bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Improved constraints increase the predictivity and applicability of a linear programming-based dynamic metabolic modeling framework

Justin Y. Lee[a] and Mark P. Styczynski[a]

a School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Background Current metabolic modeling tools suffer from a variety of limitations, from scalability to simplifying assumptions, that preclude their use in many applications. We recently created a modeling framework, LK-DFBA, that addresses a key gap: capturing metabolite dynamics and regulation while retaining a potentially scalable linear programming structure. Key to this framework's success are the linear kinetics and regulatory constraints imposed on the system. Here, we present improvements to these constraints to improve the predictivity of LK-DFBA models and their applicability to biological systems.

Method Three new constraint approaches were created to better model interactions between metabolites and the reactions they regulate. These new approaches (and the original LK-DFBA approach) were tested on several synthetic and biological systems to determine their performance when using both noiseless and noisy data. To validate our framework, we compared experimental data to metabolite dynamics predicted by LK-DFBA.

Results There was no single optimal type of constraints across all synthetic and biological systems; rather, any one of the four approaches could perform best for a given system. The optimal approach for fitting to wildtype data of a given model was consistently the best approach

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when predicting new phenotypes for that model. Furthermore, many of LK-DFBA's predictions qualitatively agreed with experimental data.

Conclusions LK-DFBA can be improved by using several kinetics constraint approaches, with the ideal one selected based on wild-type training data. LK-DFBA's ability to predict metabolic trends in experimental data further supports its potential for modeling metabolite dynamics in systems of all sizes.

Keywords

Constraint-based modeling, metabolic systems, flux balance analysis, metabolomics, metabolite dynamics

Introduction

Mathematical and computational models are often used to study metabolism, the set of reactions that supply the chemical precursors necessary for almost all cellular processes. These metabolic models are significantly cheaper and faster to run than laboratory experiments, meaning that they can be of tremendous value when they are able to predict how changes in or to a metabolic system can affect its state. While a few pathways and sections of metabolism (e.g., glycolysis and central carbon metabolism) have been modeled and characterized quite well in a few organisms (e.g., *Saccharomyces cerevisiae* and *Escherichia coli*) [1, 2], genome-scale models that capture metabolism at a systems scale have been more difficult to develop. Metabolism involves many interconnected reactions and pathways, making it critical to include as much of metabolism as possible in metabolic models to better represent the system and generate accurate predictions. Metabolomics, which is the systems-scale measurement of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

metabolites in biological systems, thus has great potential to provide the information necessary to drive systems-scale metabolic models. However, creating genome-scale metabolic models that capture critical system behaviors like metabolic dynamics remains an outstanding challenge in the field, which has prevented the value of metabolomics data in this context from being fully realized.

The most popular type of frameworks for metabolic modeling are constraint-based models, including flux balance analysis (FBA) [3, 4], and ordinary differential equation (ODE) models. FBA assumes that the metabolic system is at steady state, which allows it to be modeled as a linear program (LP) that can be efficiently solved (even at the genome scale) but precludes modeling metabolite dynamics without substantial changes to the framework. ODE models are more widely used when dynamics are important, but are typically limited to smaller-scale modeling of well-studied pathways and the best-studied organisms due to uncertainty in the mathematical form and parameter values for the reaction kinetics terms. Only a few exceptions [5-8] have approached genome-scale ODE models, and they still require lengthy parameter estimation steps, prior information about kinetic constants, or have only been shown to be useful near the reference state of the system. As a result, steady-state fluxes continue to be the almost exclusive focus of study for genome-scale models. Modeling frameworks that can predict various metabolic phenotypes at the genome scale in a computationally tractable way have great potential for understanding, predicting, and controlling metabolism.

To address this problem, we recently developed Linear Kinetics-Dynamic Flux Balance Analysis (LK-DFBA), a modeling strategy to efficiently track metabolite dynamics [9]. LK-DFBA combines advantages of both constraint-based and ODE models, unrolling the temporal aspect of the system into a larger stoichiometric matrix that captures metabolite dynamics while

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

retaining a LP structure. The most critical element to accomplishing this goal is the addition of linear kinetics constraints that model the interactions between metabolites and the reactions whose fluxes they affect, including mass action kinetics and allosteric regulatory interactions. The number of parameters in LK-DFBA that need to be estimated can be far fewer than in ODE models due to these linear kinetics constraints. This enables LK-DFBA to potentially be applied to metabolic systems of all sizes, with a smaller increase in computational burden compared to ODE models. Furthermore, because LK-DFBA retains a linear structure, it can potentially be used with many existing metabolic modeling tools that require constraint-based models, such as OptKnock [10]. We have previously shown that LK-DFBA can outperform ODE-based modeling approaches when used in conditions most relevant to metabolomics data (low sampling frequency and high noise). A framework such as LK-DFBA that can model systems at the genome scale is essential to take full advantage of metabolomics data.

In our initial description of LK-DFBA, we explored two different approaches for model parameterization. The first approach, LK-DFBA (LR), parameterizes constraints solely via linear regression of interacting metabolite concentration and flux data. The second approach, LK-DFBA (LR+), uses the parameters from the linear regressions as initial seeding values for a secondary optimization to identify the optimal constraints for each interaction. While LK-DFBA (LR+) yields better fits to training data than LK-DFBA (LR), the latter approach estimates its parameters with trivial computational effort while still producing results that are similar in error to ODE models. As a result, LK-DFBA (LR) may be the preferable approach for the efficient construction and parameterization of metabolic models at the genome scale.

However, the overall LK-DFBA framework still has some limitations in terms of how accurately it represents the underlying biology and biochemistry of the system. For example, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

linear kinetics constraints used in LK-DFBA (LR) may be viewed as crude approximations of the interactions between metabolites and fluxes, which are typically non-linear in nature. While kinetic equations found in ODE models (such as Michaelis-Menten or biochemical system theory (BST) representations [11, 12]) can capture the non-linearity of these interactions, the current linear framework in LK-DFBA cannot. Additionally, when allosteric regulatory information is considered (which LK-DFBA includes in its framework), reaction fluxes are often controlled by multiple metabolites. Currently, LK-DFBA creates separate constraints for each metabolite that controls a flux, which precludes modeling how multiple metabolites simultaneously interact with a reaction flux.

Since the linear kinetics constraints are so critical in LK-DFBA's functioning, it is likely that improving those constraints could have a substantial impact on LK-DFBA's ability to capture and predict biological phenomena. Accordingly, we devised three new types of kinetics constraints for LK-DFBA to account for biologically relevant features like non-linearity and simultaneous regulation by multiple metabolites. These new approaches were compared to the original LK-DFBA (LR) constraints by testing on synthetic model systems as well as models based on *Lactococcus lactis* and *Escherichia coli* [1, 2]. We also probed these constraint approaches for their robustness to model perturbation and their ability to predict phenomena not captured in training data. We found that these new constraint approaches can improve model performance, and that the optimal constraint approach varied depending on the system being modeled but was consistent across perturbations for any given model. We also showed that the LK-DFBA approach chosen for the *L. lactis* and *E. coli* models can be used to predict changes in several critical metabolites and fluxes in agreement with literature experimental results. These improvements to LK-DFBA and demonstration of its effectiveness on new metabolic models

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

support its attractiveness as a framework for modeling increasingly large metabolic systems in the future.

Methods

LK-DFBA

Linear Kinetics-Dynamics Flux Balance Analysis (LK-DFBA) is a recently developed modeling strategy that is both scalable and capable of capturing metabolite dynamics. The full details of this approach have been described in detail previously [9], so we only outline the most important aspects of our framework here. In brief, LK-DFBA uses an LP-based structure with temporal dynamics modeled by discretizing time and unrolling the system into a larger matrix representing each time point separately, with an objective function that reflects the unrolling of the model. Linear inequality constraints that model mass action kinetics and metabolite-dependent regulation are included in the model; they are the driving force behind metabolite accumulation and depletion by limiting the maximum flux allowed based on the availability of metabolites over time. To parameterize these constraints, the LK-DFBA (LR) approach uses linear regression on assumed available metabolomics and fluxomics data, as described in the next section. If fluxomics data are unavailable, dynamic flux estimation (DFE) can be used to infer flux values from concentration data [13]. In the LK-DFBA (LR+) method, the parameters from the LK-DFBA (LR) approach are used as initial conditions in a secondary optimization step that finds improved kinetics constraint parameters, though at the cost of computational time. Because LK-DFBA retains an LP structure, it is readily scalable and has the potential to be used with current constraint-based modeling tools.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Constraint Approaches

LK-DFBA (LR) The original LK-DFBA approach uses linear kinetics constraints to model the interaction between a metabolite and a flux, parameterized using available metabolomics and fluxomics data. These constraints take the form of $v < ax + b$, where v is the flux being constrained, x is the concentration of a metabolite that interacts with the flux, and a and b are the linear constraint parameters. These interactions may be due to mass action kinetics, where the interactions are known based on the stoichiometric topology of the system, or they may stem from allosteric regulation. While we have previously shown that these linear approximations of metabolic interactions can be effective for modeling metabolism, they are still approximations of the true non-linear and interconnected biochemical relationships in metabolism. Below, we discuss three new constraint approaches to address these potential limitations.

LK-DFBA (NLR) While the key advantage of using constraint-based models is their LP structure that enables efficient identification of the optimal solution of the problem, most metabolite-flux interactions exhibit non-linear behavior that may not be captured well by linear equations. Recently, computational solvers have improved such that quadratically constrained programs (QCPs) are not much more computationally expensive than LPs. Accordingly, we implemented quadratic constraints into the LK-DFBA framework to explore their potential for improving model accuracy with only a modest increase in computational time. One important aspect of LPs and QCPs is that all of the constraints must create a convex feasible solution space in order to guarantee that a global optimum can be found [14]. If $v < ax^2 + bx + c$ represents a quadratic constraint, where v is the flux being constrained, x is the concentration of a metabolite

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

that interacts with v, and a, b, and c are the parameters of the quadratic constraint, a must be a negative value to retain a convex solution space. If a is found to be a positive value during parameterization, we convert the quadratic constraint into its original linear form as found in LK-DFBA (LR). We refer to this overall approach as LK-DFBA (NLR).

LK-DFBA (DR) Enzymatic reactions are often controlled by more than a single metabolite that can either induce or inhibit enzyme activity, which should ideally be captured in the model constraints. To model such regulation of a reaction by multiple metabolites, LK-DFBA (LR) creates individual linear constraints for each controller metabolite that are independent of each other and are thus unable to capture the synergistic or antagonistic effects of multiple metabolites working in conjunction to regulate a flux. We implemented a new strategy that uses dimensionality reduction to consolidate information content from all controller metabolites for a flux into a single constraint. Dimensionality reduction is often used in data analysis, including analysis of metabolomics data, to more easily represent and digest datasets with many measured variables. Principal component analysis (PCA) is one of the most commonly used dimensionality reduction approaches; it linearly transforms the original variables into new, orthogonal composite variables called principal components that capture as much variance in the original variable data in as few principal components as possible [15]. Ideally, the first or first few principal components capture the majority of the variance in the original dataset, which allows one to focus only on those composite variables rather than all of the original variables at once. Here, we use PCA to capture the maximal variance of the controller metabolite data in a single principal component and use that composite variable as the regressor during linear regression with the target flux data. These new constraints are represented as $v < aPC_1 +$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

b, where v is the flux being constrained, PC_1 is the metabolite concentration data projected into the first principal component, and a and b are the constraint parameters. We refer to this dimensionality reduction approach as LK-DFBA (DR).

LK-DFBA (HP) Another approach for modeling interactions with multiple metabolites is to use hyperplane constraints. Unlike LK-DFBA (DR), which always builds constraints in two dimensions (i.e. the target flux vs. the first principal component), the hyperplane constraint exists in (n + 1) dimensions, where n is the number of metabolites that control a target flux. This approach may avoid loss of information content from metabolite data as is possible during dimensionality reduction: as the number of metabolites in an interaction increases, the likelihood of the first principal component not capturing the majority of variance in the data increases. The hyperplane constraint equation can be represented as $v < \sum_{i=1}^{n} a_i x_i + b$, where v is the flux being constrained, n is the number of metabolites that interact with v, x_i is the concentration of metabolite i, a_i is the constraint parameter for metabolite x_i, and b is another constraint parameter. We refer to the hyperplane approach as LK-DFBA (HP).

Translating constraints to contain training data

We found that translating the constraints such that all training data fall in the region under the inequality constraint decreased the possibility of the computational solver encountering infeasible solutions when simulating metabolite dynamics. Thus, for all LK-DFBA approaches, each constraint was translated to contain the training data by increasing the intercept of the constraint (i.e. the b parameter in LK-DFBA (LR), LK-DFBA (DR), and LK-DFBA (HP), and the c parameter in LK-DFBA (NLR)) until no training data were above each constraint.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Test Models

Synthetic model The first system we examine is a simple synthetic model with five metabolites and five fluxes that was derived from a branched pathway model used previously [9]. This system is based on an ODE-based modeling framework that uses power-law kinetics to represent reaction fluxes [12]. The kinetic equations for each pathway are shown in Figure 1. To create a variety of synthetic models with the same stoichiometric topology, we randomly generated *a* and *b* parameters in each kinetic equation. The parameters for each model can be found in Table S1. Time course metabolite and flux data were generated by solving the ODE system in MATLAB (2018b).

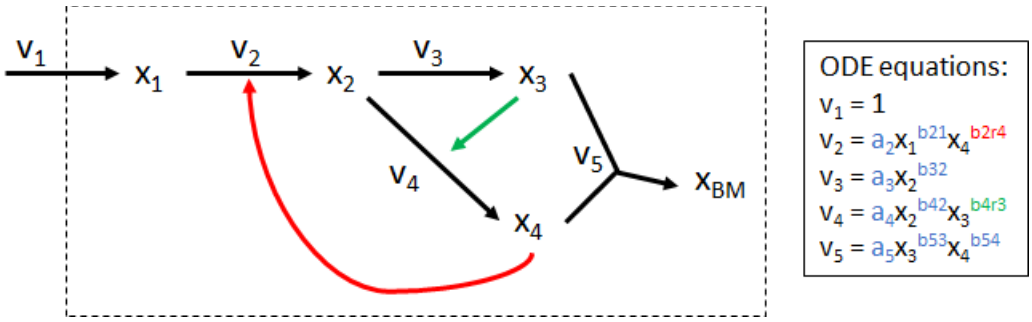

Figure 1: Synthetic model. Adapted from another branched pathway model used in previous work [9]. v_1, v_2, v_3, v_4, and v_5 are system fluxes (black arrows) and x_1, x_2, x_3, x_4, and x_{BM} are metabolites, where x_{BM} is a metabolite representing biomass. Green and red arrows represent positive and negative regulatory interactions, respectively. ODE equations for the model are shown in the inset, where blue *a* and *b* parameters are mass action kinetic parameters and green and red *b* parameters are positive and negative regulatory parameters, respectively.

***Lactococcus lactis* model** This model was created by Costa *et al*. and comprises central metabolism and production pathways for important metabolites such as mannitol and 2,3-butanediol [2]. The *L. lactis* model consists of 26 metabolites and 21 fluxes and is publicly

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

available on KiMoSys [16]. Noiseless data were generated in COPASI 4.24 (Build 197) using the default initial conditions and parameters over a simulation time of two hours.

***Escherichia coli* model** The *E. coli* model developed by Chassagnole *et al*. encompasses glycolysis and the pentose phosphate pathway [1]. This model is publicly available on KiMoSys, but was rebuilt within MATLAB to allow easy creation of new models that use the original *E. coli* model's topology and stoichiometry. Noiseless data for the original *E. coli* model were generated in MATLAB (2018b) using the default initial conditions and parameters, while random initial conditions and parameters were used for the new models with the *E. coli* topology. To be consistent with our previous work, we used a simulation time of ten seconds [9]. More information about the model parameters can be found in the Supplementary Methods.

LK-DFBA Objective Functions

Like other constraint-based methods, LK-DFBA requires an objective function, which is usually tied to some presumed goal of the system (such as maximizing biomass or ATP production) that stems from evolutionary pressure. FBA models for specific organisms commonly have a separate flux reaction dedicated to biomass, made up of precise ratios of different metabolites. While LK-DFBA models with tuned objective functions can be created, the biological models we sought to use here do not have pre-existing tuned objective functions, so we instead focused on LK-DFBA's performance using generic objective functions.

Here, we have chosen flux v_5 as the objective function for the synthetic model, as it is the only efflux out of the system. For the *L. lactis* model, we use the LDH pathway as the objective function to maximize production of lactate because it is a key metabolite in the organism (which

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is commonly used for dairy products) and was the metabolite produced at the highest levels in the original *L. lactis* model [2]. The objective function used for the *E. coli* model was to maximize all effluxes from the system, which included murein synthesis, glycerol-3-phosphate dehydrogenase, serine synthesis, PEP carboxylase, DAHP synthesis, pyruvate dehydrogenase, ribose phosphate pyrophosphokinase, glucose-1-phosphate adenyltransferase, the synthesis of murein and chorismate from PEP, and the synthesis of isoleucine, alanine, α-ketoisovalerate, and diaminopimelate from pyruvate. While we have observed that these objective functions can be further improved, and approaches have been developed for finding an optimal objective function for a model by creating a bilevel optimization problem and then leveraging the duality theorem [17, 18], our chosen objective functions were sufficient to at least qualitatively model the synthetic, *L. lactis*, and *E. coli* systems.

Pathway Perturbations

To test the ability of LK-DFBA to predict metabolic behaviors not represented in the training data, we introduced perturbations into each system either through down-regulation (indicated with a prefix 'd' in all figures) or up-regulation (indicated with a prefix 'u') of reaction fluxes. For the synthetic models, we down-regulated v_2, v_3, and v_4 by multiplying their constraint equation parameters (which restricts their maximum allowable flux value) by 0.5x and up-regulated these pathways by doubling the constraint equation parameters. The pathways and reactions to be perturbed in the *L. lactis* [2, 19-22] and *E. coli* [23-27] models were chosen based on previous literature. Reactions in the *L. lactis* model (lactate dehydrogenase, phosphofructokinase, acetate kinase, mannitol 1-phosphatase) were down-regulated to 0.1x their original parameter values (since completely knocking out reactions would often produce

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

infeasible solutions for the linear program) and up-regulated to 2x their original parameter values, magnitudes that were necessary to effect significant perturbations to the system's behavior. Reactions in the *E. coli* model (pyruvate kinase, phosphoglucose isomerase, glyceraldehyde-3-phosphate dehydrogenase, phosphofructokinase, triose-phosphate isomerase, ribulose-phosphate epimerase, phosphoglucomutase) were down-regulated to 0.1x and up-regulated to 2x their original parameter values.

Generating Noisy Data

Noise was introduced into the system by down-sampling the original noiseless data (originally 50 time points) into nT time points that are evenly spaced over the time interval of interest. Both metabolite and flux values were then replaced with a random value drawn from $N_{i,k} \sim (y_i(t_k), CoV \cdot y_i(t_k))$, where $y_i(t_k)$ is the value of species (metabolite or flux) i at time point k, and CoV is a coefficient of variance. For each sampling frequency and CoV condition, ten noisy datasets were generated.

Error Calculation

The error of the predictions made by LK-DFBA was calculated using a normalized root mean squared error (NRMSE) between the LK-DFBA predicted metabolite concentrations and the noiseless ODE concentration or experimental data. P_{ik} and R_{ik} are the predicted (e.g. results from LK-DFBA) and reference (e.g. ODE model or experimental) data from a system with m metabolites and nT time points. $\bar{R}_i$ is the mean of the concentrations of reference metabolite i across all time points to normalize the data and N is the total number of data points used in the NRMSE calculation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\boldsymbol{NRMSE} = \sqrt{\frac{\sum_{i=1}^{m}\sum_{k=1}^{nT}\left(\frac{P_{ik}-R_{ik}}{\bar{R}_i}\right)^2}{N}}$$

Pearson Correlation Calculation

The available *E. coli* knockout experimental data consisted of steady-state flux data, so to compare these to the knockout predictions made by LK-DFBA (which did not yield a steady state over the ten second time interval of the model) we used the average flux of our time course predictions. Because the average flux of our predictions and the steady-state fluxes of the experimental data are different measurements and therefore not directly comparable using NRMSE, we chose to use a Pearson correlation coefficient to evaluate our framework, which was recently used in a similar comparative analysis of metabolic models [28]. High correlations between steady-state flux experimental data and the average flux predictions would indicate that LK-DFBA can effectively predict if gene knockouts lead to an increase or decrease in flux for modeled reactions. The calculation for the Pearson correlation coefficient is shown below, where A_i is the average of the predicted flux profile for the *i*th flux, v_i is the flux value of the *i*th flux from the experimental data, $\bar{A}$ is the mean across all fluxes for the average of computationally predicted fluxes, $\bar{v}$ is the mean flux value across all fluxes for the experimental data, and *n* is double the number of fluxes that are shared between both the *E. coli* model and experimental data because it includes flux values before and after the gene knockout (n = 28).

$$\boldsymbol{Pearson\ Correlation\ Coefficient} = \frac{\sum_{i=1}^{n}(A_i-\bar{A})(v_i-\bar{v})}{\sqrt{\sum_{i=1}^{n}(A_i-\bar{A})^2\sum_{i=1}^{n}(v_i-\bar{v})^2}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Fitting and predicting phenotypes in synthetic models

We generated twenty random sets of parameters and initial conditions for the kinetic equations in the synthetic model to examine if different constraint approaches were more suitable for different models. We produced *in silico* metabolite concentration and flux data over a time interval of ten seconds by solving the ODEs in each synthetic system. The four constraint approaches were used for parameterization of LK-DFBA models to the twenty datasets. The fitted LK-DFBA models were then simulated over the same ten second interval using the initial conditions for each respective synthetic system to compare against the original ODE data. This process was performed on both noiseless (nT = 50, CoV = 0) and noise-added data with different sampling frequencies (nT = 50 or 15) and levels of noise (CoV = 0.05 or 0.15). To test the ability of each LK-DFBA approach to predict the effects of defined genetic perturbations, we down- and up-regulated the v_2, v_3, and v_4 pathways in the original kinetic equations by multiplying the kinetic coefficient parameters (*a* parameters in the inset of Figure 1) by 0.5x or 2x, respectively, and generating new ODE data. We then simulated the LK-DFBA model after adjusting the fitted LK-DFBA constraints to reflect the down- or up-regulation by multiplying the kinetics constraint parameters by 0.5x and 2x, respectively. The NRMSE between the predicted LK-DFBA metabolite concentrations and the ODE concentration data from the perturbed synthetic models was then calculated.

For the noiseless cases with no genetic perturbation (WT) as shown in the first row below each bold line in Figure 2A, the best-fitting constraint approach (dark green) varied across the different models. All four approaches performed best for at least one of the models. When fluxes were either down- or up-regulated via *in silico* genetic perturbations and LK-DFBA models

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fitted to the WT ODE data were used to predict these changes, the best constraint approach across all perturbations (dV2 through uV4) was generally consistent with the best approach in the absence of perturbations. The NRMSE heatmap with quantitative error values can be found in Figure S1.

When using noisy data, similar trends were observed (representative example in Figure 2B). While the best constraint approach for WT noisy data was not always the same as the best approach for noiseless data, the best constraint approach for a given noisy WT dataset was still generally the best for predicting the impacts of *in silico* genetic perturbations in the same model (dV2 through uV4). Interestingly, noisy data negatively affected the performance of LK-DFBA (HP) to a much greater extent than the other approaches, which caused LK-DFBA (HP) to never be identified as the best approach under the most realistic conditions (nT = 15, CoV = 0.15). NRMSE heatmaps with quantitative error values for various sampling frequencies and noise levels can be found in Figure S2-S5.

We also tested the effect of smoothing the noisy (nT = 15, CoV = 0.15) metabolite concentration and flux time course profiles by fitting to a previously described [29] impulse function (Figure S6). Smoothing the noisy data can often lead to lower error of the final model, but requires increased computation time for estimating the parameters of the impulse function and in certain cases can actually increase error if a specific dataset deviates significantly from all of the profiles that an impulse function can capture. The best constraint approach for WT smoothed data was the same as for unsmoothed data in 19 of the 20 models. As with the unsmoothed cases, the best constraint approach for smoothed data was typically consistent between WT and *in silico* genetic perturbations, and there were no cases where LK-DFBA (HP) performed the best (and it was generally the worst out of the four approaches) for smoothed data.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

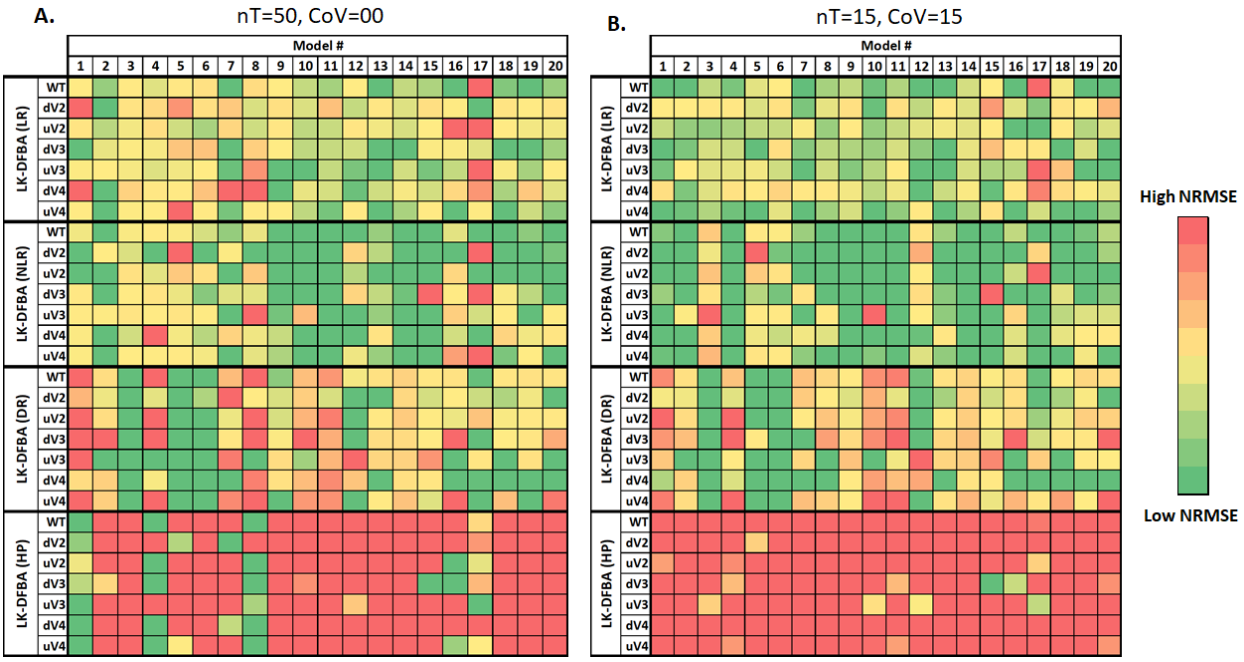

Figure 2: NRMSE heatmap of LK-DFBA approaches on different synthetic models. Each constraint approach was used to fit parameters to wild-type (WT) data and then used to simulate the WT system and *in silico* genetic perturbations with fluxes v_2, v_3, or v_4 down- or up-regulated. Dark green boxes represent the lowest NRMSE within each perturbation for each synthetic model, while dark red boxes represent the highest NRMSE (meaning that the dynamic range of the color scale varies for each perturbation for each synthetic model to better convey the relative performance of different methods). Panel A shows results for noiseless data with a sampling frequency of 50 time points. Panel B shows results for noisy datasets with a sampling frequency of 15 time points and with noise added at a CoV of 0.15. The average NRMSE of 10 noisy datasets is shown in the heatmap. The same NRMSE heatmaps with explicitly annotated error values can be found in Figure S1 and Figure S5.

Fitting and predicting phenotypes in *L. lactis* and *E. coli* models

For the *L. lactis* model, we tested the four constraint approaches on noiseless data and noisy data under different conditions (nT = 50 or 15, CoV = 0.05 or 0.15). On the noiseless data, the best constraint approach for the WT system was LK-DFBA (HP), which also had the lowest NRMSE when predicting the results of perturbations to five different pathways (Figure 3A). At

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

high sampling frequencies and low noise (nT = 50, CoV = 0.05), LK-DFBA (HP) still performed the best, but as more noise was added or lower sampling frequencies were used, LK-DFBA (NLR) was optimal. This is consistent with the findings described above for the small synthetic systems where LK-DFBA (HP) can produce low NRMSE with noiseless data, but has difficulties under more realistic conditions.

As with the *L. lactis* model, we tested all constraint approaches on both noiseless and noisy data from the *E. coli* model under different conditions (nT = 50 or 15, CoV = 0.05 or 0.15). For this model, LK-DFBA (NLR) was the best constraint approach for noiseless data (Figure 3B). Noisy *E. coli* data produced the same results: for all noisy conditions, LK-DFBA (NLR) was optimal for the WT system. It was also optimal for almost all of the *in silico* genetic perturbations, showing once again that the same constraint approach that was optimal for the WT system at a given sampling condition was generally also optimal for the perturbed systems.

We also perturbed the original parameters and initial conditions (drawing from the random normal distribution $N_i \sim (p_i, p_i)$, where p_i is the original value of the ith parameter) of the *E. coli* model to create five new models with the same topology (more information about parameter randomization can be found in the Supplementary Methods). As with the twenty different versions of the small synthetic system, we found that the best constraint approach was not conserved across models with the same topology as the original *E. coli* model when tested on noiseless data (Figure S8). Instead, the rates of individual reactions and how they affect overall model dynamics appear to be important factors in determining the optimal constraint approach.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

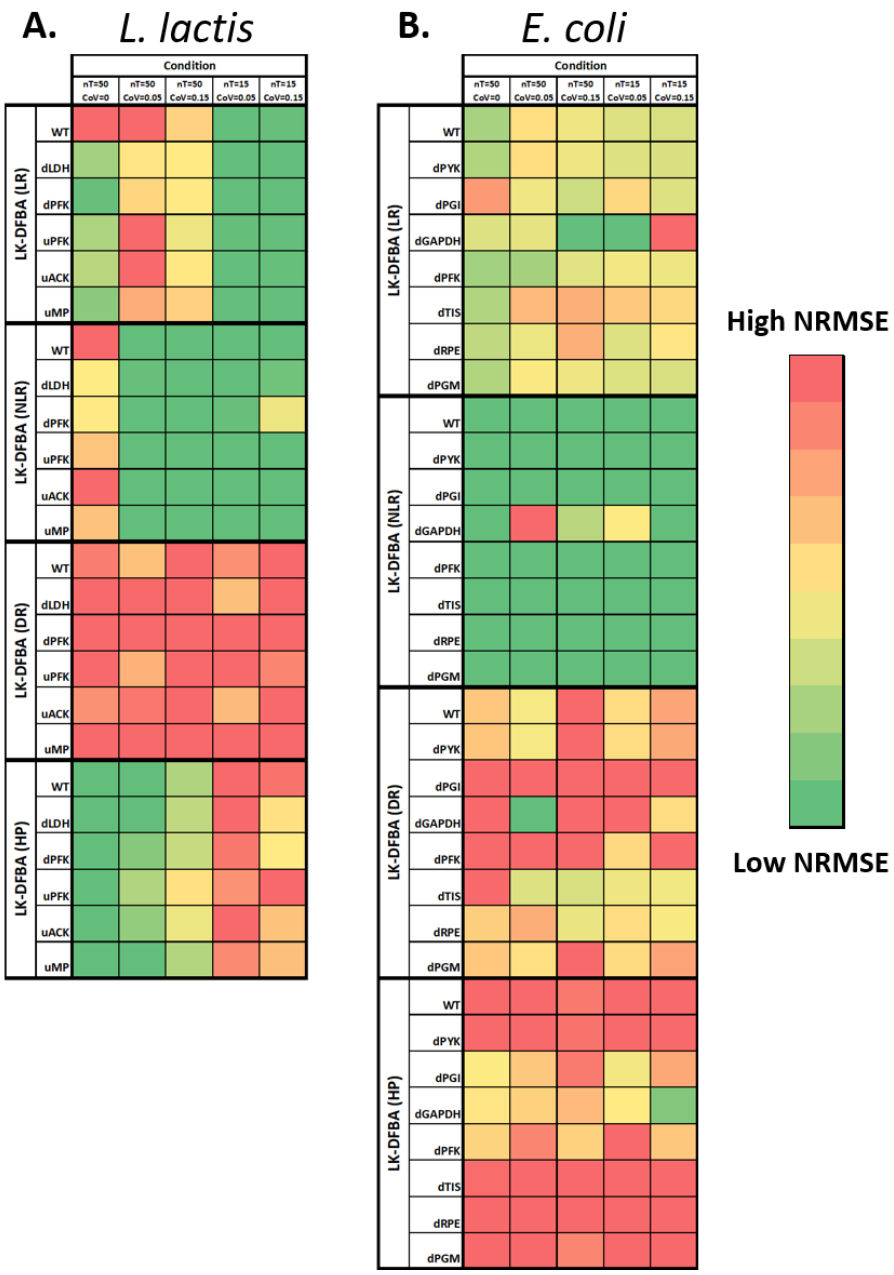

Figure 3: NRMSE heatmaps of constraint approaches on *L. lactis* and *E. coli* models. Each constraint approach was used to fit parameters to wild-type (WT) data and then used to simulate the WT system and the system with *in silico* genetic perturbations with literature-reported important pathways down- or up-regulated. Dark green boxes represent the lowest NRMSE within each phenotype for each model, while dark red boxes represent the highest NRMSE. Both the *L. lactis* (A) and *E. coli* (B) heatmaps show the mean of 10 noisy datasets, except for the noiseless condition (leftmost columns in each heatmap). The same NRMSE heatmaps with explicitly annotated error values can be found in Figure S7.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Improved LK-DFBA predictions yield qualitative consistency with experimental *L. lactis* metabolite concentration data

To further assess how well LK-DFBA performs when predicting different phenotypes, we compared the predictions of LK-DFBA to available experimental data for the first time. The previously described ODE-based *L. lactis* model was originally parameterized using experimental metabolite time course data from *L. lactis* cultures grown with an initial glucose concentration of 40 mM [2] and validated by comparison to some experimental data from cultures grown at initial concentrations of 20mM and 80 mM glucose. Here, we similarly fitted all LK-DFBA approaches to data generated by the ODE model at 40 mM glucose and then simulated the LK-DFBA model using the best constraint approach at 20 mM and 80 mM initial concentrations of glucose for validation.

Figures 4A, 4B, and 4C depict the metabolite concentrations predicted by LK-DFBA (HP) (the best approach for noiseless data in the *L. lactis* model) when trained on noiseless data. For multiple initial glucose concentrations, LK-DFBA (HP) captured the general qualitative trends of glucose (depletion) and lactate (accumulation), two key metabolites in *L. lactis* that are often studied [30, 31]. For cofactor metabolites that participate in many different reactions, such as ATP, NAD(H), and inorganic phosphate (Figure S9), it was more challenging for LK-DFBA (HP) to predict their concentration profiles over the simulation interval, which is a problem found in other modeling frameworks [5]. Although LK-DFBA's predictions were overall not as smooth or quantitatively accurate as the ODE model, this is to be expected due to the lack of a validated objective function for this constraint-based model; the objective function we used was a gross approximation that likely does not reflect the cell's true "goal", and it is known that the objective function can significantly affect the predictions of FBA approaches. Nevertheless, as

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

presented here, LK-DFBA can still qualitatively track important metabolite dynamics even when using a crude objective function. This is important to note, as many organisms that are not well-studied have no readily available objective function to use.

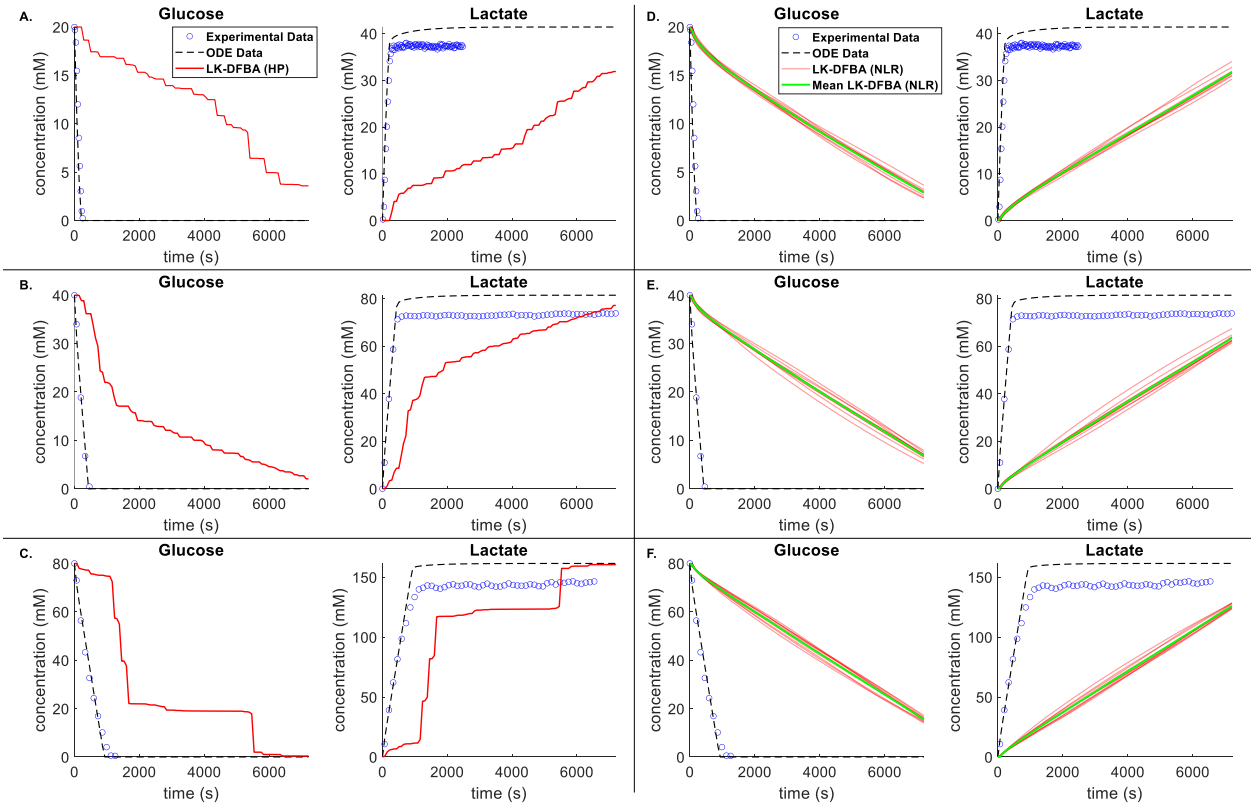

Figure 4: Comparison of LK-DFBA metabolite concentration predictions against ODE data and *L. lactis* experimental data when fitted to noiseless and noisy ODE data. Panels A, B, and C depict concentration profiles for LK-DFBA (HP) (solid red line) and the ODE model (dashed black line) compared to experimental data (blue circles) for initial glucose concentrations of 20 mM, 40 mM, and 80 mM, respectively, when LK-DFBA is fitted to noiseless data. Panels D, E, and F depict concentration profiles for LK-DFBA (NLR) on 10 noisy datasets (nT = 15, CoV = 0.15) and the ODE model compared to experimental data. The mean concentration profile (solid green line) is shown with each of the concentration profiles (solid red lines) from the 10 noisy datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures 4D, 4E, and 4F depict the glucose and lactate concentration profiles predicted by LK-DFBA (NLR) (the best approach for noisy data in the *L. lactis* model) after being fitted to 10 noisy datasets generated by the ODE model and simulated at 20 mM, 40 mM, and 80 mM initial glucose, respectively. Again, the LK-DFBA framework generally captured the qualitative trends of major metabolites such as glucose and lactate, though unsurprisingly not as accurately as when noiseless data are used and with difficulties predicting cofactor concentrations (Figure S10). Because LK-DFBA (NLR) contains quadratic constraints, its results are generally smoother compared to the other LK-DFBA approaches, which helped it predict some metabolites, such as PEP, arguably better than in the noiseless case. Furthermore, LK-DFBA (NLR) is less susceptible to noise for some metabolites, such as glucose and lactate, as observed in predicting similar time courses across the 10 noisy datasets. This could be advantageous if one is modeling a system with multiple noisy data sets and requires consistent predictions for certain metabolites. Likewise, if only using a single dataset, LK-DFBA (NLR) can ensure that these metabolic profiles would not dramatically change if a different dataset had been used. Other methods, such as the original LK-DFBA (LR) approach, can result in more varied predictions (Figure S11) depending on the noisy dataset used; some appear to produce better predictions than LK-DFBA (NLR), while others are worse (though all predictions follow the same trends). These observations reiterate that the best approach is dependent on the systems and datasets being studied, so having multiple LK-DFBA approaches available is an improvement over only using the LK-DFBA (LR) framework.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Changes in LK-DFBA flux profiles due to gene knockouts are correlated with experimental *E. coli* steady-state flux data

We also compared the predictions of the best LK-DFBA approach on the *E. coli* model to experimental steady-state flux data obtained through gene knockout experiments by Ishii *et al*. [24]. Because the Chassagnole model, which LK-DFBA is fitted to, only encompasses central carbon metabolism, we focused on 13 gene knockouts and 14 fluxes that are included in both the Chassagnole model and the Ishii steady-state flux results. We used the dilution rate of 0.2 h^{-1} for all experimental data. To emulate a gene knockout in the LK-DFBA model, we down-regulated the pathway(s) that correspond with the gene by multiplying the parameters of the relevant constraints by 0.1x instead of completely removing the reaction, as we found that this sufficiently reduced the possible flux reaction rate without causing infeasible solutions from the solver. Additionally, it is not uncommon for enzymatic activity to remain in a pathway after single gene knockouts due to paralogous enzymes and enzyme promiscuity. Because the LK-DFBA predictions do not reach steady-state for the simulation time examined in this work and our previous work (ten seconds), we instead used the average flux of the predicted time course to describe how LK-DFBA's predictions change from the wildtype to gene knockout phenotype. The average flux before and after a gene knockout should reflect whether the reaction rate generally increases or decreases across the studied time interval after a system perturbation. We used a Pearson correlation to determine if the average flux profiles predicted by LK-DFBA changed similarly to the experimental data after a gene knockout. This assessment method has been used previously by Lima et al. to compare multiple *E. coli* models, including the Chassagnole model, to the Ishii dataset [28].

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To evaluate how our framework compares to *E. coli* experimental data, we examined LK-DFBA (NLR), as it was the best approach in the case of low sampling frequency and high noise (Figure 3B). Figure 5 shows the average Pearson correlation of the LK-DFBA (NLR) flux predictions (after being fitted to ten noisy datasets with nT = 15 and CoV = 0.15) and the average correlation of the ODE model flux predictions with the experimental steady-state flux data. Of the gene knockouts and fluxes examined, LK-DFBA (NLR) generally gave reliable predictions for whether fluxes increased or decreased due to gene knockouts, with correlation values greater than 0.6 in all but two cases and correlations greater than 0.7 in 6 out of 13 cases. These correlations were very similar to the correlations yielded by the ODE-based model. In 10 out of 13 knockouts, the correlations calculated for LK-DFBA outperformed or were within 5% of the correlations calculated with the ODE-based model. These results support the significant promise of LK-DFBA approaches for predictivity comparable to that of standard models but with the additional benefits (including relative model simplicity and potential scalability) that accrue from using a LP-based formulation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

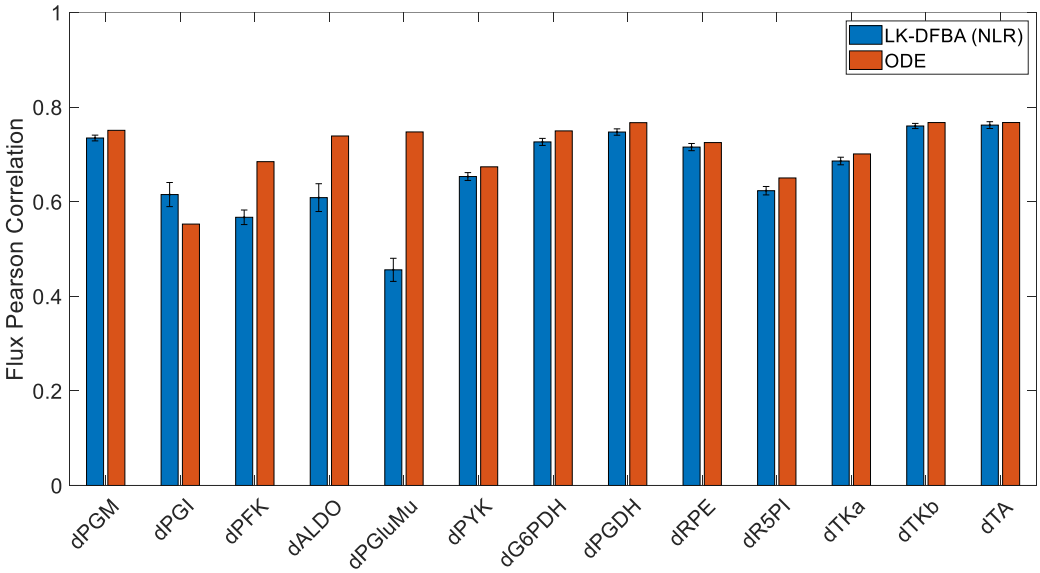

Figure 5: Pearson correlation coefficients of LK-DFBA and ODE model flux predictions with *E. coli* experimental data. LK-DFBA (NLR) was the best approach when fitting on low sampling frequency (nT = 15) and high noise (CoV = 0.15) data. Blue and red bars represent LK-DFBA (NLR) and ODE model mean correlations, respectively, between the average predicted flux profiles and experimental steady-state flux data for various gene knockout conditions. Gene knockouts in the LK-DFBA and ODE-based models were simulated by down-regulating relevant pathways. Error bars for LK-DFBA represent one standard deviation (N = 10 runs).

Discussion

At the outset of this work, we sought to find a single LK-DFBA constraint approach that would improve upon the originally published framework. Instead, we have shown that the best constraint approach is highly dependent on the system being modeled. Despite each of the 20 small synthetic models having the exact same stoichiometry and allosteric regulatory interactions, the optimal LK-DFBA approach varied for both noiseless and noisy training datasets, with one of the new constraint approaches performing the best in the majority of cases. This finding suggests that the topology of the system is less important than the emergent dynamics from the collective metabolic reactions. It also supports the importance of having

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

multiple types of constraints to choose from, as presented in this work, to allow more accurate modeling of any given system.

These conclusions are reinforced by analysis of biological systems, where LK-DFBA (HP) performs the best on *L. lactis* noiseless data and LK-DFBA (NLR) performs the best on *E. coli* noiseless data (though LK-DFBA (NLR) is superior for both systems when using low sampling frequency and high noise data). We further confirmed that topology is not the determining factor by randomizing parameters in the *E. coli* model (Figure S8): again, the best constraint approach varied across these topologically identical new models. Many metabolic pathways are conserved topologically across many species (e.g. glycolysis), though the kinetic parameters within these pathways can be vastly different. This suggests that having multiple LK-DFBA constraint approaches to choose from will improve our ability to model different systems.

While the best constraint approach varied across different model parameterizations and topologies, the best approach (in terms of predicting metabolic phenotypes) for a given model was generally consistent across a wide range of pathway perturbations. This trend remained true whether using noiseless data, data with low sampling frequency and high noise, or noisy data that had been smoothed. These results instill confidence that the best constraint approach found when fitting to a wildtype metabolic system will also be the best approach when predicting changes to that system, meaning that an approach that can select the best-fitting of multiple constraint frameworks is viable and likely to be successful. One possible reason for the success of this approach is that when pathways are down- or up-regulated, it is common for only the nearest neighboring pathways to be significantly affected if the change to the system is not drastic or the perturbed pathway is not essential for cell survival, meaning that the emergent behavior from the system would not change too greatly and thus the same constraint approach

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

would be optimal. To easily construct the optimal LK-DFBA model for a given biological system, we envision the workflow presented in Figure 6. After compiling the relevant system stoichiometry, regulatory information, and metabolomics and fluxomics data, one can fit each of the four LK-DFBA approaches to the data and determine which constraint approach most likely works best for predicting the results of different perturbations.

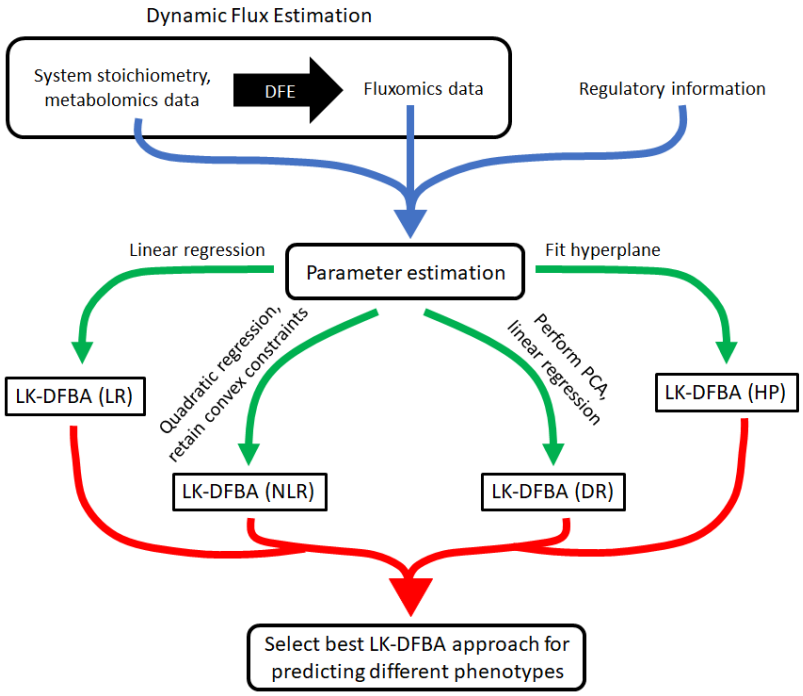

Figure 6: Workflow for selecting the best constraint approach for LK-DFBA when modeling metabolic systems. Dynamic Flux Estimation (DFE) is applied to the system stoichiometry and available metabolomics data to infer instantaneous fluxes. The system stoichiometry, metabolomics data, inferred flux data, and system regulatory information are then used to estimate parameters for each LK-DFBA approach (blue arrow). Using multiple constraint approaches (green arrows), four different LK-DFBA models are created and tested for their respective abilities to recapitulate training data. The model with the lowest error is selected and can be used for future *in silico* predictions (red arrow).

Using ODE models and experimental data from *L. lactis* and *E. coli*, we found that LK-DFBA can effectively predict qualitative trends in concentration profiles of some important

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

metabolites. While we have previously shown that LK-DFBA captures metabolite dynamics in synthetic data generated by ODE models, this is the first time LK-DFBA predictions have been validated with experimental data. For key metabolites that are important inputs of outputs of the system (e.g. carbon sources or end products), LK-DFBA can qualitatively predict if their concentration profiles are expected to decrease or increase, which is an important capability if one is using LK-DFBA to engineer organisms to efficiently produce certain metabolites. Cofactors, on the other hand, are more difficult to model using LK-DFBA but are still typically predicted to be within an order of magnitude of the experimental data in most cases. This capability could be useful when assessing levels of accumulating toxic metabolites or cofactor imbalances if exact concentrations are not necessary. We also found that LK-DFBA flux profile predictions were highly correlated with experimental flux data from genetic knockout experiments. Furthermore, these correlations were comparable to those found when using the ODE-based model. We note, though, that this comparable predictivity is limited by the fact that LK-DFBA was trained using ODE-generated data; if it had instead been fitted to actual metabolomics and fluxomics time course data used in the Ishii experiments (which is not available), these correlation values could possibly be even higher. Similarly, an improved objective function over the reasonable but arbitrary and unoptimized one used here could also lead to significant improvements in the performance of LK-DFBA.

By showing for the first time that LK-DFBA can predict changes in metabolite concentrations and flux profiles qualitatively, we have demonstrated LK-DFBA's potential as a widely-applicable metabolic modeling tool. Unlike many ODE-based modeling approaches that require specific kinetic equations for each flux reaction, LK-DFBA is generalizable. With four types of kinetics constraints that account for different biological interaction phenomena between

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

metabolites and fluxes, we have improved LK-DFBA to be amenable to many different systems. Additionally, applying the four LK-DFBA approaches to these models of *L. lactis* and *E. coli* has established that our framework can handle various biological systems of substantial size without the need for computationally taxing parameter estimation steps. Because each of the four LK-DFBA approaches maintains an easily solvable LP or QCP structure, LK-DFBA is a prime candidate for being one of the first frameworks able to model a variety of genome-scale systems while also capturing their metabolite dynamics.

While the addition of new constraint approaches has significantly improved the original LK-DFBA (LR) framework, there are still several areas where LK-DFBA can be improved. If computational resources when building the model are not a concern, a secondary optimization step can be used, as in the LK-DFBA (LR+) approach, to improve the parameters in each of the new constraint approaches. In addition, as previously noted the objective function used in LK-DFBA is also a ripe target for future efforts to improve this modeling framework. Here we have chosen objective functions that lead to the maximization of putatively important fluxes, but unlike many other constraint-based models, there was no specific biomass or other objective flux to use. Optimizing the weight of each flux or metabolite in the objective function could lead to even lower observed errors compared to experimental data and may also provide insight into what real biological systems tend to maximize.

Conclusion

In this work, we have shown that the LK-DFBA modeling framework can be improved by implementing more complex constraints with increased biological relevance. We showed that there is no single best LK-DFBA constraint approach for all models, and the optimal approach

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

depends not just on the topology of the biochemical system but also its kinetics and parameters. The constraint approach that performs the best in recapitulating training data consistently outperforms other constraint approaches at predicting the results of metabolic perturbations on the same system. With these new constraint approaches, we are able to model a variety of metabolic systems more accurately than if we were to just use the original LK-DFBA (LR) method. Moreover, based on comparisons to experimental data we showed that the improved LK-DFBA approaches can reasonably capture the qualitative dynamics of important metabolites and fluxes of interest to researchers. While these predictions may not be smooth or quantitative, the qualitative prediction of trends of metabolite dynamics in response to major perturbations is arguably the most critical aspect needed for creating metabolic models that give insight on how pathways can be further optimized or how metabolic resources can be rerouted to produce valuable chemicals: knowing that a specific knockout will increase or decrease flux is often sufficient to justify the expense of experimental implementation of such genetic perturbations. Moreover, we expect this computational framework to (with future effort) provide opportunities for computationally reasonable scale-up to the genome scale. While the acquisition of quality metabolomics and fluxomics data to build the constraints in LK-DFBA is still a challenge, the work we have presented here lays the groundwork needed to take full advantage of these types of datasets as they become increasingly more readily available.

Data availability

The code and datasets generated during the current study are available at https://github.com/gtstylab/lk-dfba-constraints.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Declaration of competing interest

The authors declare there is no conflict of interest

Funding

Funding for JYL and MPS was provided by the National Institutes of Health (www.nih.gov, R35-GM119701) and National Science Foundation (www.nsf.gov, 1254382). The funders had no role in the design of the study or collection, analysis, or interpretation of data or writing of the manuscript.

References

1. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. Dynamic modeling of the central carbon metabolism of Escherichia coli. Biotechnol Bioeng. 2002;79(1):53-73. Epub 2007/06/27. doi: 10.1002/bit.10288. PubMed PMID: 17590932.

2. Costa RS, Hartmann A, Gaspar P, Neves AR, Vinga S. An extended dynamic model of Lactococcus lactis metabolism for mannitol and 2,3-butanediol production. Mol Biosyst. 2014;10(3):628-39. Epub 2014/01/15. doi: 10.1039/c3mb70265k. PubMed PMID: 24413179.

3. Gottstein W, Olivier BG, Bruggeman FJ, Teusink B. Constraint-based stoichiometric modelling from single organisms to microbial communities. J R Soc Interface. 2016;13(124). Epub 2016/01/01. doi: 10.1098/rsif.2016.0627. PubMed PMID: 28334697; PubMed Central PMCID: PMCPMC5134014.

4. Lularevic M, Racher AJ, Jaques C, Kiparissides A. Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions. Biotechnology and Bioengineering. 2019;116(9):2339-52. doi: 10.1002/bit.27025. PubMed PMID: WOS:000478674700020.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5. Khodayari A, Maranas CD. A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nat Commun. 2016;7:13806. Epub 2016/12/21. doi: 10.1038/ncomms13806. PubMed PMID: 27996047; PubMed Central PMCID: PMCPMC5187423.

6. Smallbone K, Simeonidis E, Swainston N, Mendes P. Towards a genome-scale kinetic model of cellular metabolism. BMC Syst Biol. 2010;4:6. Epub 2010/01/30. doi: 10.1186/1752-0509-4-6. PubMed PMID: 20109182; PubMed Central PMCID: PMCPMC2829494.

7. Srinivasan S, Cluett WR, Mahadevan R. Constructing kinetic models of metabolism at genome-scales: A review. Biotechnol J. 2015;10(9):1345-59. doi: 10.1002/biot.201400522. PubMed PMID: WOS:000362724300002.

8. Stanford NJ, Lubitz T, Smallbone K, Klipp E, Mendes P, Liebermeister W. Systematic construction of kinetic models from genome-scale metabolic networks. PLoS One. 2013;8(11):e79195. Epub 2013/12/11. doi: 10.1371/journal.pone.0079195. PubMed PMID: 24324546; PubMed Central PMCID: PMCPMC3852239.

9. Dromms RA, Lee JY, Styczynski MP. LK-DFBA: a linear programming-based modeling strategy for capturing dynamics and metabolite-dependent regulation in metabolism. BMC Bioinformatics. 2020;21(1):93. Epub 2020/03/04. doi: 10.1186/s12859-020-3422-0. PubMed PMID: 32122331; PubMed Central PMCID: PMCPMC7053146.

10. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng. 2003;84(6):647-57. Epub 2003/11/05. doi: 10.1002/bit.10803. PubMed PMID: 14595777.

11. Chen WW, Niepel M, Sorger PK. Classic and contemporary approaches to modeling biochemical reactions. Genes Dev. 2010;24(17):1861-75. Epub 2010/09/03. doi: 10.1101/gad.1945410. PubMed PMID: 20810646; PubMed Central PMCID: PMCPMC2932968.

12. Voit EO. Biochemical Systems Theory: A Review. ISRN Biomathematics. 2013;2013.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

13. Goel G, Chou IC, Voit EO. System estimation from metabolic time-series data. Bioinformatics. 2008;24(21):2505-11. Epub 2008/09/06. doi: 10.1093/bioinformatics/btn470. PubMed PMID: 18772153; PubMed Central PMCID: PMCPMC2732280.

14. Maranas CD, Zomorrodi AR. Optimization Methods in Metabolic Networks. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2016.

15. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometr Intell Lab. 1987;2(1-3):37-52. doi: 10.1016/0169-7439(87)80084-9. PubMed PMID: WOS:A1987K579400006.

16. Costa RS, Verissimo A, Vinga S. KiMoSys: a web-based repository of experimental data for KInetic MOdels of biological SYStems. BMC Syst Biol. 2014;8:85. Epub 2014/08/15. doi: 10.1186/s12918-014-0085-3. PubMed PMID: 25115331; PubMed Central PMCID: PMCPMC4236735.

17. Burgard AP, Maranas CD. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng. 2003;82(6):670-7. Epub 2003/04/04. doi: 10.1002/bit.10617. PubMed PMID: 12673766.

18. Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformatics. 2008;9:43. Epub 2008/01/26. doi: 10.1186/1471-2105-9-43. PubMed PMID: 18218092; PubMed Central PMCID: PMCPMC2258290.

19. Gaspar P, Neves AR, Gasson MJ, Shearman CA, Santos H. High Yields of 2,3-Butanediol and Mannitol in Lactococcus lactis through Engineering of NAD(+) Cofactor Recycling. Appl Environ Microb. 2011;77(19):6826-35. doi: 10.1128/Aem.05544-11. PubMed PMID: WOS:000295123300012.

20. Neves AR, Ramos A, Shearman C, Gasson MJ, Santos H. Catabolism of mannitol in Lactococcus lactis MG1363 and a mutant defective in lactate dehydrogenase. Microbiology (Reading). 2002;148(Pt 11):3467-76. Epub 2002/11/13. doi: 10.1099/00221287-148-11-3467. PubMed PMID: 12427938.

21. Oh E, Lu M, Park C, Park C, Oh HB, Lee SY, et al. Dynamic Modeling of Lactic Acid Fermentation Metabolism with Lactococcus lactis. J Microbiol Biotechnol. 2011;21(2):162-9. Epub 2011/03/03. doi: 10.4014/jmb.1007.07066. PubMed PMID: 21364298.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. Wisselink HW, Moers AP, Mars AE, Hoefnagel MH, de Vos WM, Hugenholtz J. Overproduction of heterologous mannitol 1-phosphatase: a key factor for engineering mannitol production by Lactococcus lactis. Appl Environ Microbiol. 2005;71(3):1507-14. Epub 2005/03/05. doi: 10.1128/AEM.71.3.1507-1514.2005. PubMed PMID: 15746354; PubMed Central PMCID: PMCPMC1065179.

23. Asmawaty T, Kadir A, Mannan AA, Kierzak AM, McFadden J, Shimizu K. Modeling and simulation of the main metabolism in Escherichia coli and its several single-gene knockout mutants with experimental verification. Microbiol Cell Factories. 2010;9(88).

24. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, et al. Multiple high-throughput analyses monitor the response of E. coli to perturbations. Science. 2007;316(5824):593-7. Epub 2007/03/24. doi: 10.1126/science.1132067. PubMed PMID: 17379776.

25. Kurata H, Sugimoto Y. Improved kinetic model of Escherichia coli central carbon metabolism in batch and continuous cultures. J Biosci Bioeng. 2018;125(2):251-7. Epub 2017/10/22. doi: 10.1016/j.jbiosc.2017.09.005. PubMed PMID: 29054464.

26. Long CP, Gonzalez JE, Sandoval NR, Antoniewicz MR. Characterization of physiological responses to 22 gene knockouts in Escherichia coli central carbon metabolism. Metab Eng. 2016;37:102-13. Epub 2016/05/24. doi: 10.1016/j.ymben.2016.05.006. PubMed PMID: 27212692; PubMed Central PMCID: PMCPMC5845446.

27. Usui Y, Hirasawa T, Furusawa C, Shirai T, Yamamoto N, Mori H, et al. Investigating the effects of perturbations to pgi and eno gene expression on central carbon metabolism in Escherichia coli using 13C metabolic flux analysis. Microbiol Cell Factories. 2012;11(87).

28. Lima AP, Baixinho V, Machado D, Rocha I. A Comparative Analysis of Dynamic Models of the Central Carbon Metabolism of Escherichia coli. Ifac Papersonline. 2016;49(26):270-6. doi: 10.1016/j.ifacol.2016.12.137. PubMed PMID: WOS:000401256700044.

29. Dromms RA, Styczynski MP. Improved metabolite profile smoothing for flux estimation. Mol Biosyst. 2015;11(9):2394-405. Epub 2015/07/15. doi: 10.1039/c5mb00165j. PubMed PMID: 26172986.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.13.444091; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

30. Kleerebezem M, Boels IC, Groot MN, Mierau I, Sybesma W, Hugenholtz J. Metabolic engineering of Lactococcus lactis: the impact of genomics and metabolic modelling. J Biotechnol. 2002;98(2-3):199-213. Epub 2002/07/27. doi: 10.1016/s0168-1656(02)00132-3. PubMed PMID: 12141987.

31. Papagianni M. Metabolic engineering of lactic acid bacteria for the production of industrially important compounds. Comput Struct Biotechnol J. 2012;3:e201210003. Epub 2012/01/01. doi: 10.5936/csbj.201210003. PubMed PMID: 24688663; PubMed Central PMCID: PMCPMC3962192.