several homeostasis-related objective functions predicted somewhat inefficient and wasteful metabolisms that consumed ample substrate only to partially discard them as various byproducts. This hypothesis was supported by carbon balance estimations, which also suggest that consumed substrates are converted only partially towards biomass and secreted recombinant proteins (Fig S9).

These objective functions specifically hypothesized two different carbon shunting mechanisms: (1) secretion of glycerol, which is converted from the glycolytic intermediate dihydroxyacetone phosphate via glycerol-3-phosphate dehydrogenase and glycerol-3-phosphate phosphatase (Fig. 3, *B*), and (2) secretion of malate, an intermediate of the citric acid cycle and malate-aspartate shuttle (Fig. 3, *D*). The objective functions *minimize enzyme costs*, *minimize Gibbs energy dissipation* and *minimize cytosolic NADPH regeneration* predicted shunting via glycerol, with varied fluxes through glycolysis, glycerol synthesis, mitochondrial respiration and pyruvate carboxylate. The objective function *minimize mitochondrial NADH regeneration* predicted shunting via malate. This loss of flux in the citric acid cycle was then counteracted by converting carbon dioxide from respiration to bicarbonate, to eventually oxaloacetate via pyruvate carboxylate[73] (Fig. 3, *F*). The secretions of both glycerol and malate have been reported previously in CHO fed-batch systems[74,75,39], although only glycerol was observed in the present study to accumulate in the spent medium (Fig. S10).

Lastly, the *minimize ROS synthesis* objective function hypothesized negligible activities in glycolysis and citric acid cycle, deviating from known CHO biology. Instead, most consumed glucose was diverted to the folate cycle via phosphoserine transamination to produce energy and byproduct formic acid (Fig. 3, *C*), resulting in an improbable metabolic phenotype. Additional comparison of objective functions and their influence on metabolic configurations are also provided (Fig. S11).

Discussion

Parameters help constrain metabolic networks sufficiently and accurately by embedding knowledge and data onto models. We compared thousands of distinctly parameterized models for their ability to describe CHO clones with wide-ranging phenotypes. The resulting analyses lead to important insights for parameterizing metabolic network models of cultured mammalian cells. Specifically, this study (1) confirms that experimental data improve model parameterization, (2) identifies relevant parameters for mammalian metabolic models, chief among them the *objective function formulation*, and (3) challenges the popular use of the

maximize biomass production objective function for models of exponentially growing mammalian cells and explores promising alternatives.

Results from the parameter sensitivity analysis agreed with experimental observations in several notable ways. First, experimentally measured parameter values resulted in robust and accurate model predictions, confirming the value of experimental measurements in parameterization. Second, parameter settings that described one clone well also tended to describe similar clones well, recapitulating differences in cell line lineages and glucose-feeding strategies. Lastly, parameter values of *cell death rate* recapitulated experimental observations at the clone level. These broad agreements demonstrate that well-parameterized metabolic network models can describe diverse metabolic states.

The sensitivity analysis identified parameters that strongly affected model precision. Of all the investigated parameters, the *objective function formulation* had a predominant influence on model-predicted growth rates and their underlying metabolic activities. *Biomass composition* was also reconfirmed as a key parameter, agreeing with previous findings in microbial models[15,76,16]. Lastly, novel parameters accounting for time-course changes in cell size and viability improved model predictions substantially. This emphasizes the relevance of dynamic and bioprocess-specific features of mammalian cell cultures for parameterization.

Notably, the widely-used *maximize biomass production* objective function performed poorly in describing mammalian metabolism. This formulation assumes that biomass production is chiefly limited by nutrient availability, and has been reported as robustly predictive for a wide range of experimental conditions and different models[14,67]. Accordingly, several recent studies using mammalian metabolic models (e.g., CHO cells, cancer cells, stem cells, immune cells) adopted this assumption. In the present study, however, *maximize biomass production* consistently overestimated growth rates and intracellular metabolic activity. Alternatively, five other presented formulations assumed that mammalian cells limit biomass production to maintain various homeostatic conditions. These homeostasis-limited assumptions performed markedly better than *maximize biomass production* by hypothesizing a restrained CHO metabolism. Objective functions considering cytosolic redox homeostasis and enzyme capacity predicted metabolic configurations that agreed especially well with observed growth rates and exometabolomic measurements. Future studies can rigorously refine and validate these

formulations by comparing predicted intracellular flux distributions with experimental flux measurements, following the well-established footsteps of *Escherichia coli* models[13,14].

Conclusion

Proper parameter selection is essential for metabolic network models to provide accurate descriptions of cellular metabolism. Above all, the objective function parameter is highly influential in predicting metabolic and bioprocess performance phenotype. For this purpose, mammalian cells may be described as limiting their metabolic activities to maintain homeostasis. Various objective functions predicted different growth rates and metabolic pathway usage, leaving room for further validation and refinements. In addition, parameters describing biomass and time-course metabolic shifts also improved model predictions, especially when set to experimentally measured values. These results will guide future efforts to develop and improve models for mammalian cell metabolism, with applications ranging from biotherapeutic production[35,37,77] to unraveling the metabolic basis of diverse diseases[78–80].

Methods

Cell culture experiments

Two production fed batch processes were used, Fed batch 1 and Fed batch 2. Both fed batch processes used chemically defined media and feeds over the 12-day cell culture. Fed batch 1 used a glucose restricted fed batch process called HiPDOG[65]. Glucose concentration is kept low during the initial phase of the process, Day 2-7, through intermittent addition of feed medium containing glucose at the high end of pH dead-band and then glucose was maintained above 1.5 g/L thereafter. These conditions help restrict lactate production in fed batch cultures without compromising the proliferative capability of cells. In Fed batch 2 a conventional cell culture process was used where glucose was maintained above 1.5 g/L throughout the process.

For both process conditions, bioreactor vessels were inoculated at 2 x 10^6 viable cells/mL. The following bioprocess characteristics were quantified daily using a NOVA Flex BioProfile Analyzer (Nova Biomedical, Waltham, MA): viable cell density, average live cell diameter and

concentrations of glucose, lactate, glutamate, and glutamine. Viable cell density data was converted to growth rates by following equation to be compared to model-predicted growth rates.

$$(1) \qquad Growth\ rate = \frac{1}{vcd} \cdot \frac{\Delta vcd}{\Delta time}$$

In addition, cell cultures were sampled on specific days (Day 0, 3, 5, 7, 10 and 12) for cell pellets and supernatant for transcriptomic (RNA-Seq), metabolomics analyses, and titer measurements. Titers were analyzed using a protein A HPLC (model 1100 HPLC, Agilent Technologies, Inc., Santa Clara, CA, protein A column model 2-1001-00, Applied Biosystems, Foster City, CA). The RNA libraries were mapped to the CHO genome[81,82] using STAR aligner[83] and processed to quantify gene expression counts with HTSeq-count[84].

Flash-frozen cell pellets (10E6 cells) and supernatant (1 mL) were collected from bioreactor runs for A1 and A2 for each sampling day. Similarly, cell pellet and supernatant samples were collected from bioreactor runs for B1 and B2 for days 7, 10 and 12. Collected samples were sent to Metabolon (Metabolon Inc, Morrisville, NC) for metabolomics analyses. Proteins were removed by methanol precipitation and the metabolites were recovered by vigorous shaking and centrifugation. The extracted samples were run for reverse-phase Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy with negative ion mode ESI. Raw data was extracted, peak-identified and processed for quality control using Metabolon's hardware and software. The raw ion count data was normalized against the extracted proteins quantified using a Bradford assay.

Biomass measurement experiments

During cell culturing of clones Z1-Z4, 20 x 10^6 cells were sampled during days 0, 3, and 7. At day 7, an additional 1 mL of culture was sampled. The samples were sub-divided and stored at -80 °C for the following measurements.

Cellular dry weights were measured by weighing dehydrated cells, as previously described[85]. Briefly, for each sample, an aluminum weigh boat was dried at 70 °C for 48 hours and pre-weighed after being cooled down to room temperature. Stored samples (5×10^6 cells) were thawed in ice, centrifuged, washed with PBS and centrifuged again. The cell pellets were

resuspended in deionized water and transferred to the pre-weighed aluminum boats and dried for 48 hours at 70° C. Then, the samples were cooled back to room temperature and weighed in triplicates. The weighed mass was divided by cell count to calculate cellular dry weights.

Cellular protein contents were quantified by the Bicinchoninic Acid assay using a commercial kit (Pierce™ BCA Protein Assay Kit; Thermo Fisher, Waltham, MA). Stored samples (1×10^6 cells) were thawed in ice, centrifuged, and washed with PBS. Assay standards were prepared as instructed for the range of 0 – 2000 μg/mL protein. The cells were disrupted by a mixture of cell lysis agent (CelLytic™ M; Sigma Aldrich, St. Louis, MO) and protease inhibitor cocktail (Sigma Aldrich, St. Louis, MO). A negative control was prepared from the mixture without cell samples. The standards, samples and control were treated with assay reagents, incubated and measured following kit instructions. The resulting standard curve was confirmed to be linear ($R^2 = 0.988$). Measured absorbance values were converted to protein concentrations following this standard curve and considering background absorbance described by the negative control. Protein content in biomass was calculated to be 64.9% (±6.8) for days 3 and 7, agreeing well with previously published measurements.

The lipid contents of the cells were extracted using the Blight and Dyer protocol[86]. Briefly, for each bioreactor run, 1 mL stored culture samples were centrifuged and re-suspended in water. Chloroform and methanol were added sequentially in 1:2 ratio and mixed well by vortex. Then, chloroform and water are sequentially added in 1:1 ratio and mixed well in between by vortex. The resulting mixture was then centrifuged at 4000 RPM for 15 minutes at 20 °C, resulting in phase separation. The upper phase and cell debris were carefully discarded. The lower chloroform phase containing the lipids was carefully transferred to pre-weighed glass beakers. Chloroform evaporated in a fume hood for 24 hours, after which the lipid samples were weighed. The measured weights were divided by cell count and measured cellular dry weight, yielding a cellular lipid content of 21.62% (±2.46), agreeing well with previously published measurements.

Constraint-based modeling and analysis

Flux balance analyses were conducted using the COBRA Toolbox 2.0[87] and the Gurobi solver version 8.0.0 (Gurobi Optimization; Beaverton, Oregon) in MATLAB R2018b (MathWorks; Natick, Massachusetts, USA).

Constructing a context-specific model

A community-consensus genome-scale model[31] was modified to produce a cell line-specific model for the cell lines from Pfizer by qualitatively interpreting transcriptomics data for clones A1 and A2, following benchmarking results[10,58,59]. Specifically, transcript abundance data for each gene was pre-processed by equation 2 to produce a binary gene score metric (ON: >5, OFF: 0). The equation's threshold values were calculated from the mean transcript abundance across all samples. Genes with transcript abundances of top and bottom 25 percentile were binarized as ON and OFF, respectively. The binary gene scores were converted to reaction scores by considering multimeric or isozyme relationships as described by the model's gene-protein-relationship matrix. These binary reaction scores were used as input data for model extraction by the mCADRE algorithm[60].

(2) $$Gene\ Score = 5 * \log(1 + \frac{transcript\ abundance}{threshold})$$

(3) $$Reaction\ Score = \mathrm{GPR}\ (\ Gene\ Score_1 \ldots Gene\ Score_n)$$

The resulting draft model was then manually curated to be consistent with experimental data and established cell biology. Given the defined media of Pfizer's bioprocess, we allowed extracellular transport only for metabolites with experimental basis – e.g. glucose, lactate, proteinogenic amino acids, oxygen, carbon dioxide etc. We confirmed amino acid essentialities and CHO-specific auxotrophies for arginine, cysteine and proline. We also ensured the inclusion of synthesis pathways of non-essential amino acids and degradation pathways for all amino acids. We confirmed the model utilized glucose via central carbon metabolism reactions, and removed unlikely reactions such as the methylglyoxal pathway. We ensured cytosolic and mitochondrial compartmentalization of redox cofactors such as NAD/H and NADP/H.

Model implementation of parameters

The parameter value of *biomass weight* (i.e. cellular dry weight) was used to convert the units of experimental measurements from [pg·cell^{-1}·day^{-1}] to [mmol·g_{DW}^{-1}·hr^{-1}]. When *considering dynamic dry weight (I)* for data input, the dry weight values were estimated from cell diameter measurements by assuming a constant cellular density and geometry[88]. When *considering dynamic dry weight (II)* for growth rate calculations, the following equation was used.

$$\text{Adjusted Growth Rate} = \text{Growth Rate} * \frac{Future\ dry\ weight}{Current\ dry\ weight} \quad (4)$$

The *biomass composition* parameter was adjusted by altering the stoichiometric formulation of the biomass production reaction. Specifically, macromolecular protein content and lipid content were inversely varied, while assuming content of nucleotides and carbohydrates were stable. This was because protein and lipid contents were observed to vary the most[54]. Molecular makeup of amino acids and specific lipid molecules was unchanged from the genome-scale model, which was based on approximations from hybridoma cells. This was because variations in molecular composition did not affect model predictions. The *cell death rate* was implemented by the following equation for timepoints between days 7-11, in accordance to experimental observations.

$$\text{Adjusted Growth Rate} = \text{Growth Rate} - Death\ Rate \quad (5)$$

Consideration of secretion costs was implemented by joining a model of the CHO secretory pathway[62] onto our metabolic network model. The secretory pathway model included about 100 reactions for folding, transport, post-translational modification and related activities in the endoplasmic reticulum and Golgi body. We approximated the structure, folding requirements and glycosylation needs of Pfizer's monoclonal antibodies by that of Rituximab. *Biomass turnover rate* was implemented a single ATP hydrolysis reaction.

Consider byproduct synthesis was implemented for the following byproduct molecules: glycerol, formic acid, 2-hydroxybutyurate, isovalerate, acetic acid, hydroxyphenylpyruvate, hydroxyphenyl-lactate, phenyl-lactate, homocysteine, indole 3-lactate, citrate, and malate. These notably include products of amino acid catabolism as well as citric acid cycle intermediates.

Glycerol was assumed to be synthesized from dihydroxyacetone phosphate via glycerol-3-phosphate dehydrogenase and glycerol-3-phosphate phosphatase. Formic acid was assumed to be synthesized in the folate cycle during tetrahydrofolate synthesis from 10-formyltetrahydrofolate. The synthesis of isovalerate, 2-hydroxybutyurate, acetic acid, hydroxyphenylpyruvate, hydroxyphenyl-lactate, phenyl-lactate, homocysteine, and indole 3-lactate were formulated as catabolic byproducts, as previously published[63]. The model already included synthesis reactions for homocysteine, acetic acid, citrate and malate. Extracellular transport reactions were enabled for all byproducts.

Formulation of the objective function

The most widely used objective function is the maximization of biomass production, which formulates the model as a linear programming optimization problem around the biomass production reaction[89]. Alternatively, we formulated the objective function as a bilevel linear programming optimization to reflect cellular limitations to biomass production. First, possible limitations to biomass production were enumerated, as discussed above. Second, model reactions were annotated for 'penalty' metrics describing such limitations: Gibbs energy dissipation (90 reactions), associated enzyme mass (1071 reactions), and stoichiometric production of ROS (2 reactions), cytosolic NADPH (21 reactions), and mitochondrial NADH (29 reactions). Gibbs energy dissipation was approximated by reported *S. cerevisiae* values[69]. Third, according to parameter settings, the 'penalty' metric of choice was minimized based on experimentally measured substrate consumption rates. This was followed by a maximization of biomass production while maintaining the minimized 'penalty' metric. The resulting objective describes the maximum biomass production possible given some minimized cellular 'penalty'.

The presented method builds upon previous work that minimizes network flux[90] and parsimonious flux balance analysis[70] and the 'max biomass per unit flux' objective[13] with important differences. Previous methods implicitly assumed that biomass production was gated by substrate availability and therefore first maximized biomass production and then minimized total unit flux as a 'penalty' metric. Here, we assume cellular limitations restrict growth rather than substrate availability, and therefore inverse the order of optimization. Also, while previous methods penalized reactions uniformly, the presented method weighs the 'penalties' to better

reflect their underlying assumptions. For example, reactions that are catalyzed by complex multimers are 'penalized' more severely under the assumptions of enzyme capacity limitation. Like previous methods, these bilevel objective formulations require the model to be irreversible.

Parameters sensitivity analysis

4000 parameter settings were generated by randomly varying parameters to within a range based on literature or experimental measurements (Table 1). For each setting, parameter values were implemented to yield a flux balance analysis prediction. The prediction used as input data the following experimental measurements: specific productivity and the consumption rates of glucose, lactate and 20 proteinogenic amino acids. The prediction was repeated 80 times by using experimental data from 10 clones across 8 days. The resulting 80 model-predicted growth rates were compared to their respective experimental values to yield prediction errors. To facilitate human interpretation, the prediction errors were first transformed by the negative log function. Then, the transformed values were normalized so that mean metric values for all 80 predictions laid between 0 and 1. The resulting metric was called 'accuracy'.

A regression analysis between accuracy and parameter values were performed, using Python 3.7.3 in the Jupyter Notebook environment. For this analysis, parameter values were normalized to be between 0 and 1 by referencing minimum and maximum possible values (Table 1). A linear regression analysis was performed according to equation 6, where B_i and X_i represent respectively the effect size and normalized parameter values of parameter *i*. The resulting effect size of parameters and their p-values were used to identify parameter relevance.

$$\text{Accuracy} = \mathrm{B}_0 + \sum_{i=1}^{9} B_i \cdot \mathrm{X}_i \qquad (6)$$

Flux sampling analysis

We computed distributions of likely fluxes for each reaction in the model by stochastically sampling 5000 points within the solution space via the Markov chain Monte Carlo sampling via artificially centered hit-and-run algorithm, as described previously[91]. First, the metabolic network model was parameterized according to experimentally observed values. Then,

experimental measurements for all clones on day 4 were used to constrain model reactions for biomass production, monoclonal antibody secretion and consumption of glucose, lactate and proteinogenic amino acids. A set of non-uniform 'points' or flux values was generated within the feasible flux space. Each point was subsequently moved randomly, while remaining within the feasible flux space. To do this, a random direction was first chosen. Second, the limit for how far the point can travel in the randomly-chosen direction was calculated. Lastly, a new random point on this line was selected. This process was iterated until the set of points approached a uniform sample of the solution space. For further analyses, sampled fluxes were normalized by total model flux[92]. We found that 577 model reactions accounted for 99% of total metabolic flux. Flux distributions of these 577 reactions were compared and visualized by Principle Component Analysis and PHATE[93] (Potential of Heat-diffusion for Affinity-based Trajectory Embedding). These analyses were done using Python 3.7.3 in the Jupyter Notebook environment.

References

1. Hyduke DR, Lewis NE, Palsson BØ. Analysis of omics data with genome-scale models of metabolism. *Mol Biosyst*. 2013;9(2):167-174. doi:10.1039/c2mb25453k

2. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet*. 2014;15(2):107-120. doi:10.1038/nrg3643

3. Sastry AV, Gao Y, Szubin R, et al. The Escherichia coli transcriptome mostly consists of independently regulated modules. *Nat Commun*. 2019;10(1):5536. doi:10.1038/s41467-019-13483-w

4. Yizhak K, Chaneton B, Gottlieb E, Ruppin E. Modeling cancer metabolism on a genome scale. *Mol Syst Biol*. 2015;11(6):817. doi:10.15252/msb.20145307

5. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun*. 2014;5:3083. doi:10.1038/ncomms4083

6. Lewis NE, Abdel-Haleem AM. The evolution of genome-scale models of cancer metabolism. *Front Physiol*. 2013;4:237. doi:10.3389/fphys.2013.00237

7. Magnúsdóttir S, Thiele I. Modeling metabolism of the human gut microbiome. *Curr Opin Biotechnol*. 2018;51:90-96. doi:10.1016/j.copbio.2017.12.005

8. Sen P, Orešič M. Metabolic Modeling of Human Gut Microbiota on a Genome Scale: An Overview. *Metabolites*. 2019;9(2). doi:10.3390/metabo9020022

9. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. *Front Physiol*. 2012;3. doi:10.3389/fphys.2012.00299

10. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst*. 2017;4(3):318-329.e6. doi:10.1016/j.cels.2017.01.010

11. Huang Z, Yoon S. Integration of Time-Series Transcriptomic Data with Genome-Scale CHO Metabolic Models for mAb Engineering. *Processes*. 2020;8(3):331. doi:10.3390/pr8030331

12. Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol*. 2010;13(3):344-349. doi:10.1016/j.mib.2010.03.003

13. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol Syst Biol*. 2007;3:119. doi:10.1038/msb4100162

14. Costa RS, Nguyen S, Hartmann A, Vinga S. Exploring the Cellular Objective in Flux Balance Constraint-Based Models. In: Mendes P, Dada JO, Smallbone K, eds. *Computational Methods in Systems Biology*. Lecture Notes in Computer Science. Springer International Publishing; 2014:211-224. doi:10.1007/978-3-319-12982-2_15

15. Dikicioglu D, Kırdar B, Oliver SG. Biomass composition: the “elephant in the room” of metabolic modelling. *Metabolomics*. 2015;11(6):1690-1701. doi:10.1007/s11306-015-0819-2

16. Lakshmanan M, Long S, Ang KS, Lewis N, Lee D-Y. On the impact of biomass composition in constraint-based flux analysis. *bioRxiv*. Published online May 28, 2019:652040. doi:10.1101/652040

17. Dikicioglu D, Oliver SG. Extension of the yeast metabolic model to include iron metabolism and its use to estimate global levels of iron-recruiting enzyme abundance from cofactor requirements. *Biotechnol Bioeng*. 2019;116(3):610-621. doi:10.1002/bit.26905

18. Monk JM, Lloyd CJ, Brunk E, et al. iML1515, a knowledgebase that computes Escherichia coli traits. *Nat Biotechnol*. 2017;35(10):904-908. doi:10.1038/nbt.3956

19. García Sánchez CE, Torres Sáez RG. Comparison and analysis of objective functions in flux balance analysis. *Biotechnol Prog*. 2014;30(5):985-991. doi:10.1002/btpr.1949

20. Zhao Q, Stettner AI, Reznik E, Paschalidis ICh, Segrè D. Mapping the landscape of metabolic goals of a cell. *Genome Biol*. 2016;17(1):109. doi:10.1186/s13059-016-0968-2

21. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P-J, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol*. 2017;13(8):935. doi:10.15252/msb.20167411

22. Edwards JS, Palsson BO. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A*. 2000;97(10):5528-5533. doi:10.1073/pnas.97.10.5528

23. Orth JD, Conrad TM, Na J, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol Syst Biol*. 2011;7:535. doi:10.1038/msb.2011.65

24. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res*. 2003;13(2):244-253. doi:10.1101/gr.234503

25. Heavner BD, Smallbone K, Price ND, Walker LP. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database*. 2013;2013. doi:10.1093/database/bat059

26. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biol*. 2019;20(1):121. doi:10.1186/s13059-019-1730-3

27. Yilmaz S, Walhout A. Metabolic network modeling with model organisms. *Curr Opin Chem Biol*. 2017;36:32-39.

28. Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol*. 2010;4:140. doi:10.1186/1752-0509-4-140

29. Witting M, Hastings J, Rodriguez N, et al. Modeling Meets Metabolomics-The WormJam Consensus Model as Basis for Metabolic Studies in the Model Organism Caenorhabditis elegans. *Front Mol Biosci*. 2018;5:96. doi:10.3389/fmolb.2018.00096

30. Brunk E, Sahoo S, Zielinski DC, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 2018;36(3):272-281. doi:10.1038/nbt.4072

31. Hefzi H, Ang KS, Hanscho M, et al. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst*. 2016;3(5):434-443.e8. doi:10.1016/j.cels.2016.10.020

32. Calmels C, McCann A, Malphettes L, Andersen MR. Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process. *Metab Eng*. 2019;51:9-19. doi:10.1016/j.ymben.2018.09.009

33. Popp O, Müller D, Didzus K, et al. A hybrid approach identifies metabolic signatures of high-producers for chinese hamster ovary clone selection and process optimization. *Biotechnol Bioeng*. 2016;113(9):2005-2019. doi:10.1002/bit.25958

34. Chen Y, McConnell BO, Gayatri Dhara V, et al. An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. *NPJ Syst Biol Appl*. 2019;5. doi:10.1038/s41540-019-0103-6

35. Fouladiha H, Marashi S-A, Torkashvand F, Mahboudi F, Lewis NE, Vaziri B. A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. *bioRxiv*. Published online 2020:751347. doi:10.1101/751347

36. Zhuangrong H, Seongkyu Y. Identifying metabolic features and engineering targets for productivity improvement in CHO cells by integrated transcriptomics and genome-scale metabolic model. *Biochem Eng J*. Published online May 5, 2020:107624. doi:10.1016/j.bej.2020.107624

37. Huang Z, Lee D-Y, Yoon S. Quantitative intracellular flux modeling and applications in biotherapeutic development and production using CHO cell cultures. *Biotechnol Bioeng*. 2017;114(12):2717-2728. doi:10.1002/bit.26384

38. Golabgir A, Gutierrez JM, Hefzi H, et al. Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. *Biotechnol Adv*. 2016;34(5):621-633. doi:10.1016/j.biotechadv.2016.02.011

39. Sumit M, Dolatshahi S, Chu A-HA, et al. Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses. *iScience*. 2019;12:102-120. doi:10.1016/j.isci.2019.01.006

40. Richelle A, Lewis NE. Improvements in protein production in mammalian cells from targeted metabolic engineering. *Curr Opin Syst Biol*. 2017;6:1-6. doi:10.1016/j.coisb.2017.05.019

41. Mulukutla BC, Mitchell J, Geoffroy P, et al. Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. *Metab Eng*. 2019;54:54-68. doi:10.1016/j.ymben.2019.03.001

42. Wang Q, Yin B, Chung C-Y, Betenbaugh MJ. Glycoengineering of CHO Cells to Improve Product Quality. *Methods Mol Biol Clifton NJ*. 2017;1603:25-44. doi:10.1007/978-1-4939-6972-2_2

43. Baik JY, Gasimli L, Yang B, et al. Metabolic engineering of Chinese hamster ovary cells: towards a bioengineered heparin. *Metab Eng*. 2012;14(2):81-90. doi:10.1016/j.ymben.2012.01.008

44. Gupta SK, Srivastava SK, Sharma A, et al. Metabolic engineering of CHO cells for the development of a robust protein production platform. *PLOS ONE*. 2017;12(8):e0181455. doi:10.1371/journal.pone.0181455

45. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*. 2012;10(4):291-305. doi:10.1038/nrmicro2737

46. Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc*. 2019;14(3):639-702. doi:10.1038/s41596-018-0098-2

47. Zhang C, Hua Q. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Front Physiol*. 2016;6. doi:10.3389/fphys.2015.00413

48. Kim HU, Lee SY. Applications of genome-scale metabolic network models in the biopharmaceutical industry. *Pharm Bioprocess*. 2013;1(4):337-339. doi:10.4155/pbp.13.37

49. Sommeregger W, Sissolak B, Kandra K, von Stosch M, Mayer M, Striedner G. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnol J*. 2017;12(7). doi:10.1002/biot.201600546

50. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol*. 1994;60(10):3724-3731.

51. Aurich MK, Fleming RMT, Thiele I. A systems approach reveals distinct metabolic strategies among the NCI-60 cancer cell lines. *PLOS Comput Biol*. 2017;13(8):e1005698. doi:10.1371/journal.pcbi.1005698

52. Traustason B. Amino Acid Requirements of the Chinese Hamster Ovary Cell Metabolism during Recombinant Protein Production. *bioRxiv*. Published online October 7, 2019:796490. doi:10.1101/796490

53. Gerdtzen ZP. Modeling Metabolic Networks for Mammalian Cell Systems: General Considerations, Modeling Strategies, and Available Tools. *Genomics Syst Biol Mamm Cell Cult*. Published online 2011:71-108. doi:10.1007/10_2011_120

54. Szeliova D, Ruckerbauer D, Galleguillos S, et al. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metab Eng*. Published online 2020.

55. Savinell JM, Palsson BO. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J Theor Biol*. 1992;154(4):421-454. doi:10.1016/s0022-5193(05)80161-4

56. Ahn WS, Antoniewicz MR. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metab Eng*. 2011;13(5):598-609. doi:10.1016/j.ymben.2011.07.002

57. Dean J, Reddy P. Metabolic analysis of antibody producing CHO cells in fed-batch production. *Biotechnol Bioeng*. 2013;110(6):1735-1747. doi:10.1002/bit.24826

58. Richelle A, Chiang AWT, Kuo C-C, Lewis NE. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLOS Comput Biol*. 2019;15(4):e1006867. doi:10.1371/journal.pcbi.1006867

59. Richelle A, Joshi C, Lewis NE. Assessing key decisions for transcriptomic data integration in biochemical networks. *PLOS Comput Biol*. 2019;15(7):e1007185. doi:10.1371/journal.pcbi.1007185

60. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol*. 2012;6(1):153. doi:10.1186/1752-0509-6-153

61. Pan X, Dalm C, Wijffels RH, Martens DE. Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Appl Microbiol Biotechnol*. 2017;101(22):8101-8113. doi:10.1007/s00253-017-8531-y

62. Gutierrez JM, Feizi A, Li S, et al. Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat Commun*. 2020;11(1):1-10. doi:10.1038/s41467-019-13867-y

63. Mulukutla BC, Kale J, Kalomeris T, Jacobs M, Hiller GW. Identification and control of novel growth inhibitors in fed-batch cultures of Chinese hamster ovary cells. *Biotechnol Bioeng*. 2017;114(8):1779-1790. doi:10.1002/bit.26313

64. Gunn JM, Brancheau MR. Protein turnover, growth and proliferation in CHO cells. Variation within and between mutant classes for salvage pathway enzymes. *Biochem J*. 1992;282 (Pt 1):49-57. doi:10.1042/bj2820049

65. Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J Physiol Regul Integr Comp Physiol*. 2001;280(3):R695-704. doi:10.1152/ajpregu.2001.280.3.R695

66. Knorr AL, Jain R, Srivastava R. Bayesian-based selection of metabolic objective functions. *Bioinforma Oxf Engl*. 2007;23(3):351-357. doi:10.1093/bioinformatics/btl619

67. Sánchez CEG, García CAV, Sáez RGT. Predictive Potential of Flux Balance Analysis of Saccharomyces cerevisiae Using as Optimization Function Combinations of Cell Compartmental Objectives. *PLOS ONE*. 2012;7(8):e43006. doi:10.1371/journal.pone.0043006

68. Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R, Liebermeister W. The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLOS Comput Biol*. 2016;12(11):e1005167. doi:10.1371/journal.pcbi.1005167

69. Niebel B, Leupold S, Heinemann M. An upper limit on Gibbs energy dissipation governs cellular metabolism. *Nat Metab*. 2019;1(1):125-132. doi:10.1038/s42255-018-0006-7

70. Lewis NE, Hixson KK, Conrad TM, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol*. 2010;6:390. doi:10.1038/msb.2010.47

71. Yeo HC, Hong J, Lakshmanan M, Lee D-Y. Enzyme capacity-based genome scale modelling of CHO cells. *bioRxiv*. Published online December 20, 2019:2019.12.19.883652. doi:10.1101/2019.12.19.883652

72. Ray PD, Huang B-W, Tsuji Y. Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. *Cell Signal*. 2012;24(5):981-990. doi:10.1016/j.cellsig.2012.01.008

73. Nicolae A, Wahrheit J, Bahnemann J, Zeng A-P, Heinzle E. Non-stationary 13C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. *BMC Syst Biol*. 2014;8(1):50. doi:10.1186/1752-0509-8-50

74. Sellick CA, Croxford AS, Maqsood AR, et al. Metabolite profiling of recombinant CHO cells: designing tailored feeding regimes that enhance recombinant antibody production. *Biotechnol Bioeng*. 2011;108(12):3025-3031. doi:10.1002/bit.23269

75. Chong WPK, Goh LT, Reddy SG, et al. Metabolomics profiling of extracellular metabolites in recombinant Chinese Hamster Ovary fed-batch culture. *Rapid Commun Mass Spectrom*. 2009;23(23):3763-3771. doi:10.1002/rcm.4328

76. Chan SHJ, Cai J, Wang L, Simons-Senftle MN, Maranas CD. Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*. 2017;33(22):3603-3609. doi:10.1093/bioinformatics/btx453

77. Gutierrez JM, Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnol J*. 2015;10(7):939-949. doi:10.1002/biot.201400647

78. Bordbar A, Palsson BO. Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *J Intern Med*. 2012;271(2):131-141. doi:10.1111/j.1365-2796.2011.02494.x

79. Chen Y, Li G, Nielsen J. Genome-Scale Metabolic Modeling from Yeast to Human Cell Models of Complex Diseases: Latest Advances and Challenges. *Methods Mol Biol Clifton NJ*. 2019;2049:329-345. doi:10.1007/978-1-4939-9736-7_19

80. Geng J, Nielsen J. In silico analysis of human metabolism: Reconstruction, contextualization and application of genome-scale models. *Curr Opin Syst Biol*. 2017;2:29-38. doi:10.1016/j.coisb.2017.01.001

81. Lewis NE, Liu X, Li Y, et al. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat Biotechnol*. 2013;31(8):759-765. doi:10.1038/nbt.2624

82. Li S, Cha SW, Heffner K, et al. Proteogenomic Annotation of Chinese Hamsters Reveals Extensive Novel Translation Events and Endogenous Retroviral Elements. *J Proteome Res*. 2019;18(6):2433-2445. doi:10.1021/acs.jproteome.8b00935

83. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

84. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169. doi:10.1093/bioinformatics/btu638

85. Edros RZ, McDonnell S, Al-Rubeai M. Using molecular markers to characterize productivity in Chinese hamster ovary cell lines. *PloS One*. 2013;8(10):e75935. doi:10.1371/journal.pone.0075935

86. Beck AE, Hunt KA, Carlson RP. Measuring Cellular Biomass Composition for Computational Biology Applications. *Processes*. 2018;6(5):38. doi:10.3390/pr6050038

87. Schellenberger J, Que R, Fleming RMT, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc*. 2011;6(9):1290-1307. doi:10.1038/nprot.2011.308

88. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res*. 2010;38(Database issue):D750-753. doi:10.1093/nar/gkp889

89. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28(3):245-248. doi:10.1038/nbt.1614

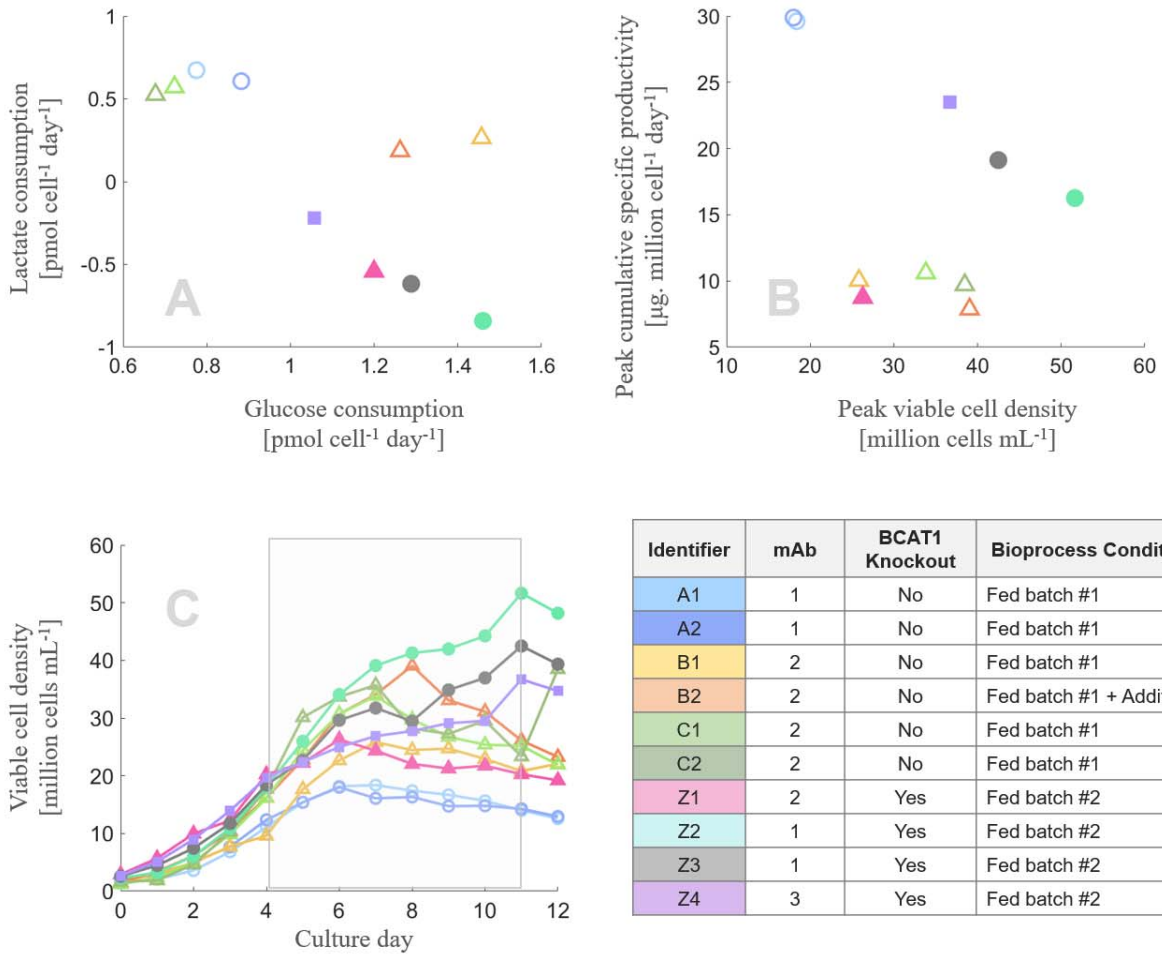
90. Klitgord N, Segrè D. The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome Inform Int Conf Genome Inform*. 2010;22:41-55.

91. Nam H, Lewis NE, Lerman JA, et al. Network Context and Selection in the Evolution to Enzyme Specificity. *Science*. 2012;337(6098):1101-1104. doi:10.1126/science.1216861

92. Lewis NE, Schramm G, Bordbar A, et al. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol*. 2010;28(12):1279-1285. doi:10.1038/nbt.1711

93. Moon KR, van Dijk D, Wang Z, et al. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol*. 2019;37(12):1482-1492. doi:10.1038/s41587-019-0336-3

Figures



Identifier	mAb	BCAT1 Knockout	Bioprocess Conditions
A1	1	No	Fed batch #1
A2	1	No	Fed batch #1
B1	2	No	Fed batch #1
B2	2	No	Fed batch #1 + Additive
C1	2	No	Fed batch #1
C2	2	No	Fed batch #1
Z1	2	Yes	Fed batch #2
Z2	1	Yes	Fed batch #2
Z3	1	Yes	Fed batch #2
Z4	3	Yes	Fed batch #2

Figure 1: 10 CHO clones with diverse metabolic phenotypes were studied – The examined clones expressed different monoclonal antibodies (1: ○, 2: △, 3: □) and were subjected to different bioprocess conditions (Fed batch #1: empty; Fed batch #2: filled). Due to these differences, the cells exhibited diverse metabolic phenotypes: (a) the cells consumed key nutrients such as glucose at varying amounts, and variously consumed or secreted lactate, (b) resulting in distinct growth and productivity performances. (c) These clones were observed between culture days 4 and 11, during which the cells traversed exponential and stationary phases (highlighted).

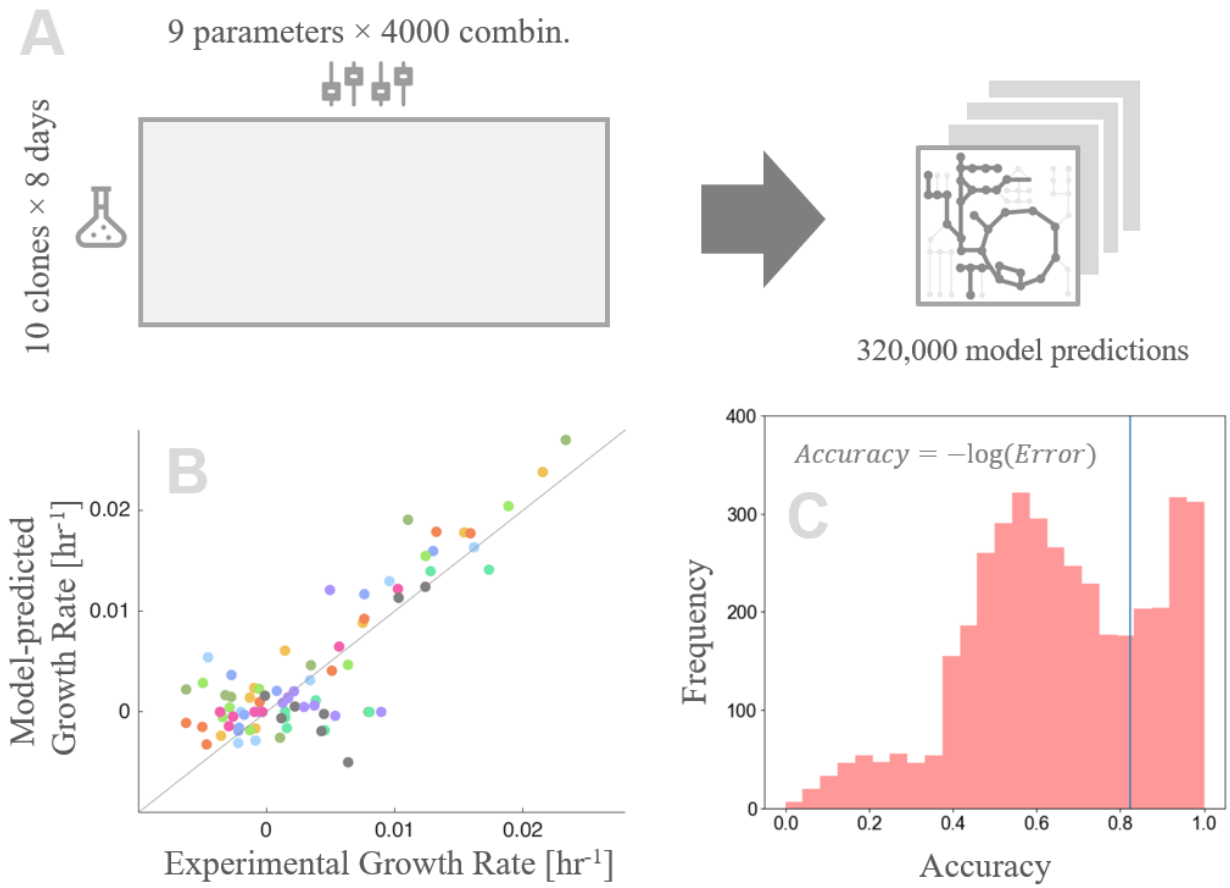


Figure 2: Parameter sensitivity analysis, workflow scheme & results – (a) 4000 parameter settings were evaluated against 80 datapoints to produce thousands of model predictions. (b) Here, we provide an example model prediction of growth rate for 80 points. These particular predictions were made by assuming parameter values equal to experimental measurements – that is, a baseline biomass dry weight of 280 picograms/cell at day 3, and 65% protein composition of biomass; biomass weight was also varied according to time-course cell diameter measurements. Biomass production was assumed to be limited by enzyme capacity. (c) Analysis results were examined by their 'accuracy', which was calculated and normalized from the negative-log of mean prediction residuals. Predictions assuming experimental values, featured in panel (b), resulted in an average accuracy of 0.82 (blue line).

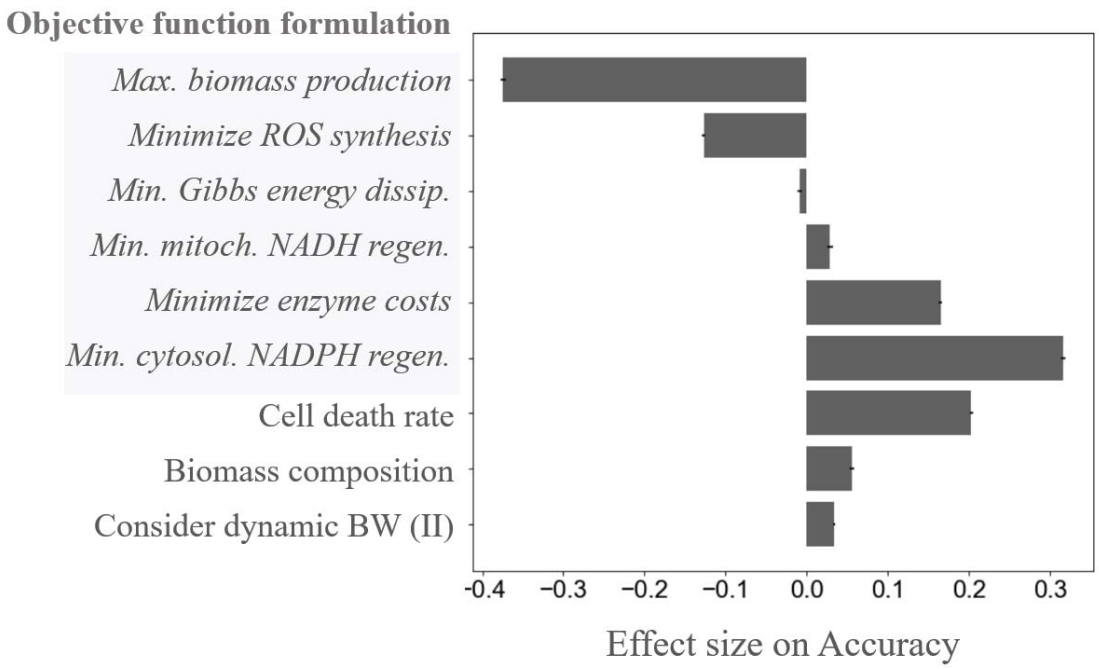


Figure 3: Sensitivity analysis reveals important parameters – A regression model estimated the importance of parameters by their effect size. *Objective function formulation* was treated as six independent binary parameters representing the various formulations (italicized and highlighted in grey). Several of these formulations were highly correlated to prediction accuracy. Other important parameters were *cell death rate*, *biomass composition* and *consideration of dynamic biomass weight* (BW)*for growth rate calculations*. The remaining parameters had negligible effects, which are detailed elsewhere (Fig. S6)

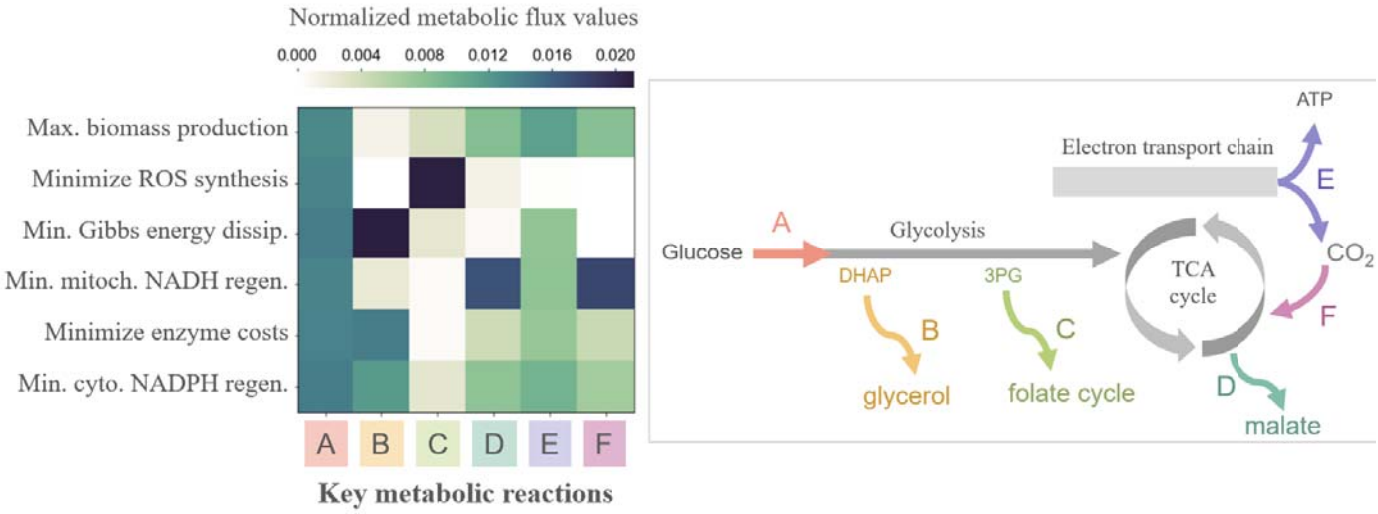


Figure 4: Objective functions vary in hypothesized metabolic activities underlying the predicted growth rates – Objective functions helped determine the metabolic activities of key reactions in the central carbon metabolism. The heatmap displays the reactions' fractional contribution to total metabolic flux. These reactions include the first step of glycolysis catalyzed by hexokinase 1 (*A*); carbon shunting variously towards glycerol (*B*), folate cycle intermediates (*C*), malate secretion (*D*); mitochondrial respiration (*E*); anaplerosis of respiratory carbon dioxide via bicarbonate and pyruvate carboxylase (*F*). The flowchart on the right contextualizes the reactions within the central carbon metabolism.

Table 1: Parameters varied in this study *– the table provides names and descriptions of the parameters, along with range of values explored in our analysis; mean experimental values of the presented clones are given in parentheses when applicable.*

#	Parameter name & description	Parameter values
1	**Biomass weight:** a baseline biomass dry weight value at day 4.	210 – 400 pg/cell (280)
2	**Biomass composition:** as a percentage of protein in cellular biomass.	45 – 75% (65)
3	**Consider dynamic biomass weight (I)** when inputting experimental data as boundary conditions.	TRUE / FALSE
4	**Consider dynamic biomass weight (II)** to refine growth rate calculations (Fig. S1).	TRUE / FALSE
5	**Cell death rate:** a constant cell death rate due to decreased viability (Fig. S3).	0 – 0.005 hr^{-1}
6	**Consider secretion costs:** to estimated energy and material costs for monoclonal antibody secretion.	TRUE / FALSE
7	**Consider amino acid catabolism byproduct synthesis** as recently described[63].	TRUE / FALSE
8	**Biomass turnover rate:** a constant rate of energy expenditure for biomass maintenance.	0 – 0.01 mmol g^{-1} hr^{-1}
9	**Objective function formulation:** the assumption of a 'cellular objective', which renders the metabolic network as a linear programming optimization problem.	• max. biomass production • min. enzyme costs • min. Gibbs energy dissipation • min. ROS synthesis • min. cytosolic NADPH regeneration • min. mitochondrial NADH regeneration