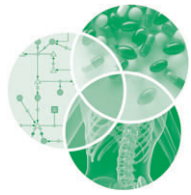
Advanced Review



The best models of metabolism

Eberhard O. Voit*

Biochemical systems are among of the oldest application areas of mathematical modeling. Spanning a time period of over one hundred years, the repertoire of options for structuring a model and for formulating reactions has been constantly growing, and yet, it is still unclear whether or to what degree some models are better than others and how the modeler is to choose among them. In fact, the variety of options has become overwhelming and difficult to maneuver for novices and experts alike. This review outlines the metabolic model design process and discusses the numerous choices for modeling frameworks and mathematical representations. It tries to be inclusive, even though it cannot be complete, and introduces the various modeling options in a manner that is as unbiased as that is feasible. However, the review does end with personal recommendations for the choices of default models.

How to cite this article:
WIREs Syst Biol Med 2017, e1391. doi: 10.1002/wsbm.1391

INTRODUCTION

Over the past decades, mathematical and computational modeling has become a widely accepted tool in biology, and its aspirations of making reliable predictions or offering explanations for complex, and sometimes counterintuitive phenomena are generally appreciated as potentially very useful. At the same time, newcomers to the field do not always seem to recognize that modeling can entail very different approaches and structures. In particular, it appears that many nonexperts seldom wonder where exactly models come from, and if they actively want to engage in basic modeling, they resort to a few defaults that are not necessarily optimal or even valid.

A prime example is the common use of the Michaelis-Menten rate law (MMRL[1,2]), which was developed for analyses of enzyme catalyzed reactions *in vitro*. Specifically, it was conceptually based on the reversible formation of an intermediate complex between substrate and enzyme and the conversion of this complex into product and enzyme, and formulated mathematically in the language of elementary chemical reaction kinetics. Over the decades, this function was frequently chosen not only for metabolic modeling *in vitro*, but also in living cells, where its underlying assumptions are not satisfied[3,4]; indeed, it has been used even in biological fields that have very little to do with enzyme kinetics.[5] The reasons of this default utilization are threefold. First, MMRL has proven to be an excellent representation of enzymatic processes *in vitro*, and thousands of articles have measured or used its characteristic parameters K_M and V_{max}. The assumption therefore has been that, barring obvious alternatives, one might be justified to extrapolate its usage to conditions *in vivo*. Second, MMRL captures a nonlinear saturation process, while allowing simple transformations to linearity, for instance, by expressing $1/V$ as a function of $1/S$, where V is the rate of product formation and S the substrate concentration.[6–8] This combination makes MMRL very appealing, especially if one considers that there are infinitely many nonlinear functions and that there is no guidance regarding optimal choices among them. Also, linear regression becomes applicable for parameter estimation, although with some distortion of the error structure, and is incomparably easier than nonlinear regression. Finally, a default choice is attractive as it seems impossible to infer valid nonlinear representations directly from experimental data, and much has been made of the fact that there are 'no true models.'

Physicists have it a little easier in this respect, because many 'laws,' such as the law of gravity, have

*Correspondence to: eberhard.voit@bme.gatech.edu

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

Conflict of interest: The author has declared no conflicts of interest for this article.

been gleaned directly from experiments or derived from first principles and are valid under common conditions, even though they may fail in the realms of astronomy and quantum physics. Biology, of course, is embedded in the physical world and must obey its laws, but biological processes are often convolutions of so many fundamental processes that a physics-based description is no longer feasible.[9] To see this discrepancy, one might just imagine all the physical processes associated with cell division or signal transduction to realize that biology must develop its own higher-level functional representations. Biology is of course not alone in this respect, and fields like medicine, psychology or sociology face the same challenges.

Even statistics, which is a shining example of mathematical rigor, is not immune against wrong model choices. In some cases, such as flipping a fair coin over and over again, the phenomenon itself may dictate which statistical distribution or process description best captures its random features. However, if biological or medical data have an observed distribution with an extended right tail, many probability distributions—with very different mathematical formats—may approximately fit the data, and the identification of a particular distribution becomes a somewhat troubling matter of unguided choice. Sometimes one may be able to use arguments like the applicability of the central limit theorem in a multiplicative space, which can help provide support for a log-normal distribution, but there are no true guarantees when it comes to real data. We will return to this issue later in the review, but the situation demonstrates how much more complicated the choice of model representations may become in the analysis of biological systems.

To address the issue of model choice, let's start at the beginning. If one is not convinced already, some pondering will lead to the conclusion that there are no true models in biology and related fields. The reasons are manifold but collectively simple: Every biological phenomenon, no matter how small, contains so many components and occurs in such a complicated environment that we cannot even list—let alone represent—all factors that directly or indirectly affect it. At least, that is the current state of the art. Also, the purpose of a model is often to distill, understand, or explain the essence of a phenomenon, which suggests omitting, simplifying, or abstracting nonessential details. But if a detail is omitted, the model is bound to fail if this particular detail becomes important. This conundrum has no real solution, and Ockham's razor is of little help, because it advises against redundancy which is widespread in biology. Instead, modelers should require that a good model be driven by crisp biological questions, which in turn determine the structure of the model.[10] As an example, consider the growth of a bacterial culture. If we plan to investigate how many bacteria to expect at a given point in time, a simple exponential function with an appropriate growth rate might be an appropriate solution for relatively small and well-fed populations. However, the function breaks down for large numbers, does not tell us anything about the variability among several populations that *should* all grow with the same characteristics, does not take into account spatial considerations, and certainly does not reveal molecular mechanisms. Thus, if it is important to determine which genes affect the speed of growth, we obviously require a much more complex model. In the end, the clearest and most concise formulations of questions have the best chance of being answered by mathematical and computational models.

The typical construction of a metabolic systems model from scratch consists of four steps (Figure 1):

1. Identification of the constituents of the system.
2. Identification of the topology and regulation of the system.
3. Choice of mathematical representations for all processes.
4. Estimation of parameter values for the process representations.

These construction steps are followed by diagnostics and analyses of consistency and model appropriateness, which assess technical details, such as model stability and sensitivity, and more globally try to ascertain that the model 'makes sense.' The latter is often tested with a series of simulations whose results are compared to data or expectations. The final modeling steps pertain to various model uses, which may include explanations of observed phenomena, predictions of untested scenarios, various manipulations, perturbations, and interventions, or optimizations toward desirable goals.

The emphasis of this review is on Step 3. Steps 1 and 2 are briefly discussed, but they are primarily a matter of the biological subject area. Step 4 has been reviewed numerous times in recent years and will only be discussed very coarsely. The steps following the model construction fall outside the scope of this review, although the diagnostic steps often lead to a revisiting of Steps 1 through 4.

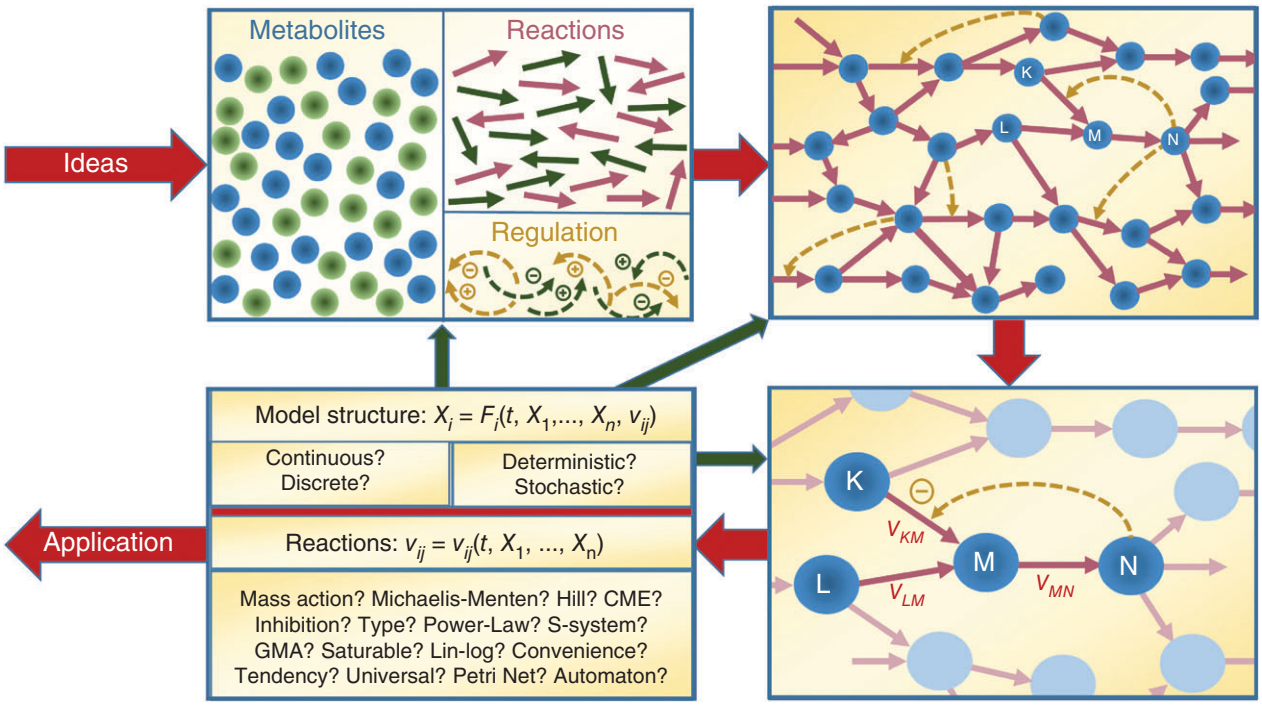


FIGURE 1 | Model design process. Top left: Modeling ideas lead to the selection of (blue) metabolites, (red) reactions, and (gold) regulatory signals, as far as they are known. Top right: These components are arranged in the format of a dynamic system, consisting of a static metabolic network of reactions and their regulation. Bottom right: Each reaction in this dynamic system needs to be mathematically formulated, with account of regulatory signals. Bottom left: The identification of the most appropriate model structure and representations faces many difficult choices. The full characterization of all reactions, including the determination of parameter values, completes the model design. Green arrows indicate that several iterations of the model design process may be necessary.

STEPS 1 AND 2: IDENTIFICATION OF CONSTITUENTS, TOPOLOGY AND REGULATION OF THE SYSTEM

For most of the century of metabolic modeling, the design of a model has focused on a specific metabolite or pathway of interest, or on a pathway system that contains several linear, branched, or cyclic pathways. This focus may have derived from an interesting research question or from a hypothesis generated by earlier investigations. The modeling process begins with two steps, namely, identifying what is to be included in the model, and how the components interact with each other through mass flow or regulation. The two steps go hand in hand and are guided by the biological hypotheses or questions to be answered by the model. They should involve the most parsimonious lists of components, processes, and regulators that still retain the integrity of the system or phenomenon. Such a compromise is easily stated but often very difficult to implement, because assessing the importance of all possible factors would require almost complete knowledge of the system and its environment, which is seldom the case. For this reason, Steps 1 and 2 are arguably the most important and most difficult steps of the modeling process. They may look deceivingly simple at first glance, because 'one intuitively knows' what the model is supposed to be about and who the main players are. However, the selection of components naturally incurs bias, because the modeler does not have full information about the pathway system but must make decisions to include or exclude certain aspects. The hidden challenge is the following: If the lists are missing important components, many model results will be compromised or flat-out wrong. But if they contain too many components, the model becomes unwieldy, over-parameterized, and unreliable in predictions and extrapolations.

Generically, the system definition in Step 1 follows the etymology of the word: to *define* means to *set boundaries*. The challenge is that this boundary setting is often complicated and seldom unique, because every system in biology is embedded in a larger system, which could affect the system of interest. As a rule of thumb, Savageau proposed selecting

components and processes such that there is high connectivity within the system while the number of processes crossing the boundaries of the system is minimal (p. 80 of Ref 11). Of great help are databases like KEGG[12] and BioCyc,[13] which contain comprehensive maps of a large number of pathways, along with ample other information. However, they do not solve the problem.

Step 2 has been approached in two distinct ways. When the model is constructed from the bottom up, which used to be the case for almost all metabolic models until quite recently, one collects information regarding the production and degradation of every component and thereby establishes the system topology and possibly the regulatory structure. Clearly, this procedure requires substantial *a priori* knowledge of the details of the system, but rich information is available as the result of over 100 years of biochemical research and has been documented in an enormous body of literature. The exact determination of the regulatory structure is often murkier, and if it is not known from biochemical experimentation, it is also difficult to infer with computational methods (see below), although this is sometimes possible (e.g., Refs 14–16). In any event, rich information on metabolic reactions has been amassed over time, and the database BRENDA[17] is a treasure trove for kinetic parameter values, including information regarding regulation.

An entirely different approach toward constructing models is based on metabolic time series or data from several input–output steady-state experiments. If such data are available, it is in principle possible to *infer* the most likely connectivity of a system.[18,19] The necessary data have become available in recent years as the result of targeted or untargeted metabolomics.[20–24] In this approach, very many metabolite concentrations are measured, typically with mass spectrometry or nuclear magnetic resonance, for the system under investigation and a control system. Often the goal is to identify and characterize significant differences in the peak pattern of the spectrograms. The differing peaks point to metabolites with different concentrations in the two situations. If these metabolites can be identified, which however is not always the case, they become the focus of further analysis, for instance with dynamic models. To some degree it is also possible to evaluate genome data with respect to expression differentials in genes coding for enzymes.[4,25–27]

The methods of analysis for these data are very different from the bottom-up methods. Rather than scanning the literature and databases for possible connections between pairs of components A and B, and subsequently determining functional descriptions, the data are collectively subjected to a statistical machine learning analysis that optimizes the system connectivity so that it matches the observed data as closely as possible. Specifically, one defines a space of feasible models, which are formulated as graphs where conditional probabilities are associated with the connections between variables. A learning procedure then selects the model that best fits the actual observations.[28] Early methods used Bayesian inference[29] or mutual information,[30] but the topic is generically so complicated, and there are so many solution proposals—some effective, others less so—that recurring crowdsourcing competitions are held to determine the best inference algorithms; they are called DREAM competitions (Dialogue on Reverse-Engineering Assessment and Methods[31,32]). Many of the proposed DREAM algorithms have focused on co-occurrence patterns of changes in the expression of particular genes, transcriptomic responses, proteins, or metabolites. The vast majority of these have targeted gene regulatory or protein–protein interaction networks (e.g., Refs 33–36), but the results of these genome- and proteome-based inferences can be translated, under certain assumptions, into inferences regarding specific metabolic networks.[37–46] It is also sometimes feasible, and of course desirable, to combine the metabolomics top-down approaches with more traditional bottom-up approaches.[47,48] While most reverse-engineering methods and applications have targeted gene networks, some have been applied to metabolic networks as well. Depending on the data, these inferences have addressed static networks[49–52] or dynamic systems.[53–62]

The machine learning inferences may target not only the connectivity of a network but also the distribution of metabolic flux magnitudes at a steady state. For the latter, two methodological frameworks have been developed. The first is metabolic flux analysis, which is based on supplying the metabolic system with a labeled substrate, waiting until the substrate has been converted into a variety of other metabolites, and then using sophisticated computational methods for determining the fluxes that drive the system.[63–69] The second is flux balance analysis (FBA), which is a very popular extension of stoichiometric analysis[70,71] and estimates flux distributions, sometimes genome-wide, under the assumption of some optimality criterion, such as maximal growth or ATP production.[72–75]

Quite a different, graph-based inference method is the metabolic pathway reconstruction from molecular structures of compounds and knowledge about enzymes that possibly convert these compounds into

others. In these approaches, reactions are considered transfers of atoms between metabolic compounds, and computational graph methods determine feasible or most likely paths.[76–78]

In all these approaches, no matter how different they are, the ultimate outcome is a directed network of metabolic fluxes, ideally with an indication of its regulatory control structure. An exception is the definition of reactions as nodes and metabolites as edges that represent 'shared resources among modules.'[79,80] This option is rather counterintuitive but has the advantage of natural clustering with a reduced number of nodes. In a similar vein, one can mathematically represent a traditional metabolic network as a model where the reactions are the dependent variables.[81]

STEP 3: CHOICE OF MATHEMATICAL REPRESENTATIONS FOR ALL PROCESSES

The great advantage of graph and machine learning methods is that only minimal *a priori* knowledge about a system is needed, as long as enough suitable data are available. Alas, this advantage is directly connected to the greatest limitation of these methods, namely, that the final result consists only of the connectivity of the system and, in some cases, the amount of material flowing through each connection. For some purposes, this information is sufficient, and corresponding models have been used to predict the consequences of gene knock-outs or changes in substrate availability. However, these predictions implicitly assume that the organism does not call up its multi-level regulatory machinery to mount compensatory mechanisms. Such responses are nonlinear and therefore not always well modeled by linear methods such as graph analyses or FBA. For instance, if a gene is knocked out that codes for some enzyme, the corresponding enzymatic step is also eliminated and no product is generated. However, if the organism needs this product, it will express other genes with the goal of redirecting alternative pathways toward the desired metabolite. This problem is significant for minimalistic FBA models, but does not entirely disappear for genome-wide models either, because the analysis makes inferences in the absence of regulation. In a beautiful demonstration, Ishii et al.[82] studied the responses of *Escherichia coli* to numerous environmental and genetic perturbations. While the expression of genes close to the perturbations often changed dramatically, the induced disruptions led to surprisingly small changes in mRNAs and proteins. Moreover, the metabolite levels remained unexpectedly stable, and the authors showed that this stability was achieved through wide-spread rerouting of fluxes throughout the metabolic system.

If complex responses such as compensation or adaptation are to be understood, the modeling effort needs to be stepped up toward fully dynamic models that permit true extrapolations and predictions of untested scenarios. For these models, the connectivity of components is not sufficient, but every process needs to be formulated as a function that appropriately captures the roles of all system components that affect this particular process. Needless to say, this step is challenging. Furthermore, the need to identify suitable functional representations arises whether a model is constructed from the bottom up or from the top down.

The Challenge of Choosing Suitable Functions

Even for single processes, the choice of an explicit function can be difficult. As a comparatively simple example, consider the selection of a statistical distribution function for representing a sample or process. As discussed before, the stochastic features underlying the distribution may suggest a formulation, as for flipping a coin, but such guidance is rare. To demonstrate the problem, Sorribas et al.[83] generated 25 samples with 160 drawings each from normal, gamma and Weibull distributions and then used an optimization algorithm to fit one of the standard distribution functions (normal, log-normal, gamma, logistic, Weibull, or Gumbel) to these artificial data. Surprisingly, in about 75% of all cases, the best fit was obtained with a different distribution than the one used to create the sample. This unexpected finding causes concern for selecting functions for biological processes, especially if data are noisy.

The wrong identification of functions raises the question whether it is really problematic if the selected function is wrong, as long as the data are represented well. For the statistical example it may actually not even be a major problem for data analyses because distributions are used for concise descriptions, sampling, decision making, and maybe for computing some test statistics. However, wrong functions can be detrimental in dynamical systems, which are commonly used for extrapolations toward new scenarios. As a simple example, consider two very similar models of the small pathway in Figure 2. Model A is formulated as

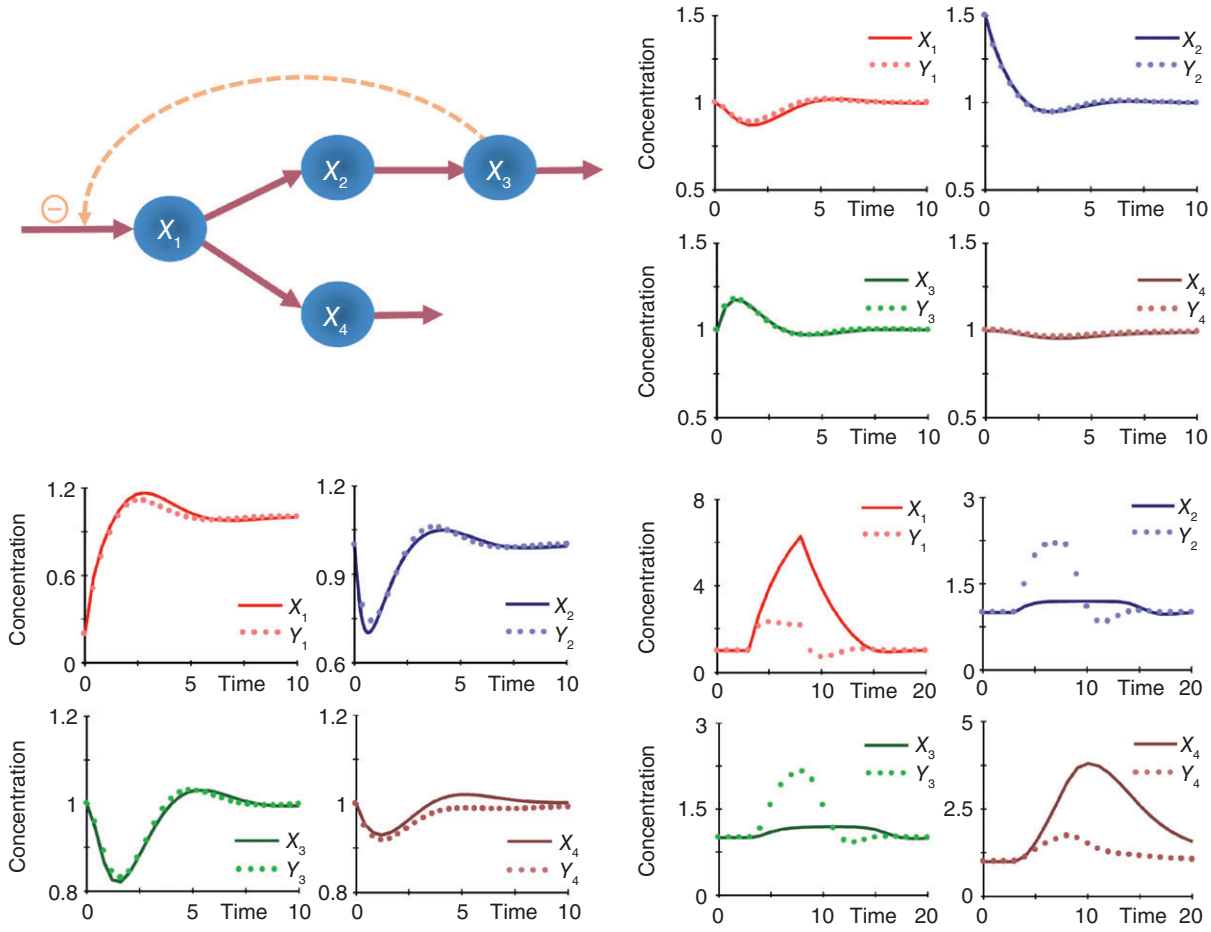


FIGURE 2 | The diagram in the top-left panel was modeled with two slightly differing models (see Text). The responses to moderate perturbations (top-right: $X_2(0) = Y_2(0) = 1.5$ and bottom-left ($X_1(0) = Y_1(0) = 0.2$) are quite similar. However, a bolus of 2 units added to X_1 and Y_1 during the time period $t \in [3, 8]$ triggers very different response trends (bottom-right).

$$\dot{X}_1 = 1.2\,X_3^{-1} - F(X_1) - 0.2\,X_1$$

$$\dot{X}_2 = F(X_1) - X_2$$

$$\dot{X}_3 = X_2 - X_3$$

$$\dot{X}_4 = 0.2\,X_1 - 0.2\,X_4 \qquad (1)$$

where, the material flux between X_1 and X_2 is represented with the Hill function $F(X_1) = \frac{1.2 \cdot X_1^4}{0.2 + X_1^4}$. In Model B, all variables are called Y_i, and the process corresponding to $F(X_1)$ is instead formulated as the mass action function $G(Y_1) = cY_1$ with $c = 1$. Otherwise the two models are identical. One could argue that a Hill model and a mass-action model are quite different, which is certainly true, and that's the point of the example: In a real-world situation, it may be entirely unclear how to formulate $F(X_1)$ mathematically, and the modeler will have to rely entirely on available experimental data to determine a well-fitting model. Without *a priori* knowledge, the modeler may try the mass-action formulation or a Hill function (or a variety of other functions), which in both cases leads to excellent results (Figure 2). The modeler is presumably satisfied, and using the argument of parsimony, s/he may choose the simple mass-action format. This choice does not encounter problems until a new scenario is modeled, as we will discuss next. It appears that this scenario is generically much more frequent than one would like to admit.

Both models have the same steady state of (1, 1, 1, 1), and the differences in responses to reasonable perturbations are small, especially if one imagines that either model in reality should match data with a bit of noise. Figure 2 shows responses to changing the initial values from the steady state to (0.2, 1, 1, 1) or (1, 1.5, 1, 1). The resulting fits are certainly satisfactory, and one may conclude that either model choice is perfectly suitable.

However, substantial differences arise if the models are used for other types of simulations. For instance, the panel on the bottom right of Figure 2 exhibits the results of a simulation where an external bolus of 2 units is added, quasi as a second input, to the first differential equation during the time period $t \in [3, 8]$, whereas it is 0 at other times. For larger dynamical systems, where error compensation easily happens within or among terms of the same equation or of different equations, it is possible that wrong model choices lead to utterly faulty results. Note that the example was not constructed around a bifurcation point, where slight changes in parameter values may lead to different qualitative behavior, but that it addresses typical alternative functions within their normal operation ranges.

Generic Strategies for Identifying Appropriate Functions

Clearly, the consequence of choosing a wrong model should be a cause for concern that requires attention. For inspiration, let's return to statistics to explore documented strategies for selecting distribution functions. A first, intuitive strategy may be fitting a more or less comprehensive set of candidate functions to the data under investigation, one at a time, and judging the quality of fit by residual error. Sorribas et al.[83] pursued this strategy to analyze growth distributions of girls between ages 5 and 17. Needless to say, this strategy is cumbersome and time-consuming, and any choice based on residual errors becomes questionable if the various candidate distributions contain different numbers of parameters: On the one hand, a function with more parameters should be expected to yield better accuracy, but if one is interested in a good fit, should a highly parameterized (or over-parameterized) function be penalized?

A second strategy is the formulation of a generalized distribution function that contains a few or many candidate functions as special cases.[84–88] Driven to the extreme, Savageau's *suprasystem of probability distributions*, which consists of a large system of nonlinear ordinary differential equations (ODEs), is so comprehensive that one can prove mathematically that all continuous distributions are exact special cases.[89] Problem solved? No, because, first, it is *a priori* unclear how many ODEs should be involved, and second, increasing numbers of ODEs demand so many parameters that it becomes impossible to determine them from data with any degree of reliability. In a similar vein, one could propose a generic polynomial which, with sufficiently many terms, provides a perfect representation to any dataset. In all these cases, the compromise obviously lies in the complexity of the generalization, which strongly affects the number of parameters to be estimated.

A third option is in some sense a compromise between the first two approaches. It consists of the generic representation of a distribution family in an approximate manner that strikes a reasonable balance between accuracy and the number of parameter values. Examples in statistics are the four-parameter S-distribution and the five-parameter GS-distribution.[90–92] These distributions contain only a few statistical distributions as exact special cases, but approximate very many of them, including complicated noncentral and discrete distributions, in a relatively simple, streamlined and quite accurate manner. Owing to their simplicity, these distributions permit interesting analyses, especially with respect to fitting data of ill-characterized origins, where they can even suggest which traditional distribution might be well suited to represent a dataset. Similar strategies have been proposed for growth processes.[93–96]

Selecting Metabolic Rate Functions from a Smorgasbord of Options

The issues associated with model choices are exacerbated in metabolic modeling, because experimental data are typically the output of complex systems, rather than explicit functions. As a consequence, the selection problem is confounded by compensation of errors within and among the equations of the ODE system that is supposed to represent the data, as we saw before (Figure 2). The good news is that metabolic models tend to be quite robust, as long as all variables remain in relatively small ranges. This is especially so in large models with ample regulation, which tends to buffer the variables in the vicinity of a homeostatic state. As a consequence, different models may be appropriate for modeling systems close to their normal operating state. Direct comparisons between alternative modeling frameworks are relative rare but do exist and demonstrate similar outputs for different models in response to small perturbations.[97–111] Nonetheless, models are often designed for explorations of new realms, where differences among alternative representations can become significant. For instance, the goal of a model analysis may be to understand or intervene in a disease, or to manipulate a microbe toward the high-yielding production of some metabolite, such as biodiesel, insulin, or citric acid.[112–115] These types of manipulations often require larger changes, where

differences between alternative modeling formats become significant. Thus, the search for the best models must not be abandoned.

The field of metabolic modeling is dominated by a few functional formats that have been used time and again. In some way, many of them may be traced back to mass action kinetics, which was proposed over 150 years ago for explaining the kinetics of elementary chemical reactions.[116–119] Indeed, mass action functions are the most prevalent defaults. For a simple irreversible reaction that converts A into B, the mass-action formulation is $\dot{A} = -k \cdot A$, $\dot{B} = k \cdot A$. Thus, degradation and production are linear functions of A with a rate k. If A and B are converted into C, the model is represented as $\dot{A} = \dot{B} = -\kappa \cdot A \cdot B$, $\dot{C} = \kappa \cdot A \cdot B$, where κ is again a rate constant. It is important to note that many models in biology are direct derivatives of these formulations[119]; they include Michaelis-Menten and Hill functions,[1,2,120] SIR models[121] for the spread of infectious diseases, and Lotka-Volterra models.[122–125] For single-variable processes, the differential equation for the loss of A is linear, which is also the case for many elementary processes in physics, such as exponential decay, heating and cooling, and simple transport processes.

Outside the mass-action formulation, the MMRL[1,2] has been the undisputed workhorse of metabolic modeling.[4] Initially formulated as a set of differential equations, describing the binding of substrate to enzyme and the generation of product, as discussed before, the power of the rate law came from assumptions that are mostly true for experiments *in vivo*. These permitted the reformulation of the overall rate of conversion of substrate into product as an explicit function of the form

$$v = \frac{V_{\max} S}{K_M + S}, \tag{2}$$

where $V_{\max}$ is the maximal rate, and the Michaelis constant K_M reflects the affinity between substrate and enzyme. The simplest generalization of MMRL is the Hill function,[120] which associates the Hill coefficient n with the substrate and with K_M:

$$v = \frac{V_{\max} S^n}{K_M^n + S^n}. \tag{3}$$

The Hill coefficient is typically 2 or 4, as it was originally meant to represent molecular subunits. Nonetheless, there are no mathematical reasons for these settings, and indeed, noninteger coefficients have been used in metabolic modeling (e.g., Refs 126).

In their pure, unmodulated form, MMRL and the Hill function are easy to use and parameterize,[6–8] and while we already discussed limitations with respect to valid applicability, they have been and will remain to be a mainstay in metabolic modeling for a long time. For larger pathway systems, the seeming simplicity evaporates, and standard analyses, such as the computation of a steady state or of parameter sensitivities, become very cumbersome.[104]

The most important generalization of these functions is the account of regulation, which can make these functions complicated, if not unwieldy. In the simple case of competitive inhibition by inhibitor I, the result is still rather simple:

$$v = \frac{V_{\max} S}{K_M (1 + I/K_I) + S}. \tag{4}$$

However, different inhibition mechanisms mandate different formulae. As a result, the mathematical formats for rate laws with competitive, noncompetitive, uncompetitive, mixed, or allosteric inhibition are all different, which implies that the type of inhibition must be known before the appropriate model can be formulated. Numerous books have published details of the various types of inhibition.[11,127–130]

For reactions that involve several inhibitors and modulators, the results can become very complicated, leading to a feeling of frustration and concern. An interesting example is the phosphofructokinase (PFK) reaction of glycolysis, which phosphorylates fructose 6-phosphate through a transfer of ATP. Because PFK is a critical control point in glycolysis, the reaction is affected by several metabolites, as well as pH. Extensive investigations of these modulators have led to a large variety of rate functions that were formulated to capture the reaction kinetics. The following collection of examples is certainly not complete, but the reason for showing at least a subset is to indicate the growing confusion the newcomer experiences when studying rate functions. The examples are presented essentially in the notation in which they were originally published.

One of the simpler rate laws was presented by Eicher and colleagues,[126] who formulated it as a bi-substrate, irreversible Hill process of the form

$$v = V_f \frac{\mathrm{F6P}^h \, \mathrm{ATP}^h}{\left(\mathrm{F6P}^h + \mathrm{ATP}^h\right) \cdot \left\{1 + \left(\frac{1 + \mathrm{PEP}^h + 1 + \alpha^{2h} \mathrm{PEP}^h}{1 + \alpha^{4h} \mathrm{PEP}^h}\right)\right\}}, \tag{5}$$

where the variables F6P, ATP and PEP represent metabolite concentrations that are scaled by their half-saturation constants. Thus, in addition to V_f, α and h, the model contains three further parameters. Blangy et al.[131] proposed a model based on the so-called *concerted transition theory* proposed by Monod et al.[132] This model has over twenty parameters. A few years later, Otto et al.[133] proposed a model in the format

$$v = \frac{V\max\_PFK \times \frac{\mathrm{F6P}}{K\_\mathrm{F6P} + \mathrm{F6P}} \times \frac{\mathrm{MgATP}}{K\_\mathrm{MgATP} + \mathrm{MgATP}}}{1 + L}, \quad (6a)$$

where

$$L = L_0\_\mathrm{PFK} \times \frac{\left(1.0 + \frac{\mathrm{MgATP}}{K_{\mathrm{MgATP}}}\right)^4 \times \left(1.0 + \frac{\mathrm{Mg}}{K\_\mathrm{Mg}}\right)^4}{\left(1.0 + \frac{\mathrm{F6P}}{K\_\mathrm{F6P}}\right)^4 \times \left(1.0 + \frac{\mathrm{AMP}}{K\_\mathrm{AMP}}\right)^4}. \quad ((6b))$$

Mulquiney et al.[134] suggested the alternative

$$v = \frac{\left(\frac{K_{\mathrm{catf}} \times \mathrm{MgATP} \times \mathrm{F6P}}{K\_\mathrm{MgATP} \times K\_\mathrm{F6P}} - \frac{K_{\mathrm{catr}} \times \mathrm{F16BP} \times \mathrm{MgADP}}{K\_\mathrm{F16BP} \times K\_\mathrm{MgADP}}\right)}{1 + \frac{\mathrm{MgATP}}{K\_\mathrm{MgATP}} + \frac{\mathrm{F6P}}{K\_\mathrm{F6P}} + \frac{\mathrm{MgATP} \times \mathrm{F6P}}{K\_\mathrm{MgATP} \times K\_\mathrm{F6P}} + \frac{\mathrm{F16BP}}{K\_\mathrm{F16BP}} + \frac{\mathrm{MgADP}}{K\_\mathrm{MgADP}} + \frac{\mathrm{F16BP} \times \mathrm{MgADP}}{K\_\mathrm{F16BP} \times K\_\mathrm{MgADP}}} \times \frac{1}{1 + L} \quad (7a)$$

where

$$L = \frac{\left(\frac{[\mathrm{H}^+]}{K_a}\right)^4 \left(1 + \frac{\mathrm{ATP}}{K\_\mathrm{ATP}}\right)^4 \times \left(1 + \frac{\mathrm{Mg}}{K\_\mathrm{Mg}}\right)^4 \left(1 + \frac{\mathrm{23BPG}}{K\_\mathrm{23BPG}}\right)^4}{\left(1 + \frac{\mathrm{F6P}}{K\_\mathrm{F6P}} + \frac{\mathrm{F16BP}}{K\_\mathrm{F16BP}}\right)^4 \times \left(1 + \frac{\mathrm{AMP}}{K\_\mathrm{AMP}}\right)^4 \times \left(1 + \frac{P_i}{K\_P_i}\right)^4 \times \left(1 + \frac{\mathrm{G16BP}}{K\_\mathrm{G16BP}}\right)^4}. \quad (7b)$$

One of the most complex formulations to date was proposed by Peskov et al.,[135] again based on the ideas of Monod et al.[132] This model involves about 40 parameters. Numerous other models for the PFK reaction have been described, including.[131,136–142] Granted, some of these formulations were developed by enzyme biochemists interested in characterizing the catalytic mechanisms underlying an enzymatic reaction, so that the resulting format and numbers of parameters were not really an issue. However, Peskov et al. explicitly state that their model 'can be used in the kinetic modeling of biochemical pathways containing phosphofructokinase-1,' so that a quantitative determination of parameter values is unavoidable.

The co-existence of such drastically different models for the same reaction is disconcerting: How should one decide on the most appropriate model formulation? Is it even possible to obtain information regarding all kinetic parameters, and if so, is it valid to mix parameter values measured under different conditions and maybe even in different organisms? Are questionable parameter values and a confusing, overwhelming diversity of formats a fact of life, or are there alternatives? Savageau (p. 75 of Ref 11) responded to these questions over forty years ago: 'It must be concluded that the complete kinetic characterization of more complex regulatory enzymes is impossible for practical reasons. Even if such a characterization were available, it would hardly be useful. In the simplest case, for which the highest power of any concentration variable is one, a reaction with eight reactants and modifiers will have on the order of 500 terms in its rate law.' Schulz[130] came to a similar conclusion after having discussed the overwhelming complexity of some rate functions that attempted to capture the correct mechanism.

The solution Savageau proposed ushered in the modeling framework of *Biochemical Systems Theory*, whose core feature is the streamlined representation of all processes as products of power-law functions.[143–145] Thus, if the reaction producing a metabolite X_3 is affected by two substrates, X_1 and X_2, inhibited by X_i, and activated by X_a, the mathematical formulation is

$$\dot{X}_3 = \alpha_3 X_1^{g_{31}} X_2^{g_{32}} X_i^{g_{3i}} X_a^{g_{3a}}. \quad (8)$$

Here, the rate constant α_3 and the kinetic orders g_{31}, g_{32}, g_{3a} are positive, while the inhibition parameter g_{3i} is negative. Since the inception of BST, hundreds of articles have successfully used this straightforward formalism; some comprehensive reviews are.[11,96,114,146–149]

BST comes in two main variants. In the *Generalized Mass Action* (GMA) representation, each term is represented with a product of power-law functions, as in Eq. (8), which makes intuitive sense as a direct generalization of mass action systems. In the *S-system* variant, by contrast, the focus is on metabolite pools rather than on reactions. Thus, all reactions entering a pool are aggregated into a single power-law function and the same is done with all fluxes leaving a pool. The result is an ODE system that contains at most one positive and one negative power-law term in each equation. While less intuitive to the biochemist, this formulation has the enormous advantage that steady states and their associated features like stability, sensitivities, and gains can be computed straightforwardly with algebraic means.[143] This aspect can be crucial for a variety of analyses that require the frequent computation of steady states, such as the search for design principles or steady-state optimization (e.g., Refs 104,112–114,150–155). While BST has been very successful, it has been clear from the beginning that the power-law formulation is the result of Taylor approximation in logarithmic space. As such, this representation is exact at an operating point, excellent close to this point, but of unknown quality if one moves farther away from this point. In particular, power-law functions with positive exponents do not saturate. Also, if an inhibitor is included as in Eq. (8), one may find it acting like an activator if its concentration falls below 1. However, simply rescaling the inhibitor concentration or using $(1 + I)$ instead of I easily remedies this situation.[156] Finally, the aggregation of fluxes leading to the S-system format introduces slight inaccuracies at branch points. Interestingly, it was shown that this inaccuracy compensates for the error of the power-law approximation and that this format is actually more accurate than the GMA form, if hyperbolic functions such as MMRL are represented in this form.[111] Overall, power-law representations are approximations, but so are all other functions discussed in this review. They have their shortcomings, but they are terrific defaults, especially if the metabolic system is not all that well characterized numerically.

A later proposal for streamlined representations was the *lin-log model*,[101,157–159] which was inspired by *Metabolic Control Analysis*,[160–164] and takes the form

$$\frac{v_i}{J_i^0} = \frac{e_i}{e_i^0}\left(1 + \sum_{j=1}^{n+m} \varepsilon_{ij}^0 \log\left(\frac{X_j}{X_j^0}\right)\right). \tag{9}$$

Here the variables are denoted by $X_1, \ldots, X_{n+m}$ and the reaction rates are denoted by $v_1, \ldots, v_r$. X_j^0 is a reference concentration of species X_j, J_i^0 is a reference flux of the i^{th} reaction, e_i is the enzyme activity, e_i^0 is the reference level of the enzyme activity, and ε_{ij}^0 are elasticities that correspond to the exponents in BST. A troubling problem with this formulation is that the rate becomes negative for small concentrations and approaches $-\infty$ for substrates approaching zero.[99,100]

Searching for a better compromise between mathematical tractability and modeling accuracy, Sorribas and collaborators generalized BST by using a Hill function as the core element for each variable, so that all terms in the model are guaranteed to saturate.[165,166] The result is the *Saturable and Cooperative* (SC) formalism, where each reaction has the format

$$v = \frac{\prod_j X_j^{n_j}}{a\prod_j X_j^{n_j} + \sum_i b_i \prod_{j,j\neq i} X_j^{n_j}} \tag{10}$$

and the parameters a, b_i, and n_j are related to the kinetic parameters of the Hill functions. This model is often more accurate than a power-law model, but contains quite a few more parameters.

In a similar vein, Wayman et al.[167] proposed *Multiple Saturation Kinetics* with reaction terms of the Michaelis-Menten form

$$r_j = V_j^{\max} E_i \left(\prod_s \frac{X_s}{K_{js} + X_s}\right) v_j. \tag{11}$$

Here, $V_j^{\max}$ is the maximal rate, E_i is the enzyme activity, each K_{js} is a Michaelis constant, and v_j is an allosteric regulatory term, which is typically expressed as a Hill function.

Yet another proposal is *Convenience Kinetics*. For a reaction that converts $A_1, A_2, \ldots, A_n$ (with concentrations $a_1, a_2, \ldots, a_n$) into $B_1, B_2, \ldots, B_m$ (with concentrations $b_1, b_2, \ldots, b_m$), the proposed rate function is

$$v(a,b) = E_{\text{tot}} f_{\text{reg}} \frac{k_+^{cat}\prod_i \tilde{a}_i^{\alpha_i} - k_-^{cat}\prod_j \tilde{b}_j^{\beta_j}}{\prod_i\left(1 + \tilde{a}_i + \ldots + \tilde{a}_i^{\alpha_i}\right) + \prod_j\left(1 + \tilde{b}_j + \ldots + \tilde{b}_j^{\beta_j}\right) - 1.} \tag{12}$$

In this formulation, the variables with tildes denote concentrations that are scaled by the corresponding Michaelis constants, E_{tot} is the total enzyme activity,

and f_{reg} is a regulatory term. Owing to the scaling, this model contains more parameters than an initial impression suggests.

In the spirit of BST, Rohwer et al.[168] argued that 'in systems biology, … the precise mechanism of an enzyme is less important; what is required is a description of the kinetics of enzymes that takes into account the systemic context in which each enzyme is found.' As a solution they proposed a generalized reversible Hill equation of the generic form

$$v = \frac{V_f \alpha \left(1 - \frac{\Gamma}{K_{eq}}\right)(\alpha + \pi)^{h-1}}{\frac{1+\mu^h}{1+\sigma^{2h}\mu^h} + (\alpha + \pi)^h}, \qquad (13)$$

where α is the substrate concentration divided by its half-saturation constant, π is the correspondingly scaled product concentration, Γ is the mass-action ratio, K_{eq} is the equilibrium constant, h the Hill coefficient, μ a scaled modifier concentration, and σ an interaction factor. The authors proclaimed this function as 'a universal rate equation for systems biology' that 'should lay the groundwork for a "new" enzyme kinetics for systems biology.' Other authors proposed different 'universal canonical forms' for modeling dynamic systems, based on power laws.[169–171] Visser et al.[172] suggested a combination of Michaelis-Menten type reactions, stoichiometric analysis, and model reduction based on pseudo-steady-states, which they termed *tendency modeling*. Marino devised an iterative method for determining the necessary model complexity.[173]

Similar to power-law models in BST and representations in the SC formalism, these latter formulations have a fixed structure, but it is easy to see that they quickly become unwieldy. For instance, the great advantage of S-systems within BST, namely the algebraic computability of steady states[143] and of features like stability and sensitivities,[146] is gone in SC, multiple saturation, and convenience-kinetic models, as well as Rohwer's universal rate equation. As a compromise, piecewise S-system formulations have been proposed, which improve accuracy, but require more parameters.[174–176] Furthermore, Löwe et al.[177] proposed a BST model reduction based on hierarchies of time-scales, which indeed can span several orders of magnitude in actual applications.[178]

An entirely different type of metabolic model makes use of typical control strategies in cybernetics. These models can be highly predictive. However, the control does not occur through metabolic processes, but is based on optimal decisions with respect to hypothetical physiological goals, so that the control strategies are difficult to translate into a specific biological mechanism.[167,179]

Space, Delays, and Stochasticity

All ODE models of metabolic pathway systems make the implicit assumptions that the reactions occur under homogeneous, well-mixed conditions and that very many molecules are involved. The first assumption is often—if not almost always—violated. Nevertheless, very few attempts have been made to develop formalisms not needing the assumption of spatial homogeneity. The reasons are that the mathematics immediately becomes much more complicated and that data very rarely exist that would allow a more appropriate model representation. For instance, as soon as a metabolic pathway is distributed over the cytosol, Golgi and endoplasmic reticulum, a modeler would need to know volumes of each, just to compute concentrations, as well as transport rates across membranes, and possibly a clear picture of the spatial arrangements of organelles within a cell.[180] Rudimentary attempts have been made to accommodate different compartments through different variables, an 'ecosystem of organelles,' or variables on different types of grids.[181–187] However, generally effective models and solutions are yet to be developed.

Several authors addressed an important subproblem within the class of space issues, namely the formulation of kinetic rate laws in crowded cellular environments, where the movement of molecules is impeded.[188–194] The conclusion was that power-law functions are good representations, but require different parameter values than under noncrowded conditions.

The second assumption in ODE modeling is that all processes occur without delays. This assumption is often, but certainly not always true. While delay-differential equations are cumbersome, approximation methods can introduce delays into ODE models in a straightforward fashion (e.g., Refs 195–199).

The third implicit assumption of ODE models is that sufficiently many molecules are present. If so, enough averaging occurs in time and space to justify a continuous model. However, if it is not the case, averaging over individual catalytic events may not be valid. A more accurate description in this situation is a stochastic process, in which each conversion of a substrate molecule into a product molecule is considered a separate event. The strategy to capture this situation is to compute the probability that the state of the system (in terms of numbers of molecules of a particular type) changes during a given time interval.

The base formulation for this purpose is the chemical master equation (CME), which is a finite-state, continuous-time Markov process that can be expressed as a Chapman-Kolmogorov differential equation.[200,201] CME describes the probability P_k that the chemical system is in state k at time t. Suppose **A** is a matrix of rate constants of possible transitions between species, where the destination is given by the first subscript and the source by the second subscript of each element. Each transition rate is characterized by a so-called *propensity function* $\alpha(X)$, where $\alpha(X)dt$ represents the probability that a system variable will change during the infinitesimal interval (t, dt) due to the activity of a reaction. Generally, the propensity function is the product of a stochastic rate constant with a combinatorial factor representing the number of different combinations of reactant molecules that are available just before the reaction.[202] For instance, for a reaction between X and Y, where N_X and N_Y molecules are available, respectively, the propensity is $c \cdot N_X \cdot N_Y$.

Any increase in the probability P_k is driven by transitions from other states P_m into P_k and can be written as

$$\sum_m A_{km} P_m. \tag{14}$$

The contribution of P_k to other states is formulated analogously. CME can therefore be written as

$$\frac{dP_k}{dt} = \sum_{m \neq k} (A_{km} P_m - A_{mk} P_k). \tag{15}$$

Because k is not a metabolite, but a possible state, which is composed of the numbers of metabolite molecules in each pool, the dimension of this system may be very large even for relatively small systems.

In some chemical systems, some reactions are very fast, whereas others are slow. Like for any other system of ODEs, this difference in time scale may lead to stiffness, which slows down typical integration algorithms.[203] In both, stochastic and deterministic simulations, the apparent difference in time scales may actually be due to very different concentrations among reactants. At any rate, if the fast reactions are not of interest, they may be removed by grouping states that are connected through fast transitions, which possibly remedies the problem.[204] Another speed-up was proposed by López-Caamal et al.[205] who reduced the order of the model by reformulating the propensities as Michaelis-Menten terms. Wu et al.[206] showed how the propensity of a stochastic simulation is to be derived from the continuous analog. Hahl and Kremling compared CME and ODE methods[207] and Ao incorporated stochasticity into deterministic metabolic models.[208]

The typical approach to working with CME is a Monte-Carlo simulation, which is often implemented as the *Stochastic Simulation Algorithm* (SSA). This algorithm constructs numerical realizations of the state of the system over time and averages the results of many such realizations.[201,209,210] Specifically, it computes the probability of the time to the next reaction and the probability that the next reaction is a particular reaction r_i. Based on these probabilities, the state of the system is updated. While this algorithm captures the process appropriately, it is computationally rather slow and can only be applied to small systems, unless one uses high-performance computing. The solution can be sped up, by maybe one order of magnitude, with the *τ-leaping method* which, instead of determining exactly which next reaction fires when, performs all expected reactions within an entire interval of length τ. Many implementations of this concept have been proposed.[211–213] A disadvantage of this method is that the error of the approximation is difficult to assess.

A modeling language that organizes deterministic and stochastic approaches under a unifying umbrella is Petri net theory.[214,215] This theory was originally proposed in the 1960s as a formal language for assessing network graphs,[216] by characterizing their topology, invariants, reachability, spaces of possible states, and possible structural reduction. To model metabolic networks, a Petri net model was formulated as discrete-event system describing how the state of the system changed from one time point to the next.[217–219] Nowadays, Petri nets may be continuous, in which form they are quite similar to typical kinetic ODE models that use as default rates mass-action, Michaelis-Menten, or Hill representations. They may also be stochastic, where they fundamentally implement Gillespie's ideas of randomly timed reaction steps, or they can combine stochastic and deterministic aspect in the form of *Functional Hybrid Petri Nets*.[198,218,220–226] This aspect is of particular appeal, as continuous systems are often subject to stochastic perturbations or delays, which can be subsumed under the format of hybrid functional Petri nets.

Outside Petri nets and stochastic models, relatively few discrete approaches have been proposed to model metabolism. Watson[227] formulated metabolism as a discrete optimization model. Manca et al.[228] developed a conceptual approach based on membrane computing.[229] Asenjo et al.[230] combined gene regulation with a metabolic network model in a

discrete manner. Velez-Cuba and collaborators[231] showed that Boolean, logical, and Petri net models can be represented as time-discrete dynamical systems that may be analyzed with methods of computer algebra. Nehaniv et al.[232] formulated metabolic and other biological systems in the language of automata.

In addition to typical simulations and steady-state analyses, some types of ODE models permit a very different type of exploration, namely, the application of automated methods for *algebraic model checking*.[233–237] A good example is the Simpathica software that explores the time-trajectories of models with an automaton-based semantic language.[238,239] This language permits answering questions about the logical properties of the temporal evolution of a system, such as: *is the system able to reach a steady state?* Or: *what are the possible bounds for the trajectories of a particular dependent variable in the system?* The model checking language also contains qualifiers such as *eventually* and *always*, as well as their negation, which leads to the qualifiers *never* and *sometimes*. As a specific example, a computer algebra algorithm may check the truth of the expression *Eventually(Always(zero-derivatives))*, which corresponds to the situation that the system will certainly approach a steady state for t going toward infinity. In this manner, the system can qualitatively reason about features of the system by using propositional temporal logic that succinctly and unambiguously addresses ordered sequences of events. A different type of metabolic model checking was proposed by Gevorgyan et al.,[240] who proposed algorithms for the verification of the stoichiometric consistency of a model.

A Glimpse of the Truth

Taking account of the great variety of representations of metabolic processes, the question arises if it is at all possible to identify the true functions governing metabolism *in vivo*? The answer is *probably not*. Nonetheless, it is possible to get a glimpse of what such representations might look like. An approach toward this aspect is *Dynamic Flux Estimation* (DFE[241]). In a nutshell, DFE works by separating linear and nonlinear aspects of metabolic models (Figure 3). The typical generic representation of a metabolic model is the stoichiometric equation

$$\dot{\mathbf{X}} = \mathbf{N} \cdot \mathbf{V}, \qquad (16)$$

where the metabolites X_i are the state variables and $\dot{\mathbf{X}}$ is the corresponding vector of derivatives. The

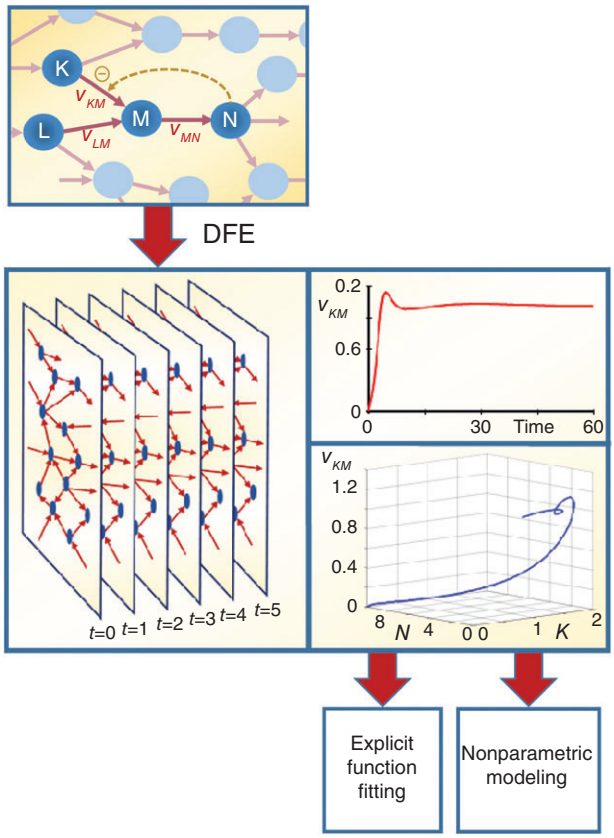


FIGURE 3 | Dynamic flux estimation (DFE). Beginning with the diagram of regulated fluxes (Figure 1, bottom right), DFE separates the linear stoichiometry from the nonlinear fluxes. Specifically, at a series of time points, the distributions of fluxes are given by linear algebraic systems. These systems are solved and flux values are plotted either against time or against the variables that affect them. One may attempt to find explicit functions for these plots or pursue nonparametric modeling. The subsystem of K, L, M, and N is demonstrated here with the model: $\dot{K} = 1 - v_{KM}$; $\dot{L} = 0.2 - L/(4 + L)$; $\dot{M} = v_{KM} + L/(4 + L) - 1.2\ M^{0.2}$; $\dot{N} = 1.2\ M^{0.2} - 1.2\ N^{0.8}$; $v_{KM} = K^{0.4}\ N^{-1}$; $(K_0, L_0\ M_0, N_0) = (0, 8, 0.1, 10)$.

stoichiometric matrix $\mathbf{N}$ captures the connectivity between the fluxes and metabolite pools.

Given time series of metabolite concentrations, the data trends are smoothed (e.g., Refs 242–244), and it is possible to obtain estimates $\mathbf{S}_{t_1}, \mathbf{S}_{t_2}, \ldots, \mathbf{S}_{t_k}$ of the slopes $\dot{\mathbf{X}}_{t_1}, \dot{\mathbf{X}}_{t_2}, \ldots, \dot{\mathbf{X}}_{t_k}$ at K time points, which are substituted in Eq. (16). The result,

$$\mathbf{S} = \mathbf{N} \cdot \mathbf{V}, \qquad (17)$$

with numerical values on the left-hand side, is a system of linear algebraic equations in the flux values $\mathbf{V}$. The fluxes are unknown functions of metabolites, but at each time point, Eq. (17) is a typical algebraic matrix equation. Thus, DFE is an extension of stoichiometric flux balancing into the dynamic,

nonsteady-state realm. It works quite well, if the available time series data capture the true trends in the data.

Under opportune conditions, **N** is square, so that Eq. (17) can be solved at every time point; we will discuss nonsquare matrices in a moment. The solution is a numerical value of every flux in the system at every time point. Collecting the values for a particular flux V_i over time results in a plot of V_i *versus* time. Alternatively, V_i may be plotted against all variables on which it depends, which include substrates and modulators (see Figure 3). Again, such a plot is a collection of points in a two- or higher-dimensional space. Under the assumption that all pertinent metabolites are included in this plot, the plot is a representation of the true flux that is essentially free of assumptions. One cannot tell from the plot which mathematical formulation would optimally capture it, but the plot itself is in some sense a true image of the actual flux.

The stoichiometric matrix **N** usually has more columns than rows, which prevents a direct matrix inversion. One could use a pseudo-inverse to overcome the issue, but the solution typically contains negative flux values, which are not appropriate if the system was set up with fluxes pointing in the right directions. As alternatives, several methods have been proposed to obtain external information regarding a few fluxes, which makes the system invertible.[245–250] Similar issues of underdeterminedness have been addressed in different ways with methods of metabolic flux analysis[251–253] and with the characterization of elementary modes,[254–256] although these analyses had different goals.

The result of a DFE analysis not only shows the shapes of all fluxes, either plotted against time or against their contributing variables, but may also provide hints that something is missing in the model. As an example, Dolatshahi et al.[249] discovered plots that were not true functions of variables, but bent back so that some substrate values were associated with two flux values. This situation led to the targeted, and ultimately successful, search of modulators that had not been known for the species under investigation, although they had been documented for other species.[15,16]

A DFE analysis usually includes the attempt to convert the metabolite-flux plots into explicit mathematical representations. Sometimes the shape of the plot suggests such a representation, but this is not always the case (Figure 3). Nonetheless, this type of fitting is much easier than for the entire ODE system, as it involves one explicit function at a time. As an intriguing novel alternative, which at first glance may appear to be impossible, one can retool DFE for *nonparametric dynamic modeling*.[257] Specifically, one forgoes the last step of flux identification and instead enters the metabolite-flux relationships into a library. This call-up library henceforth replaces explicit functions, but still permits nonparametric model analysis. This approach not only permits dynamic simulations of what-if scenarios, but even steady-state, sensitivity, and stability analyses.

STEP 4: ESTIMATION OF PARAMETER VALUES FOR THE PROCESS REPRESENTATIONS

It is quite evident that the parameter estimation step is tightly connected to Step 3, because parameter values can only be estimated if one has functional representations of all processes that contain parameters. Similar to Step 3, parameter estimation depends very much on the manner in which the model is constructed.[19] In the bottom-up approach, where individual processes are modeled and then assembled into the comprehensive model of interest, parameter values are to be obtained for explicit functions. If data were available for each process individually, the determination of parameter values would be quite simple. However, this is almost never the case, and one has to rely on literature values or on kinetic parameters that had been obtained from *in vitro* experiments. To what degree this *in vitro–in vivo* extrapolation is justified and valid has been discussed widely in the literature (e.g., Refs 258–261).

If the model was constructed top-down from series of metabolite concentrations, all parameter values must be estimated simultaneously, which can cause compensation of errors, as discussed before. In these cases, one can easily identify quite a wrong model which, however, only becomes evident when new data are fitted with the same model, and the model may fail quite spectacularly. Presently the only systematic rescue from this trap of compensation appears to be DFE. In all other cases, error compensation and numerous technical challenges can make the parameter estimation step very difficult. Correspondingly, a Google Scholar search for *Parameter Estimation* yields more than 3.6 million hits. Uncounted reviews have been published describing these challenges, even within the field of biology alone, along with a wide variety of possible solutions.[148,262–280] Unfortunately, powerful reverse engineering and parameter estimation algorithms from other fields are often not easily applicable as they require long, dense time series.[281,282] In spite of strong, concerted efforts, the scientific community is

still awaiting effective solutions that work in most applications. As an aside, parameters in biology are often affected by the lab executing the experiments, so that they are not absolute and should not be over-interpreted. Instead, the more important questions in many cases should be whether a model offers new insights and explanations and whether it is capable of capturing novel situations.

In addition to not being able to obtain good solutions, parameter estimation can lead to entire domains of combinations of parameter values that yield essentially equivalent solutions. This issue of *sloppiness* has been discussed frequently in recent times (e.g., Refs 283–296). Thus, in some cases, there are no good solutions, and in other cases, there are arguably too many. In yet other cases, one obtains a good fit, but this fit is inappropriate for other reasons.[297] One solution to this conundrum is to accept the fact that many parameter combinations may be equally good and search not for one optimal solution, but for an entire ensemble of well-fitting models.[298–303]

DISCUSSION

The purpose of this review was to present the many options the metabolic modeler has for structuring a model and setting up process representations. The challenge of model selection may sound almost philosophical, but it is very real and has a direct impact on the practice of metabolic modeling.

The title of the review alluded to the identification of the best models of metabolism, and given the diversity presented here, there is indeed a good chance that the best models available today are somewhere mentioned in this report. But which ones are they? Before selecting models that may serve at least as the best defaults, one must keep in mind that *all* descriptions of enzymatic processes are approximations. There are no true representations, and the question is therefore which type of approximation is in some sense better than others. Also, one must not forget that the quality or superiority of a model in comparison to others does not necessarily lie in its simplicity or complexity, but is to be judged by how well the model answers the questions that triggered the modeling process in the first place.[10] As a consequence, there will never be a single model type that is best for all purposes, because different modeling goals impose contradictory demands: A model for determining the dose of a drug presumably needs to be as comprehensive as possible, whereas a model revealing the essence of a phenomenon should probably be as simple as possible.

Keeping these fundamental caveats in mind, I would like to proffer some suggestions, based on several decades in the field. In most real-world systems, the numbers of substrate and enzyme molecules are presumably sufficient to allow for the type of averaging that is common to ODE models. Furthermore, throughout almost half a century, the use of power-law functions has proven to strike a good balance between their ability to capture complex phenomena, their mathematical tractability, and their numbers of parameters. For theoretical studies, such as the search for design and operating principles, the S-system format is an excellent starting point, because it permits algebraic computations of steady states, stability, static and dynamic sensitivities,[304–306] and more complex features such as bifurcations to limit cycles,[307,308] and structural design space analysis.[309–313] To judge this advantage as trivial would be a mistake, because this computability opens further analytical avenues, which are not available if the steady state for every new model setting has to be determined with a search algorithm.[104] For applied simulation studies, the S-system's close cousin, the GMA model, may be the preferable default. This format is more intuitive, because every process in the system corresponds uniquely to a term in the ODEs, and every quantity within a term has a unique meaning as 'the strength of the effect of a variable on a specific process.' It is easy to account for modulators in GMA models, and when more variables are included in the model, the number of parameters grows only modestly, at least in comparison with other options. Of course, power-law frameworks do not shed light on reaction mechanisms and they do not really allow spatial or stochastic aspects to be analyzed. But taking all arguments together, power-law models are so transparent, have so many advantages, and make so few assumptions, that they offer a great balance between validity and tractability. Finally, if good time-series data are available, DFE offers insights unmatched by any other formalism. The required datasets are still rare, but will be much more commonplace in the near future. Analyzing many of such datasets with DFE might indeed offer us insights into the true functional shapes of simple or complex metabolic processes.

ACKNOWLEDGMENTS

The author is very grateful to Ms. Mojdeh Faraji and two anonymous reviewers for critically reading the manuscript and offering valuable feedback. This work was supported in part by the following grants: NSF (MCB-0958172, MCB-0946595, and MCB-1517588; PI: EOV; MCB 1411672; PI: Diana Downs; DEB-1241046; PI: Kostas Konstantinides); NIH (1P30ES019776-01A1, Gary W. Miller, PI); and DOE-BESC (DE-AC05-00OR22725; PI: Paul Gilna). BESC, the BioEnergy Science Center, is a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This project was furthermore supported in part by Federal funds from the US National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract HHSN272201200031C, which supports the Malaria Host–Pathogen Interaction Center (MaHPIC). The funding agencies are not responsible for the content of this article.

REFERENCES

1. Michaelis L, Menten ML. Die Kinetik der Invertinwirkung. *Biochem Z* 1913, 49:333–369.
2. Henri MV. *Lois générales de l'action des diastases*. Paris: Hermann; 1903.
3. Savageau MA. Enzyme kinetics in vitro and in vivo: Michaelis-Menten revisited. In: Bittar EE, ed. *Principles of Medical Biology*, vol. 4. Greenwich, CT: JAI Press Inc.; 1995, 93–146.
4. Tummler K, Lubitz T, Schelker M, Klipp E. New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. *FEBS J* 2014, 281:549–571.
5. Voit EO, Sands PJ. Modeling forest growth. I. Canonical approach. *Ecol Model* 1996, 86:51–71.
6. Lineweaver H, Burk D. The determination of enzyme dissociation constants. *J Am Chem Soc* 1934, 56:658–666.
7. Scatchard G. The attractions of proteins for small molecules and ions. *Ann N Y Acad Sci* 1949, 51:660–672.
8. Woolf B. In: Haldane JBS, Stern KG, eds. *Allgemeine Chemie der Enzyme*. Dresden and Leipzig: Steinkopf Verlag; 1932, 19.
9. Voit EO. Modelling metabolic networks using power-laws and S-systems. *Essays Biochem* 2008, 45:29–40.
10. Voit EO. *A First Course in Systems Biology*. Garland Science: New York, NY; 2012.
11. Savageau MA. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Reading, MA: Addison-Wesley Pub. Co; 1976, Advanced Book Program (reprinted 2009).
12. KEGG. The Kyoto encyclopedia of genes and genomes; 2017. Available at: http://www.genome.jp/kegg/pathway.html.
13. BioCyc. BioCyc database collection; 2017. Available at: https://biocyc.org/.
14. Voit EO, Neves AR, Santos H. The intricate side of systems biology. *Proc Natl Acad Sci USA* 2006, 103:9452–9457.
15. Dolatshahi S, Fonseca LL, Voit EO. New insights into the complex regulation of the glycolytic pathway in *Lactococcus lactis*. II. Inference of the precisely timed control system regulating glycolysis. *Mol Biosyst* 2016, 12:37–47.
16. Dolatshahi S, Fonseca LL, Voit EO. New insights into the complex regulation of the glycolytic pathway in *Lactococcus lactis*. I. Construction and diagnosis of a comprehensive dynamic model. *Mol Biosyst* 2016, 12:23–36.
17. BRENDA. The comprehensive enzyme information system; 2017. Available at: http://www.brenda-enzymes.org/.
18. Voit EO. Models-of-data and models-of-processes in the post-genomic era. *Math Biosci* 2002, 180:263–274.
19. Voit EO. The dawn of a new era of metabolic systems analysis. *Drug Discov Today BioSilico* 2004, 2:182–189.
20. Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, Jones DP, Pulendran B. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* 2013, 9:e1003123.
21. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. *Curr Protoc Mol Biol*. 98:30.2:30.2.1–30.2.24.
22. Krumsiek J, Suhre K, Evans AM, Mitchell MW, Mohney RP, Milburn MV, Wagele B, Romisch-Margl W, Illig T, Adamski J, et al. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet* 2012, 8:e1003005.
23. Neves AR, Ramos A, Nunes MC, Kleerebezem M, Hugenholtz J, de Vos WM, Almeida J, Santos H. *In*

vivo nuclear magnetic resonance studies of glycolytic kinetics in *Lactococcus lactis*. *Biotechnol Bioeng* 1999, 64:200–212.

24. Neves AR. Metabolic strategies to reroute carbon fluxes in *Lactococcus lactis*: kinetics of intracellular metabolite pools by *in vivo* NMR. PhD Dissertation. *Instituto de Tecnologia Química e Biológica, Lisboa, Portugal*, 2001.
25. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BO. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci USA* 2013, 110:20338–20343.
26. Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BO. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol* 2010, 4:140.
27. Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol* 2005, 5:8.
28. Friedman N. Inferring cellular networks using probabilistic graphical models. *Science* 2004, 303:799–805.
29. Stone JV. *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*. Sheffield, UK: Sebtel Press; 2013.
30. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput* 2000, 5:415–426.
31. DREAM. Dialogue on reverse-engineering assessment and methods. Available at: http://dreamchallenges.org/about-dream/.
32. Stolovitzky G, Monroe D, Califano A. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann N Y Acad Sci* 2007, 1115:1–22.
33. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005, 37:382–390.
34. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform* 2006, 7(suppl 1):S7.
35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, 13:2498–2504.
36. Zahiri J, Bozorgmehr JH, Masoudi-Nejad A. Computational prediction of protein-protein interaction networks: algorithms and resources. *Curr Genomics* 2013, 14:397–414.
37. Yamanishi Y, Vert J-P, Kanehisa M. Supervised enzyme network inference from the integration of genomic data and chemical information. *Bioinformatics* 2005, 21:1468–1477.
38. Mounet F, Moing A, Garcia V, Petit J, Maucourt M, Deborde C, Bernillon S, Le Gall G, Colquhoun I, Defernez M, et al. Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. *Plant Physiol* 2009, 149:1505–1528.
39. Osorio S, Alba R, Nikoloski Z, Kochevenko A, Fernie AR, Giovannoni JJ. Integrative comparative analyses of transcript and metabolite profiles from pepper and tomato ripening and development stages uncovers species-specific patterns of network regulatory behavior. *Plant Physiol* 2012, 159:1713–1729.
40. Durot M, Bourguignon P-Y, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev* 2009, 33:164–190.
41. Villaverde AF, Ross J, Banga JR. Reverse engineering cellular networks with information theoretic methods. *Cells* 2013, 2:306–329.
42. Ebenhöh O, Kempa S. Genome-scale metabolic network inference. In: Dubitzky W, Wolkenhauer O, Cho K-H, Hiroki Y, eds. *Encyclopedia of Systems Biology*. New York: Springer Verlag; 2013, 832–833.
43. Džeroski S, Todorovski L. Equation discovery for systems biology: finding the structure and dynamics of biological networks from time course data. *Curr Opin Biotechnol* 2008, 19:360–368.
44. Gennemark P, Wedelin D. ODEion—a software module for structural identification of ordinary differential equations. *J Bioinform Comput Biol* 2014, 12:1350015.
45. Okamoto M, Tominaga D, Koga N, Maki Y. Efficient numerical optimization algorithm based on genetic algorithm for inverse problem. In: *Proceedings of 1st International Conference on Systems Biology*; 2000.
46. Voit EO, Radivoyevitch T. Biochemical systems analysis of genome-wide expression data. *Bioinformatics* 2000, 16:1023–1037.
47. Çakir T, Khatibipour MJ. Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation. *Front Bioeng Biotechnol* 2014, 2:62.
48. Alves R, Vilaprinyo E, Sorribas A. Integrating bioinformatics and computational biology: Perspectives and possibilities for in silico network reconstruction in molecular systems biology. *Curr Bioinform* 2008, 3:98–129.
49. Kim HU, Kim TY, Lee SY. Framework for network modularization and Bayesian network analysis to

investigate the perturbed metabolic network. *BMC Syst Biol* 2011, 5(suppl 2):S14.

50. Shafiei M, Dunn KA, Chipman H, Gu H, Bielawski JP. BiomeNet: a Bayesian model for inference of metabolic divergence among microbial communities. *PLoS Comput Biol* 2014, 10:e1003918.
51. Çakir T, Hendriks MM, Westerhuis JA, Smilde AK. Metabolic network discovery through reverse engineering of metabolome data. *Metabolomics* 2009, 5:318–329.
52. Frainay C. *Jourdan F*. Brief Bioinform: Computational methods to identify metabolic sub-networks based on metabolomic profiles; 2016.
53. Albert R. Network inference, analysis, and modeling in systems biology. *Plant Cell* 2007, 19:3327–3338.
54. Hendrickx DM, Hendriks MM, Eilers PH, Smilde AK, Hoefsloot HC. Reverse engineering of metabolic networks, a critical assessment. *Mol Biosyst* 2011, 7:511–520.
55. Steinway SN, Biggs MB, Loughran TP Jr, Papin JA, Albert R. Inference of network dynamics and metabolic interactions in the gut microbiome. *PLoS Comput Biol* 2015, 11:e1004338.
56. Villaverde AF, Banga JR. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J R Soc Interface* 2014, 11:20130505.
57. Crampin EJ, Schnell S, McSharry PE. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog Biophys Mol Biol* 2004, 86:77–112.
58. Voit EO, Almeida J. Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics* 2004, 20:1670–1681.
59. Voit EO, Savageau MA. Power-law approach to modeling biological systems; III. Methods of analysis. *J Ferment Technol* 1982, 60:223–241.
60. Voit EO, Savageau MA. Power-law approach to modeling biological systems; II. Application to ethanol production. *J Ferment Technol* 1982, 60:229–232.
61. Liu PK, Wang F-S. Inference of biochemical network models in S-system using multiobjective optimization approach. *Bioinformatics* 2008, 24:1085–1092.
62. Searson DP, Willis MJ, Horne SJ, Wright AR. Inference of chemical reaction networks using hybrid S-system models. *Chem Prod Process Model* 2007, 2:1–21.
63. Kruger NJ, Ratcliffe RG. Fluxes through plant metabolic networks: measurements, predictions, insights and challenges. *Biochem J* 2015, 465:27–38.
64. Nikoloski Z, Perez-Storey R, Sweetlove LJ. Inference and prediction of metabolic network fluxes. *Plant Physiol* 2015, 169:1443–1455.
65. Wiechert W. 13C metabolic flux analysis. *Metab Eng* 2001, 3:195–206.
66. Wiechert W, Mollney M, Petersen S, de Graaf AA. A universal framework for 13C metabolic flux analysis. *Metab Eng* 2001, 3:265–283.
67. Wiechert W, Noh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Curr Opin Biotechnol* 2013, 24:979–986.
68. Toya Y, Shimizu H. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol Adv* 2013, 31:818–826.
69. Kelleher JK. Flux estimation using isotopic tracers: common ground for metabolic physiology and metabolic engineering. *Metab Eng* 2001, 3:100–110.
70. Gavalas GR. *Nonlinear Differential Equations of Chemically Reacting Systems*. Berlin: Springer-Verlag; 1968.
71. Heinrich R, Schuster S. *The Regulation of Cellular Systems*. New York: Chapman and Hall; 1996.
72. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Curr Opin Biotechnol* 2003, 14:491–496.
73. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol* 2010, 28:245–248.
74. Raman K, Chandra N. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform* 2009, 10:435–449.
75. Palsson BØ. *Systems Biology: Properties of Reconstructed Networks*. New York: Cambridge University Press; 2006.
76. Arita M. Metabolic reconstruction using shortest paths. *Simul Model Pract Theory* 2000, 8:109–112.
77. Boyer. *Ab Initio* reconstruction of metabolic pathways. *Bioinformatics* 2003, 9:26–34.
78. Planes F, Beasley J. Path finding approaches and metabolic pathways. *Discrete Appl Math* 2009, 157:2244–2256.
79. El Kaissi M, Jia M, Reiners D, Dickerson J, Wuertele E. Reaction centric layout for metabolic networks. In: Bebis G, Boyle R, Parvin B, Koracin D, Kuno Y, Wang J, Pajarola R, Lindstrom P, Hinkenjann A, Encarnação ML, et al., eds. *Advances in Visual Computing (ISVC 2009), Lecture Notes in Computer Science*, vol. 5876. Berlin, Heidelberg: Springer; 2009, 81–91.
80. Sridharan GV, Hassoun S, Lee K. Identification of biochemical network modules based on shortest retroactive distances. *PLoS Comput Biol* 2011, 7: e1002262.
81. Marin-Sanguino A, Mendoza E, Voit EO. Flux duality in non-linear GMA systems: implications for metabolic engineering. *J Biotechnol* 2009, 149:166–172.
82. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, et al. Multiple high-throughput analyses monitor the

response of *E. coli* to perturbations. *Science* 2007, 316:593–597.

83. Sorribas A, March J, Voit EO. Estimating age-related trends in cross-sectional studies using S-distributions. *Stat Med* 2000, 19:697–713.
84. Pearson K. Contributions to the mathematical theory of evolution, II: skew variation in homogeneous materia. *Philos Trans R Soc* 1895, 186:443–459.
85. Pearson K. Mathematical contributions to the theory of evolution, XIX: second supplement to a memoir on skew variation. *Philos Trans R Soc* 1916, 216:429–457.
86. Savageau MA. A suprasystem of probability distributions. *Biom J* 1982, 24:323–330.
87. Johnson ME, Tietjen GL, Beckman RJ. A new family of probability distributions with applications to Monte Carlo studies. *J Am Stat Assoc* 1980, 75:276–279.
88. Johnson N, Kotz S. *Continuous Univariate Distributions*. Boston: Houghton Mifflin; 1970.
89. Savageau MA, Voit EO. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Mathem Biosci* 1987, 87:83–115.
90. Voit EO. The S-distribution. A tool for approximation and classification of univariate, unimodal probability distributions. *Biom J* 1992, 34:855–878.
91. Voit EO, Yu S. The S-distribution. Approximation of discrete distributions. *Biom J* 1994, 36:205–219.
92. Sorribas A, Muiño JM, Voit EO. GS-distributions: a new family of distributions for continuous unimodal variables. *Comput Stat Data Anal* 2006, 50:2769–2798.
93. Savageau MA. Growth of complex systems can be related to the properties of their underlying determinants. *Proc Natl Acad Sci USA* 1979, 76:5413–5417.
94. Savageau MA. Growth equations: a general equation and a survey of special cases. *Math Biosci* 1980, 48:267–278.
95. Voit EO. Recasting nonlinear models as S-systems. *Math Comput Model* 1988, 11:140–145.
96. Voit EO, ed. *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*. New York: Van Nostrand Reinhold; 1991, xi +365 pp.
97. Curto R, Voit EO, Sorribas A, Cascante M. Mathematical models of purine metabolism in man. *Math Biosci* 1998, 151:1–49.
98. Alvarez-Vasquez F, Sims KJ, Hannun YA, Voit EO. Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models. *J Theor Biol* 2004, 226:265–291.
99. del Rosario RC, Mendoza E, Voit EO. Challenges in lin-log modelling of glycolysis in *Lactococcus lactis*. *IET Syst Biol* 2008, 2:136–149.
100. Wang F-S, Ko C-L, Voit EO. Kinetic modeling using S-systems and lin-log approaches. *Biochem Eng J* 2007, 33:238–247.
101. Heijnen JJ. Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng* 2005, 91:534–545.
102. Dräger A, Kronfeld M, Ziller MJ, Supper J, Planatscher H, Magnus JB, Oldiges M, Kohlbacher O, Zell A. Modeling metabolic networks in *C. glutamicum*: a comparison of rate laws in combination with various parameter optimization strategies. *BMC Syst Biol* 2009, 3:5.
103. Savageau MA. Biochemical systems theory: operational differences among variant representations and their significance. *J Theor Biol* 1991, 151:509–530.
104. Shiraishi F, Savageau MA. The tricarboxylic acid cycle in *Dictyostelium discoideum*. I. Formulation of alternative kinetic representations. *J Biol Chem* 1992, 267:22912–22918.
105. Sorribas A, Savageau MA. Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways. *Math Biosci* 1989, 94:239–269.
106. Sorribas A, Savageau MA. A comparison of variant theories of intact biochemical systems. II. Flux-oriented and metabolic control theories. *Math Biosci* 1989, 94:195–238.
107. Sorribas A, Savageau MA. A comparison of variant theories of intact biochemical systems. I. Enzyme-enzyme interactions and biochemical systems theory. *Math Biosci* 1989, 94:161–193.
108. Sorribas A, Curto R, Cascante M. Comparative characterization of the fermentation pathway of Saccharomyces cerevisiae using biochemical systems theory and metabolic control analysis: model validation and dynamic behavior. *Math Biosci* 1995, 130:71–84.
109. Cascante M, Curto R, Sorribas A. Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: steady-state analysis. *Math Biosci* 1995, 130:51–69.
110. Hadlich F, Noack S, Wiechert W. Translating biochemical network models between different kinetic formats. *Metab Eng* 2009, 11:87–100.
111. Voit EO, Savageau MA. Accuracy of alternative representations for integrated biochemical systems. *Biochemistry* 1987, 26:6869–6880.
112. Alvarez-Vasquez F, Gonzalez-Alcon C, Torres NV. Metabolism of citric acid production by *Aspergillus niger*: model definition, steady-state analysis and constrained optimization of citric acid production rate. *Biotechnol Bioeng* 2000, 70:82–108.
113. Garcia J, Torres N. Mathematical modelling and assessment of the pH homeostasis mechanisms in

Aspergillus niger while in citric acid producing conditions. *J Theor Biol* 2011, 282:23–35.

114. Torres NV, Voit EO. *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge: Cambridge University Press; 2002.
115. Torres NV, Rodríguez F, Glez-Alcón C, Voit EO. A novel approach to designing an overexpression strategy for metabolic engineering. Application to the carbohydrate metabolism in the citric acid producing mould *Aspergillus niger*. *Food Technol Biotechnol* 1998, 36:177–184.
116. Guldberg CM, Waage P. Studier i affiniteten. *Forh i Vidensk Selsk Christiania* 1864, 35:35–45.
117. Guldberg CM, Waage P. *Études sur les affinites chimiques*. Christiania: Brøgger & Christie; 1867.
118. Guldberg CM, Waage P. Über die chemische Affinität. *Erdmann's J Prakt Chem* 1879, 127:69–114.
119. Voit EO, Martens HA, Omholt SW. 150 years of the mass action law. *PLoS Comput Biol* 2015, 11: e1004012.
120. Hill AV. Possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol* 1910, 40:iv–viii.
121. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics. *R Stat Soc J* 1927, 115:700–721.
122. Lotka A. *Elements of Physical Biology*. Baltimore: Williams and Wilkins; reprinted as 'Elements of Mathematical Biology'. Dover, New York, 1956; 1924.
123. Peschel M, Mende W. *The Predator–prey Model: Do we Live in a Volterra World?* Berlin: Akademie-Verlag; 1986.
124. May RM. *Stability and Complexity in Model Ecosystems*. Princeton, NJ: Princeton University Press; 1973.
125. Volterra V. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem Rend Accad Lincei* 1926, 2:31–113.
126. Eicher JJ, Snoep JL, Rohwer JM. Determining enzyme kinetics for systems biology with nuclear magnetic resonance spectroscopy. *Metabolites* 2012, 2:818–843.
127. Segel LE. *Biological Kinetics*. Cambridge: Cambridge University Press; 1991.
128. Cornish-Bowden A. *Fundamentals of Enzyme Kinetics*. 4th ed. Wiley-Blackwell: Weinheim; 2012.
129. Sauro HM. *Enzyme Kinetics for Systems Biology*. Seattle, WA: Ambrosius Publishing and Future Skill Software; 2012.
130. Schulz AR. *Enzyme Kinetics: From Diastase to Multienzyme Systems*. Cambridge; New York: Cambridge University Press; 1994.
131. Blangy D, Buc H, Monod J. Kinetics of the allosteric interactions of phosphofructokinase from *Escherichia coli*. *J Mol Biol* 1968, 31:13–35.
132. Monod J, Wyman J, Changeux J-P. On the nature of allosteric transitions: a plausible model. *J Mol Biol* 1965, 12:88–118.
133. Otto M, Heinrich R, Kuhn B, Jacobasch G. A mathematical model for the influence of fructose 6-phosphate, ATP, potassium, ammonium and magnesium on the phosphofructokinase from rat erythrocytes. *Eur J Biochem* 1974, 49:169–178.
134. Mulquiney PJ, Kuchel PW. Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement. *Biochem J* 1999, 342(Pt 3):581–596.
135. Peskov K, Goryanin I, Demin O. Kinetic model of phosphofructokinase-1 from *Escherichia coli*. *J Bioinform Comput Biol* 2008, 6:843–867.
136. Pham AS, Reinhart GD. Pre-steady state quantification of the allosteric influence of *Escherichia coli* phosphofructokinase. *J Biol Chem* 2001, 276:34388–34395.
137. Symcox MM, Reinhart GD. A steady-state kinetic method for the verification of the rapid-equilibrium assumption in allosteric enzymes. *Anal Biochem* 1992, 206:394–399.
138. Auzat I, Le Bras G, Garel JR. Hypercooperativity induced by interface mutations in the phosphofructokinase from *Escherichia coli*. *J Mol Biol* 1995, 246:248–253.
139. Johnson JL, Reinhart GD. MgATP and fructose 6-phosphate interactions with phosphofructokinase from *Escherichia coli*. *Biochemistry* 1992, 31:11510–11518.
140. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol Bioeng* 2002, 79:53–73.
141. Galazzo JL, Bailey JE. Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *Saccharomyces cerevisiae*. *Enzyme Microb Technol* 1990, 12:162–172.
142. Curto R, Sorribas A, Cascante M. Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: model definition and nomenclature. *Math Biosci* 1995, 130:25–50.
143. Savageau MA. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *J Theor Biol* 1969, 25:370–379.
144. Savageau MA. Biochemical systems analysis. I. Some mathematical properties of the rate law for the

component enzymatic reactions. *J Theor Biol* 1969, 25:365–369.

145. Savageau MA. Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol* 1970, 26:215–226.
146. Savageau MA. The behavior of intact biochemical control systems. *Curr Top Cell Regul* 1972, 6:63–129.
147. Voit EO. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. Cambridge: Cambridge University Press; 2000.
148. Voit EO. Biochemical systems theory: a review. *ISRN Biomathematics* 2013, Article ID 897658, 53 pages. doi:10.1155/2013/897658.
149. Voit EO. Canonical modeling: a review of concepts with emphasis on environmental health. *Environ Health Perspect* 2000, 108(suppl 5):895–909.
150. Savageau MA, Voit EO, Irvine DH. Biochemical systems theory and metabolic control theory. I. Fundamental similarities and differences. *Math Biosci* 1987, 86:127–145.
151. Savageau MA, Voit EO, Irvine DH. Biochemical systems theory and metabolic control theory. II. The role of summation and connectivity relationships. *Math Biosci* 1987, 86:147–169.
152. Irvine DH, Savageau MA. Network regulation of the immune response: modulation of suppressor lymphocytes by alternative signals including contrasuppression. *J Immunol* 1985, 134:2117–2130.
153. Irvine DH, Savageau MA. Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J Immunol* 1985, 134:2100–2116.
154. Voit EO. Optimization in integrated biochemical systems. *Biotechnol Bioeng* 1992, 40:572–582.
155. Torres NV, Voit EO, Glez-Alcón C, Rodríguez F. An indirect optimization method for biochemical systems: description of method and application to the maximization of the rate of ethanol, glycerol, and carbohydrate production in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 1997, 55:758–772.
156. Berg PH, Voit EO, White RL. A pharmacodynamic model for the action of the antibiotic imipenem on *Pseudomonas aeruginosa* populations in vitro. *Bull Math Biol* 1996, 58:923–938.
157. Visser D, Heijnen JJ. The mathematics of metabolic control analysis revisited. *Metab Eng* 2002, 4:114–123.
158. Visser D, Heijnen JJ. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab Eng* 2003, 5:164–176.
159. Wu L, Wang W, van Winden WA, van Gulik WM, Heijnen JJ. A new framework for the estimation of control parameters in metabolic pathways using linlog kinetics. *Eur J Biochem* 2004, 271:3348–3359.
160. Fell DA. *Understanding the Control of Metabolism*. London: Portland Press; 1997.
161. Heinrich R, Rapoport TA. A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector. *Eur J Biochem* 1974, 42:97–105.
162. Kacser H, Burns JA. The control of flux. *Symp Soc Exp Biol* 1973, 27:65–104.
163. Moreno-Sanchez R, Saavedra E, Rodriguez-Enriquez S, Olin-Sandoval V. Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways. *J Biomed Biotechnol* 2008, 2008:597913.
164. Westerhoff HV, van Heeswijk W, Kahn D, Kell DB. Quantitative approaches to the analysis of the control and regulation of microbial metabolism. *Antonie Van Leeuwenhoek* 1991, 60:193–207.
165. Sorribas A, Hernandez-Bermejo B, Vilaprinyo E, Alves R. Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnol Bioeng* 2007, 97:1259–1277.
166. Alves R, Vilaprinyo E, Hernadez-Bermejo B, Sorribas A. Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways. *Biotechnol Genet Eng Rev* 2008, 25:1–40.
167. Wayman JA, Sagar A, Varner JD. Dynamic modeling of cell-free biochemical networks using effective kinetic models. *Processes* 2015, 3:138–160.
168. Rohwer JM, Hanekom AJ, Hofmeyr J-H. A universal rate equation for systems biology. In: *2Nd Beilstein ESCEC Symposium (Experimental Standard Conditions of Enzyme Characterization)*, Rüdesheim/Rhein, Germany, 2006.
169. Brenig L, Goriely A. Universal canonical forms for time-continuous dynamical systems. *Phys Rev A* 1989, 40:4119–4122.
170. Kerner EH. Universal formats for nonlinear ordinary differential systems. *J Math Phys* 1981, 22:1366–1371.
171. Hernández-Bermejo B, Fairén V. Lotka-Volterra representation of general nonlinear systems. *Math Biosci* 1997, 140:1–32.
172. Visser D, van der Heijden R, Mauch K, Reuss M, Heijnen S. Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. *Metab Eng* 2000, 2:252–275.
173. Marino S, Voit EO. An automated procedure for the extraction of metabolic network information from

time series data. *J Bioinform Comput Biol* 2006, 4:665–691.

174. Ponosov A, Machina A. *Tafinseva V*. Georgian Math J: Piecewise synergetic systems and applications in biochemical systems theory; 2016.
175. Savageau MA. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* 2001, 11:142–159.
176. Machina A, Ponosov A, Voit EO. Automated piecewise power-law modeling of biological systems. *J Biotechnol* 2009.
177. Löwe H, Kremling A, Marin-Sanguino A. Time hierarchies and model reduction in canonical non-linear models. *Front Genet* 2016, 7:166.
178. Ferreira AE, Ponces Freire AM, Voit EO. A quantitative model of the generation of N(epsilon)-(carboxymethyl)lysine in the Maillard reaction between collagen and glucose. *Biochem J* 2003, 376:109–121.
179. Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT. A cybernetic view of microbial growth: modeling of cells as optimal strategists. *Biotechnol Bioeng* 1985, 27:1–9.
180. Sweetlove LJ, Fernie AR. The spatial organization of metabolism within the plant cell. *Annu Rev Plant Biol* 2013, 64:723–746.
181. Chen J, Gomez JA, Hoffner K, Phalak P, Barton PI, Henson MA. Spatiotemporal modeling of microbial metabolism. *BMC Syst Biol* 2016, 10:21.
182. Cowan AE, Moraru II, Schaff JC, Slepchenko BM, Loew LM. Spatial modeling of cell signaling networks. *Methods Cell Biol* 2012, 110:195–221.
183. Henson MA. Genome-scale modelling of microbial metabolism with temporal and spatial resolution. *Biochem Soc Trans* 2015, 43:1164–1171.
184. Klitgord N, Segre D. The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome Inform* 2010, 22:41–55.
185. Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, Rocha I. Modeling formalisms in systems biology. *AMB Express* 2011, 1:45.
186. Alvarez-Vasquez F, Riezman H, Hannun YA, Voit EO. Mathematical modeling and validation of the ergosterol pathway in *Saccharomyces cerevisiae*. *PLoS One* 2011, 6:e28344.
187. Agmon E, Gates AJ, Churavy V, Beer RD. Exploring the space of viable configurations in a model of metabolism–boundary co-construction. *Artif Life* 2016, 22:153–171.
188. Neff KL. Biochemical reaction kinetics in dilute and crowded solutions: Predictions of macroscopic and mesoscopic models and experimental observations. Doctoral Dissertation, *Mayo Clinic, Rochester, MN*, 2010.
189. Neff KL, Offord SP, Caride AJ, Strehler EE, Prendergast FG, Bajzer Ž. Validation of fractal-like kinetic models by time-resolved binding kinetics of dansylamide and carbonic anhydrase in crowded media. *Biophys J* 2011, 100:2495–2503.
190. Savageau MA. Michaelis-Menten mechanism reconsidered: implications of fractal kinetics. *J Theor Biol* 1995, 176:115–124.
191. Savageau MA. Fractal kinetic effects on equilibrium. In: Yamakawa T, Matsumoto G, eds. *Methodologies for the Conception, Design, and Application of Intelligent Systems*, vol. 1. Singapore: World Scientific; 1996, 187–190.
192. Savageau MA. Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *Biosystems* 1998, 47:9–36.
193. Schnell S, Turner TE. Reaction kinetics environments with macromolecular crowding: laws. *Prog Biophys Mol Biol* 2004, 85:235–260.
194. Bajzer Ž. Mathematical analysis of models for reaction kinetics in intracellular environments. *Math Biosci* 2008, 215:35–47.
195. elik E, Aykut A, Bayram M. The ordinary successive approximations method and Padé approximants for solving a differential equation with variant retarded argument. *Appl Math Comput* 2003, 44:173–180.
196. Mocek WT, Rudnicki R, Voit EO. Approximation of delays in biochemical systems. *Math Biosci* 2005, 198:190–216.
197. Yin W, Jo H, Voit EO. Systems analysis of the role of bone morphogenic protein 4 in endothelial inflammation. *Ann Biomed Eng* 2010, 38:291–307.
198. Wu J, Qi Z, Voit EO. Investigation of delays and noise in dopamine signaling with hybrid functional Petri nets. *In Silico Biol* 2010, 10:0005.
199. Zhang J, Knospe CR, Tsiotras P. New results for the analysis of linear systems with time-invariant delays. *Int J Robust Nonlinear Control* 2003, 13:1149–1175.
200. Ullah M, Wolkenhauer O. *Stochastic Approaches for Systems Biology*. New York: Springer-Verlag; 2011.
201. Gillespie DT. A rigorous derivation of the chemical master equation. *Physica A* 1992, 188:404–425.
202. Roussel MR. Modelling biochemical reaction networks; Lecture 14: Stochastic theory of reaction kinetics; 2017. Available: people.uleth.ca/~roussel/C4000biochemnet/slides/14CME.pdf.
203. Burchard H, Deleersnijder E, Meister A. A high-order conservative Patankar-type discretisation for stiff systems of production–destruction equations. *Appl Num Math* 2003, 47:1–30.
204. Khammash M. The chemical master equation in gene networks: complexity and approaches; 2017. Available at: https://www.cds.caltech.edu/~murray/wiki/

images/d/d9/Khammash_master-15aug06.pdf (Accessed 2017).

205. López-Caamal F, Marquez-Lago TT. Order reduction of the chemical master equation via balanced realisation. *PLoS One* 2014, 9:e103521.
206. Wu J, Vidakovic B, Voit EO. Constructing stochastic models from deterministic process equations by propensity adjustment. *BMC Syst Biol* 2011, 5:187.
207. Hahl SK, Kremling A. A comparison of deterministic and stochastic modeling approaches for biochemical reaction systems: On fixed points, means, and modes. *Front Genet* 2016, 7:157.
208. Ao P. Metabolic network modelling: Including stochastic effects. *Comp Chem Eng* 2005, 29:2297–2303.
209. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 1976, 22:403–434.
210. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 1977, 81:2340–2361.
211. Cao Y, Petzold L. Slow scale tau-leaping method. *Comput Methods Appl Mech Eng* 2008, 197:3472–3479.
212. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys* 2001, 115:1716–1733.
213. Gillespie DT. Stochastic simulation of chemical kinetics. *Ann Rev Phys Chem* 2007, 58:35–55.
214. Blättke MA, Heiner M, Marwan W. Biomodel engineering with petri nets. In: Robeva R, ed. *Algebraic and Discrete Mathematical Methods for Modern Biology*. Amsterdam: Academic Press; 2015.
215. Koch I, Reisig W, Schreiber F. *Modeling in Systems Biology. The Petri Net Approach*. London: Springer Verlag; 2011.
216. Petri CA. Kommunikation mit Automaten. PhD Dissertation, *Technische Hochschule Darmstadt, Germany*; 1962.
217. Kohn MC, Letzkus WJ. A graph-theoretical analysis of metabolic regulation. *J Theor Biol* 1983, 100:293–304.
218. Reddy VN, Mavrovouniotis ML, Liebman MN. Petri net representations in metabolic pathways. *Proc Int Conf Intell Syst Mol Biol* 1993, 1:328–336.
219. Barjis I, Gehlot V. Petri net based description and modeling of metabolic pathways. In: *Summer Computer Simulation Conference (SCSC'07)*. San Diego, CA: The Society for Modeling and Simulation International (SCS); 2007.
220. Goss PJE, Peccoud J. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc Natl Acad Sci USA* 1998, 95:6750–6755.
221. Hardy S, Robillard PN. Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. *J Bioinform Comput Biol* 2004, 2:595–613.
222. Murata T. Petri nets: properties, analysis and applications. *Proc IEEE* 1989, 77:541–579.
223. Matsuno H, Tanaka Y, Aoshima H, Doi A, Matsui M, Miyano S. Biopathways representation and simulation on hybrid functional Petri net. *In Silico Biol* 2003, 3:389–404.
224. David R, Alla H. *Discrete, Continuous, and Hybrid Petri Nets*. Berlin: Springer-Verlag; 2005.
225. Wu J, Voit E. Hybrid modeling in biochemical systems theory by means of functional petri nets. *J Bioinform Comput Biol* 2009, 7:107–134.
226. Marin-Sanguino A, Mendoza ER. Hybrid modeling in computational neuropsychiatry. *Pharmacopsychiatry* 2008, 41(Suppl 1):S85–S88.
227. Watson MR. A discrete model of bacterial metabolism. *Comput Appl Biosci* 1986, 2:23–27.
228. Manca V, Castellini A, Franco G, Pagliarini R. Metabolic P systems: a discrete model for biological dynamics. *Chinese J Electron* 2013, 22:717–723.
229. Păun G. Computing with membranes. *J Comput Syst Sci* 2000, 61:108–143.
230. Asenjo AJ, Ramirez P, Rapaport I, Aracena J, Goles E, Andrews BA. A discrete mathematical model applied to genetic regulation and metabolic networks. *J Microbiol Biotechnol* 2007, 17:496–510.
231. Veliz-Cuba A, Jarrah AS, Laubenbacher R. Polynomial algebra of discrete models in systems biology. *Bioinformatics* 2010, 26:1637–1643.
232. Nehaniv CL, Rhodes J, Egri-Nagy A, Dini P, Morris ER, Horváth G, Karimi F, Schreckling D, Schilstra MJ. Symmetry structure in discrete models of biochemical systems: natural subsystems and the weak control hierarchy in a new model of computation driven by interactions. *Philos Trans A Math Phys Eng Sci* 2015, 373:20140223–2014268.
233. Donaldson R, Gilbert D. A model checking approach to the parameter estimation of biochemical pathways. *Proc. CMSB*; 2008, 5307 of LNCS/LNBI. Berlin: Springer.
234. Calzone L, Chabrier-Rivier N, Fages F, Soliman S. Machine learning biochemical networks from temporal logic properties. *Trans Comput Syst Biol VI* 2006, 4220:68–94.
235. Mishra B. Algebraic systems biology: theses and hypotheses. In: Anai H, Horimoto K, Kutsia T, eds. *Algebraic Biology*, vol. 4545. Berlin, Heidelberg: Springer-Verlag; 2007, 1–14 (Lecture Notes in Computer Science).
236. Antoniotti M, Piazza C, Policriti A, Simeoni M. Taming the complexity of biochemical models through

bisimulation and collapsing: theory and practice. *Theory Comp Sci* 2004, 325:45–67.

237. Antoniotti M, Policriti A, Ugel N, Mishra B. XS-systems: eXtended S-systems and algebraic differential automata for modeling cellular behavior In: *High performane computing - HIPC 2002*. Vol. 2552/2002: Lecture Notes in Computer Science; 2002, 431–442.
238. Antoniotti M, Park F, Policriti A, Ugel N, Mishra B. Foundations of a query and simulation system for the modeling of biochemical and biological processes. *Pac Symp Biocomput* 2003, 8:116–127. ISBN: 981-238-217-8.
239. Mishra B, Antoniotti M, Paxia S, Ugel N. Simpathica: a computational systems biology tool within the valis bioinformatics environment. In: Kriete A, Eils R, eds. *Computational Systems Biology*. Burlington, MA: Elsevier Academic Press; 2006, 79–102.
240. Gevorgyan A, Poolman MG, Fell DA. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics* 2008, 24:2245–2251.
241. Goel G, Chou IC, Voit EO. System estimation from metabolic time-series data. *Bioinformatics* 2008, 24:2505–2511.
242. Eilers PHC. A perfect smoother. *Anal Chem* 2003, 75:3631–3636.
243. Vilela M, Borges CC, Vinga S, Vasconcelos AT, Santos H, Voit EO, Almeida JS. Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinform* 2007, 8:305.
244. Dolatshahi S, Vidakovic B, Voit EO. A constrained wavelet smoother for pathway identification tasks in systems biology. *Comput Chem Eng* 2014, 71:728–733.
245. Chou I-C, Voit EO. Estimation of dynamic flux profiles from metabolic time series data. *BMC Syst Biol* 2012, 6:84.
246. Iwata M, Shiraishi F, Voit EO. Coarse but efficient identification of metabolic pathway systems. *Int J Syst Biol* 2013, 4:57–72.
247. Voit EO. Characterizability of metabolic pathway systems from time series data. *Math Biosci* 2013, 246:315–325.
248. Voit EO, Goel G, Chou I-C, da Fonseca L. Estimation of metabolic pathway systems from different data sources. *IET Syst Biol* 2009, 3:513–522.
249. Dolatshahi S, Voit E. Identifiation of dynamic fluxes from metabolic time series data. *Front Genet* 2013, 7:6.
250. Faraji M, Voit EO. Stepwise inference of likely dynamic flux distributions from metabolic time series data. *(in review)*; 2017.
251. Bonarius HPJ, Schmidt G, Tramper J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotchnol* 1997, 15:308–314.
252. Zamorano F, Wouwer AV, Bastin G. A detailed metabolic flux analysis of an underdetermined network of CHO cells. *J Biotechnol* 2010, 150:497–508.
253. Klamt S, Schuster S. Calculating as many fluxes as possible in underdetermined metabolic networks. *Mol Biol Rep* 2002, 29:243–248.
254. Llaneras F, Picó J. An interval approach for dealing with flux distributions and elementary modes activity patterns. *J Theor Biol* 2007, 246:290–308.
255. Trinh CT, Wlaschin A, Srienc F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* 2009, 81:813–826.
256. Poolman MG, Venakatesh KV, Pidcock MK, Fell DA. A method for the determination of flux in elementary modes, and its application to *Lactobacillus rhamnosus*. *Biotechnol Bioeng* 2004, 88:601–612.
257. Faraji M. *Voit EO*. Math Biosci: Nonparametric dynamic modeling; 2016.
258. Albe KR, Butler MH, Wright BE. Cellular concentrations of enzymes and their substrates. *J Theor Biol* 1990, 143:163–195.
259. Blaauboer BJ. Biokinetic modeling and in vitro-in vivo extrapolations. *J Toxicol Environ Health B Crit Rev* 2010, 13:242–252.
260. Yang Y, Himmelstein MW, Clewell HJ. Kinetic modeling of beta-chloroprene metabolism: Probabilistic in vitro-in vivo extrapolation of metabolism in the lung, liver and kidneys of mice, rats and humans. *Toxicol In Vitro* 2012, 26:1047–1055.
261. Houston JB. Utility of in vitro drug metabolism data in predicting in vivo metabolic clearance. *Biochem Pharmacol* 1994, 47:1469–1479.
262. Mendes P, Kell D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 1998, 14:869–883.
263. Villaverde AF, Banga JR. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Interface* 2013, 11:20130505.
264. Chou I-C, Voit EO. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci* 2009, 219:57–83.
265. Gennemark P, Wedelin D. Efficient algorithms for ordinary differential equation model identification of biological systems. *IET Syst Biol* 2007, 1:120–129.
266. Gennemark P, Wedelin D. Benchmarks for identification of ordinary differential equations from time series data. *Bioinformatics* 2009, 25:780–786.
267. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models. *FEBS J* 2009, 276:886–902.

268. Banga JR, Balsa-Canto E. Parameter estimation and optimal experimental design. *Essays Biochem* 2008, 45:195–209.
269. Vilela M, Chou IC, Vinga S, Vasconcelos AT, Voit EO, Almeida JS. Parameter optimization in S-system models. *BMC Syst Biol* 2008, 2:35.
270. Clermont G, Zenker S. The inverse problem in mathematical biology. *Math Biosci* 2015, 260:11–15.
271. Penas DR, González P, Egea JA, Doallo R, Banga JR. Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy. *BMC Bioinform* 2017, 18:52.
272. Baker SM, Schallau K, Junker BH. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J Integr Bioinform* 2010, 7:133–141.
273. Gugushvili S, Klaassen CAJ. √n-consistent parameter estimation for systems of ordinary differential equations : bypassing numerical integration via smoothing. *Bernoulli* 2012, 18:1061–1098.
274. Kunna MA, Kadir TAA, Jaber AS, Odili J. Large-scale kinetic parameter identification of metabolic network model of *E. coli* using PSO. *Adv Biosci Biotechnol* 2015, 6:120–130.
275. Tucker W, Kutalik Z, Moulton V. Estimating parameters for generalized mass action models using constraint propagation. *Math Biosci* 2007, 208:607–620.
276. Guillen-Gosalbez G, Miro A, Alves R, Sorribas A, Jimenez L. Identification of regulatory structure and kinetic parameters of biochemical networks via mixed-integer dynamic optimization. *BMC Syst Biol* 2013, 7:113.
277. Brunel NJ-B. Parameter estimation of ODE's via nonparametric estimators. *Electron J Stat* 2008, 2:1242–1267.
278. Goel G, Chou IC, Voit EO. Biological systems modeling and analysis: a biomolecular technique of the twenty-first century. *J Biomol Tech* 2006, 17:252–269.
279. Ko C-L, Voit E, Wang FS. Estimating parameters for generalized mass action models with connectivity information. *BMC Bioinform* 2009, 10:140–149.
280. Polisetty PK, Voit EO, Gatzke EP. Identification of metabolic system parameters using global optimization methods. *BMC Theoret Biol Med Model* 2006, 3:4.
281. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. *J R Stat Soc Ser B* 2007, 69:741–796.
282. Campbell D, Steele RJ. Smooth functional tempering for nonlinear differential equation models. *Stat Comput* 2012, 22:429–443.
283. Němcová J. Structural identifiability of polynomial and rational systems. *Math Biosci* 2010, 223:83–96.
284. Raue A, Kreutz C, Theis F, Timmer J. Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *J Philos Trans R Soc A* 2013, 371:20110544.
285. Liu Y, Gunawan R. Parameter estimation of dynamic biological network models using integrated fluxes. *BMC Syst Biol* 2014, 8:127–137.
286. Jia G, Stephanopoulos G, Gunawan R. Parameter estimation of kinetic models from metabolic profiles. *Bioinformatics* 2011, 27:1964–1970.
287. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 2007, 3:1871–1878.
288. Tafinseva V, Tøndel K, Ponosov A, Martens HA. Global structure of sloppiness in a nonlinear model. *J Chemomet* 2014, 28:645–655.
289. Tönsing C, Timmer J, Kreutz C. Cause and cure of sloppiness in ordinary differential equation models. *arXiv:1406.1734v1 [q-bio.MN]*; 2014.
290. Sands PJ, Voit EO. Flux-based estimation of parameters in S-systems. *Ecol Model* 1996, 93:75–88.
291. Vilela M, Vinga S, Maia MA, Voit EO, Almeida JS. Identification of neutral sets of biochemical network models from time series data. *BMC Syst Biol* 2009, 3:47.
292. Miao H, Xia X, Perelson AS, Wu H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev Soc Ind Appl Math* 2011, 53:3–39.
293. Eisenberg MC, Hayashi MA. Determining identifiable parameter combinations using subset profiling. *Math Biosci* 2014, 256:116–126.
294. Srinath S, Gunawan R. Parameter identifiability of power-law biochemical system models. *J Biotechnol* 2010, 149:132–140.
295. Chris O-T, Villaverde AF, Banga JR, Balsa-Canto E. On the relationship between sloppiness and identifiability. *Math Biosci* 2016, 282:147–161.
296. Vanlier J, Tiemann CA, Hilbers PA, van Riel NA. Parameter uncertainty in biochemical models described by ordinary differential equations. *Math Biosci* 2013, 246:305–314.
297. Voit EO. What if the fit is unfit? Criteria for biological systems estimation beyond residual errors. In: Dehmer M, Emmert-Streib F, Salvador A, eds. *Applied Statistics for Biological Networks*. New York: John Wiley & Sons; 2011, 183–200.
298. Battogtokh D, Asch DK, Case ME, Arnold J, Schüttler HB. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of *Neurospora crassa*. *Proc Natl Acad Sci USA* 2002, 99:16904–16909.
299. Tan Y, Liao JC. Metabolic ensemble modeling for strain engineers. *Biotechnol J* 2012, 7:343–353.

300. Faraji M, Fonseca LL, Escamilla-Trevino L, Dixon RA, Voit EO. Computational inference of the structure and regulation of the lignin pathway in *Panicum virgatum*. *Biotechnol Biofuels* 2015, 8:151.

301. Yu Y, Dong W, Altimus C, Tang X, Griffith J, Morello M, Dudek L, Arnold J, Schuttler HB. A genetic network for the clock of *Neurospora crassa*. *Proc Natl Acad Sci USA* 2007, 104:2809–2814.

302. Alves R, Savageau MA. Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics* 2000, 16:534–547.

303. Alves R, Savageau MA. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics* 2000, 16:527–533.

304. Shiraishi F, Hatoh Y. Dynamic sensitivities in chaotic dynamical systems. *Appl Math Comput* 2007, 186:1347–1359.

305. Shiraishi F, Hatoh Y, Irie T. An efficient method for calculation of dynamic logarithmic gains in biochemical systems theory. *J Theor Biol* 2005, 234:79–85.

306. Schwacke JH, Voit EO. Computation and analysis of time-dependent sensitivities in generalized mass action systems. *J Theor Biol* 2005, 236:21–38.

307. Lewis DC. A qualitative analysis of S-systems: Hopf bifurcations. In: Voit EO, ed. *Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity*. New York, NY: Van Nostrand Reinhold; 1991, 304–344.

308. Yin W, Voit EO. Construction and customization of stable oscillation models in biology. *J Biol Syst* 2008, 16:463–478.

309. Fasani RA, Savageau MA. Automated construction and analysis of the design space for biochemical systems. *Bioinformatics* 2010, 26:2601–2609.

310. Savageau MA. Biomedical engineering strategies in system design space. *Ann Biomed Eng* 2011, 39:1278–1295.

311. Savageau MA, Coelho PM, Fasani RA, Tolla DA, Salvador A. Phenotypes and tolerances in the design space of biochemical systems. *Proc Natl Acad Sci USA* 2009, 106:6435–6440.

312. Savageau MA, Fasani RA. Qualitatively distinct phenotypes in the design space of biochemical systems. *FEBS Lett* 2009, 583:3914–3922.

313. Lomnitz JG, Savageau MA. Design space toolbox V2: automated software enabling a novel phenotype-centric modeling strategy for natural and synthetic biological systems. *Front Genet* 2016, 7:118.