

Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/molecularbiosystems

New Insights into the Complex Regulation of the Glycolytic Pathway in *Lactococcus lactis*

I. Construction and diagnosis of a comprehensive dynamic model

Sepideh Dolatshahi, Luis L. Fonseca and Eberhard O. Voit*

Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA

* Corresponding author

E-mail Addresses: sepideh@gatech.edu, lfonseca@gatech.edu, eberhard.voit@bme.gatech.edu

Key Words: Biochemical Systems Theory (BST), Dynamic Flux Estimation (DFE), Glycolytic Pathway, *Lactococcus lactis*, Metabolic Pathway Analysis, Systems Biology

ABSTRACT

This article and the companion paper use computational systems modeling to decipher the complex coordination of regulatory signals controlling the glycolytic pathway in the dairy bacterium *Lactococcus lactis*. In this first article, the development of a comprehensive kinetic dynamic model is described. The model is based on *in vivo* NMR data that consist of concentration trends in key glycolytic metabolites and cofactors. The model structure and parameter values are identified with a customized optimization strategy that uses as its core the method of dynamic flux estimation. For the first time, a dynamic model with a single parameter set fits all available glycolytic time course data under anaerobic operation. The model captures observations that had not been addressed so far and suggests the existence of regulatory effects that had been observed in other species, but not in *L. lactis*. The companion paper uses this model to analyze details of the dynamic control of glycolysis under aerobic and anaerobic conditions.

1 Introduction

The results of most experiments in biology are snapshots that shed light on one or a few aspects of a phenomenon of interest. These snapshots are usually singular, sometimes comparative, contrasting two different conditions, or increasingly in the form of time series that represent cellular responses following a stimulus. The snapshots may be very complicated, for instance, by consisting of detailed images, capturing the expression levels of whole genomes, or revealing thousands of peaks in a proteomic or metabolomics mass spectrogram. Yet, each snapshot is separately frozen in time, and inferring a comprehensive cellular strategy from these data requires an additional and genuinely different cognitive process in the form of a conceptual or formal model. Particularly well suited for this purpose are dynamic models, which have the unparalleled capacity of permitting the weaving of multiple, often heterogeneous data into chains of events with which a cell or organism responds to a change in its milieu.

This article and its companion ¹ describe this process of connecting snapshots into cohesive storylines within the context of a complex task that the bacterium *Lactococcus lactis* must solve. Namely, the bacterium has to survive periods of starvation and at the same time be optimally positioned to take up substrate once it becomes available again, which is not a trivial matter, as we will discuss. Expressed differently, it must reach a “ready-to-respond” state while substrate is running out. The goal of these two papers is to decipher through dynamical modeling how the cells coordinate a complex control program to ensure this state. The first paper focuses on the construction of the model, while the second paper ¹ utilizes this model to shed light on the complex regulatory program, with which the organism controls glycolysis and ensures a viable ready-to-respond state.

The construction of a dynamic model of the type envisioned here has become possible with the recent availability of comprehensive time series data that characterize the dynamics of genomic, proteomic, metabolic, and physiological responses. Of particular interest for biochemical pathway modeling is the availability of relatively dense time profiles of metabolites through measurement techniques such as mass spectrometry (MS^{2,3}), nuclear magnetic resonance (NMR⁴), protein kinase phosphorylation, or mass cytometry (CyTOF⁵). These data contain valuable, but implicit, information about the structure and dynamics of the biological system under study and permit the use of top-down modeling approaches. These approaches consist of minimizing the residual error between the measured data, *i.e.*, the time profiles, and the assumed model, which typically consists of a system of nonlinear ordinary differential equations (ODE) that are to be parameterized⁶.

In an effort to build a kinetic-dynamic model using time-series data, the most challenging steps are the identification of suitable mathematical formats for all flux representations and the estimation of their unknown kinetic parameters. Central to these tasks is the methodology of dynamic flux estimation (DFE)⁷.

In contrast to previous models, which captured the dynamics of single datasets (*e.g.*,^{8,9}), this paper describes a single aggregate model that combines three datasets for different input glucose concentrations. This single model enables the assessment of several intriguing observations, which to date had not been explained convincingly.

A particular observation that has been puzzling for a long time is the following: if glucose is supplied to the medium in different amounts, one would expect a corresponding increase in the peak level of the downstream metabolite fructose-1,6-bisphosphate (FBP). However, the data clearly show that the peak concentration is largely independent of the external glucose concentrations. Specifically, comparing the different series of experimental results with increasing glucose concentrations, the FBP accumulation shows a progressively more noticeable plateau, whose duration, but not height, varies with substrate availability (Fig. 1). A previous model for aerobic conditions⁹ was able to fit glycolytic data for a single glucose concentration of 20 mM, but extrapolating the model toward different glucose inputs, such as 40 or 80 mM, predicted almost a doubling or quadrupling of the FBP peak height, which is in obvious contrast to the experimental measurements under anaerobic conditions. Here, we analyze the situation in detail and extract constraints that the model and its parameters must satisfy to yield a good fit to the data.

Throughout the analysis here and in the companion paper, such sometimes non-intuitive observations will be discussed, and the roles of numerous precisely timed regulatory mechanisms will be explained through controlled simulations with and without such mechanisms.

2 Data

The time series data on which this study is based were collected through *in vivo* nuclear magnetic resonance (NMR) spectroscopy¹⁰ from *L. lactis* cells that were initially starved and then offered a pulse of labeled glucose at time zero. These data contain extensive information about the structure, dynamics and regulation of the organism's metabolism, which in this method is essentially unadulterated by cell disruption, purification, centrifugation or other harsh

experimental methods. The technical aspects of the non-invasive determination of the concentrations of the intracellular pools of intermediate metabolites using *in vivo* NMR (^{13}C - and ^{31}P -NMR) experiments were thoroughly discussed in ¹⁰⁻¹². A brief summary of the data follows below.

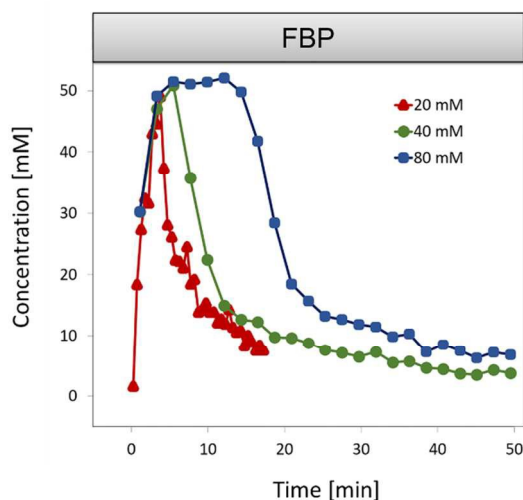


Figure 1: FBP data for three external glucose concentrations under anaerobic conditions. The peak level appears to be independent of the available substrate concentrations. One also notes that FBP does not vanish completely and instead maintains some residual concentration. Data summarized from ^{12, 13}

The measured time series include the external concentrations of glucose and of the end product lactate in the medium, along with several of the more abundant intermediate metabolites, namely glucose 6-phosphate (G6P), fructose 1,6-bisphosphate (FBP), phosphoenol pyruvate (PEP), 3-phosphoglycerate (3PGA). The time series reflect three experiments performed with 20, 40, or 80 mM of glucose input, respectively. Additional time series are available for some ubiquitous cofactors, such as NAD^+ and NADH , which are detectable with somewhat limited detection capability due to the required NMR acquisition time ¹². Finally, the level of NTP was measured using ^{31}P -NMR. The datasets used for modeling are summarized in Table 1.

Table 1- Overview of data used for modeling

Experiment	Condition	Technique	Metabolites Measured	Time Interval
1: 20mM [6- ¹³ C] glucose	Anaerobic pH = 6.5	¹³ C-NMR	glc, G6P, FBP, 3PGA, lac	30 sec
2: 40mM [1- ¹³ C] glucose and [5- ¹³ C] nicotinate	Anaerobic pH = 6.5	¹³ C-NMR	glc, FBP, 3PGA, PEP, lac, NAD ⁺ , NADH	2.2 min
		³¹ P-NMR	ATP, P _i	2.75 min
3: 80mM [1- ¹³ C] glucose and [5- ¹³ C] nicotinate	Anaerobic pH = 6.5	¹³ C-NMR	glc, FBP, 3PGA, PEP, lac, NAD ⁺ , NADH	2.2 min
		³¹ P-NMR	ATP, P _i	2.75 min

In Experiment 1, 20 mM of [6-¹³C] glucose were supplied to the cell suspension, and time series of concentrations were recorded for glucose, G6P, FBP, 3PGA, and lactate with a resolution of 30 seconds. In Experiments 2 and 3, cells were supplied beforehand with labeled [5-¹³C] nicotinic acid, a precursor of NADH, thus ensuring this pool to be 100% labeled¹². To start the *in vivo* NMR experiments, cells were supplied with 40 and 80 mM of [1-¹³C] glucose, respectively, and time series of glucose, FBP, PEP, 3PGA, lactate, NAD⁺ and NADH were recorded. The time resolution in these experiments was 2.2 minutes, which was needed to accommodate the determination of NAD⁺ and NADH with a reasonable signal-to-noise ratio. In a separate, comparable experiment, cellular metabolism was investigated with ³¹P-NMR, which allowed the measurement of NTP (mostly ATP), P_i and pH with a time resolution of 2.75 min. Usage of [1-¹³C] glucose prevents the determination of G6P in these datasets, due to the similarity in chemical shifts of [1-¹³C] glucose and [1-¹³C] G6P. However, some measurements of G6P are available from experiments with 20mM [6-¹³C] glucose. Although the data described

above exhibit experimental noise and are incomplete, with some metabolites or time points missing in each set, they are as good as a modeler can presently hope for. Duplicates are available for Experiments 1 and 2. These replicate datasets have a highly similar behavior to the modeled datasets, which shows that the data obtained with these NMR methods are highly reproducible. The additional datasets are not used for the modeling for the same reason of similarity and the fact that they contain time series data for fewer intermediate metabolites.

Under the given experimental conditions, cells are not able to synthesize ATP (nor ADP) *de novo*, because they are suspended in phosphate buffer and supplied only with glucose, which is insufficient for *L. lactis* to grow. Once the culture is harvested and washed prior to the NMR experiment, the total amount of ATP+ADP remains constant. This fact enables us to infer ADP from the ATP data where these are available (for Experiments 2 and 3), which is beneficial for the initial parameter estimation. By the same token, the cells do not synthesize proteins and they do not alter the expression of their genes.

In all datasets, the concentrations of 3PGA and PEP are highly covariant, due to the fact that these two metabolites can be converted into 2PGA by two enzymatic steps (phosphoglycerate mutase and enolase). These reactions are fast and reversible with equilibria that favor 3PGA and PEP¹⁴⁻¹⁶, thus maintaining 2PGA in a concentration range below the *in vivo* ¹³C-NMR detection level. Since this covariance is maintained even during high glycolytic flux, we decided to aggregate these two pools into one dependent variable (X_4). Nonetheless, we are still able to calculate the concentration of each intermediate from the constant proportionality of ~ 0.6444 (3PGA/PEP).

The data seem to indicate an apparent absence of PEP and FBP at the beginning of the experiment. However, these metabolites are present, but just not labeled initially and therefore missed by the NMR detection. The concentrations of these metabolites were measured in a control experiment ¹¹, supporting the reasonable assumption that the cells re-enter a state of starvation at the end of the experiment and that the residual values of now labeled PEP and FBP constitute a state that is similar to the state at the beginning of the experiment. This conjecture is further supported by a tandem experiment in which glucose was presented twice after periods of starvation ¹⁰.

3 Model Design

Generically, a mathematical model of a metabolic system consists of a system of ordinary differential equations with three components: (1) a stoichiometric matrix; (2) a vector of fluxes; and (3a) the functional forms of these fluxes and (3b) their corresponding parameter values. The stoichiometric matrix represents the essentially time-invariant wiring diagram of the pathway and shows which fluxes enter or leave each pool. It is often assumed to be known from biochemical experimentation, and the results of uncounted such studies are collected in databases like KEGG ¹⁷ and MetaCyc ¹⁸. Kinetic details and regulatory features are represented by parameters that are incorporated in appropriate functional forms representing the fluxes.

The structure of the pathway at hand is presented in Figure 2. A dynamic model for a similar type of experiments with 20 mM external glucose, but under aerobic conditions, was developed previously ⁹. We started out with this model attempting to adjust it toward anaerobic conditions by including additional regulatory mechanisms from the biochemical literature, accounting for ubiquitous metabolites including NAD⁺, NADH, ATP, and ADP as dependent variables, and

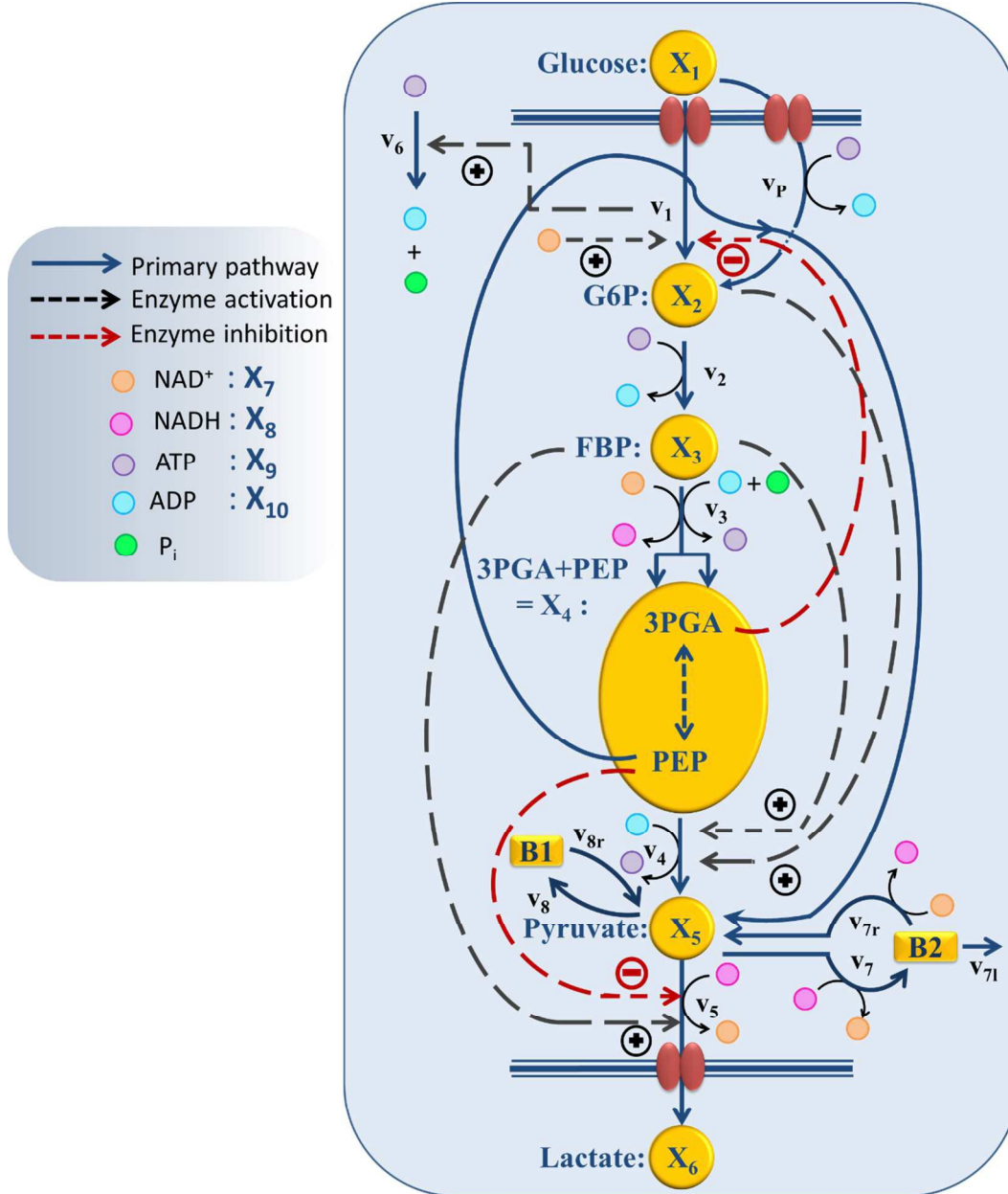


Figure 2: Model structure of the glycolytic pathway in *L. lactis*. Of particular importance are the PTS mechanism, which uses PEP for the initial phosphorylation of glucose, and several regulatory signals, indicated here with dashed arrows.

making other necessary adjustments to the pathway based on biological literature. Unfortunately, the final task of parameter estimation did not succeed when the three datasets were to be

captured simultaneously. Thorough analysis suggested a revisiting of the assumptions that were made originally regarding the pathway regulation and of the subtle, but genuine differences between operation under aerobic and anaerobic conditions. The amendments and alterations that became necessary will be explained in the following sections.

3.1 Dynamic Flux Estimation

The numerical characterization of an appropriate metabolic pathway model consists of the identification of mathematical formats for all process steps, and the estimation of optimal numerical parameter values in these process formulations. The determination of optimal representations for the processes in the model is by no means trivial, as no guidelines are available, but it is very important, because inadequate representations, even if they fit a target dataset, run the risk of error compensation among flux terms and of incurring problems during extrapolations^{7, 19, 20}.

We use for this identification step Dynamic Flux Estimation (DFE)⁷. In a nutshell, DFE works in the following manner. First, the time courses of all metabolites are smoothed, so that slopes can be estimated. These slopes are substituted for the derivatives on the left-hand sides of the ordinary differential equations (ODEs) at many time points, thereby replacing each ODE with a set of algebraic equations. At each time point, the algebraic equations can be solved by matrix inversion, with the help of pseudo-inverses, or with auxiliary methods²¹⁻²⁴. Combining these solutions yields graphical or numerical representations of all fluxes, either plotted against time or against the variables on which they depend. These representations can secondarily be fitted with

functions that might be appropriate, which directly permits an assessment of the appropriateness and quality of the chosen functions.

The main attraction of DFE is the fact that this method does not presuppose a functional form for any of the flux representations. This feature allows us to test in an objective manner whether particular functions, such as power-laws, Michaelis-Menten rate laws, or Hill functions, are capable of appropriately modeling a specific flux, or if other formulations should be considered. Importantly, careful analyses of all fluxes in this manner may suggest the existence of regulatory signals that had been missing from the assumed pathway structure. Such a suggestion corresponds to a novel hypothesis that is in principle testable with lab experiments and may lead to biological discoveries. An example is the glucose uptake process, which the companion paper¹ discusses in detail. In this case, slopes can directly be obtained through numerical differentiation of the uptake profile. The result of the first phase of DFE, which is independent of any model assumptions, reveals that this process must be regulated and suggests candidate regulators. These insights are used to refine the model construction.

In addition to its diagnostic capacities, DFE allows for a much more efficient parameter estimation strategy in terms of computation costs that are associated with the integration of ODEs and with global optimization. Namely, the parameters can be estimated one flux at a time. This decoupling is very advantageous, as one uses explicit functions, thereby avoiding—or at least reducing—the need to integrate ODEs numerous times, and because there is no risk of error compensation among terms or equations.

The first, model-free phase of DFE requires the estimation of slopes from the metabolic time course data. Several methods have been proposed for this purpose (*e.g.*, see discussion in²⁵; as a

further alternative, a smoothing algorithm was introduced in ²⁶ based on the wavelet transforms that ensures conservation of mass.

Not all processes in the model can be directly assessed with DFE because; for instance, some time courses are missing. Nonetheless, DFE provides a beneficial initial step that significantly constrains the subsequent parameter estimation.

Our generic strategy thus is the following. As far as feasible, we use DFE to identify candidates for possibly missing regulatory effects. Next, we computationally select reasonable functional forms in a process that is ideally guided by biology known from the literature. Then we use DFE to obtain starting points for as many parameters as possible. For the remaining parameters, we use multi-start and randomized search algorithms, implemented with generous search intervals, to make sure that the candidate parameter sets are not unduly restricted. We do not impose limits on the rate constants, except that they have to be non-negative, and allow the kinetic orders to vary within boundaries of [-3,3], which are considered in the field as rather wide. Once a working parameter set is obtained, we use evolutionary and randomized optimization methods, followed by local searches, to attain the best parameter set according to the criterion of minimizing the sum of squared errors between data and model simulations. Finally, a search in the vicinity of the optimal solution will indicate the sloppiness of the solution, and large-scale random searches outside the vicinity indicate whether the solution is more or less unique.

One should note that the model construction process is not always cut and dry. In particular, some of the insights obtained in the companion paper fed back into the estimation process, as they constrained the parameter space. More generally, the state of the art is at this point not advanced enough to permit a fully automated parameter estimation of complex models.

3.2 Specific Issues of Parameter Estimation in the Glycolytic Pathway

Once functional forms for all processes are determined through DFE and appropriate additional assumptions and settings, the model is to be fitted per optimization of parameter values. This optimization generically calls for minimizing a cost function, calculated as the sum of squared differences between the experimental data points and simulated values, summed over all time points and metabolites, and over all three data sets simultaneously. Although straightforward in principle, this optimization is almost always difficult ⁶.

In our specific case, several factors render the estimation of system parameters particularly challenging. First, the regulatory structure and reaction mechanisms of the model are *a priori* unknown, and their identification is by itself non-trivial, even if aided by DFE. Second, the error surface is embedded within a large parameter space of 44 dimensions. This surface is complicated, and one has to expect numerous local minima that severely confound steepest descent, evolutionary, or randomized optimization techniques that are typically used for parameter estimation. Third, datasets for some of the intermediate metabolites are not available, due to experimental limitations and, in particular, of the NMR-technique, which include higher detection limits for the various metabolites than would be desirable. Also, the ¹³C- and ³¹P-NMR experiments were executed separately and therefore both exclude some of the dependent variables. Missing data include time series for pyruvate for all three datasets, G6P for Datasets 2 and 3, and ATP, NAD⁺, and NADH for Dataset 1, as well as F6P, G3P, DHAP, 1,3GBP, 2PGA, and ADP for all datasets. Fourth, stiffness and other numerical issues associated with the model equations can lead to substantial algorithmic difficulties while integrating the differential equations. While these issues may not even be present in the ultimate, optimized model, they are

frequently encountered when an estimation algorithm determines inopportune combinations of settings during its scanning of a high-dimensional parameter space²⁷. Finally, even if a good fit has been determined, one cannot be sure of its uniqueness and must assume that the system might have a certain degree of sloppiness²⁸⁻³². In the present case, incorporating three datasets in conjunction with DFE can alleviate the identifiability issues³⁰ and help reveal regulatory information about the pathway.

3.3. Functional Formats of Fluxes: Biochemical Systems Theory

Biochemical systems theory (BST) is a mathematical and computational framework of ordinary differential equations (ODEs), which was originally developed for modeling and simulating biochemical pathways but has been widely applied to other biological systems ever since³³⁻³⁵. BST is considered “canonical,” which implies that the construction of the system of ODEs, its analysis, and its diagnosis follow relatively strict, well-structured guidelines. The power-law representation of each reaction is the key ingredient of BST. It constitutes a multi-variate, linear approximation in a logarithmic space and expresses a process as a product of power-law functions of all variables that directly affect the process^{35,36}. It has been shown that these power-law models are highly effective representations of biochemical kinetics^{34,37}. Power-laws offer the flexibility of non-integer kinetic orders, which enable a representation of situations commonly found in real biological systems. Additional support for the richness of power-laws in presenting complex nonlinear dynamic behavior comes from work showing that essentially any set of continuous nonlinear differential equations can be recast equivalently as a power-law system³⁸.

A generic reaction v_{ik} within the i^{th} equation of a BST model is represented as $v_{ik} = \alpha_{ik} \prod_{j=1}^n X_j^{h_{ikj}}$, $k = 1, 2, \dots, m$, where the rate constant α_{ik} and the kinetic orders h_{ikj} are fundamental characteristics. The rate constant describes the turn-over of the process, while the kinetic orders quantify the strengths with which reactants and regulators affect the process. These parameters need to be estimated to provide a full description of the metabolic pathway under study.

3.4. Details of the Mathematical Representation of the *L. lactis* Model

Our default for the functional forms of fluxes in the *L. lactis* model is a product of power-law functions, so that the model is mostly in the generalized mass action (GMA) format within the framework of BST. However, we allow for deviations from this format if they are suggested by the DFE analysis. The result is presented in Eqs. (1) and (2). In addition to the typical BST parameters, namely rate constants and kinetic orders, the equations contain the intracellular and extracellular volumes V_{in} and V_{out} . The extracellular volume of the bioreactor is 50 ml in all three experiments, while the intracellular volume is calculated from the measured biomass concentration, M_B , in each case, using a conversion value of 2.9 $\mu\text{L}/\text{mg}$ of protein for the intracellular volume³⁹. Thus, $V_{in} = 0,0029 V_{out} M_B$. The differential equations need to account for these volume differences, because the amounts of biomass differ among the experiments. B_1 and B_2 in Eqs. (2.h) and (2.i) are temporary buffers.

$$\dot{X}_1 = -\frac{v_1}{V_{out}} - \frac{v_P}{V_{out}} \quad (1.a)$$

$$\dot{X}_2 = \frac{v_1}{V_{in}} + \frac{v_P}{V_{in}} - \frac{v_2}{V_{in}} \quad (1.b)$$

$$\dot{X}_3 = \frac{v_2}{V_{in}} - \frac{v_3}{V_{in}} \quad (1.c)$$

$$\dot{X}_4 = 2 \frac{v_3}{V_{in}} - \frac{v_1}{V_{in}} - \frac{v_4}{V_{in}} \quad (1. d)$$

$$\dot{X}_5 = \frac{v_4}{V_{in}} + \frac{v_1}{V_{in}} - \frac{v_5}{V_{in}} - \frac{(v_7 - v_{7r})}{V_{in}} - \frac{(v_8 - v_{8r})}{V_{in}} \quad (1. e)$$

$$\dot{X}_6 = \frac{v_5}{V_{out}} \quad (1. f)$$

$$\dot{X}_7 = \frac{v_5}{V_{in}} - 2 \frac{v_3}{V_{in}} + \frac{(v_7 - v_{7r})}{V_{in}} \quad (1. g)$$

$$\dot{X}_8 = -\frac{v_5}{V_{in}} + 2 \frac{v_3}{V_{in}} - \frac{(v_7 - v_{7r})}{V_{in}} \quad (1. h)$$

$$\dot{X}_9 = 2 \frac{v_3}{V_{in}} - \frac{v_2}{V_{in}} + \frac{v_4}{V_{in}} - \frac{v_6}{V_{in}} - \frac{v_P}{V_{in}} \quad (1. i)$$

$$\dot{X}_{10} = -2 \frac{v_3}{V_{in}} + \frac{v_2}{V_{in}} - \frac{v_4}{V_{in}} + \frac{v_6}{V_{in}} + \frac{v_P}{V_{in}} \quad (1. j)$$

$$\dot{B}_1 = \frac{(v_7 - v_{7r})}{V_{in}} - \frac{v_{7l}}{V_{in}} \quad (1. k)$$

$$\dot{B}_2 = \frac{(v_8 - v_{8r})}{V_{in}} \quad (1. l)$$

While power-law functions appear to be adequate for most flux terms, they are in general not defined for variables with values of below zero, if they have non-integer exponents. While concentrations are obviously non-negative in reality, algorithmic integration methods for ODEs work in a step-wise manner, which can result in solution steps crossing zero into the negative domain. To avoid this numerical shortcoming, which in MATLAB returns complex numbers, it is beneficial to introduce automatic flags that set flux terms equal to zero just before a variable becomes zero. The model equations (2.a)-(2.i) below contain such flags in the PTS flux (v_1) and the permease flux (v_P), both of which remove glucose, which eventually runs out. In MATLAB, such flags are easily implemented with the inclusion of a term like ($X_1 > 10^{-4}$), which renders a

flux to become zero if glucose drops below the concentration of 10^{-4} mM. The model was numerically integrated with the stiff ODE solver ode15s in MATLAB. Specifically, the fluxes are defined as follows:

$$v_1 = \alpha_1 M_B X_1^{h_{1,1}} \left(\frac{X_4}{X_4 + K_{M4}} \right) (1 + X_4)^{h_{1,4}} X_7^{h_{1,7}} (1 - e^{-t/T}) (X_1 > 10^{-4}) \quad (\text{Eq. 2a})$$

$$v_P = \alpha_0 M_B X_1^{h_{0,1}} X_9^{h_{0,9}} (X_1 > 10^{-4}) \quad (\text{Eq. 2b})$$

$$v_2 = \alpha_2 M_B X_2^{h_{2,2}} X_9^{h_{2,9}} \quad (\text{Eq. 2c})$$

$$v_3 = \alpha_3 M_B X_3^{h_{3,3}} X_7^{h_{3,7}} X_{10}^{h_{3,10}} \quad (\text{Eq. 2d})$$

$$v_4 = \alpha_4 M_B X_4^{h_{4,4}} X_2^{h_{4,2}} X_{10}^{h_{4,10}} X_3^{h_{4,3}} \quad (\text{Eq. 2e})$$

$$v_5 = \alpha_5 M_B X_5^{h_{5,5}} X_8^{h_{5,8}} X_3^{h_{5,3}} (1 + X_4)^{h_{5,4}} \quad (\text{Eq. 2f})$$

$$v_6 = \alpha_6 M_B X_9^{h_{6,9}} v_1^{h_{6,1}} \quad (\text{Eq. 2g})$$

$$v_7 = \alpha_7 M_B X_5^{h_{7,5}} X_8^{h_{7,8}}, \quad v_{7r} = \alpha_{7r} M_B B_1^{h_{7r,1}} X_7^{h_{7r,7}}, \quad v_{7l} = \alpha_{7l} M_B B_1^{h_{7l,1}} \quad (\text{Eq. 2h})$$

$$v_8 = \alpha_8 M_B X_5^{h_{8,5}}, \quad v_{8r} = \alpha_{8r} M_B B_2^{h_{8r,2}} \quad (\text{Eq. 2i})$$

A detailed analysis during the second phase of DFE suggested that the PEP: Carbohydrate Phosphotransferase System (PTS; v_{PTS}) is a function not only of its substrates glucose and PEP, but that it also needs to be regulated by additional effectors. This analysis, which is described in detail in the companion paper ¹, leads to the specific conclusion that NAD^+ or the redox system

are potential activators, that 3PGA is an inhibitor, and that either one or both effectors are present. These regulatory terms are included in the model respectively as $X_7^{h_{1,7}}$ and $(1 + X_4)^{-h_{1,4}}$ in Eq. (2.a).

3.4.1. Modeling the First Two Minutes of Glucose Uptake

The initial rate of glucose uptake increases in the first two minutes, while the availability of substrate, *i.e.* glucose, decreases. This direct observation is structurally incompatible with any function whose rate monotonically increases with substrate availability, which is the case for typical Michaelis-Menten, Hill, and power-law functions. The reason is that the glucose concentration is the highest at the beginning, which would suggest the highest uptake rate in the first minutes. Galazzo and Bailey⁴⁰ suggested that G6P could be an inhibitor of this step in yeast. But even if this were true in *L. lactis*, the initial rise could not be explained, because the concentration of G6P is initially very small.

Several options are available to address the initial brief rise in PTS. First, one could attempt to identify the true mechanisms leading to the initial increase in glucose uptake, which corresponds to the sigmoidal shape of the glucose concentration curve. For instance, one could explain this observation with the fact that the experimental set-up causes slight delays, which could affect the uptake profile. It has also been argued⁹ that the cells might recover from starvation with some variability, which could be due to cell-to-cell variation in the speed of glucose uptake or slight differences in glucose availability to individual cells, or could be caused by the process of mixing of glucose throughout the medium or other extraneous factors. Indeed, it was shown with simulations that a narrowly distributed uptake profile can directly convert the monotonic trend in glucose consumption, as predicted by the model structure, into a sigmoidal trend, as it is

observed⁹. As an alternative, one could try to identify some fitting function without being constrained by mechanistic considerations⁴¹. One could furthermore use the data directly as (“off-line”) inputs instead of representing them functionally in the model⁴². Finally, one could imagine that insufficient amounts of FBP accumulate during the first two minutes, thereby limiting the material flux into the pool of PEP and 3PGA. As a consequence, the initial PEP concentration might not be able to produce the high level of v_1 that we directly computed through numerical differentiation of the glucose consumption profile. This short-term discrepancy can be resolved when we account for a small auxiliary permease flux (v_p), which has been observed in this strain of *L. lactis*⁴³, but was never used in earlier models.

Given the variety of speculations, we decided on a hybrid option, where we use a black-box adjustment function for the first two minutes and dynamically model glucose uptake afterwards, starting at $t = 2$ [min]. Specifically, we multiply a term of the form $\left(1 - e^{-\frac{t}{T}}\right)$ to the modeled PTS flux. This term, which is borrowed from physics and engineering⁴⁴, where it typically represents a delayed response of a linear time-invariant system to a constant input starting at time zero, does not convey biological meaning. It is simply useful since it exponentially approaches 1 and loses its effect after a short period of time. For instance, at $3T$ minutes, it is equal to $1 - e^{-3} = 0.9502$. Thus, the term is ineffective for most of the experimental period. Different time constants T were allowed for different experiments.

3.5 Summary of the Model Design Procedure

In broad strokes, the modeling strategy employed here consisted of the following main steps: (1) Initially, the first phase of DFE was used, which resulted in the identification of a subset of fluxes. However, because the stoichiometric matrix is not full rank and some of the

measurements were missing, some fluxes could not be identified in this step. (2) For the identifiable subset of fluxes, diagnostic methods were utilized in the second phase of DFE to ensure the validity of power-law representation as well as the reasonableness of each of the hypothesized underlying regulatory mechanisms. This step provided starting points for parameters that were used in the later optimization algorithms. (3) Loose boundaries on parameter values were imposed and additional constraints were taken into account including those resulting from the detailed analysis of the FBP dynamics (see Section 5.2). (4) The information from Steps 2 and 3 was used as input to a suite of several global and local optimization algorithms, which ultimately led to one parameter set that minimizes the error between the experimental measurements and the simulation results. Finally, Monte-Carlo methods and randomization schemes were utilized to search the parameter space as thoroughly as possible to ensure the optimality of the parameter set and to characterize a compact ensemble of close-by, good solutions. While these strategies reduced the chance of other, distinctly different solutions, they do not provide a mathematical proof.

5 Results

5.1 General Features of the Model

Based on literature information, we established an initial model diagram (Fig. 2), determined fluxes with methods of DFE, and subsequently performed parameter estimation. The fully parameterized model was diagnosed with standard methods of stability, sensitivity, and robustness analysis, and subsequently used for representative simulations.

The analysis led to several slight amendments of the original diagram. First, we observed that the actually measured total carbon mass decreases over time for all three datasets, with about 7% - 10% of the mass being unaccounted (see *Fig. 6A* in a later section), in spite of the fact that, outside glycolysis, the organism is metabolically inactive under the given conditions¹⁰. The loss is biologically not very significant, but it is immediately inconsistent with the model structure of the initial diagram, where glucose is the only input substrate and lactate is the only output. The most reasonable option for remedying this inconsistency is the addition of a minor efflux out of one or more metabolite pools. A good candidate is pyruvate, because several reactions may use pyruvate as a substrate and thereby be responsible for the diversion of material⁴⁵. Because some of these effluxes potentially affect the NAD^+/NADH balance, we included two types of leakage from pyruvate, one with and one without the consumption of NADH. Optimization determined the very low capacities of these fluxes. We also considered alternative locations of leakage, such as PEP and G6P, but did not find them beneficial under the given experimental conditions. Other amendments are discussed in the following sections.

The model contains twelve dependent variables. Six represent the main metabolites glucose, G6P, FBP, the aggregated pool of 3PGA and PEP, pyruvate, and lactate, four variables represent the cofactors ATP, ADP, NAD^+ , and NADH, and the remaining two represent temporary buffers⁴⁶. The buffers were used for the following reason. When we studied both carbon flow and electron flow, we observed a temporal mismatch that required some NADH to be consumed but later returned to the pathway. We rationalized this observation with the possibility that one or more pools must exist in connection with some glycolytic intermediates that allow NADH to be consumed and later returned. These reactions could include processes such as the mannitol-1P dehydrogenase or glycerol-3P dehydrogenase steps. The fact that there are several possible pools

might explain why no other metabolite was actually measured by the NMR, because if the missing mass is distributed among several pools, none would have a concentration high enough to be detectable by the NMR. We therefore allowed for two buffers to exist in the model, one out of G6P and one out of pyruvate. No parameter set was ever obtained, by optimization, which used the buffer out of G6P, which suggested removal of this buffer. The best data consistency was always obtained with the two buffers presented, one NADH dependent and one NADH independent. These allowed the model to accommodate the carbon-electron temporal mismatch.

Eq. 1 of the previous section presents ordinary differential equations describing the system connectivity, while Eq. 2 shows the functional representations. The model is more complicated than earlier models of glycolysis in *L. lactis*, as it addresses the pathway dynamics under anaerobic conditions. In contrast to aerobic conditions, the organism cannot easily recycle NAD^+ under anaerobic conditions, and the balance between NAD^+ and NADH therefore changes dynamically. These changes must be expected to affect the dynamics of glycolysis and are therefore considered important for the model.

5.2 Temporarily Steady FBP Peak

An intriguing observation among the three experiments is that FBP reaches more or less the same peak level as long as a certain minimal level of glucose is available to the system (Fig. 1). This observation has been puzzling for a long time but can now be explained with a detailed analysis of the model structure and the quantitative features of its fluxes. The mathematical and biological details of the explanation are discussed below. In a nutshell, the shapes of the peaks are driven by saturation of the PTS flux (v_1), while the subsequent production flux of FBP (v_3) operates below saturation and while the FBP concentration results in a v_3 flux equal to the

saturated PTS flux v_1 . The proposed relationship between v_1 and v_3 yields strong mathematical constraints that restrict the admissible parameter space. In particular, the analysis reveals that glucose must have a low kinetic order in the PTS flux. Interestingly, similar considerations of relationships among kinetic Michaelis-Menten parameters have recently helped explain ultrasensitivity in cascaded signaling systems⁴⁷.

5.2.1. Analysis of FBP Dynamics

To rationalize the counterintuitive peak dynamics of FBP, let us at first consider a slightly simplified scenario where G6P is omitted as an explicit intermediate between glucose (X_1) and FBP (X_2) and where the PTS flux v_1 solely depends on X_1 and v_2 on X_2 . The dependence on NADH is not of importance here. This simplified diagram is shown in Figure 3A.

Figure 3B visualizes the dynamics of fluxes of accumulation (v_1) and consumption (v_2) of FBP relative to one another for the representative case of Experiment 3 with 80 mM of initial glucose. The horizontal top and bottom axes for the two curves are different and color-coded with blue (glucose) and black (FBP). In the first phase of the experiment, glucose is abundant and FBP accumulates (shown with a curved green upward arrow) while glucose is being consumed at a more or less constant rate (shown with the horizontal green arrow indicating that glucose is consumed at the same time as FBP accumulates). The intersection of the curves represents a temporarily steady state for FBP with a concentration of about 50 mM. At this state, accumulation and consumption of FBP are in balance: $v_1 = v_2$, while other processes in the system are not. As the experiment proceeds, FBP remains constant, whereas glucose is being consumed from about 50 mM down to about 10 mM (color-coded with straight orange arrow). Since the value of v_1 is essentially constant, FBP remains at its temporary plateau during this

period. As soon as the glucose concentration drops below about 10 mM, FBP becomes depleted while glucose is being used up as the two red arrows indicate. Panel C shows the same three phases on a concentration vs. time plot of glucose and FBP for further clarification.

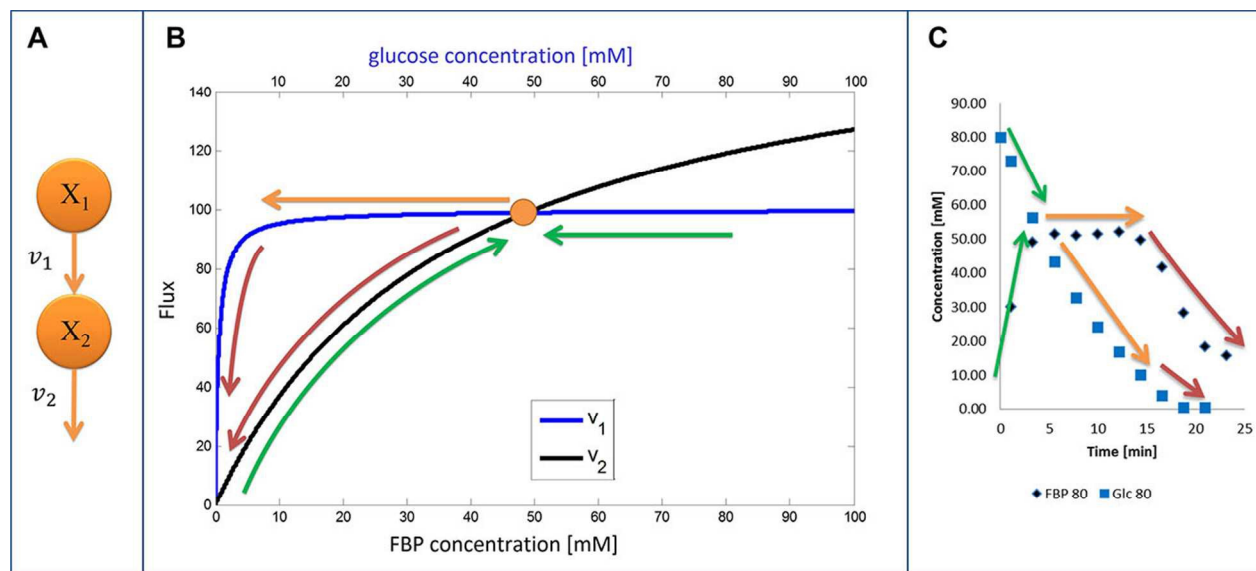


Figure 3: (A) Simplified diagram of glucose (X_1) conversion into FBP (X_2). Panel (B) depicts the relative shapes of v_1 and v_2 as functions of their substrate concentrations. Colored arrows show different phases of the dynamic behavior of FBP. Green: FBP accumulation and decrease in glucose (first 5 minutes); Orange: FBP constancy at peak level and further decrease in glucose (~15 minutes for 80 mM glucose); Red: Glucose and FBP depletion. Panel (C) shows the same phases on a concentration vs. time plot. The color coding is consistent between B and C.

The concepts outlined above were implemented in a simple model. We start by representing the flux terms with Michaelis-Menten rate laws. In order to obtain the same peak for different amounts of glucose, v_1 needs to be saturated and equal to V_{max1} for glucose with a value between about 10 to 80 mM for Experiment 1 and between about 5 to 40 mM for Experiment 2. These numerical settings permit the temporarily steady state and extended peak for FBP if v_2 has a high V_{max2} and a low K_{m2} . In fact, these values must be such that for the peak FBP concentration (about 50 mM), v_2 equals v_1 for the appropriate glucose concentrations.

These quantitative considerations can be converted into constraints for the model parameters. Specifically, the equation $v_2(\text{FBP} = 50) = v_1 (5 \leq \text{glucose} \leq 50)$ mandates the following: First, K_{m1} must be such that v_1 is saturated for high glucose values (above about 5mM); this ensures a similar FBP peak for all pertinent glucose concentrations in the three experiments. Second, the FBP accumulation and consumption fluxes v_1 and v_2 need to intersect. This requires $K_{m2} > K_{m1}$ and $V_{max2} > V_{max1}$, and the values must be set such that $v_1 = v_2$ when FBP is at the observed plateau of about 50 mM. A solution to these constraints is:

$$V_{max1} = \frac{V_{max2} \cdot (\text{FBP peak conc.})}{K_{m2} + (\text{FBP peak conc.})}$$

An additional constraint pertaining to V_{max1} can be derived from the speed with which glucose is depleted. Taken together, we obtain the following set of conditions:

$$\begin{aligned} K_{m2} &> K_{m1} \\ V_{max2} &> V_{max1} \\ V_{max1} &= \frac{V_{max2} \cdot (\text{FBP peak conc.})}{K_{m2} + (\text{FBP peak conc.})} \\ V_{max1} &\approx 4 \end{aligned}$$

Figure 4A shows the glucose and FBP dynamics resulting from a representative example for which the inferred sets of constraints hold. The concentration curves are very similar to the measurement data (Fig 1).

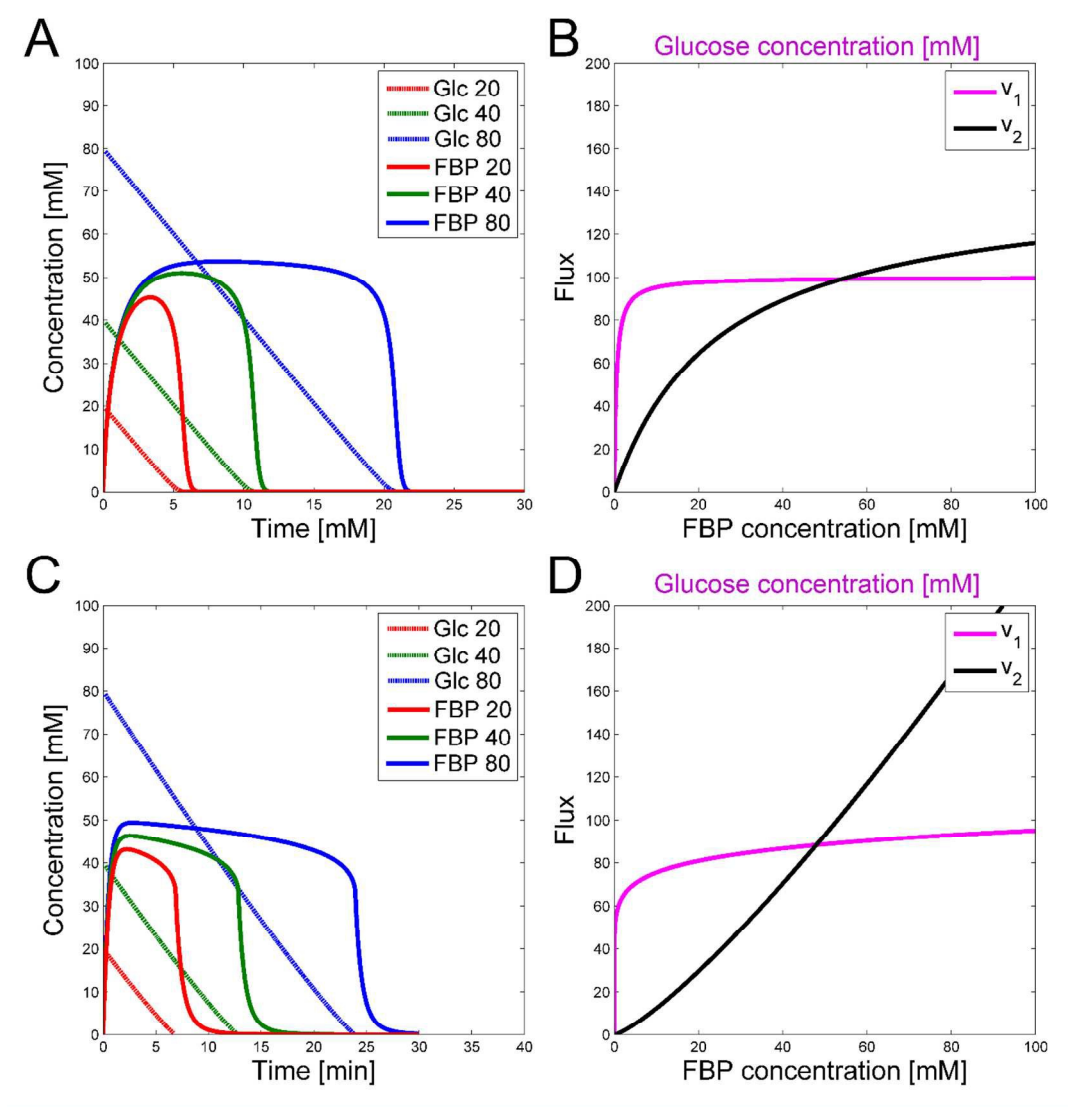


Figure 4: (A) Computed concentrations of glucose (more or less linear) and FBP (humped) for 20 (red), 40 (green), and 80 mM (blue) of initial glucose, using the constraints on parameters in a Michaelis-Menten formulation with representative values of $V_{max1} = 4$, $V_{max2} = 145$, $K_{m1} = 0.5$, $K_{m2} = 25$ and an external to internal volume ratio of 25. (B) Corresponding glucose and FBP trends vs. time for the Michaelis-Menten formulation in (A). (C) Plots of flux vs. concentration for the v_1 and v_2 fluxes in power-law format. v_1 has a low kinetic order. (D) Corresponding glucose and FBP trends vs. time for the power-law formula in (C). Experiment 3 with the highest amount of glucose input results in the most extended FBP peak.

The relationships between fluxes are not dependent on the choice of the Michaelis-Menten framework. To reproduce similar conditions for power-law representations, one must simply

require a low kinetic order and rate constant for glucose that results in the correct dynamics for v_1 . The result is a similar flux value for different glucose concentrations between 5 and 50 mM. Furthermore, v_2 needs to have a higher kinetic order so that the two flux curves intersect. Rate constants are calculated such that $v_1 = v_2$ for the peak FBP concentration of about 50 mM. Similar to the Michaelis-Menten case, the kinetic order and rate constant for v_1 need to be set such that they satisfy the speed for glucose depletion. Results for a representative, feasible parameter set with power-law functions are illustrated with the fluxes in Figure 4C. Figure 4D shows the corresponding glucose and FBP concentrations for the three experiments with different initial glucose concentrations.

The question arises of whether the above considerations are affected by the existence of G6P as an intermediate between glucose and FBP. The short answer is *no*. Unfortunately, the [1-¹³C] NMR experiments did not permit measurements of G6P for the three experiments. However, it is likely that the G6P dynamics is similar to the FBP dynamics, although with a lower peak value or a very low saturation threshold as it is observed for Experiment 1 with initial glucose of 20 mM.

Supposing that v_3 is a function of FBP only, we find: (i) at times when FBP is constant, v_3 is constant as well; and (ii) FBP being constant requires that $v_2 = v_3$. (i) and (ii) are possible in the following two scenarios only when G6P and FBP are constant simultaneously, or when v_2 is saturated for the amount of G6P during that time period, and therefore has a very low K_m . Furthermore, no matter whether G6P is constant or v_2 has a low K_m (high affinity) for G6P, $v_1 = v_2 = v_3$ must hold at the FBP peak, and v_1 must thus be saturated for glucose values

between about 5 and 50 mM as reasoned before. Thus, the earlier conclusions regarding the FBP dynamics hold, whether or not G6P is explicitly presented.

5.2 Fully Parameterized Model and Model Fits

Methods of DFE, combined with numerous parameter estimation techniques, led to the following set of parameter values, which lead to simultaneous fits for the three available experiments. They are displayed below with one row per flux:

$$\begin{aligned}
 P = & [\alpha_1 \quad h_{1,1} \quad h_{1,4} \quad K_{M4} \quad h_{1,7} \quad \dots \\
 & \alpha_2 \quad h_{2,2} \quad h_{2,9} \quad \dots \\
 & \alpha_3 \quad h_{3,3} \quad h_{3,7} \quad h_{3,10} \quad \dots \\
 & \alpha_4 \quad h_{4,4} \quad h_{4,2} \quad h_{4,10} \quad h_{4,3} \quad \dots \\
 & \alpha_5 \quad h_{5,5} \quad h_{5,4} \quad h_{5,8} \quad h_{5,3} \quad \dots \\
 & \alpha_6 \quad h_{6,9} \quad h_{6,1} \quad \dots \\
 & \alpha_0 \quad h_{0,1} \quad h_{0,9} \quad \dots \\
 & \alpha_7 \quad h_{7,5} \quad h_{7,8} \quad \dots \\
 & \alpha_{7r} \quad h_{7r,2} \quad h_{7r,7} \quad \dots \\
 & \alpha_{7l} \quad h_{7l,2} \quad \dots \\
 & \alpha_8 \quad h_{8,9} \quad \dots \\
 & \alpha_{8r} \quad h_{8r,1} \quad]
 \end{aligned}$$

Eq. 3

$$\begin{aligned}
 P = & [12.5290 \quad 0.1422 \quad -0.2547 \quad 0.2250 \quad 0.6469 \quad \dots \\
 & 1.1208 \quad 1.9844 \quad 0.2742 \quad \dots \\
 & 0.2403 \quad 1.0079 \quad 0.7114 \quad 0.0603 \quad \dots \\
 & 0.1600 \quad 0.6192 \quad 0.8661 \quad 0.3715 \quad 0.7479 \quad \dots \\
 & 0.6522 \quad 3.0000 \quad -0.5013 \quad 0.4867 \quad 0.8638 \quad \dots \\
 & 0.3432 \quad 1.2558 \quad 0.4101 \quad \dots \\
 & 0.1228 \quad 0.1720 \quad 0.0252 \quad \dots \\
 & 0.3293 \quad 1.3664 \quad 1.4756 \quad \dots \\
 & 0.1504 \quad 0.3026 \quad 0.1657 \quad \dots \\
 & 0.0061 \quad 0.0764 \quad \dots \\
 & 1.0528 \quad 0.4075 \quad \dots \\
 & 0.1950 \quad 0.5500 \quad];
 \end{aligned}$$

Eq. 4

In addition, the model uses the following initial values for the dependent variables and settings:

$$X_{0_{80}} = [80; 0.1; 4; 14.8; 0.1; 0.1; 5.74; 0.1; 0.1; 8.815; 0.1; 0.1];$$

$$X_{0_{40}} = [40; 0.1; 4; 14.8; 0.1; 0.1; 5.74; 0.1; 0.1; 8.815; 0.1; 0.1];$$

$$X_{0_{20}} = [20; 0.1; 4; 14.8; 0.1; 0.1; 5.74; 0.1; 0.1; 8.815; 0.1; 0.1];$$

$$M_{B-20} = 13.92 \text{ mg protein/ml}$$

$$M_{B-40} = 17.11 \text{ mg protein/ml}$$

$$M_{B-80} = 19.53 \text{ mg protein/ml}$$

$$T_{20} = 0.7124 \text{ min}$$

$$T_{40} = 0.7450 \text{ min}$$

$$T_{80} = 1.6661 \text{ min}$$

Note that all initial conditions are identical, except for the external glucose concentration, the measured biomass concentrations in the three experiments (M_B), and the time constants (T) capturing delayed glucose uptake in the beginning of the experiments.

These parameter settings lead to good fits of the data, which are displayed in Panels A-C of Fig. 5. The capacity of a single parameter set to capture different conditions is important, because it significantly increases the predictive power of the model. Furthermore, since this parameter set was derived from DFE, its extrapolation reliability is increased, because the risk of error compensation among flux terms within the same or in different equations is greatly reduced ⁷.

While the simultaneous fitting with one parameter set is important, one should consider that, in reality, natural systems obviously exhibit a certain degree of variability. We therefore allowed the parameters in the common set, which fits the three experiments simultaneously, to vary slightly in order to account for this variability among the different cell populations. The results are shown in Panels D-F of Figure 5. These model instantiations used the following parameter sets for the 20, 40, and 80 mM experiment. A parameter-by-parameter comparison demonstrates how close the three sets are.

$$\begin{aligned}
 P_{20} = & [14.3646 \quad 0.0722 \quad -0.0858 \quad 0.1703 \quad 0.4805 \quad \dots \\
 & 1.0071 \quad 1.4924 \quad 0.1773 \quad \dots \\
 & 0.2032 \quad 1.0207 \quad 1.0117 \quad 0.1416 \quad \dots \\
 & 0.1888 \quad 0.6645 \quad 0.7712 \quad 0.6590 \quad 0.7766 \quad \dots \\
 & 0.5633 \quad 3.0000 \quad -0.6826 \quad 0.4946 \quad 1.0296 \quad \dots \\
 & 0.3627 \quad 1.1649 \quad 0.3731 \quad \dots \\
 & 0.0335 \quad 0.7391 \quad 0.1038 \quad \dots \\
 & 0.3761 \quad 1.9952 \quad 0.9710 \quad \dots \\
 & 0.1119 \quad 0.5980 \quad 0.7281 \quad \dots \\
 & 0.0282 \quad 0.1696 \quad \dots \\
 & 1.1080 \quad 0.5652 \quad \dots \\
 & 0.1998 \quad 0.6500]; , T_{20} = 0.7391 \text{ min}
 \end{aligned}$$

Eq. 5

$$\begin{aligned} P_{40} = & [11.6912 \quad 0.1790 \quad -0.3313 \quad 0.3439 \quad 0.4664 \quad \dots \\ & 0.9740 \quad 1.7155 \quad 0.4049 \quad \dots \\ & 0.2059 \quad 0.9700 \quad 0.8164 \quad 0.1410 \quad \dots \\ & 0.1830 \quad 0.4572 \quad 0.7985 \quad 0.4100 \quad 0.9091 \quad \dots \\ & 0.5838 \quad 2.9162 \quad -0.6379 \quad 0.3987 \quad 0.8822 \quad \dots \\ & 0.3335 \quad 1.1774 \quad 0.3556 \quad \dots \\ & 0.0231 \quad 0.8343 \quad 0.1026 \quad \dots \\ & 0.3624 \quad 2.0246 \quad 1.2299 \quad \dots \\ & 0.0702 \quad 0.5273 \quad 0.5884 \quad \dots \\ & 0.0238 \quad 0.2526 \quad \dots \\ & 1.0248 \quad 0.2501 \quad \dots \\ & 0.2021 \quad 0.6321]; \quad , T_{40} = 0.5990 \text{ min} \end{aligned}$$

Eq. 6

$$\begin{aligned} P_{80} = & [12.4598 \quad 0.1422 \quad -0.2547 \quad 0.2250 \quad 0.7444 \quad \dots \\ & 1.2357 \quad 1.9157 \quad 0.2879 \quad \dots \\ & 0.2403 \quad 1.0079 \quad 0.7090 \quad 0.0603 \quad \dots \\ & 0.1600 \quad 0.6110 \quad 0.8661 \quad 0.3715 \quad 0.7479 \quad \dots \\ & 0.6522 \quad 2.9076 \quad -0.5013 \quad 0.4635 \quad 0.8638 \quad \dots \\ & 0.3432 \quad 1.2558 \quad 0.4101 \quad \dots \\ & 0.2000 \quad 0.1720 \quad 0.0252 \quad \dots \\ & 0.3293 \quad 1.4836 \quad 1.4756 \quad \dots \\ & 0.1475 \quad 0.3104 \quad 0.1740 \quad \dots \\ & 0.0061 \quad 0.0764 \quad \dots \\ & 1.0528 \quad 0.4717 \quad \dots \\ & 0.1950 \quad 0.5500]]; \quad , T_{80} = 1.5985 \text{ min} \end{aligned}$$

Eq. 7

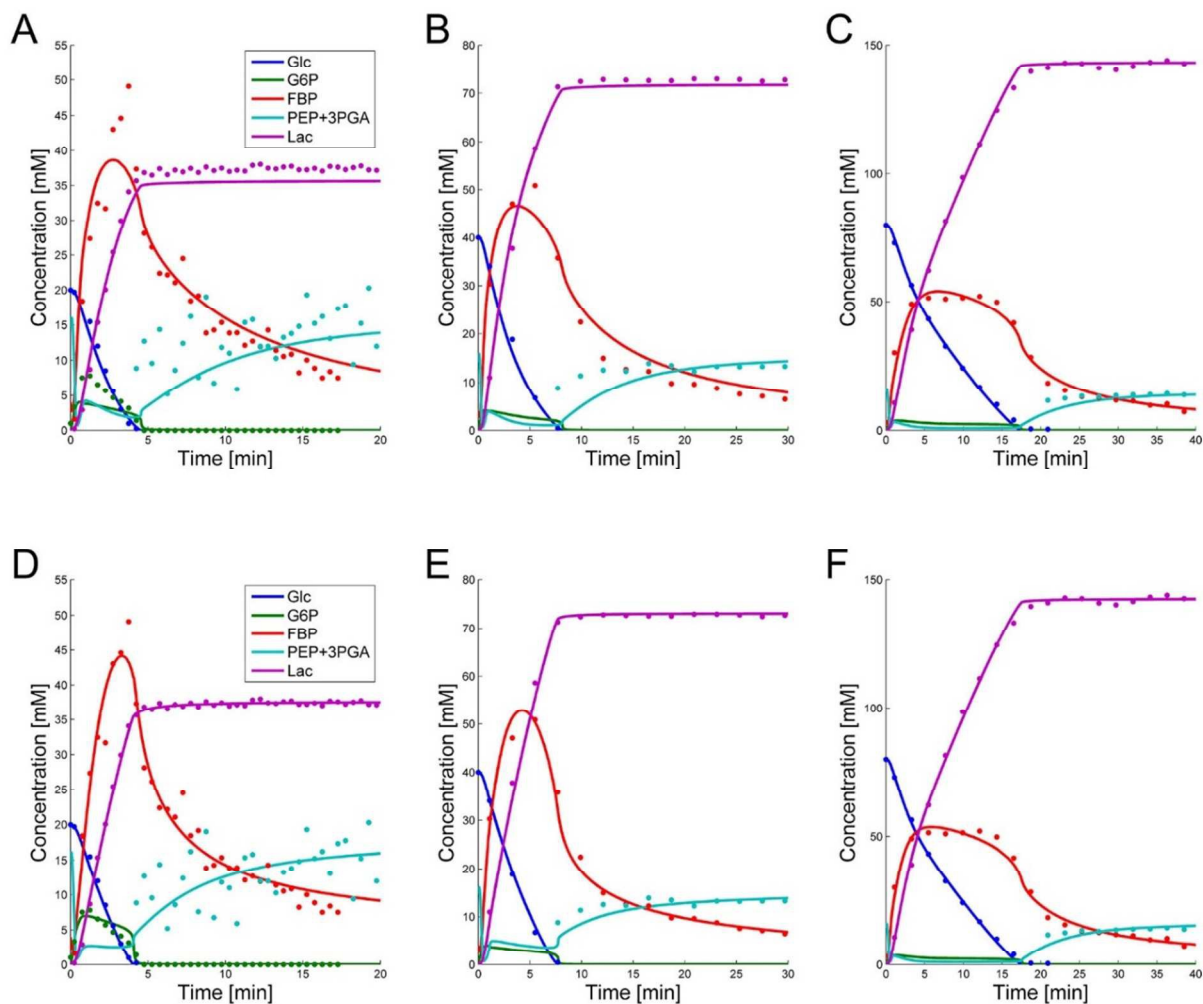


Figure 5: Simulation results for glucose, lactate, G6P, FBP and PEP+3PGA, superimposed on the corresponding data, for Experiment 1 (Panel A), Experiment 2 (Panel B), and Experiment 3 (Panel C). Note the different Y-scales. Accounting for modest variability among cell populations, the parameters in the common set of Panels A-C were allowed to vary slightly among experiments. The resulting fits are depicted in Panels D, E, and F.

5.3 Simulation Results for Secondary Metabolites

NAD^+ , NADH , and ATP data are available only for Experiments 2 and 3. The simulation results for these metabolites as well as pyruvate are shown in Figures 6B and 6C. Although the fits are not as good as those for the main metabolites, they capture the trends and timing. Three aspects

are of note. First, these secondary metabolites all have concentrations below 10mM and are often close to the detection limit. Second, the ATP measurements had to be performed in a separate experiment, as phosphate was labeled, and while an attempt was made to match the experimental conditions as closely as possible to the 40 and 80 mM experiments, the data did not come from the same experiments. Finally, NADH was not observed until the time point when glucose was depleted in these experiments. With these limitations in mind, emphasis with respect to ATP was given to the general trend of an increase to high concentrations in the beginning of the experiment, remaining high for the duration of experiment until glucose was depleted, then being depleted. For NADH, it was important to remain at its high concentration for the duration of time after the glucose depletion. Furthermore, no pyruvate was detected in the experiments. Consistently, the simulation results are below the detection limit of each experiment, except for the first few minutes where unlabeled PEP and 3PGA are converted into unlabeled pyruvate that remains undetected in NMR experiments. The simulation results show this spike in the beginning; shortly afterwards, pyruvate decreases to below 2 mM for all experiments and stays below the detection limit until it approaches zero toward the end of the experiment.

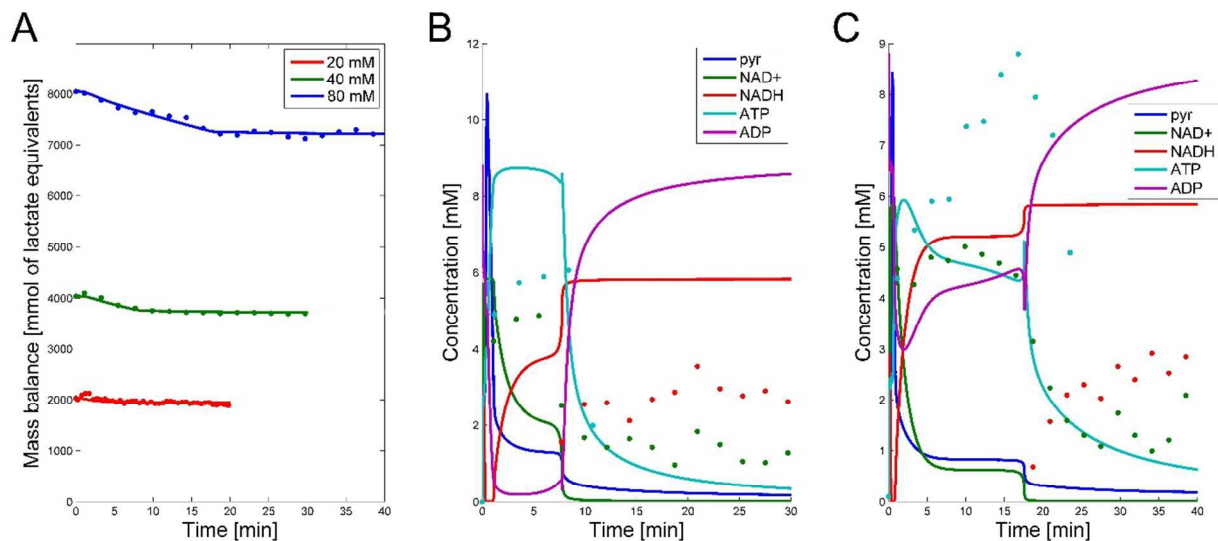


Figure 6: (A) Mass balance in mmol of lactate equivalents vs. time, calculated by taking into account the appropriate stoichiometry and volume conversions. The mass is plotted for the three datasets. The dots show the total carbon mass calculated from the measured data, and the lines show the calculated amounts resulting from the simulated model with leakage terms. (B) Data of NAD⁺, NADH and ATP are shown as dots. Simulation results for NAD⁺, NADH and ATP are superimposed. No data are available for pyruvate; simulation results are shown in dark blue.

7 Conclusions and Future Directions

A dynamic mathematical model was devised that accounts for the key glycolytic metabolites in *L. lactis*, as well as the dynamics of such cofactors as NAD⁺, NADH, ATP, and ADP. As with many modeling studies, the most difficult step of model development was the estimation of parameter values. Here, this estimation was based on experimental time series of glycolytic intermediates from three experiments with different substrate availability. The technical difficulties of the estimation process were directly related to the high dimensionality of the parameter space, the enormous complexity of the landscape of residual errors between model and data, and the often ignored fact that we do not really know what functions are best suited to represent each process within a biological system. To address these issues, a combination of mathematical and computational techniques were developed, including a custom-tailored Monte-

Carlo algorithm, various optimization techniques and, most importantly, methods of dynamic flux estimation (DFE), which enabled us to reduce the admissible parameter space and prevent flux terms from compensating errors in their representations.

For the first time, a single model fits all available metabolic time courses reasonably well with the same parameter set and also captures the counterintuitive FBP dynamics across the three experiments. Because this model reflects three independent datasets, one might expect that it has a higher extrapolation potential than earlier models that were based on single datasets. The companion paper ¹ discusses key aspects of the control of the pathway, and explains in detail how the organism terminates glycolysis at the end of glucose availability such that it is ready to take up now glucose as soon as substrate becomes available again.

Acknowledgments

The authors are grateful to Dr. Helena Santos and Dr. Ana Rute Neves of the Instituto de Tecnologia Química e Biológica (ITQB) at the New University of Lisbon (Portugal) for allowing them to use the *in vivo* NMR data presented here. The NMR spectrometers are part of The National NMR Facility (RECI/BBB-BQB/0230/2012), supported by Fundação para a Ciência e a Tecnologia. This work was funded in part by the National Science Foundation (Projects MCB-0958172 and MCB-0946595; PI: EO). The funding agency is not responsible for the content of this article.

References

1. S. Dolatshahi, L. L. Fonseca and E. O. Voit, *Molecular BioSystems*, 2015.
2. A. El-Aneed, A. Cohen and J. Banoub, *Applied Spectroscopy Reviews*, 2009, **44**, 210-230.
3. S. Li, Y. Park, S. Duraisingham, F. H. Strobel, N. Khan, Q. A. Soltow, D. P. Jones and B. Pulendran, *PLoS computational biology*, 2013, **9**, e1003123.
4. A. R. Neves, W. A. Pool, J. Kok, O. P. Kuipers and H. Santos, *FEMS Microbiol Rev*, 2005, **29**, 531-554.
5. R. V. Bruggner, B. Bodenmiller, D. L. Dill, R. J. Tibshirani and G. P. Nolan, *Proceedings of the National Academy of Sciences of the United States of America*, 2014, **111**, E2770-2777.

6. I.-C. Chou and E. O. Voit, *Mathematical biosciences*, 2009, **219**, 57-83.
7. G. Goel, I. C. Chou and E. O. Voit, *Bioinformatics*, 2008, **24**, 2505-2511.
8. M. H. Hoefnagel, A. van der Burgt, D. E. Martens, J. Hugenholtz and J. L. Snoep, *Molecular biology reports*, 2002, **29**, 157-161.
9. E. O. Voit, J. Almeida, S. Marino, R. Lall, G. Goel, A. R. Neves and H. Santos, *Syst. Biol. (Stevenage)*, 2006, **153**, 286-298.
10. A. R. Neves, A. Ramos, M. C. Nunes, M. Kleerebezem, J. Hugenholtz, W. M. de Vos, J. Almeida and H. Santos, *Biotechnol Bioeng*, 1999, **64**, 200-212.
11. A. R. Neves, Ph.D. Ph.D. Dissertation, Universidade Nova de Lisboa, 2001.
12. A. R. Neves, A. Ramos, H. Costa, S. van, II, J. Hugenholtz, M. Kleerebezem, W. de Vos and H. Santos, *Appl. Environ. Microbiol.*, 2002, **68**, 6332-6342.
13. A. R. Neves, R. Ventura, N. Mansour, C. Shearman, M. J. Gasson, C. Maycock, A. Ramos and H. Santos, *The Journal of biological chemistry*, 2002, **277**, 28088-28098.
14. J. B. Clarke, M. Birch and H. G. Britton, *Biochemical Journal*, 1974, **139**, 491-497.
15. B. G. Malmstrom, ed. P. D. Boyer, Lardy, H., and Myrbäck, K., Academic Press, New York and London, 2nd edn., 1961, vol. V, ch. 471.
16. F. Marcel and S. Elmer H, in *Comprehensive Biochemistry*, Elsevier, 1969, vol. Volume 17, p. 136.
17. M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe, *Nucleic Acids Research*, 2014, **42**, D199-D205.
18. R. Caspi, T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp, *Nucleic Acids Research*, 2014, **42**, D459-D471.
19. P. Gennemark and D. A. G. Wedelin, *Journal of bioinformatics and computational biology*, 2013, **12**, 1350015.
20. E. O. Voit, in *Applied Statistics for Biological Networks*, eds. M. Dehmer, F. Emmert-Streib and A. Salvador, J. Wiley and Sons, New York, 2011, pp. 183-200.
21. I. C. Chou and E. O. Voit, *BMC systems biology*, 2012, **6**, 84.
22. M. Iwata, F. Shiraishi and E. O. Voit, *International Journal of Systems Biology*, 2013, **4**, 57-72.
23. E. O. Voit, *Mathematical biosciences*, 2013, **246**, 315-325.
24. E. O. Voit, G. Goel, I.-C. Chou and L. da Fonseca, *IET Systems Biol.*, 2009, **3**, 513-522.
25. M. Vilela, C. C. Borges, S. Vinga, A. T. Vasconcelos, H. Santos, E. O. Voit and J. S. Almeida, *BMC Bioinformatics*, 2007, **8**, 305.
26. S. Dolatshahi, B. Vidakovic and E. O. Voit, *Computers & Chemical Engineering*, 2014, **71**, 728-733.
27. E. O. Voit and J. Almeida, *Bioinformatics*, 2004, **20**, 1670-1681.
28. R. N. Gutenkunst, F. P. Casey, J. J. Waterfall, C. R. Myers and J. P. Sethna, *Ann. N. Y. Acad. Sci.*, 2007, **1115**, 203-211.
29. R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers and J. P. Sethna, *PLoS Comput Biol*, 2007, **3**, 1871-1878.
30. S. Srinath and R. Gunawan, *J. Biotechnol.*, 2010, **149**, 132-140.
31. M. Vilela, S. Vinga, M. A. Maia, E. O. Voit and J. S. Almeida, *BMC Syst Biol*, 2009, **3**, 47.
32. P. J. Sands and E. O. Voit, *Ecological Modelling*, 1996, **93**, 75-88.
33. E. O. Voit, *ISRN Biomathematics*, 2013, **2013**, 53.
34. M. A. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley Pub. Co. Advanced Book Program (reprinted 2009), Reading, Mass, 1976.
35. E. O. Voit, *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*, Cambridge University Press, Cambridge; New York, 2000.
36. M. A. Savageau, *J Theor Biol*, 1969, **25**, 365-369.

37. E. O. Voit, ed., *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*, Van Nostrand Reinhold, NY, 1991.
38. M. A. Savageau and E. O. Voit, *Mathem Biosci*, 1987, **87**, 83-115.
39. B. Poolman, K. J. Hellingwerf and W. N. Konings, *J. Bacteriol.*, 1987, **169**, 2272-2276.
40. J. L. Galazzo and J. E. Bailey, *Enzyme Microb. Technol.*, 1990, **12**, 162-172.
41. A. Hartmann, J. M. Lemos and S. Vinga, *Computers in Biology and Medicine*, 2015, **63**, 301-309.
42. S. Marino and E. O. Voit, *Journal of bioinformatics and computational biology*, 2006, **4**, 665-691.
43. R. Castro, A. R. Neves, L. L. Fonseca, W. A. Pool, J. Kok, O. P. Kuipers and H. Santos, *Mol. Microbiol.*, 2009, **71**, 795-806.
44. B. G. Lipták *Instrument Engineers' Handbook- Process Control and Optimization*, Taylor & Francis, 4th edn., 2006.
45. P. Gaspar, A. L. Carvalho, S. Vinga, H. Santos and A. R. Neves, *Biotechnology Advances*, 2013, **31**, 764-788.
46. E. O. Voit and A. E. N. Ferreira, *Journal of Theoretical Biology*, 1998, **191**, 429-438.
47. J. E. J. Ferrell and S. H. Ha, *Trends in Biochem Sci*, 2014, **39**, 496-503.