

Systems biology

System estimation from metabolic time-series data

Gautam Goel, I-Chun Chou and Eberhard O. Voit*

Integrative BioSystems Institute and The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 313 Ferst Drive, Atlanta, GA 30332, USA

Received on June 26, 2008; revised on August 27, 2008; accepted on August 29, 2008

Advance Access publication September 4, 2008

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: At the center of computational systems biology are mathematical models that capture the dynamics of biological systems and offer novel insights. The bottleneck in the construction of these models is presently the identification of model parameters that make the model consistent with observed data. Dynamic flux estimation (DFE) is a novel methodological framework for estimating parameters for models of metabolic systems from time-series data. DFE consists of two distinct phases, an entirely model-free and assumption-free data analysis and a model-based mathematical characterization of process representations. The model-free phase reveals inconsistencies within the data, and between data and the alleged system topology, while the model-based phase allows quantitative diagnostics of whether—or to what degree—the assumed mathematical formulations are appropriate or in need of improvement. Hallmarks of DFE are the facility to: diagnose data and model consistency; circumvent undue compensation of errors; determine functional representations of fluxes uncontaminated by errors in other fluxes and pinpoint sources of remaining errors. Our results suggest that the proposed approach is more effective and robust than presently available methods for deriving metabolic models from time-series data. Its avoidance of error compensation among process descriptions promises significantly improved extrapolability toward new data or experimental conditions.

Contact: eberhard.voit@bme.gatech.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The construction of a mathematical model occurs in five stages. The first consists of collecting ideas, data and information, which are converted into a conceptual model that is often visualized as a diagram with nodes and arrows. The second stage includes the choice of a mathematical modeling framework and the formulation of suitable equations. The goal of the third stage is the determination of numerical parameter values that make the model consistent with observations, while the fourth and fifth stages are dedicated to diagnostics and to model use, respectively. In most cases the process is iterative, requiring the return to earlier stages.

Arguably the most challenging task is the estimation of parameter values. Until recently, this task was typically pursued from the

bottom up, by characterizing model components and processes one at a time and subsequently merging all ‘local’ descriptions into one comprehensive model. This procedure often failed, for unknown or speculative reasons, and if it succeeded, it was the product of excruciatingly slow and cumbersome effort if models of even moderate size were considered.

Recent advances in molecular and systems biology have provided us with a strikingly different estimation strategy, which is based on experimentally determined time series of observations at the genomic, proteomic or metabolic levels. These time profiles contain enormous information about the structure, dynamics and regulatory mechanisms that govern the biological systems of interest. However, extraction and integration of this information into fully functional, explanatory models is a daunting task, and about one hundred articles have appeared within the past 10 years, each improving certain aspects of the estimation process. Most of them used regression, genetic algorithms, simulated annealing or different evolutionary approaches (Cho *et al.*, 2006; Daisuke and Horton, 2006; Gonzalez *et al.*, 2007; Kikuchi *et al.*, 2003; Kim *et al.*, 2006; Kimura *et al.*, 2004, 2005; Noman and Iba, 2007) to attack the main problem of optimizing parameter values against the observed time-series data. Other papers developed support algorithms, for instance, for smoothing overly noisy data, characterizing basins of attractions containing solutions with minimal error, or circumventing the costly integration of differential equations (Almeida and Voit, 2003; Kimura *et al.*, 2004; Kutalik *et al.*, 2007; Maki *et al.*, 2002; Tsai and Wang, 2005; Vilela *et al.*, 2007; Voit and Almeida, 2004; Voit and Savageau, 1982).

All of the proposed estimation methods developed up-to-date face significant problems in four distinctly different classes:

- (1) Computational issues, including: slow algorithmic progress toward the error minimum or lack of convergence; very complicated error surfaces with numerous local minima; substantial time requirements for integration of differential equations.
- (2) Data-related issues, including: overly noisy data; missing data; missing time series; collinearity between time series; solution spaces with equal error; non-informative, e.g. essentially constant, time profiles.
- (3) Mathematical issues, including: distinctly different, yet equivalent solutions; non-equivalent solutions with similar error; invalid assumptions regarding the chosen process descriptions; error compensation within and among flux

*To whom correspondence should be addressed.

descriptions and within and among equations (see illustrations in the Supplementary Material).

- (4) Issues of model quality beyond goodness of fit, including: lack of diagnostic tools beyond the residual error; lack of model fit for data not used in the estimation; model failure in extrapolations; lack of criteria for optimality of the obtained parameters; lack of criteria for determining the appropriateness of the chosen mathematical representations; lack of methods for assessing whether residual errors are due to idiosyncrasies or noise in the data, an invalid model structure, inadequate computational methods or a combination thereof.

Many articles have acknowledged and discussed various computational issues in great detail and some have addressed issues related to data and models. However, there has been little if any substantial discussion of model validity and quality beyond residual errors, except for the common statement that the estimated parameter set may not be unique.

Here, we propose a novel approach to estimating metabolic pathway systems, called dynamic flux estimation (DFE), which resolves several of the issues mentioned above. The approach consists of two distinct phases. The first consists of an entirely model-free and assumption-free data analysis that reveals inconsistencies within the data, and between data and the alleged system topology. The second phase addresses the mathematical formulation of the processes in the biological system. In contrast to all currently available methods, this phase allows quantitative diagnostics of whether—or to what degree—the assumed mathematical formulations are appropriate or in need of improvement. DFE builds upon the tenets of stoichiometric (Gavalas, 1968; Heinrich and Schuster, 1996; Stephanopoulos *et al.*, 1998) and flux balance analysis [FBA; for a review see (Palsson, 2006)] in that it focuses on the stoichiometry at all nodes in the investigated system to ensure conservation of mass and to estimate flux distribution at each instant in time. However, in DFE the system is typically not in a steady state or quasi steady state (Ishii *et al.*, 2007b; Maki *et al.*, 2002; Okamoto, 2008; Sekiyama and Kikuchi, 2007; Teixeira *et al.*, 2008; Wittmann, 2007; Yang *et al.*, 2002), and its transient dynamics is utilized as a crucial indicator of the regulation within the system.

Because DFE consists of two phases that include several steps, some of which are new, some computational, some logistic (e.g. the choice of mathematical representations in the second phase) and some using any of a variety of existing methods, its exact computational time requirements and accuracy of solution are difficult to assess against currently available methods. Nonetheless, our results suggest that the proposed approach is more effective and robust than presently available methods for deriving metabolic models from time-series data. Specifically, its combined model-free and model-based analyses avoid compensation among and within equations and therefore promise significantly improved extrapolability toward new data or experimental conditions (see Supplementary Material). Its diagnostic tools pinpoint causes of inadequate fits between model and data, and suggest either changes in assumptions related to model choice or the use of data as unmodeled ‘off-line data’.

In the following, we describe DFE and demonstrate its features with a series of successively more complicated (and more realistic)

situations, beginning with an idealized, yet representative case, and ending with actual experimental observations describing fermentation in the bacterium *Lactococcus lactis*.

The proposed method requires time-series data that characterize the dynamics of the system variables. Such data are still relatively rare, but are being generated with increased frequency and quality. Some suitable datasets that exist already have been obtained with an *in vivo* NMR (Neves *et al.*, 2005), mass spectrometry (Ishii *et al.*, 2007a) and other methods (Du *et al.*, 2008). Furthermore, the prospect of the availability of efficacious methods of analysis may inspire experimentalists to generate more of these types of data, which is technically possible and probably worth the effort, even if it is more expensive. Since much of the advantage of DFE is the result of natural constraints among fluxes, DFE is particularly useful for metabolic systems, but less so for gene expression and protein interaction systems (see Section 4).

2 METHODS

DFE is a phased approach with well-defined outcomes for each step and rigorous checks and balances that ensure consistency of the solution (Table 1). We first present the overall concepts and then discuss each step in greater detail.

Each phase facilitates incremental development and analysis of the metabolic target model. Phase I, which is entirely model free, consists of two distinct sets of activities yielding slope estimates and dynamic flux profiles. First, the experimental data are analyzed for mass/material balance and smoothed as necessary. Slope estimates can be derived using different numerical techniques. Next, the pathway structure (i.e. the system topology) is used to generate a system of symbolic equations describing the dynamics of the system. Substituting slope estimates in this system of equations results in a system of fluxes that is linear at each time step t . This linear set of equations can be solved at each time step to obtain dynamic (time-series) profiles of all fluxes in the system. These dynamic flux profiles can be checked for flux balances at the overall system level and at the level of each metabolite pool. Phase II is model-based. Here, based on the flux profiles from the previous phase, one evaluates each plot of a flux versus its alleged substrates and modulators to analyze and choose between possible mathematical representations for each flux. Once decided, the parameters of the chosen functional form are fitted easily with some regression technique to obtain a fully parameterized kinetic model for the system. The fitness of

Table 1. Phases and steps of dynamic flux-based parameter estimation from metabolic time-series data

Phase	Steps/Activities	Outcomes	Checks and balances
I	<ul style="list-style-type: none"> • Estimate missing data (if any) • Smooth and optimize data to achieve mass balance (if necessary) • Estimate slopes 	<ul style="list-style-type: none"> • Smooth, balanced time-series data • Slope estimates 	<ul style="list-style-type: none"> • Mass/material balance
	<ul style="list-style-type: none"> • Formulate system of fluxes • Solve linear system of fluxes at each time point 	<ul style="list-style-type: none"> • Linear system of fluxes • Dynamic flux profiles 	<ul style="list-style-type: none"> • Flux balance • Integrateable flux time series
II	<ul style="list-style-type: none"> • Evaluate flux–substrate plots to choose representative functional forms • Fit parameters of kinetic function to flux profiles 	<ul style="list-style-type: none"> • Parameterized kinetic model 	<ul style="list-style-type: none"> • Check fit of functional forms to flux profiles • Simulate system

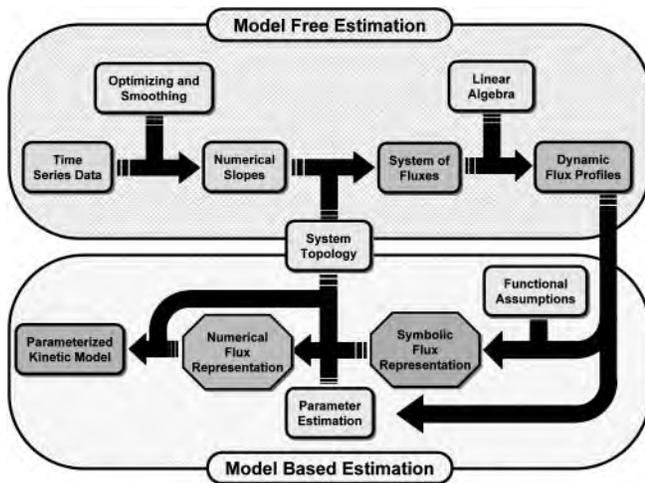


Fig. 1. DFE approach to metabolic system estimation from an *in vivo* time-series data. Starting with experimental time series, the data are simultaneously balanced and smoothed for constant total mass throughout the time series. Then the slopes are estimated using published methods. Combined with the knowledge of the system topology, the slope information yields a linear system of fluxes. The system is solved, using linear algebra techniques, yielding dynamic profiles of all extra- and intra-cellular fluxes in the system. Next, functional assumptions are formulated on how to best represent the processes mathematically. These functions result in symbolic flux representations that can be independently fitted with regression methods to the respective dynamic flux profiles. When combined with knowledge of the system topology, the numerical flux functions are integrated as a single unified system model to obtain time courses.

the parameters for each flux function can be evaluated independently and the same can be done for the overall system performance.

Wide arrays of robust numerical techniques are available for the computational aspects of each component of DFE, including data smoothing, slope estimation, the assessment of linear flux systems and linear/non-linear regression methods for parameter estimation. The proposed DFE workflow (Fig. 1) consists of distinct steps.

- (1) (Phase I: model-free estimation) If necessary, smooth and balance the data in the sense that there should be no gain or loss of material over time. This balance is readily checked against the system stoichiometry. We developed for this purpose a combined non-linear programming and moving-average algorithm to remove noise, while simultaneously balancing the time-series data for constant total mass. The smoothing and slope estimation aspects can be accomplished with finite difference approximations, cubic splines or more sophisticated methods (Vilela *et al.*, 2007).
- (2) (Phase I: model-free estimation) Substitute differentials with estimated slopes for each variable and at each time point (Voit, 2000; Voit and Almeida, 2004; Voit and Savageau, 1982) and construct a linear system of the form ‘Slope Vector(t) = [Stoichiometric Matrix] \times [Flux Vector(t)]’, where the matrix is directly derived from the known (or hypothesized) topology of the system. Solve the system with methods of linear algebra. The result is a (discrete) set of dynamic profiles (time series) of all extra- and intra-cellular fluxes in the system. Over-determined systems require the pooling of fluxes or the use of pseudo-inverse methods. Several constraint-based optimization techniques have been proposed for flux analysis of underdetermined metabolic networks (Bonarius *et al.*, 1997). These approaches have become a mainstay of FBA and served well under steady-state and quasi steady-state conditions (Palsson, 2006; Reed

and Palsson, 2003). Analogous methods may be developed for DFE by using these established approaches as the starting point. Also, Ishii and collaborators (2007b) recently proposed a hybrid method for modeling metabolic systems. This novel approach distinguishes between dynamic and static enzyme activities based on the estimation of time-dependent enzyme reaction rates. The system is split into dynamic and static modules such that a quasi steady-state is attained in the static module at each instant, while the complete system acts dynamically. The transient dynamics of the system is regenerated by interactions between kinetic-based dynamic models and metabolic flux analysis-based static models. A similar separation in dynamic and static modules could be applied to DFE as well. In addition, underdetermined systems may be complemented with information from steady-state FBA, concentration measurements using mass spectrometry or NMR and traditional enzyme kinetics. Finally, it is possible to pool sequential and collinear variables (Vilela *et al.*, 2008) and to combine DFE with methods of structure identification (Chou *et al.*, 2006; Vilela *et al.*, 2008) that are to be applied to select portions of the system. Details of these variations of DFE will be discussed elsewhere (G. Goel *et al.*, manuscript in preparation).

- (3) (Phase II: model-based estimation) Up to this point no assumptions have been made with respect to the mathematical formulation of the flux terms. The next step is now to plot each flux against time and also against the variables affecting this flux (possible in two or three dimensions). As a default, assume that each flux V_k is representable as a product of power-law functions of form $R_k X_i^{f_{ki}} \dots X_n^{f_{kn}}$ as it is done in biochemical systems theory [BST (Savageau, 1976; Voit, 2000)]. Regress V_k in logarithmic coordinates against the contributing variables to obtain the rate constant R_k and the kinetic orders f_{ki}, \dots, f_{kn} , etc. Analyze the quality of fit visually and/or with methods of linear regression diagnostics (Neter and Wasserman, 1974). For non-power-law flux representations (e.g. Michaelis–Menten or Hill functions), it might be possible to execute the analysis with inverse quantities, as in Lineweaver–Burk analysis, or one has to resort to methods of non-linear regression.

3 RESULTS

We applied DFE to four case studies that were inspired by data describing how the bacterium *L.lactis* converts glucose into lactate via the pathway shown in Figure 2a (see Supplementary Material for details). The data (Gaspar *et al.*, 2004; Neves *et al.*, 1999, 2000, 2002a, b, c; Ramos *et al.*, 2002, 2004), provided by our collaborators Dr. Santos and Dr. Neves at ITQB, Portugal, show how a bolus of external glucose is gradually converted into lactate and a few secondary products (Figs 2–5; for better visibility, these figures are reproduced in color in the Supplementary Material). Immediately after glucose addition, the initial metabolite pools (G6P and FBP) increase, while the trioses 3PGA and PEP decrease, because they are utilized for glucose phosphorylation. Once the external glucose pool is depleted, G6P and FBP decrease, while 3PGA and PEP approach high levels that decrease only very slowly afterwards.

3.1 Idealized situation

We applied DFE first to idealized data (Fig. 2b), which we constructed per simulation with an earlier model (Voit *et al.*, 2006) (see Supplementary Material). These data are by design smooth and balanced (see Section 2 and Supplementary Fig. S1), and permit error-free estimation of slopes directly from the equations. Following the guidelines of DFE, we set up and solved the

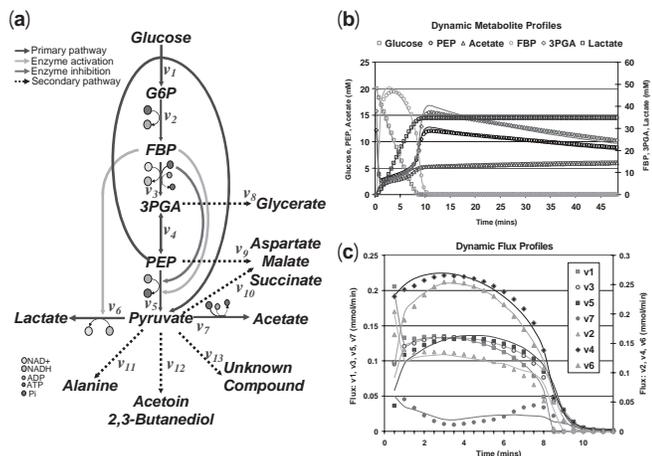


Fig. 2. Results of Case study 1. (a) Fermentation pathway in *L.lactis*. Dark arrows show flow of material. Dashed arrows indicate leakage of material into secondary pathways. Enzyme activation and inhibition are indicated by light gray arrows. G6P, glucose 6-phosphate; FBP, fructose 1,6-bisphosphate; 3-PGA, 3-phosphoglycerate; PEP, phosphoenolpyruvate; ATP, adenosine triphosphate; ADP, adenosine diphosphate; Pi, inorganic phosphate; NAD⁺, nicotinamide adenine dinucleotide (oxidized); NADH, nicotinamide adenine dinucleotide (reduced). (b) Dynamic metabolic profiles. Time-series data of major metabolites in the primary pathway (symbols). Solid lines indicate fits with a model derived using DFE. (c) Dynamic flux profiles. The symbols show the time series of flux profiles estimated solely from data and the system stoichiometry using DFE. The solid lines indicate fitting of a power-law model to the dynamic flux data.

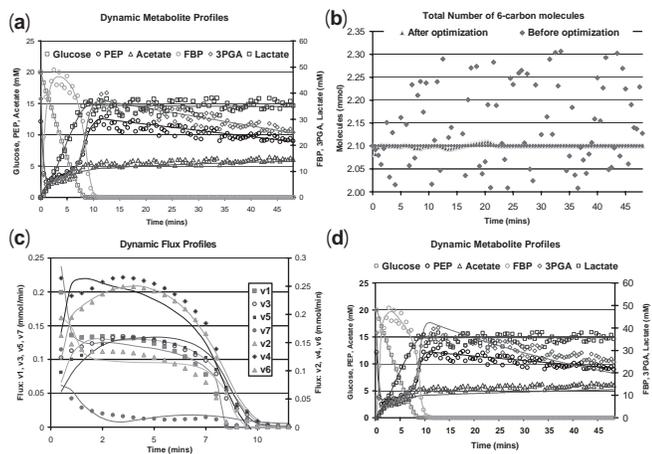


Fig. 3. Results of Case study 2. (a) Dynamic metabolic profiles. Metabolic time-series data with added artificial noise (symbols). The solid lines represent the smoothed and balanced time series. (b) Dynamic mass balance. The random noise leads to mass imbalance which is successfully restored after optimization and smoothing. (c) Dynamic flux profiles. The linear system of fluxes is solved to obtain unique flux profiles (symbols). Power-law models are fitted to each flux time series independently (solid lines). (d) Results from the numerical model. Using DFE, a fully parametric kinetic model is derived from noisy metabolic time-series data (symbols). The results of the model (solid lines) closely match the original dynamic metabolic data.

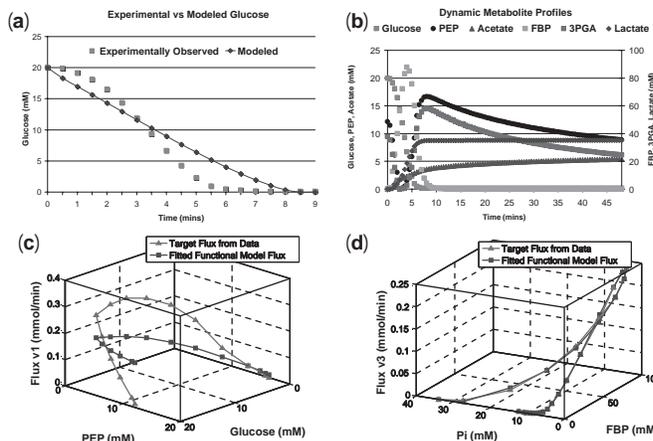


Fig. 4. Results of Case Study 3. (a) Sigmoidal glucose uptake. This type of uptake dynamics has been observed in experiments (symbols) and is difficult to represent with a simple power-law function (solid line). (b) Dynamic metabolic profiles. Time-series data of the major metabolites that result from sigmoidal glucose uptake. (c) and (d) Flux substrate plots. The ‘experimental’ flux profile (gray), obtained using DFE, is plotted against the corresponding flux obtained by fitting a power-law model (black). (c) shows systematic error when flux v1 is fitted with a power-law model. On the other hand, a power-law model accurately reproduces other fluxes like v3 in the same system (d).

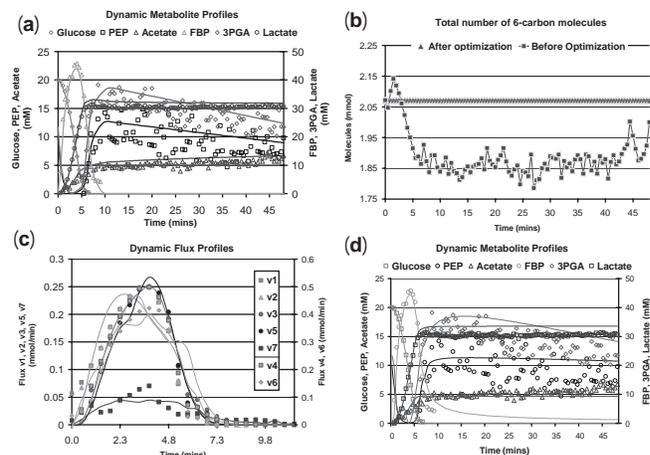


Fig. 5. Results of Case Study 4. (a) Dynamic metabolic profiles. Measured dynamics of metabolite pools in *L.lactis* following a 20 mM [6-13C] glucose bolus (symbols). (b) Dynamic mass balance. Systematic mass imbalance in the experimental data was attributable to missing information about secondary metabolites. The balance was successfully restored by accounting for secondary fluxes. (c) Dynamic flux profiles. The linear system of fluxes is solved to obtain the unique flux profiles (symbols). Power-law models are independently fitted to each flux time series, using linear and non-linear regression (solid lines). (d) Results from the numerical model. Using DFE, a fully parametric kinetic model is derived from the actual metabolic time-series data (symbols). The results of the model (solid lines) closely match the data.

stoichiometric, time-dependent matrix equation (see Supplementary Material), using computed slopes on the left-hand side of this equation, and thus obtained flux values at each time point t (Fig. 2c; see Section 2).

Note that these dynamic flux profiles were obtained purely from knowledge of the system topology and our ‘experimental data’, yet without any assumptions regarding an underlying functional model. Mimicking a realistic situation, we were then interested in a numerical model and made the default assumption that all fluxes could be validly modeled with products of power-law functions, as it is customary in BST. Thus using a symbolic power-law representation for each flux that included all contributing variables, the estimation of the kinetic orders and rate constant was straightforward, since each flux term becomes linear when represented in logarithmic coordinates. The dynamic model with these flux representations was integrated and its behavior closely matched that of the experimental time-series data (Fig. 2b).

3.2 Simulated data with noise

To test the robustness of the DFE approach against noise, we added 10% artificial pseudo-random noise (drawn from a uniform distribution) to the ideal dataset from Case 1 (Section 3.1). Due to the noise, the total mass in the system was no longer constant and required balancing, along with smoothing (see Section 2 and Fig 3a and b). Substituting slopes in the stoichiometric equation, we solved for the fluxes at each time point t (Fig 3c) and estimated parameters for each of the power-law functional forms. The result was a fully parametric kinetic model (Supplementary Fig. S2) that captured the dynamic behavior of the noisy experimental data well (Fig. 3c and d).

3.3 Simulated data with non-power-law terms

In the first two cases, the data-generating system was implemented with power-law representations. To test and demonstrate the diagnostic capabilities of DFE, we simulated the same system (without noise) with a non-power law, sigmoidal glucose uptake function (Fig. 4a and b), which has been observed in some experiments (e.g. see Case 4; Section 3.4).

Next, we estimated slopes and solved the dynamic stoichiometric system as before. The estimated fluxes were notably different from those obtained in the earlier studies, especially at the initial time points (Supplementary Fig. S3a). Attempts to model this system of fluxes exclusively with power-law functions failed. Other methods would have had to stop at this point, simply concluding that the fit was sub-optimal. Even worse in some sense, the simultaneous fitting of all equations or of all terms within each equation would have led to error compensation between terms, thereby not only mis-fitting the sigmoidal flux but other fluxes as well (see Supplementary Material for general discussion). The overall fit might actually have been acceptable, but attempts to extrapolate the resulting numerical model to other datasets or conditions would have become problematic (see Supplementary Material for general discussion). In contrast to this ‘system-wide distribution of error’, DFE prevented such distribution of error and allowed us to pinpoint the source of error accurately by enabling us to test every flux individually against any hypothesized functional representations. We executed this analysis with power laws, using linear regression in log space. The result

was encouraging: All fluxes were reasonably well represented with power laws except for the uptake process (Flux v_1). Evaluation of the flux plots for this reaction step (Fig. 4c) confirmed that the flux in glucose and PEP deviated systematically from the experimental flux when it was modeled by a product of power-law functions. More importantly, even though this flux was not well represented by power laws, we obtained excellent power-law fits for the other fluxes, such as flux v_3 (Fig. 4d), which clearly demonstrated that errors in one flux were not compensated anywhere else in the system. To the best of our knowledge, these error localization and diagnostic capabilities of DFE are unmatched by all existing algorithms for this type of biological time-series analysis.

3.4 Real data

Many methods seem to function well for artificial data, yet break down in the real world. We therefore used actual experimental NMR data from the *L.lactis* pathway (Figs 2 and 5a); they are described elsewhere in detail (Neves *et al.*, 2005; Voit *et al.*, 2006) (also see Supplementary Material).

As a first check, we assessed the total mass in the raw experimental data at each time point and detected that they were significantly unbalanced (Fig. 5b). None of the current parameter or system estimation algorithms, including our own (Chou *et al.*, 2006; Vilela *et al.*, 2008; Voit *et al.*, 2006), check for overall mass balance. As a consequence, these algorithms model something different from what is implicitly expected, which casts doubt on the ultimate estimation results and is likely to lead to problems with new datasets or extrapolations. We first attributed the imbalance to measurement noise. However, mass balancing was not possible within acceptable noise limits. Consultation with our collaborators revealed that several secondary metabolites and fluxes had not been included in the main dataset (Fig. 2a). Accounting for these enabled us to balance the system (Fig. 5b). We proceeded to compute slopes, estimated flux values at each time point t (Fig. 5c), and found that all fluxes except for glucose uptake were well represented by power-law functions. Instead of trying to fit the uptake with some sigmoidal function, we left this flux unmodeled and incorporated it into the model as off-line data (Voit *et al.*, 2005, 2006). The result was a fully parametric kinetic model that closely reproduced the dynamics of the metabolite pools (Fig. 5d).

It is worth noting that the residual error of this model may be larger than the error in a model that is optimized with standard methods, because a standard estimator has the freedom of distributing errors throughout some or all fluxes, which DFE does not permit. As a consequence, the total error in DFE may be higher, but the fit to each individual flux is more reliable.

4 DISCUSSION

Biological time-series data that characterize trends in gene expression, protein prevalence or the accumulation of metabolites *in vivo* are being generated with increased frequency and quality. They contain valuable information about the structure, dynamics and regulatory mechanisms that govern the behavior of cellular systems. However, this information is not explicit and requires extraction methods that are by no means straightforward. While many methods have been proposed over the past years, none of them is effective

in all cases. Furthermore, the existing methods have not addressed questions of diagnostics beyond CPU time and goodness of fit.

We have here proposed DFE as a new approach that resolves at least some of the open issues in the estimation of metabolic pathway systems. The first, model-free and essentially assumption-free phase of DFE permits consistency checks within the metabolic time-series data and leads to numerical representations of fluxes as functions of the variables affecting them. The second, model-based phase allows the objective testing of functional forms for fluxes and is not within the repertoire of any of the existing methods. The two-phased approach thus permits rigorous, quantitative diagnoses of the metabolic data, the alleged pathway structure, the assumptions made in the choice of flux representations and the causes of residual errors. DFE eliminates compensation of error among terms and among variables, which has been a tremendously complex problem with other methods, especially when it comes to extrapolations with the estimated model (see Supplementary Material).

While DFE very significantly reduces error compensation between equations and between flux terms, it still admits error compensation among the parameters within a given flux, independent of what representation is chosen. In the context of BST, this type of compensation between a rate constant and the kinetic orders is well known (Berg *et al.*, 1996; Chou *et al.*, 2007; Sands and Voit, 1996). For reliable extrapolations, the within-flux compensation should also be removed. This removal seems to require data covering wide ranges of variation, multiple datasets or additional information about some of the parameter values, for instance, from traditional enzyme kinetics. Illustrations and discussion of different types of error compensation are presented in the Supplementary Material.

It has been observed in related work that the strategy of replacing differentials with slopes may lead to good fits for the dynamics of each variable in isolation, yet cause problems when all estimated parameter values are entered into the differential equation model (Voit and Almeida, 2004). The reason is that even small deviations between data and model results in one variable that can lead to an amplification of error in other equations. This issue occurs in DFE as well. However, in contrast to other methods, DFE allows diagnostic analyses of the solution. For instance, it turned out in Case 4 (Section 3.4) that Flux v2, which determines the degradation of G6P to FBP, was fitted quite well in isolation with a power-law function. Yet, embedded within the system of ordinary differential equations, the deviations in its variables, G6P and ATP, were sufficient to cause notably different flux values (Supplementary Fig. S4b and c). In response to such a situation, one may ignore the differences, search for causes of the deviations, or substitute smoothed data for a troublesome flux in the form of an off-line process (Voit *et al.*, 2005, 2006).

A key feature of DFE is the requirement of time-series data that are sufficient to capture the dynamics of the system. It is in general difficult to say how many data points are needed for reliable estimations. The key reason is that there is no good, quantitative criterion for the complexity of a time course. In simple dynamic responses, such as monotonically saturating functions, a few data points may be enough to characterize a time trend with sufficient reliability. In other cases, such as the example demonstrated here, the number of time points needed is higher. It seems quite evident that the number very much depends on the complexity of the time course and the noise in the data. Importantly, the types of data required for

DFE are becoming more commonplace because modern methods of molecular biology permit their measurement with a variety of already existing experimental methods.

DFE is an estimation approach particularly geared towards metabolic pathway systems, which are better suited for this type of estimation than genomic or proteomic systems because of conservation of mass at all nodes. Furthermore, DFE focuses on parameter estimation rather than on the identification of structure and regulation in ill-characterized pathway systems. Issues needing further development are related to missing data, missing flux information, underdetermined stoichiometric matrices and ill-characterized systems topologies.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Santos and Dr. Neves at ITQB, Portugal for sharing their *in vivo* NMR data on *Lactococcus*.

Funding: National Heart, Lung and Blood Institute Proteomics Initiative (Contract N01-HV-28181; D. Knapp, PI); a Molecular and Cellular Biosciences (Grant MCB-0517135; E.O.V., PI) from the National Science Foundation; Biological Energy Science Center grant from the Department of Energy (M. Keller, PI).

Conflict of Interest: none declared.

REFERENCES

- Almeida, J.S. and Voit, E.O. (2003) Neural-network-based parameter estimation in S-system models of biological networks. *Genome Inform.*, **14**, 114–123.
- Berg, P.H. *et al.* (1996) A pharmacodynamic model for the action of the antibiotic imipenem on *Pseudomonas aeruginosa* populations in vitro. *Bull. Math. Biol.*, **58**, 923–938.
- Bonarius, H.P.J. *et al.* (1997) Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.*, **15**, 308–314.
- Cho, D.Y. *et al.* (2006) Identification of biochemical networks by S-tree based genetic programming. *Bioinformatics*, **22**, 1631–1640.
- Chou, I.-C. *et al.* (2006) Parameter estimation in biochemical systems models with alternating regression. *Theor. Biol. Med. Model.*, **3**, 25.
- Chou, I.-C. *et al.* (2007) Parameter estimation of S-distributions with alternating regression. *Stat. Oper. Res. Trans.*, **31**, 55–74.
- Daisuke, T. and Horton, P. (2006) Inference of scale-free networks from gene expression time series. *J. Bioinform. Comput. Biol.*, **4**, 503–514.
- Du, X. *et al.* (2008) A computational strategy to analyze label-free temporal bottom-up proteomics data. *J. Proteome Res.*, **7**, 2595–2604.
- Gaspar, P. *et al.* (2004) Engineering *Lactococcus lactis* for production of mannitol: high yields from food-grade strains deficient in lactate dehydrogenase and the mannitol transport system. *Appl. Environ. Microbiol.*, **70**, 1466–1474.
- Gavalas, G.R. (1968) *Nonlinear Differential Equations of Chemically Reacting Systems*. Springer-Verlag, Berlin.
- Gonzalez, O.R. *et al.* (2007) Parameter estimation using Simulated Annealing for S-system models of biochemical networks. *Bioinformatics*, **23**, 480–486.
- Heinrich, T. and Schuster, S. (1996) *The Regulation of Cellular Systems*. Chapman and Hall, New York.
- Ishii, N. *et al.* (2007a) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, **316**, 593–597.
- Ishii, N. *et al.* (2007b) Distinguishing enzymes using metabolome data for the hybrid dynamic/static method. *Theor. Biol. Med. Model.*, **4**, 19.
- Kikuchi, S. *et al.* (2003) Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, **19**, 643–650.
- Kim, K.-Y. *et al.* (2006) Multi-stage evolutionary algorithms for efficient identification of gene regulatory networks. *EvoWorkshops 2006*. Springer, pp. 45–56.
- Kimura, S. *et al.* (2004) Inference of S-system models of genetic networks from noisy time-series data. *Chem-Bio Inform. J.*, **4**, 1–14.
- Kimura, S. *et al.* (2005) Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, **21**, 1154–1163.

- Kutalik,Z. *et al.* (2007) S-system parameter estimation for noisy metabolic profiles using newton-flow analysis. *IET Syst. Biol.*, **1**, 174–180.
- Maki,Y. *et al.* (2002) Inference of genetic network using the expression profile time course data of mouse P19 cells. *Genome Inform.*, **13**, 382–383.
- Neter,J. and Wasserman,W. (1974) *Applied Linear Statistical Models*. Richard D. Irwin, Homewood, IL.
- Neves,A.R. *et al.* (1999) In vivo nuclear magnetic resonance studies of glycolytic kinetics in *Lactococcus lactis*. *Biotechnol. Bioeng.*, **64**, 200–212.
- Neves,A.R. *et al.* (2000) Metabolic characterization of *Lactococcus lactis* deficient in lactate dehydrogenase using in vivo ¹³C-NMR. *Eur. J. Biochem.*, **267**, 3859–3868.
- Neves,A.R. *et al.* (2002a) Catabolism of mannitol in *Lactococcus lactis* MG1363 and a mutant defective in lactate dehydrogenase. *Microbiology*, **148**, 3467–3476.
- Neves,A.R. *et al.* (2002b) Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge? Kinetics of NAD(+) and NADH pools determined in vivo by ¹³C NMR. *J. Biol. Chem.*, **277**, 28088–28098.
- Neves,A.R. *et al.* (2002c) Effect of different NADH oxidase levels on glucose metabolism by *Lactococcus lactis*: kinetics of intracellular metabolite pools determined by in vivo nuclear magnetic resonance. *Appl. Environ. Microbiol.*, **68**, 6332–6342.
- Neves,A.R. *et al.* (2005) Overview on sugar metabolism and its control in *Lactococcus lactis* - the input from in vivo NMR. *FEMS Microbiol. Rev.*, **29**, 531–554.
- Noman,N. and Iba,H. (2007) Inferring gene regulatory networks using differential evolution with local search heuristics. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **4**, 634–647.
- Okamoto,M. (2008) System analysis of acetone-butanol-ethanol fermentation based on time-sliced metabolic flux analysis. In *Symposium on Cellular Systems Biology*. National Chung Cheng University, Taiwan.
- Palsson,B.O. (2006) *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, New York.
- Ramos,A. *et al.* (2002) Metabolism of lactic acid bacteria studied by nuclear magnetic resonance. *Antonie Van Leeuwenhoek*, **82**, 249–261.
- Ramos,A. *et al.* (2004) Effect of pyruvate kinase overproduction on glucose metabolism of *Lactococcus lactis*. *Microbiology*, **150**, 1103–1111.
- Reed,J.L. and Palsson,B.O. (2003) Thirteen years of building constraint-based in silico models of *Escherichia coli*. *J. Bacteriol.*, **185**, 2692–2699.
- Sands,P.J. and Voit,E.O. (1996) Flux-based estimation of parameters in S-systems. *Ecol. Model.*, **93**, 75–88.
- Savageau,M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Advanced Book Program. Addison-Wesley, Reading, MA.
- Sekiyama,Y. and Kikuchi,J. (2007) Towards dynamic metabolic network measurements by multi-dimensional NMR-based fluxomics. *Phytochemistry*, **68**, 2320–2329.
- Stephanopoulos,G. *et al.* (1998) *Metabolic Engineering: Principles and Methodologies*. Academic Press, San Diego, CA.
- Teixeira,A.P. *et al.* (2008) Combining metabolic flux analysis tools and ¹³C NMR to estimate intracellular fluxes of cultured astrocytes. *Neurochem. Int.*, **52**, 478–486.
- Tsai,K.Y. and Wang,F.S. (2005) Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics*, **21**, 1180–1188.
- Vilela,M. *et al.* (2007) Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinformatics*, **8**, 305.
- Vilela,M. *et al.* (2008) Parameter optimization in S-system models, *BMC Syst. Biol.*, **2**, 35.
- Voit,E.O. (2000) *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, Cambridge, UK.
- Voit,E.O. and Almeida,J. (2004) Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*, **20**, 1670–1681.
- Voit,E.O. *et al.* (2005) Challenges for the identification of biological systems from in vivo time series data. *In Silico Biol.*, **5**, 83–92.
- Voit,E.O. *et al.* (2006) Regulation of glycolysis in *Lactococcus lactis*: an unfinished systems biological case study. *IEE Proc. Syst. Biol.*, **153**, 286–298.
- Voit,E.O. and Savageau,M.A. (1982) Power-law approach to modeling biological systems; III. Methods of analysis. *J. Ferment. Technol.*, **60**, 223–241.
- Wittmann,C. (2007) Fluxome analysis using GC-MS. *Microb. Cell Fact.*, **6**, 6.
- Yang,C. *et al.* (2002) Quantitative analysis of intracellular metabolic fluxes using GC-MS and two-dimensional NMR spectroscopy. *J. Biosci. Bioeng.*, **93**, 78–87.