



An energetic reformulation of kinetic rate laws enables scalable parameter estimation for biochemical networks

John C. Mason^{a,b}, Markus W. Covert^{a,b,*}

^a Department of Bioengineering, Stanford University, 443 Via Ortega Road, Stanford, CA 94305, United States

^b Allen Discovery Center for Systems Modeling of Infection, 443 Via Ortega Road, Stanford, CA 94305, United States



ARTICLE INFO

Article history:

Received 14 June 2018

Revised 20 September 2018

Accepted 19 October 2018

Available online 24 October 2018

Keywords:

Metabolism

Whole-cell modeling

Parameter estimation

Dynamics

ABSTRACT

The technology for building functionally complete or ‘whole-cell’ biological simulations is rapidly developing. However, the predictive capabilities of these simulations are hindered by the availability of parameter values, which are often difficult or even impossible to obtain experimentally and must therefore be estimated. Using *E. coli*'s glycolytic network as a model system, we describe and apply a new method which can estimate the values of all the system's 102 parameters – fit to observations from studies of proteomics, metabolomics, enzyme kinetics and chemical energetics – and find that the resulting metabolic models are not only well-fit, but also dynamically stable. An analysis of how well parameter values in the network were determined by the training data revealed that over 80% of the parameter values were not well-specified. Moreover, the distribution of well-determined values was biased to a specific part of the network and against certain types of experimental data. Our results also suggest that perturbing the functional, energetic space of parameters (rather than traditional metabolic parameters) is a superior strategy for exploring the space of biological dynamics. The estimated parameter values matched both training data and previously withheld validation data within an order of magnitude for over 85% of the data points; notably, the area of greatest frustration in the network was also the most fully determined. Finally, our estimation method showed that fidelity to physiological observations such as network response time is enforced at the cost of fit to molecular parameter values. In summary, our reformulation enables estimation of accurate, biologically relevant parameters, generates insight into the biology of the simulated network, and appears generalizable to any biochemical network – potentially including whole-cell models.

© 2018 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

In 2012, the construction and output of a ‘whole-cell’ model was reported for the simplest known culturable bacterium, *Mycoplasma genitalium* (Karr et al., 2012). In brief, a whole-cell model is an integrated simulation of several ‘processes’ or dynamical submodels, which attempts to account for the known function of every gene in a given organism over at least one cell cycle. The separation of the model in this way – for example, into one submodel that represents transcription and another that represents carbon and energy metabolism – enables a process-specific selection of modeling paradigms in order to best capture the available data relevant to that process as well as the behavior which is to be mod-

eled. Subsequent analysis of the *M. genitalium* model showed that it was able to predict the results of previously unperformed experiments, both at the level of cell physiology as well as enzyme kinetic parameters (Sanghvi et al., 2013). A number of follow-up studies have built on this foundation, providing new applications (Purcell et al., 2013), databases (Karr et al., 2014; 2013), and visualizations (Lee et al., 2013).

As large-scale *in silico* proxies for real cells, whole-cell models have the potential to serve as platforms for integrating diverse sources of biological data and generating new, testable hypotheses consequent of observed model behavior (Carrera and Covert, 2015). However, the assembly of a complete whole-cell model is a daunting task (Macklin et al., 2014). Throughout the iterative refinement of submodels, progress is restricted by the lack of specific quantitative data needed to parameterize a known dynamical behavior. Further, quantitative data may be inconsistent; this is particularly a concern when data is gathered from different stud-

* Corresponding author at: Department of Bioengineering, Stanford University, 443 Via Ortega Road, Stanford, CA 94305, United States.

E-mail address: mcovert@stanford.edu (M. W. Covert).

ies, on different organisms, or in non-native environments (e.g. *in vitro* rather than *in vivo*). An excellent review of the particular challenges posed by whole-cell model parameterization was published last year (Babtie and Stumpf, 2017). These challenges limit the scope of whole-cell models, and subsequently, their predictive power. To support the continued refinement of whole-cell models, it is clear that we need to adopt and develop strategies for whole-cell parameter estimation.

Towards this end, we sought to develop parameter estimation techniques applicable to whole-cell like models, starting first with the intention of expanding the metabolic submodel. To date, the whole-cell metabolic submodels have utilized variants (Birch et al., 2014) of flux-balance analysis (FBA), a constraint-based optimization technique that strictly enforces stoichiometric limitations but does not directly incorporate kinetic or energetic information (Orth et al., 2010). FBA is advantageous in that it can simultaneously capture the qualitative functional role of thousands of genes with relatively few model parameters. However, the top-down, optimization-based nature of FBA limits its generalizability, as the constraints must be manually tuned for each modeled condition. Alternatively, a traditional metabolic model composed of kinetic rate laws in a system of ordinary differential equations (ODEs) would allow for bottom-up, mechanistic predictions of metabolic activity.

As a model system for the development of parameter estimation techniques, we selected the glycolytic pathway of *E. coli* both for its detailed characterization in the literature and its immediate relevance to our ongoing work on the *E. coli* whole-cell model. ODE models of biochemical pathways require exquisite, complete parameterization of reaction activity; our model, composed of kinetic rate laws assembled using a procedure analogous to convenience kinetics (Liebermeister and Klipp, 2006), is no exception. To inform our parameter values, we utilized observations from studies of proteomics, metabolomics, chemical energetics and enzyme kinetics. However, as previously discussed, the available data is neither complete nor is it wholly applicable to *in vivo* models. Therefore obtaining a complete, consistent set of parameter values requires us to formulate and solve an explicit parameter estimation problem, with the goal of matching some desired dynamical behavior while simultaneously minimizing disagreement with the available observations.

These parameter estimation problems typically take the form of multi-modal optimization problems (i.e. problems with multiple local ‘solutions’); consequently they require global optimization algorithms. Global optimization algorithms can be quite computationally expensive (i.e. running for hours on hundreds or thousands of processors), and furthermore there is no guarantee for convergence (Moles et al., 2003). Despite these challenges, a number of successful large-scale metabolic parameter estimations have been performed. Most analogous to the work presented here are several detailed kinetic models of *E. coli* central carbon metabolism (Jahan et al., 2016; Kurata and Sugimoto, 2018; Millard et al., 2017), optimized using traditional approaches like genetic algorithms. Other approaches (such as the ORACLE Chakrabarti et al., 2013 and k-OptForce Chowdhury et al., 2014 platforms) utilize FBA to generate flux profiles in concert with kinetic model parameterization. Additionally, an ‘ensemble modeling’ approach (Khodayari et al., 2014) eschews typical Michaelis–Menten-like kinetic rate laws in favor of parameterizing elementary reactions. Beyond parameterizing these kinetic models, these large-scale parameter estimation efforts have yielded insights into the systems-level control of metabolic activity. However, indicated throughout these works is a consistent need for powerful, efficient estimation of parameters. Parameter estimation is itself an active area of research, including both statistical (e.g. likelihood maximization, Bayesian inference (Hug et al., 2013; Liepe et al., 2015)) and optimization-based approaches (Fröhlich

et al., 2017; Jagiella et al., 2017; Penas et al., 2017). Because our eventual goal involves whole-cell scale estimation, which is currently beyond the capacity of statistical approaches (Babtie and Stumpf, 2017), we focused on optimization for the work described here.

In the development of our own parameter estimation techniques, designed with needs of whole-cell-like models in mind, it was found that a reinterpretation of kinetic rate laws (our underlying dynamic models) in terms of energetics facilitated several downstream procedures and provided a new conceptual basis for interpreting the space of metabolic dynamics. The details of this reinterpretation and its consequences are the subject of this study.

2. Results

In Section 2.1 we introduce our energetic reformulation of standard kinetic rate laws, used throughout the study. In Section 2.2 this new formulation is used to connect observations to parameter values, elucidating the limited extent of existing literature in parameterizing a whole-cell-like kinetic model. We further apply this formulation to derive a new approach for exploring the parameter space, dubbed ‘parsimonious’ perturbation, in Section 2.3. The resulting model parameters and their fit to training data are interpreted in Section 2.4, with the dynamics of the parameterized models following in Section 2.5. We conclude with a discussion about the ramifications of this new formulation as well as the biological consequences of our optimized parameter values. Short explanations of our approaches are provided in the Appendix. Full mathematical derivations, explicit methods, secondary results and a complete record of all curated data are in the provided Supplementary Materials.

2.1. Standard kinetic rate laws can be reformulated to an energetic basis

For this study we developed an energetic reformulation of Michaelis–Menten-like kinetic rate laws, which greatly facilitates downstream parameter estimation procedure and unifies several energetic ideas (e.g. transition energies and binding energies) into one Boltzmann distribution-like equation. The most immediate utility of this reformulation is its decomposition of several coupled parameters (e.g. forward and reverse catalytic rate constants, equilibrium constants) into uncoupled (i.e. independent) parameters (e.g. standard energies of formation). We also discovered several additional benefits to this reformulation; a handful of these will be described in subsequent sections, with many others expanded in the Supplementary Materials.

Our reformulation begins with a simple metabolic model expressed as a system of ordinary differential equations (ODEs). These equations describe the time evolution of intracellular metabolite concentrations \dot{c} with respect to the stoichiometries of modeled reactions S and kinetic rate laws for each reaction $v(c)$ (which are themselves functions of the metabolite concentrations). Additionally, we account for the dilution of metabolites (consequence of cell growth) at the specific growth rate μ . Together this system of ODEs can be compactly represented as

$$\dot{c} = Sv(c) - \mu c \quad (1)$$

The kinetic rate laws $v(c)$ are assembled using a procedure that yields equations comparable to convenience kinetics, an approach developed by Liebermeister et al. just over a decade ago (Liebermeister and Klipp, 2006). Briefly, every kinetic rate law is designed to be ‘complete’ in the sense that it is fully reversible (subject to the limitations imposed by chemical energetics) and exhibits proper asymptotic behavior (i.e. saturating to some maximum rate at high substrate concentrations, and likewise diminishing to a rate of zero at low substrate concentrations). We use

this procedure to assemble an ODE model of 102 parameters for eleven kinetic rate laws and eighteen metabolites. The full kinetic rate laws for the model system can be found in the Supplementary Materials.

For the sake of demonstration, we consider the kinetic rate law associated with the simple case of a reversible reaction $A \rightleftharpoons B$:

$$v = \frac{k^{\text{cat},f} c_E}{K_A^M} \frac{c_A - \frac{1}{K_{\text{eq}}} c_B}{1 + \frac{c_A}{K_A^M} + \frac{c_B}{K_B^M}} \quad (2)$$

This is a standard formulation of Michaelis–Menten-like kinetics for a unimolecular, reversible reaction (i.e. an isomerization). This formulation is appealing in that it is naturally expressed in terms of easily interpreted parameters (e.g. concentrations, catalytic rate constants). In contrast, our reformulation expresses these kinetic rate laws in terms of energetic parameters. Under this transformation, developed in Appendix A.1, Eq. (2) becomes

$$v = k^* \frac{E(\Delta G^{\text{MAK},f}) - E(\Delta G^{\text{MAK},r})}{1 + E(\Delta G_A^b) + E(\Delta G_B^b)} \quad (3)$$

where $E(\Delta G) = \exp(-\Delta G/RT)$, k^* is an arbitrary constant for correcting units, and

$$\Delta G^{\text{MAK},f} = G^\ddagger - G^E - G_A^\circ - G_A^c \quad (4)$$

$$\Delta G^{\text{MAK},r} = G^\ddagger - G^E - G_B^\circ - G_B^c \quad (5)$$

$$\Delta G_A^b = G_A^b - G_A^c \quad (6)$$

$$\Delta G_B^b = G_B^b - G_B^c \quad (7)$$

are the differences in energy levels (ΔG terms) as linear combinations of other energy or energy-like parameters. The relationships between these basic and derived parameters are depicted schematically in Fig. 1.

Our reformulated parameters can be interpreted in terms of existing kinetic and energetic concepts; for example, $\Delta G^{\text{MAK},f}$ is the transition energy of the forward reaction in the limit that the rate of reaction is well approximated by mass-action kinetics (i.e. when substrate concentrations are relatively low), and ΔG_A^b is the energy associated with the propensity for molecule A to bind the enzyme. Likewise the new basic parameters are related to well understood terms; G_A° is the standard Gibbs energy of formation for molecule A, and G_A^c is the Gibbs energy contribution associated with concentration (i.e. $G_A^c = RT \ln c_A$). (More thorough descriptions of these terms and their relationships with other energetic or kinetic parameters are provided in the Supplementary Materials.)

In the following sections we will describe how this reformulation facilitates both basic analysis of our training data as well as the development of our novel perturbation approach for parameter estimation.

2.2. The energetic reformulation readily links data availability to model parameters

The parameter reformulation presented in Section 2.1 allows us to ascertain the degree to which our model parameter values can be determined from training data, even when observed parameters and model parameters do not have a one-to-one relationship (e.g. in the case of chemical equilibrium constants and standard energies of formation). The data used in this study is described in Appendix A.2, while the description of our ‘determinacy analysis’ procedure is given in Appendix A.3. In brief, the training data were

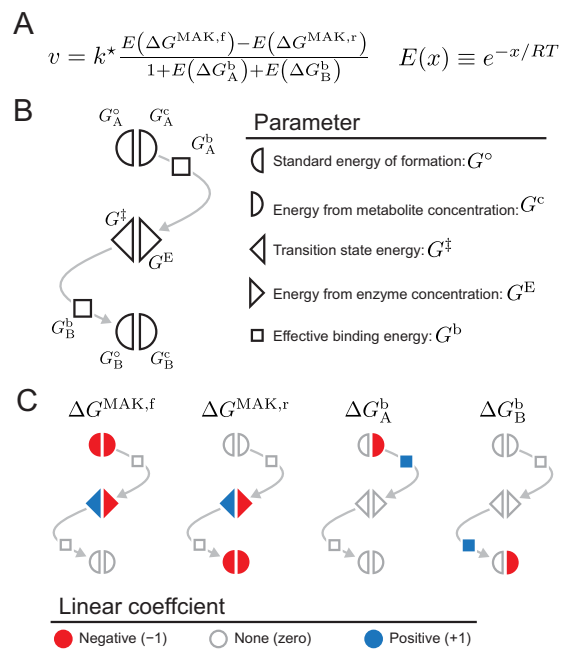


Fig. 1. Energetic reformulation of a kinetic rate law. (A) The energetic reformulation of a kinetic rate law for the reaction $A \rightleftharpoons B$, as presented in the text, in terms of differences between energy levels. (B) The new Gibbs energy (or energy-like) parameterization, overlaid upon a reaction diagram. (C) The differences between energy levels illustrated as linear combinations of the basic model parameters.

transformed into the energetic parameter space (as a vector b) and associated with the basic parameter values (as vector x) via the linear expression $b = Fx$; the matrix F can then be analyzed to determine where a parameter value is *undetermined* (the training data provides no information to help determine the parameter value) *fully determined* (the value can be completely determined from the training data), or *partially determined* (the value is informed by the training data, but requires further information to be fixed). The determinacy of each parameter is depicted on our network model of glycolysis in Fig. 2.

In applying this procedure to our model system, we find that, of the 102 model parameters, 19 are fully determined, 32 are partially determined (i.e. only determined relative to other parameters), and 51 are wholly undetermined. Thus, over 80% of the parameters in our model are underdetermined.

Beyond the total determinacy of parameter values, given the curated literature, we also observe biases in the availability of certain types of data. First, we observe that part of the network, centered around the fructobisphosphate aldolase (fba) reaction in upper glycolysis, has been investigated more extensively than the rest. Additionally, the majority of underdetermined parameter values (35/51 or 69%) correspond to effective binding energies (i.e. Michaelis–Menten saturation constants), particularly for substrates of the reactions opposite the canonical (usually glycolytic) direction of reaction.

We also see that both enzyme concentrations and standard energies of formation are never fully determined. This is readily explained in the context of the training data. In the case of standard energies of formation, these parameter values are only determined *relative* to one another, due to the fact that our most relevant training data is chemical equilibria constants (each of which is a composition of two or more standard energies of formation in their corresponding stoichiometric ratios). The underdeterminacy of the enzyme concentrations arises from the fact that proteomics data were expressed as *relative* abundances, eliminating the need to scale this data but somewhat reducing its information content.

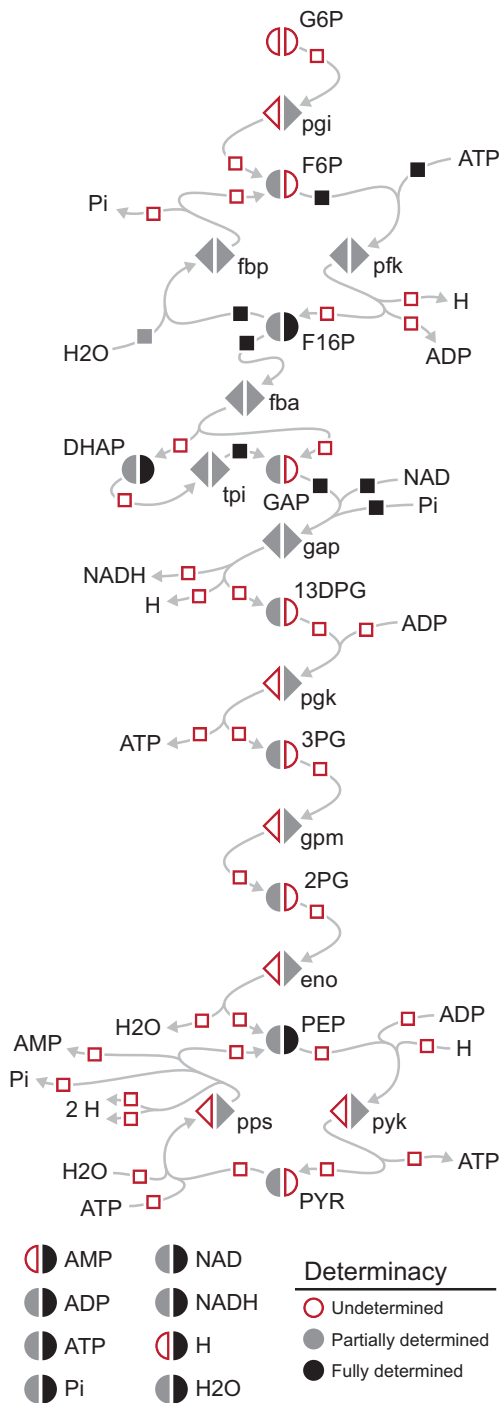


Fig. 2. Network-parameter diagram. Diagram of the modeled network (*E. coli*'s glycolytic pathway), overlaid with one symbol for each of the 102 model parameters (key provided in Fig. 1B). Each symbol is styled according to its *determinacy*, i.e. the extent to which its value can be inferred from the training data (parameter value observations).

While not all independent parameters need to be established to concretely define model dynamics, the extent of non-fully determined parameters, as well as the non-uniformity of where determinacy occurs throughout the network, demonstrates a serious need for a parameter estimation system that can estimate a large number of unknown parameter values simultaneously. In the next section we will perform such a parameter estimation, facilitated by our energetic parameter reformulation.

2.3. A novel perturbation approach leads to more successful parameter estimation

Next, we recognized that the energetic reformulation of the kinetic rate laws enabled the development of a novel perturbation approach that leads to successful estimation of parameter values, even when applied to a rudimentary optimization algorithm. We began with a 'naive' approach in which the basic model parameters were perturbed individually. This approach struggled to converge to a dynamically viable, well fit set of parameter values, as confirmed by inspecting objective values as well as performing linear stability analysis (Strogatz, 1994) (LSA) on the parameterized models (see Appendix A.6) (Fig. 3). This result was illustrative of the technical challenges posed by these parameter estimation problems. Several strategies have been employed in the context of biological networks, with varying degrees of success (Moles et al., 2003). In practice, strategies such as genetic algorithms (Jahan et al., 2016) and particle swarm optimization (Millard et al., 2017) have been successfully applied to estimate parameter on models of large biochemical networks, albeit at significant computational costs.

These results led us to suspect that a perturbation strategy which was sensitive to the context of the parameters (and therefore sensitive to the structure of the modeled biological system) could more effectively search the parameter space. For example, consider a parameter such as the concentration of ATP, a metabolite utilized in several reactions. Somewhat near an optimal set of parameter values, it would be difficult to perturb the concentration of ATP without upsetting the dynamics in many other parts of the network, presumably causing the optimization to become trapped in a 'corner' of the high-dimensional parameter space.

To address this concern we developed an approach that allowed us to perturb the parameters in a fashion that was minimally disruptive to the dynamical behavior of the system as a whole – hence, a 'parsimonious' perturbation in terms of system behavior. Parsimonious perturbations can be achieved by perturbing derived model parameters related to dynamical behavior (e.g. the differences in energy levels corresponding to a maximum rate of reaction v^{\max} or a saturation ratio c/K^M), and selecting a search direction that minimally disrupts the other derived model parameters. The mathematical procedure for obtaining a parsimonious perturbation vector is given in Appendix A.4.

For either perturbation approach, our parameter estimation problem is formalized as the simultaneous optimization of two objective terms; first, to minimize the disagreement with reported parameter values, expressed as the 'misfit cost' function $f(x)$; and second, to ensure that the simulations achieve a steady-state, which we express as the 'non-steady-state cost' function $g(x)$. Formally we write this as the optimization problem

$$\min_x f(x) + w \cdot g(x) \quad (8)$$

where w is some weight that we increase until $g(x)$ is sufficiently small. (The complete expression of these terms is provided in Appendix A.5.)

To solve this optimization problem, we employ an algorithm (described extensively in the Supplementary Materials) in which we perturb our current best parameter values and search for an improved (smaller) objective value. The resulting values of these objective terms are shown in Fig. 3 for 300 attempts using each perturbation approach.

Fig. 3 shows that, in contrast with the 'naive' approach, the majority of attempts using the 'parsimonious' perturbation approach successfully minimize the non-steady-state cost (234 or 78%). These attempts also exhibit misfit costs that are at the lowest part of the range. Our results suggest that perturbing the functional, energetic space of parameters (rather than traditional

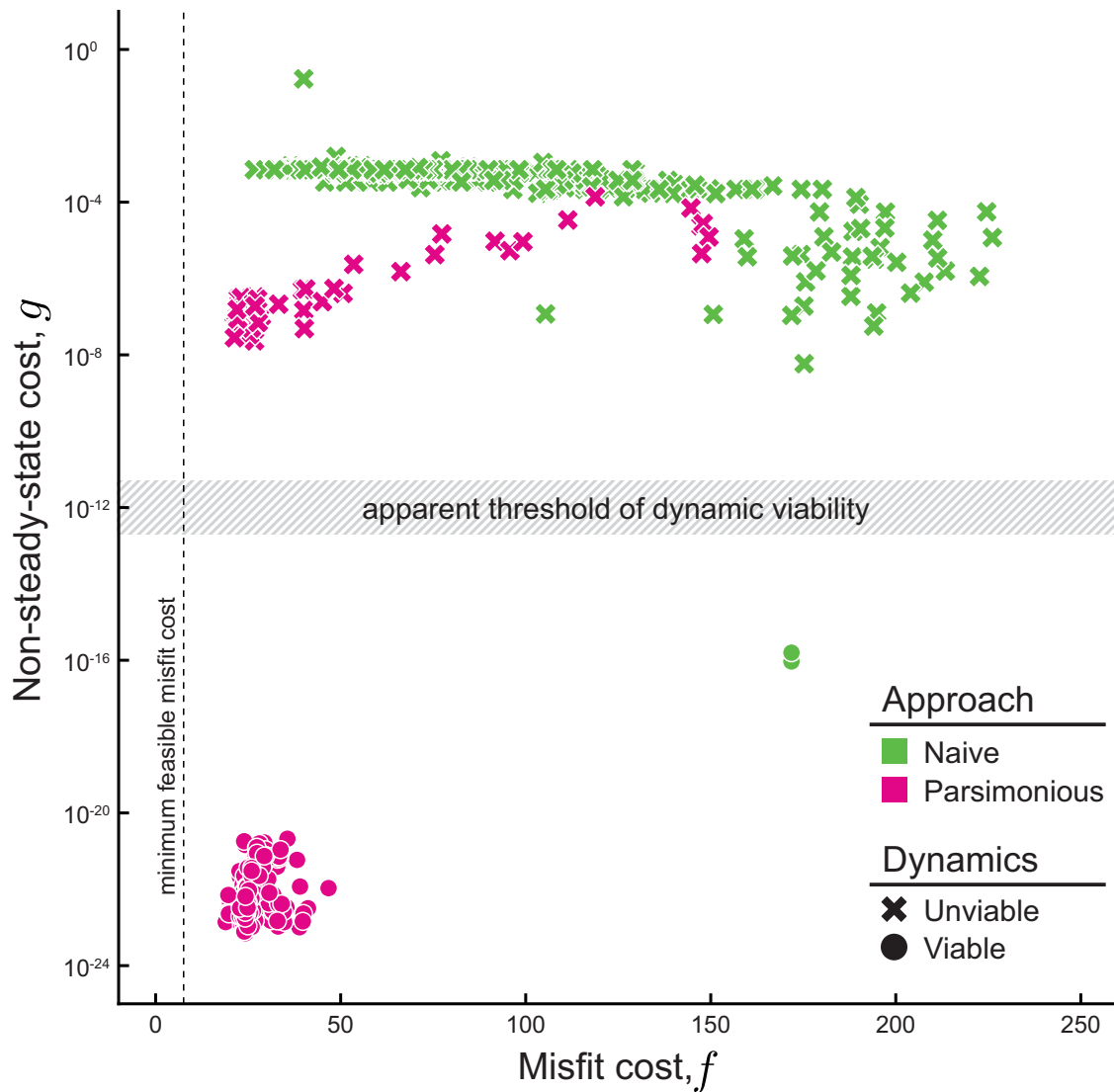


Fig. 3. Post-optimization objective values. The values of the two major objective components following three-hundred (300) optimization attempts, using the 'naive' and 'parsimonious' perturbation approaches. The dynamic 'viability' of each parameter set was determined via simulation of the parameterized models and linear stability analysis. 234 (78%) of the attempts using the 'parsimonious' approach were found to be viable, compared to just 2 (0.67%) of those using the 'naive' approach.

metabolic parameters) is a superior strategy for exploring the space of biological dynamics.

2.4. Estimated parameter value distributions compare well with training and validation data

Our next objective was to evaluate the quality of the dynamically viable sets of estimated parameter values, using three metrics: (1) fit to the training data, (2) fit to a set of withheld validation (test) data, and (3) consistency of the output simulation dynamics with physiological observations.

Fig. 4B compares the estimated parameter values against the 144 training data values (corresponding to 45 basic or derived model parameters) utilized in the expression of the misfit cost function. We find that the median estimated parameter values are, as desired, generally well fit to their training (40 of 45 or 88% falling within ten-fold of the median observed value), and moreover that the distributions of the parameter values are typically quite narrow.

Beyond the degree of fit to our training data, we also sought to evaluate whether our parameter estimation system could make

reasonable predictions of withheld parameter values (i.e. test or validation data). Fig. 4C shows the distributions of predicted parameter values against catalytic rate constants estimated from specific activity data. Here we find that our predictions are 'well fit' (again, within ten-fold) with the observed values for five of the six reactions, where the one poor fit is for the phosphoglycerate mutase (gpm) reaction.

The main point of frustration in these comparisons appears to be centered around the triosephosphate isomerase (tpi) reaction, a topologically unusual reaction for this network in that it interconverts between the two upstream products of the fructosebisphosphate aldolase (fba) reaction to feed into the glyceraldehyde phosphate dehydrogenase (gap) reaction. Parameter values related to these three reactions are poorly fit to the training data, suggesting one or more conflicts between the experimental data and the desired behavior of the dynamical model. In particular we find that the chemical equilibria constants for the fba and tpi reactions, along with the concentration of the fba enzyme and Michaelis-Menten saturation constant for glyceraldehyde phosphate in the gap reaction are all shifted in a fashion that would facilitate

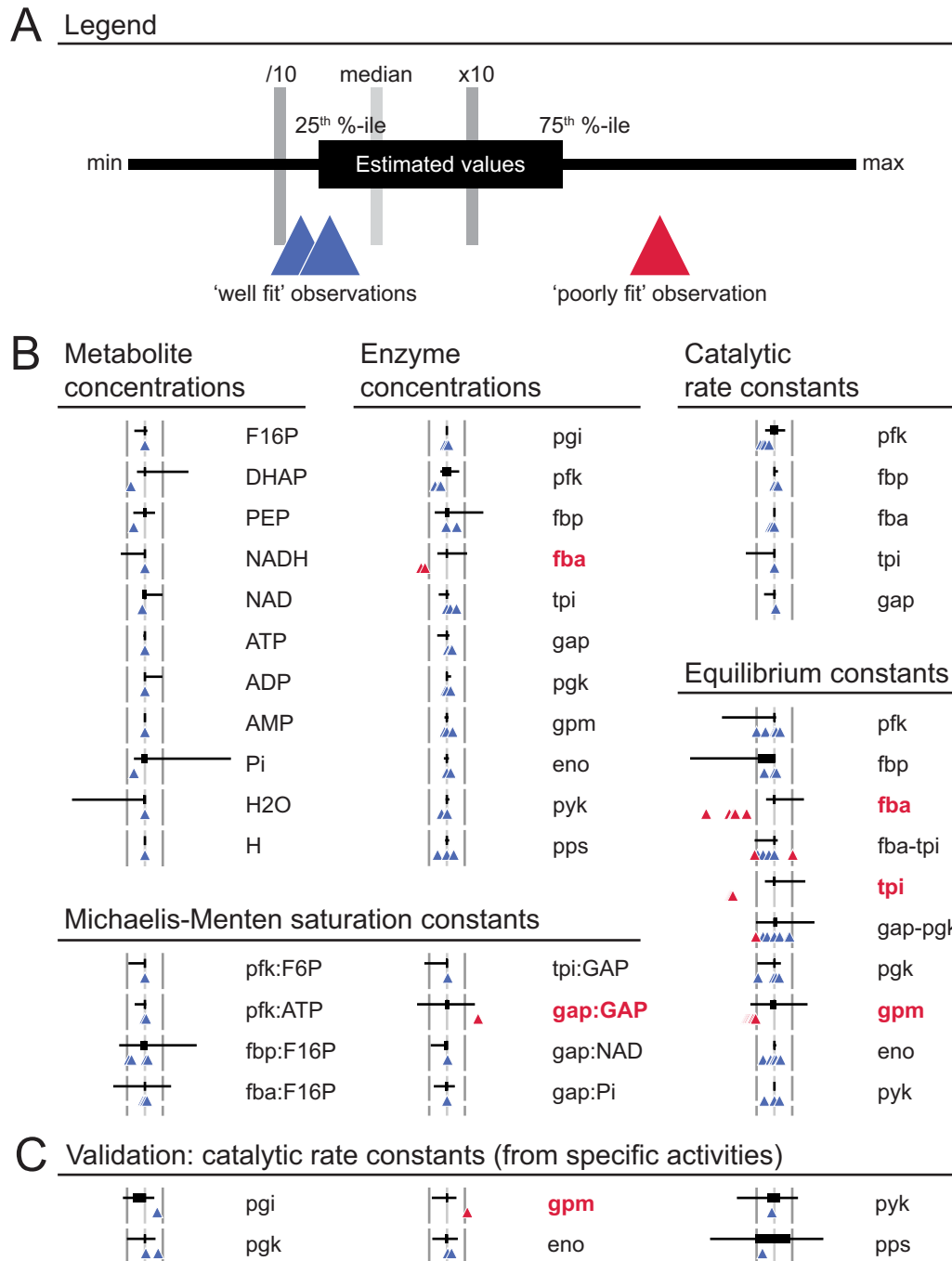


Fig. 4. Distributions of estimated parameter values versus observed values. The 234 dynamically viable sets of estimated parameter values are illustrated here as simplified box plots, centered by their median estimated value and scaled logarithmically. (A) Legend. We denote observations more than ten-fold away from the median estimated parameter value as poorly fit (emphasized in bold, red text). (B) Comparison with training data. (C) Comparison with test data (catalytic rate constants, estimated from reported specific activities). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

greater flux through glycolysis. Interestingly, this portion of the network is also the most fully determined with regard to model parameters.

2.5. Identifying a trade-off between conformity to physiological observations and fit to molecular-level measurements

We next considered our third metric listed above: the consistency of the output simulation dynamics to physiological observations. We first verified that all sets of estimated parameter values with sufficiently low non-steady-state costs (very small values of

$g(x)$, as shown in Fig. 3) were also dynamically stable, reaching a steady-state over time (top row of Fig. 5). However, this characterization also revealed that the recovery times were very slow, on the order of hours. This was far longer than the anticipated perturbation recovery times for real cells, which other studies suggested should be on the order of seconds (Schaub and Reuss, 2008).

By performing LSA on the original model equations (Eq. 1), under the assumption that all metabolite concentrations were in excess of their corresponding Michaelis–Menten saturation constants (K^M values), we found that all characteristic recovery times (CRTs) converged to $\tau = 1/\mu \approx 1.4$ h, consistent with the re-

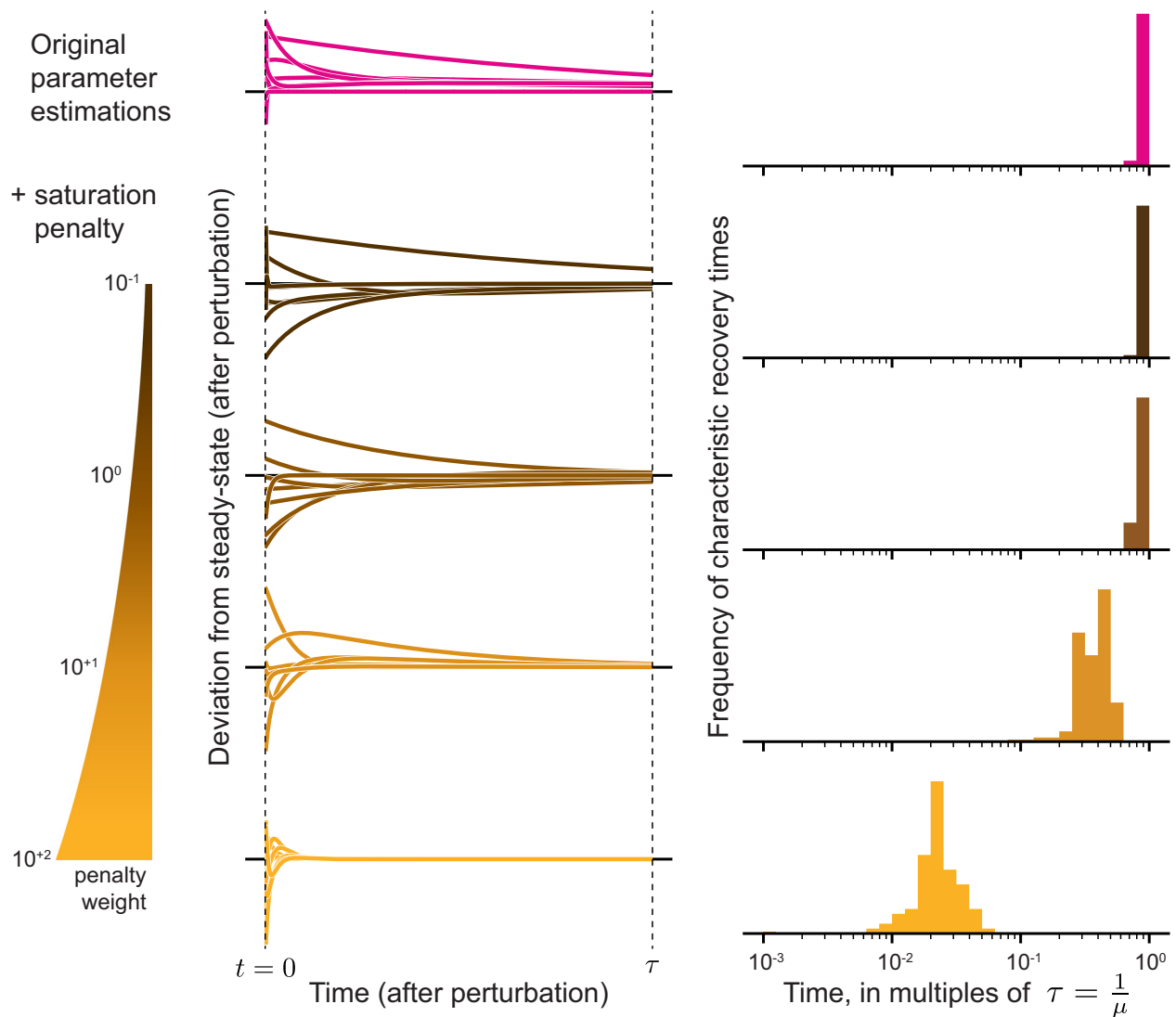


Fig. 5. Dynamics of parameterized models. The dynamically viable sets of estimated parameter values were characterized by their stability under a perturbation, using linear stability analysis to determine a characteristic recovery time. Left: A simulated perturbation recovery for a selected set of parameter values. The deviation of the eight dynamic concentrations from their steady-state is plotted over $\tau = 1/\mu \approx 1.4$ h. Right: Distributions of characteristic recovery times (CRTs), normalized by τ . Top row: Dynamics of the original parameter estimations. Subsequent rows: The dynamics consequent of parameter estimation problems augmented to include a penalty for high enzyme saturation (i.e. $c > K^M$), with increasing weight.

covery times observed by simulating the ODEs. This suggested that the cause of the slow recovery times was high enzyme saturation.

To test this hypothesis, we augmented the original parameter estimation problem to include terms in the misfit cost function $f(x)$ that penalized for excess enzyme saturation. While the misfit cost function was originally designed to penalize for order-of-magnitude differences between training data and parameter values, our energetic reformulation of model equations allows us to also penalize for any combination of parameters values against some other expected value. We chose to penalize for metabolite concentrations c in excess of their corresponding Michaelis-Menten saturation constants K^M (i.e. when $c/K^M > 1$). The parameter estimation procedure was repeated on this augmented problem, using a range of weights on the new terms.

As the weighting on these terms was increased, we found that the rates of recovery following a perturbation increased dramatically (as shown in the remaining panels in Fig. 5). This is partic-

ularly apparent when the weight on the terms is greater than the weight on the training data (10^{+1} and higher).

We also found that reduced enzyme saturation (higher weights on the saturation penalty) led to other effects on our parameter estimation, summarized in Table 1. First, we noted that the overall determinacy of the network was increased considerably by introducing our penalty terms for enzyme saturation. In particular, whereas 50% of the parameters were undetermined previously, our new strategy reduced this percentage to 9%. The inclusion of enzyme saturation constraints also led to more realistic simulations in terms of pathway utilization, as the canonically glycolytic reaction pyruvate kinase (pyk) was favored over the canonically gluconeogenic reaction, phosphoenolpyruvate synthetase (pps). In contrast, the low- and no-weight parameter estimates tended to favor the pps reaction.

The benefits of higher network determinacy and more realistic simulations with respect to response time and pathway utilization were tempered by an increase in the misfit cost. Examining the parameter misfit in more detail, we found that seven param-

Table 1

Summary of augmented (saturation penalized) parameter estimation results. Where appropriate, values are reported as medians and median absolute deviations.

Penalty weight ^a	% undetermined ^b	Misfit cost ^c (f)	pyk flux ^d
Not used (0)	50	24.7 ± 1.1	1.0 ± 0.5
10 ⁻¹	9	24.2 ± 0.9	0.9 ± 0.6
10 ^{±0}	9	41.3 ± 7.2	1.2 ± 0.7
10 ⁺¹	9	39.7 ± 2.4	79.8 ± 0.8
10 ⁺²	9	40.9 ± 1.5	80.1 ± 1.1

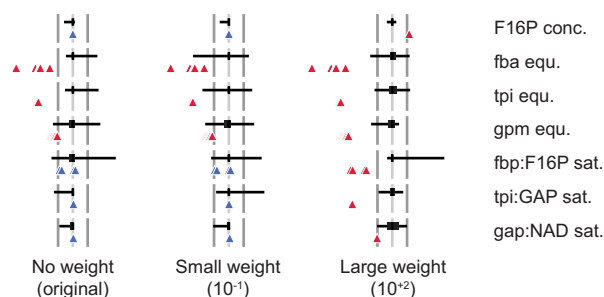
^a Weight on the enzyme saturation penalty (if used).

^b Out of 102 parameters.

^c The misfit cost f to the original (unaugmented) training data.

^d Percent of net flux relative to the target overall pyruvate production rate.

A Major shifts in estimated parameter fit to training data



B Validation: catalytic rate constants

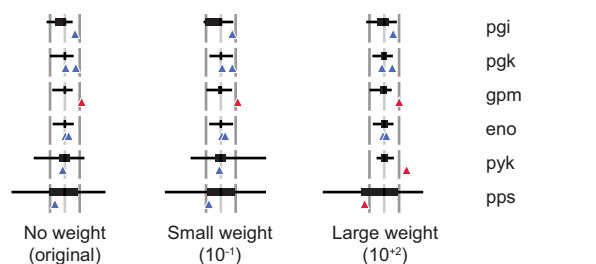


Fig. 6. Fitness of estimated parameter values following problem augmentation. (A) The major changes (ten-fold or greater) to the training data fit, following problem augmentation, at the small (10⁻¹) and large (10⁺¹) saturation penalty weights. (B) Predicted k^{cat} values compared to test data. Original distributions are reproduced from Fig. 4 for clarity; also see Fig. 4A for the legend.

eter values were impacted, again centered about upper glycolysis (Fig. 6A). The degree of fit to the catalytic rate constants for pyk and pps in our validation data set were also negatively impacted (Fig. 6B). This was unsurprising, given that the training data had originally favored the pps reaction over pyk as described above.

3. Discussion

In summary, we found that a reformulation of kinetic rate laws in terms of energetic parameters was necessary for developing parameter estimation techniques applicable to whole-cell modeling. Such a reformulation was partially considered in the development of convenience kinetics (Liebermeister and Klipp, 2006); our approach builds on those concepts by also reformulating metabolite and enzyme concentrations as energy or energy-like quantities. While care should be taken to not overly interpret these quantities as true energies (e.g. G^b should be most properly thought of as an effective binding energy), this energetic reformulation nonetheless provides a theoretical basis for exploring the space of biological

dynamics in the derived, functional coordinate space of differences between energy levels.

An interesting insight that arose from the energetic reformulation came as we analyzed the determinacy of the network, using a technique that can link parameter value observations to model parameters, even when the relationship between the two is complex. In the context of this study, we found a considerable amount of underdetermined parameter values, consequent of a lack of applicable experimental observations. In some cases the observations may have been made at some time but were excluded by our somewhat selective curation process (e.g. our measurements were required to come from studies of *E. coli*; see Appendix A.2 for more requirements). For other parameters, no measurements have been made to date. In any event, we found that the determinacy was not only biased by experimental method but also by location in the network. Thus, our determinacy analysis should help in prioritizing future measurement technologies to be developed and experiments to be performed. However, considering that our system is one of the most well-established pathways in perhaps the world's most characterized organism, this finding underscores the need for large-scale parameter estimation techniques to better understand and model biological systems.

Most critically, we discovered that our energetic reformulation enabled us to develop a new perturbation approach, which enabled a rudimentary optimization algorithm to reliably estimate well fit, dynamically viable model parameters. In particular, we found that perturbing the functional, energetic space of parameters (rather than traditional metabolic parameters) is a superior strategy for exploring the space of biological dynamics, leading to sets of parameter values that fit the training and validation data well and were also dynamically viable.

We were surprised to observe a slow response time in our original simulations (i.e., based on the first parameter value estimates). We addressed this issue by motivating our parameter estimation system to find solutions with reduced enzyme saturation. While we may naturally expect enzymes to be heavily saturated, thereby achieving the highest possible rate of reaction per enzyme, our results suggest that this comes at the cost of reduced adaptability to fluctuations in metabolite concentrations. This further implies a competing evolutionary pressure between enzyme efficiency and enzyme sensitivity, i.e. a trade-off between efficient growth under stable conditions and flexible growth under variable conditions. Whether this trade-off is a generalizable property of (unregulated) metabolic pathways remains to be seen.

Our reduced enzyme saturation approach led to more realistic system-level behaviors, including faster recovery following a perturbation as well as utilization of the more canonical pathway for pyruvate production. However, these improvements came at the cost of worse fits to training and test data. These results can be explained in three ways: first, that kinetics data for glycolysis – typically generated *in vitro* – is not consistent with expected *in vivo* behavior, due to the influence of ions, pH, and molecular crowding in the cytoplasm that could lead to different effective parameter values (Bennett et al., 2009). Alternatively, we could speculate that our expectations about pathway utilization and perturbation recovery times are incorrect. A third explanation, which requires discounting neither the parameter value observations nor our behavioral expectations, is that our saturation penalty partially captures the effects of kinetic (e.g. allosteric) regulation (effects which are not explicitly included in this model). Like the increased sensitivity conferred by the reduced levels of saturation, allosteric regulation would allow modulation of reaction rates subject to perturbations in the system (however these would likely be *global* signals, rather than the *local* inputs and outputs of a given reaction). Glycolysis is in fact known to be allosterically regulated (Frankel, 1996), lending credence to this possibility.

Finally, our parameter estimation also proved capable of identifying and reconciling points of frustration in the training data. Of particular note was our observation that the chemical equilibria constants associated with the fructobisphosphate aldolase (fba) and triosephosphate isomerase (tpi) reactions had to be shifted to favor the forward (glycolytic) direction of reaction to achieve sufficient flux through the pathway. This result is parallel to our findings regarding the catalytic rate constants for pyk and pps, in which adding the constraint on enzyme saturation unexpectedly led to simulations that used the canonical glycolytic pathway – but at the cost of agreement with some validation data. We expect that further targeted measurements may be able to resolve these frustration points.

We believe that the approach described here should generalize to any biochemical network (as it is informed by the equations, which are in turn informed by the network structure) as well as to other optimization algorithms. In fact, we see no reason why this technique could not generalize to domains beyond biology, so long as systems of equations can be expressed as sums of polynomials, rational functions, or Boltzmann distributions.

Ultimately, a *complete* whole-cell model should simulate the functional role of every gene and gene-product. This poses a tremendous parameter estimation challenge, on a scale far beyond the simple pathway we have modeled in this study. Even neglecting regulatory details, we estimate that a complete kinetic model of metabolism, including hundreds of metabolites and reactions, would easily require thousands of parameters. Future work will be required to conclusively determine whether the methods described here can truly be expanded to that scale, and even beyond metabolism to other biological submodels. If so, we anticipate that one of the most challenging bottlenecks in whole-cell modeling can be largely resolved.

4. Data declaration

All code and data used in this study are available at <https://github.com/CovertLab/parest>.

Acknowledgments

We thank the members of the Covert Lab for advice and critical commentary on the manuscript, in particular Eran Agmon for his suggestions on the text and figures. We gratefully acknowledge funding from several sources, including the Agilent Expendable Fellowship and Stanford NIST JIMB Training Grant (70NANB15H192) to JCM and an Allen Distinguished Investigator Award and Allen Discovery Center Award to MWC. Our funding sources played no role in the generation or analysis of our data. Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

Appendix A. Methods

A1. Energetic reformulation of the kinetic rate law

We begin by observing that the parameter values for the real biological systems we intend to consider are strictly non-negative, and vary over orders of magnitude. Thus it is possible and practical to work with logarithmically rescaled parameters. First consider the chemical equilibrium constant for a given reaction, K^{eq} , which can be expressed in terms of the Gibbs standard energy of reaction ΔG° , the gas constant R , and temperature T :

$$K^{\text{eq}} = \exp -\Delta G^\circ / RT \quad (\text{A.1})$$

Then, by solving for ΔG° , we obtain

$$\Delta G^\circ = -RT \ln K^{\text{eq}} \quad (\text{A.2})$$

While K^{eq} (and subsequently ΔG°) values are defined on a per-reaction basis, they are not independent, in that any two reactions or sets of reactions with the same net stoichiometry (e.g., isoenzymes, or alternate pathways) will reach the same equilibrium. For the sake of parameter estimation, we need these parameters to be independent. This is resolved by expressing the Gibbs standard energies of reaction in terms of the stoichiometry of the corresponding reaction s and the Gibbs standard energies of formation G° for each metabolite. Unlike the standard energies of reaction, the standard energies of formation are independent, and are related by the linear equation

$$s^\top G^\circ = \Delta G^\circ \quad (\text{A.3})$$

$$= -RT \ln K^{\text{eq}} \quad (\text{A.4})$$

Thus we have obtained our first set of independent, logarithmically rescaled parameters. Similar transformations are performed for the other parameters in our study. In the case of the parameters with dimensions of concentration, that is, the metabolite concentrations (e.g. c_A), enzyme concentrations (e.g. c_E), and Michaelis–Menten saturation constants (e.g. K_A^M), we define

$$G_A^c = RT \ln c_A \quad (\text{A.5})$$

where G_A^c is the Gibbs energy from the concentration of metabolite A,

$$G^E = RT \ln c_E \quad (\text{A.6})$$

where G^E is the Gibbs energy from the concentration of the enzyme, and

$$G_A^b = RT \ln K_A^M \quad (\text{A.7})$$

where G_A^b is the effective Gibbs energy for the binding of metabolite A to the enzyme. (For expressive convenience, and without loss of generality, we have selected a standard reference energy of 0 kcal mol⁻¹ for a concentration of 1M.) Substituting these transformations into Eq. (2), the kinetic rate law becomes

$$v = k^{\text{cat},f} \frac{E(-G^E - G_A^c + G_A^b) - E(-G^E - G_B^c + G_A^b - G_B^\circ + G_A^\circ)}{1 + E(G_A^b - G_A^c) + E(G_B^b - G_B^c)} \quad (\text{A.8})$$

where we define $E(x) = \exp -x/RT$ for the sake of brevity. We can further rearrange this expression to make the forward and reverse reaction rate terms in the numerator symmetric:

$$v = k^{\text{cat},f} E(G_A^b - G_A^c) \frac{E(-G^E - G_A^c - G_A^\circ) - E(-G^E - G_B^c - G_B^\circ)}{1 + E(G_A^b - G_A^c) + E(G_B^b - G_B^c)} \quad (\text{A.9})$$

Finally, we redefine the forward catalytic rate constant as

$$k^{\text{cat},f} = k^* E(G^\ddagger + G_A^\circ - G_A^b) \quad (\text{A.10})$$

where G^\ddagger is defined as the energy of the transition state, and k^* is simply a proportionality constant that could be estimated from the Eyring equation (Eyring, 1935). Together these transformations give us Eq. (3). Additional relations, as well as a glossary of symbols, can be found in the Supplementary Materials.

A2. Data curation

The pathway reaction substrates, products, enzymes and stoichiometry were obtained from EcoCyc (Keseler et al., 2017). A complete description of the curated training data (sources and values) are provided Supplemental Materials. In brief: Intracellular metabolite concentrations were mostly taken from Bennett et al. (2009). Some data were excluded; for example,

the hexose-6-phosphate concentration, which would include both glucose- and fructose-6-phosphate, was not used because of its ambiguity. The proton concentration was computed assuming a pH of 7.4 (Wilks and Slonczewski, 2007). The water concentration was computed assuming a water content of 70% (Neidhardt et al., 1990). Altogether, concentration training data for eleven of eighteen metabolites were used. In addition to the concentration data, we also utilized 61 chemical equilibrium observations curated in the same metabolomics study (Bennett et al., 2009).

Kinetic parameters were adapted from fourteen sources (Berry and Marshall, 1993; Brown et al., 2009; Eyschen et al., 1999; Fenton et al., 2003; Hao and Berry, 2004; Hines et al., 2007; Iancu et al., 2005; Kelley-Loughnane et al., 2002; Kruger, 1989; Mainfroid et al., 1993; Plater et al., 1999; Wang and Kemp, 2001; Zgiby et al., 2000; Zheng and Kemp, 1992). The curation of kinetic parameters was facilitated in part through the BRENDA enzyme database (Placzek et al., 2017). Only studies of *E. coli* enzymes were curated, and only studies that reported catalytic rate constants (k^{cat}) were included. In general, only the kinetics of the dominant (most highly expressed) isozyme were used.

Proteomics data were taken from three sources to estimate cellular concentrations (Schmidt et al., 2016; Taniguchi, 2010; Wiśniewski and Rakus, 2014). Again, only the dominant isozyme was included. All enzymes had at least two reported values.

Fluxomics data were not utilized as training data in this study. However, based on the work by Toya et al. (2010), an output pyruvate production rate of 0.14 mM s^{-1} was selected as a constraint. Further, to capture the small albeit nontrivial effects of metabolite dilution due to cellular growth, we assume a cellular doubling time of one hour (Woldringh et al., 1977). Finally, molecular weights are taken from EcoCyc (Keseler et al., 2017). These molecular weights are not used directly in the dynamic model, but are utilized in scaling the optimization objective function.

A3. Determinacy analysis

The *determinacy* of a given parameter can be ascertained by inspecting F , the matrix that relates the basic parameters x to the observed parameter values b via $b = Fx$. We define an *undetermined* parameter as one that is wholly uninformed by the training data; these are reflected by empty columns in F . A *partially determined* parameter is one that can only be explicitly defined in the context of fixing one or more other partially determined parameters. In the linearized form $b = Fx$, this is equivalent to asking whether there exists a vector Δx such that $Fx = F(x + \Delta x)$, or equivalently $F\Delta x = 0$, meaning that the perturbed parameter value vector $x + \Delta x$ has the same fit as the original set of parameters x . Such vectors Δx are said to be in the null space of F , and therefore are composed by linear combinations of basis vectors for the null space.

This idea can be expressed compactly by forming a matrix N where the columns form a complete basis for the null space. Such a matrix can be found by singular value decomposition (SVD), or else by forming the null space orthogonal projection matrix

$$N = I - F^\dagger F \quad (\text{A.11})$$

where the symbol \dagger indicates the Moore–Penrose pseudoinverse; in this case, N is an $n \times n$ matrix where n is the number of parameters (the size of x). Regardless of how N is formed, we can use this matrix to find vectors η such that $\Delta x = N\eta$ for any Δx , where $0 = F\Delta x$. If a row in N has a nonzero element, it means that the corresponding basic parameter can be perturbed in a fashion that does not change the overall fit to the data. Thus we call these parameters partially determined. The remaining parameters must then be fully determined.

A4. Parsimonious perturbation approach

The unit parsimonious perturbation vectors, used for exploring the parameter space in our optimization problem, can be calculated using the formula

$$u = (I - N(AN)^\dagger A)(d^T)^\dagger \quad (\text{A.12})$$

where d is some desired perturbation direction (i.e. the basic or derived variable $z = d^T x$ that we wish to perturb, such that $d^T u = 1$), A is a matrix which defines the basic or derived variables $y = Ax$ that we wish to hold constant, N is a null space basis on the vector d^T (see Eq. (A.11) for one way to obtain such a basis), and I is an identity matrix. Again, the symbol \dagger indicates the Moore–Penrose pseudoinverse. The complete derivation of this expression can be found in the Supplementary Materials.

For the variables y (those which we generally wish to hold constant under a perturbation) we choose the energy differences associated with the forward and reverse maximum rates of reaction ($\Delta G^{\text{max},f}$ and $\Delta G^{\text{max},r}$) as well as the binding energy differences for all substrates ΔG^b . We also include the energies from concentration G^c for all metabolites with *dynamic* concentrations. Together these derived variables can fully parameterize the ODE system given in Eq. (1). For the variables $z = d^T x$ (those which we wish to perturb) we choose all variables in y . We also include all basic variables such that the set of unit parsimonious perturbation vectors is full rank (discarding any redundant u vectors found in the process). These unit vectors are calculated once, and then reused throughout the optimization procedure.

A5. Objective formulation

Our parameter estimation objective, similar to that used by Jahan et al. (2016), is thoroughly motivated and developed in the Supplementary Methods. In brief, the original problem may be stated as

$$\min_x f(x) \quad (\text{A.13})$$

$$\text{s.t. } g(x) = 0 \quad (\text{A.14})$$

where $f(x)$ is (again) our misfit cost term, and $g(x)$ is a constraint term that is defined to be zero when we have obtained some non-trivial steady state. Since $g(x)$ is nonlinear, $g(x) = 0$ is difficult if not impossible to enforce, and thus we choose instead to solve the problem

$$\min_x f(x) + w \cdot g(x) \quad (\text{A.15})$$

where we increase the weight w until we are satisfied that $g(x)$ is sufficiently small.

We express $f(x)$ as the *sum of the absolute values* of the *differences in magnitude* between parameter values and corresponding observations. The difference-in-magnitude is expressed as the difference between the logarithms of the estimated and observed values (or, equivalently, the logarithm of the ratio). Under our energetic reformulation, this simply becomes

$$f(x) = \sum_i |F_i^T x - b_i| \quad (\text{A.16})$$

The difference-in-magnitude penalty is preferable because it is dimensionless (in contrast with, say, an absolute difference). We use the sum of absolute values, rather than the sum of squares, because the former tends to extract a few poor fits to the training data (i.e. points of frustration) rather than distributing the error more or less evenly across all observations.

We express $g(x)$ as the sum of three terms:

$$g(x) = \sum_i \dot{M}_i^2 + \sum_i \dot{G}_i^2 + \left(\frac{r_{\text{PYR}}(x)}{r_{\text{PYR, observed}}} - 1 \right)^2 \quad (\text{A.17})$$

The first term is the sum of squares over the terms in the ODE \dot{c}_i , scaled by their molecular weights. The second term is the sum of squares of the ODE terms normalized by their concentrations (i.e. \dot{c}_i/c_i), which under a scaling factor is equivalent to the sum of squares of the rate of change in molar free energies \dot{G} (where $G = G^\circ + G^c = G^\circ + RT \ln c$). These terms may be thought of as the *absolute* and *relative* deviations from steady state, respectively. In practice we find that both terms are necessary to approach a true steady state. The last term penalizes for any deviation from our target production rate of pyruvate, $r_{\text{PYR, observed}}$. Without this term, the system inevitably converges to the trivial steady state in which all rates of reaction are zero.

A6. Stability analysis

The linear response of the system about its (supposed) steady state is given by the Jacobian on the system of ODEs at the steady state point, $J\dot{c}|_{ss}$. As this matrix is a first-order approximation to the behavior of the system, its (local) dynamical modes can be decomposed by extracting the eigenvalues of $J\dot{c}|_{ss}$. The real part of an eigenvalue dictates whether it converges to or diverges from 0. A positive real component of an eigenvalue is divergent and therefore unstable, while a negative value is convergent (stable). If any of the eigenvalues have a non-negative real component, then the system is generally unstable. Regardless, the largest real eigenvalue (LRE) dominantly dictates the dynamics of the system, and therefore is used first to judge whether the system corresponding to a given set of parameters is stable, and secondly, to quantitatively characterize the rate of convergence if it is stable. For those parameterizations that are found to be dynamically stable, we calculate the characteristic recovery time (CRT) as the negative-inverse of the LRE.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.jtbi.2018.10.041](https://doi.org/10.1016/j.jtbi.2018.10.041).

References

- Babtie, A.C., Stumpf, M.P.H., 2017. How to deal with parameters for whole-cell modelling. *J. Royal Soc. Interface* 14 (133), 20170237.
- Bennett, B., Kimball, E., Gao, M., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5 (8), 593–599.
- Berry, A., Marshall, K.E., 1993. Identification of zinc-binding ligands in the class II fructose-1,6-bisphosphate aldolase of *Escherichia coli*. *FEBS Lett.* 318 (1), 11–16.
- Birch, E.W., Udell, M., Covert, M.W., 2014. Incorporation of flexible objectives and time-linked simulation with flux balance analysis. *J. Theor. Biol.* 345, 12–21.
- Brown, G., Singer, A., Lunin, V.V., Proudfoot, M., Skarina, T., Flick, R., Kochinyan, S., Sanishvili, R., Joachimiak, A., Edwards, A.M., Savchenko, A., Yakunin, A.F., 2009. Structural and biochemical characterization of the type II fructose-1,6-bisphosphatase glpx from *Escherichia coli*. *J. Biol. Chem.* 284 (6), 3784–3792.
- Carrera, J., Covert, M.W., 2015. Why build whole-Cell models? *Trends Cell Biol.* 25 (12), 719–722.
- Chakrabarti, A., Miskovic, L., Soh, K.C., Hatzimanikatis, V., 2013. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8 (9), 1043–1057.
- Chowdhury, A., Zomorodi, A.R., Maranas, C.D., 2014. K-Optforce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* 10 (2), e1003487. doi:10.1371/journal.pcbi.1003487.
- Eyring, H., 1935. The activated complex in chemical reactions. *J Chem Phys* 3 (2), 107–115.
- Eyschen, J., Vitoux, B., Marraud, M., Cung, M.T., Branlant, G., 1999. Engineered glycolytic glyceraldehyde-3-phosphate dehydrogenase binds the *anti* conformation of NAD⁺nicotinamide but does not experience A-specific hydride transfer. *Arch. Biochem. Biophys.* 364 (2), 219–227.
- Fenton, A.W., Paricharttanakul, N.M., Reinhart, G.D., 2003. Identification of substrate contact residues important for the allosteric regulation of phosphofructokinase from *Escherichia coli*. *Biochemistry* 42 (21), 6453–6459.
- Frankel, D.G., 1996. Glycolysis. In: Neidhardt, F.C. (Ed.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. American Society for Microbiology Publishers, Washington, DC, pp. 189–198.
- Fröhlich, F., Kaltenbacher, B., Theis, F.J., Hasenauer, J., 2017. Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS Comput. Biol.* 13 (1), 1–18.
- Hao, J., Berry, A., 2004. A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. *Protein Eng. Des. Selection* 17 (9), 689–697.
- Hines, J.K., Fromm, H.J., Honzatko, R.B., 2007. Structures of activated fructose-1,6-bisphosphatase from *Escherichia coli*: coordinate regulation of bacterial metabolism and the conservation of the R-state. *J. Biol. Chem.* 282 (16), 11696–11704.
- Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmüller, U., Timmer, J., Theis, F.J., 2013. High-dimensional bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. *Math Biosci.*
- Iancu, C.V., Mukund, S., Fromm, H.J., Honzatko, R.B., 2005. R-State AMP complex reveals initial steps of the quaternary transition of fructose-1,6-bisphosphatase. *J. Biol. Chem.* 280 (20), 19737–19745.
- Jagiella, N., Rickert, D., Theis, F.J., Hasenauer, J., 2017. Parallelization and high-Performance computing enables automated statistical inference of multi-scale models. *Cell Syst.* 4 (2), 194–206.e9.
- Jahan, N., Maeda, K., Matsuoka, Y., Sugimoto, Y., Kurata, H., 2016. Development of an accurate kinetic model for the central carbon metabolism of *Escherichia coli*. *Microb. Cell Fact.* 15 (1).
- Karr, J.R., Phillips, N.C., Covert, M.W., 2014. Wholecellsimdb: a hybrid relational/HDF database for whole-cell model predictions. *Database J. Biol. Database Curation* 2014 (May), 1–8.
- Karr, J.R., Sanghvi, J.C., Macklin, D.N., Arora, A., Covert, M.W., 2013. Wholecellkb: model organism databases for comprehensive whole-cell models. *Nucleic Acids Res.* 41 (D1), 787–792.
- Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.L., Covert, M.W., 2012. A whole-cell computational model predicts phenotype from genotype. *Cell* 150 (2), 389–401.
- Kelley-Loughnan, N., Biolsi, S.A., Gibson, K.M., Lu, G., Hehir, M.J., Phelan, P., Kantrowitz, E.R., 2002. Purification, kinetic studies, and homology model of *Escherichia coli* fructose-1,6-bisphosphatase. *Biochimica et Biophysica Acta – Protein Structure and Molecular Enzymology* 1594 (1), 6–16.
- Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñoz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D.A., Weaver, D., Collado-Vides, J., Paulsen, I., Karp, P.D., 2017. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45 (D1), D543–D550.
- Khodayari, A., Zomorodi, A.R., Liao, J.C., Maranas, C.D., 2014. A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab. Eng.* 25, 50–62.
- Kruger, N.J., 1989. Effects of temperature on the kinetic properties of phosphofructokinase from *Escherichia coli*. *Biochem. Soc. Trans.* 17.
- Kurata, H., Sugimoto, Y., 2018. Improved kinetic model of *Escherichia coli* central carbon metabolism in batch and continuous cultures. *J. Biosci. Bioeng.* 125 (2), 251–257.
- Lee, R., Karr, J.R., Covert, M.W., 2013. Wholecellviz: data visualization for whole-cell models. *BMC Bioinf.*
- Liebermeister, W., Klipp, E., 2006. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Modell.* 3.
- Liepe, J., Holzhütter, H.G., Bellavista, E., Kloetzel, P.M., Stumpf, M.P., Mishtro, M., 2015. Quantitative time-resolved analysis reveals intricate, differential regulation of standard- and immuno-proteasomes. *Elife*.
- Macklin, D.N., Ruggero, N.A., Covert, M.W., 2014. The future of whole-cell modeling. *Curr. Opin. Biotech.* 28, 111–115. doi:10.1016/j.copbio.2014.01.012.
- Mainfroid, V., Goraj, K., Rentierdelrue, F., Houbrechts, A., Loiseau, A., Gohimont, A.C., Noble, M.E.M., Borchert, T.V., Wierenda, R.K., Martial, J.A., 1993. Replacing the ($\beta\alpha$)-unit 8 of *E. coli* TIM with its chicken homologue leads to a stable and active hybrid enzyme. *Protein Eng.* 6 (8), 893–900.
- Millard, P., Smallbone, K., Mendes, P., 2017. Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*. *PLoS Comput. Biol.* 13 (2).
- Moles, C.G., Mendes, P., Banga, J.R., 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13 (11), 2467–2474.
- Neidhardt, F.C., Ingraham, J.L., Schaechter, M., 1990. Physiology of the bacterial cell: a molecular approach, 20. Sinauer Associates, Inc., Sunderland, Massachusetts, pp. 3–4.
- Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? *Nat. Biotechnol.* 28 (3), 245–248.
- Penas, D.R., González, P., Egea, J.A., Doallo, R., Banga, J.R., 2017. Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy. *BMC Bioinf.* 18 (1), 1–24.
- Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., Schomburg, D., 2017. BRENDA In 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res.* 45 (D1), D380–D388.

- Plater, A.R., Zgiby, S.M., Thomson, G.J., Qamar, S., Wharton, C.W., Berry, A., 1999. Conserved residues in the mechanism of the *E. coli* class II FBP-aldolase. *J. Mol. Biol.* 285 (2), 843–855.
- Purcell, O., Jain, B., Karr, J.R., Covert, M.W., Lu, T.K., 2013. Towards a whole-cell modeling approach for synthetic biology. *Chaos* 23 (2).
- Sanghvi, J.C., Regot, S., Carrasco, S., Karr, J.R., Gutschow, M.V., Bolival, B., Covert, M.W., 2013. Accelerated discovery via a whole-cell model. *Nat. Methods* 10 (12), 1192–1195.
- Schaub, J., Reuss, M., 2008. In vivo dynamics of glycolysis in *Escherichia coli* shows need for growth-rate dependent metabolome analysis. *Biotechnol. Prog.* 24 (6), 1402–1407.
- Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., Heinemann, M., 2016. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* 34 (1), 104–110.
- Strogatz, S.H., 1994. *Nonlinear Dynamics and Chaos: With applications to physics, biology, chemistry, and engineering*. Westview Press, Boulder, Colorado, pp. 24–25.
- Taniguchi, Y., 2010. Quantifying *E. coli* proteome and transcriptome with single-Molecule sensitivity in single cells. *Science* 329 (July), 533–538.
- Toya, Y., Ishii, N., Nakahigashi, K., Hirasawa, T., Soga, T., Tomita, M., Shimizu, K., 2010. ¹³C-Metabolic Flux analysis for batch culture of *Escherichia coli* and its *pyk* and *pgi* gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnol. Prog.* 26 (4), 975–992.
- Wang, X., Kemp, R.G., 2001. Reaction path of phosphofructo-1-kinase is altered by mutagenesis and alternative substrates. *Biochemistry* 40 (13), 3938–3942.
- Wilks, J.C., Slonczewski, J.L., 2007. Ph of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J. Bacteriol.* 189 (15), 5601–5607.
- Wiśniewski, J.R., Rakus, D., 2014. Quantitative analysis of the *Escherichia coli* proteome. *Data Brief* 1, 7–11.
- Woldringh, C.L., De Jong, M.A., van den Berg, W., Koppes, L., 1977. Morphological analysis of the division cycle of two *Escherichia coli* substrains during slow growth. *J. Bacteriol.* 131 (1), 270–279.
- Zgiby, S.M., Thomson, G.J., Qamar, S., Berry, A., 2000. Exploring substrate binding and discrimination in fructose 1,6- bisphosphate and tagatose 1,6-bisphosphate aldolases. *Eur. J. Biochem.* 267 (6), 1858–1868.
- Zheng, R.L., Kemp, R.G., 1992. The mechanism of ATP inhibition of wild type and mutant phosphofructo-1-kinase from *Escherichia coli*. *J. Biol. Chem.* 267 (33), 23640–23645.