Published in IET Systems Biology
Received on 30th December 2008
Revised on 31st March 2009
doi: 10.1049/iet-syb.2008.0180

Special Issue – Selected papers from The 2nd International Symposium on Optimization and Systems Biology (OSB 2008)


ISSN 1751-8849

# Estimation of metabolic pathway systems from different data sources

*E.O. Voit G. Goel I.-C. Chou L.L. Fonseca*

*Integrative BioSystems Institute and The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, 313 Ferst Drive, Atlanta, GA 30332-0535, USA*
*E-mail: eberhard.voit@bme.gatech.edu*

**Abstract:** Parameter estimation is the main bottleneck of metabolic pathway modelling. It may be addressed from the bottom up, using information on metabolites, enzymes and modulators, or from the top down, using metabolic time series data, which have become more prevalent in recent years. The authors propose here that it is useful to combine the two strategies and to complement time-series analysis with kinetic information. In particular, the authors investigate how the recent method of dynamic flux estimation (DFE) may be supplemented with other types of estimation. Using the glycolytic pathway in *Lactococcus lactis* as an illustration example, the authors demonstrate some strategies of such supplementation.

## 1 Introduction

The grand challenge of systems biology is a deeper understanding of biological systems, and the crucial step towards solving this challenge is the translation of biological systems into reliable mathematical and computational models. Meeting this challenge will constitute a very significant step towards solutions of more widely recognised grand challenge problems of the 21st Century, such as cancer and neurodegenerative disease. The promise of computational systems biology is great, because it is much easier to diagnose, interrogate and manipulate computational models than actual health and disease systems in mice or men.

To support these bold claims, consider the case of cancer and suppose a comprehensive and reliable model of carcinogenesis had been successfully developed and validated. This model would certainly reveal correlations between suspected or unsuspected risk factors, as well as the time of onset and severity of disease. The model would also allow us to follow the gradual or switch-like sequences of transitions from normal into cancerous tissue. It would provide us with tools to screen exhaustively all possible intervention schemes, independent of whether these would presently be implementable in reality. Sensitivity and robustness analysis of the model would identify the components and processes in the system with the highest impact on carcinogenesis per unit change. Although non-linear optimisation tasks are certainly not trivial, they would still be incomparably easier to execute than clinical studies aimed at identifying optimal cancer treatments. A mathematical model could be personalised by accounting for deviations between individual parameter values and the corresponding averages in the population [1] and it would, in principle, be possible to establish personalised health and disease trajectories, risk profiles and customised treatments [2]. Beyond analysis, manipulation and optimisation of previously observed systems, the model could allow predictions of yet untested situations and aid in the discovery of biological design and operating principles.

It is obvious that we are far from being able to rely on mathematical models for cancer treatment and that it will often be difficult to implement model solutions in a living body. However, given the rapidly increasing amounts of personalised information from modern medical diagnostics and from biomarkers, there is no doubt that the mind of the physician will soon be overwhelmed and that computational methods will have to come to the rescue, lest available information is lost or misinterpreted. Furthermore, if we look at the analogy of space exploration, any attempt to send a rover to Mars would have seemed

entirely ludicrous a century ago, and yet we have succeeded, based on computer models and the ingenuity in engineering. Similarly, one might reasonably expect that models and computer algorithms of the not-too-distant future will be capable of integrating hundreds or thousands of personalised data in a valid manner, thereby generating well-founded recommendations to the attending physician.

The successful translation of a biological system into a mathematical or computational model occurs in two steps: the selection of a suitable mathematical description, which entails the formulation of simplifying, conceptual and subsequently technical representations; and the determination of parameter values with which these representations match the available data. The choice of a model structure depends on the data and the purpose of the model (see [3, 4] for recent reviews), but ultimately follows one of two paths.

As the first alternative, an attempt is made to capture the mechanistic details of the biological system with accuracy and precision. A notable example is the Hodgkin–Huxley model of action potentials in neurons, which is directly derived from insights into electric circuits and thus on laws of the theory of electricity [5]. But even in this case, lack of detailed knowledge forced the non-linear control of ion channels in the cell membrane to be modelled with black box approximations that fitted observations sufficiently well. Outside this example, even the representation of a simple bi-substrate enzyme catalysed reaction becomes essentially unmanageable if it is formulated as a physico-chemical system of mass action kinetic steps [6].

The second alternative for choosing a suitable mathematical representation is the use of a non-linear canonical model, whose structure is predetermined. Canonical models are based on specific, yet generic approximations that always lead to the same mathematical structures, while their customisation towards a given application is accomplished exclusively through the choice of parameter values. A simple, pertinent analogy is a linear regression model, which always has the same structure in a given dimension (namely a straight line, plane or hyperplane), but is characterised by slope and intercept parameters that are derived from, or for, the specific application. Canonical models that have been discussed extensively in the literature include Lotka–Volterra systems and their generalisations [7–10], power-law systems within the modelling framework of biochemical systems theory (BST; [11–16]), lin-log models [17, 18] that are extensions of metabolic control analysis [19–21] and saturable-cooperative models that are based on S-shaped functional modules [22]. Two great advantages of canonical models are that knowledge of the connectivity and regulation of a pathway system is sufficient for setting up a symbolic model [13, 16, 23, 24] and that the types of model parameters are predetermined and have a clear meaning, which facilitates estimation.

Parameter values can be determined from many types of biological data. For metabolic pathway estimation, two classes of data have been dominant. The first class consists of characterisations of enzymes, inhibitors, cofactors, metabolite concentrations and other, similar pieces of information quantifying 'local' processes, such as enzyme catalysed reactions. Almost all metabolic modelling studies of the past century have been developed from such local data, and many insights have been, and will henceforth be, gained from this approach. The general procedure of this bottom-up modelling strategy is to represent individual metabolic steps with a suitable function or rate law and to merge these functional descriptions into a more comprehensive pathway model (e.g. [16, 25]). The advantage of this procedure is its straightforward nature and the direct use of available information. The biggest drawback is that models integrated from individual process descriptions often do not work as expected or observed. Reasons are manifold and include the use of data from different organisms or from experiments that were executed under different conditions and often in vitro.

The second, distinctly different class of data consists of time series that measure the responses of a pathway system to some stimulus, such as a change in substrate availability or in experimental or environmental conditions (e.g. heat stress or the addition of an inhibitor to the medium). The great appeal of these data is that the measurements are taken on exactly the same biological system under exactly the same conditions. The main challenge of time-series data is the computational difficulty of extracting reliable parameter values from the data. For BST models alone, roughly 100 articles have been published in recent years describing methods of parameter extraction from dynamic data [26].

This paper focuses on system estimation in metabolic pathway models that combines the use of time-series data with kinetic information from the literature.

## 2 Metabolic system estimation from time-series data

The extraction of parameter values from time-series data is complicated because of challenges of three distinct types, which in real systems are often superimposed [27]. The first is related to the data, which may be noisy, incomplete, correlated with each other or non-informative. The second issue pertains to the model, which may be based on wrong assumptions or contain more parameters than can be identified from the data. For instance, it is impossible to identify the parameters $p$ and $q$ if they always occur as the product $p \cdot q$. The model structure may also permit compensation of errors between different terms or equations [27]. The hideous aspect of such compensation is that the datasets used for estimation are modelled with sufficient accuracy, but that extrapolations may woefully fail, because




doi: 10.1049/iet-syb.2008.0180

the terms no longer compensate for each other in the new situation. The third source for estimation failure comes from the computational algorithms, which may not converge at all, converge to locally – but not globally – acceptable solutions or are so slow that each estimation task takes unduly long.

Recently we proposed a method, called dynamic flux estimation (DFE [27]) that resolves many of the issues listed above, if all aspects of the estimation task are ideal. DFE is executed in two phases. The first phase consists of an entirely model-free and assumption-free data analysis in three steps.

*Step 1:* This assures that no mass is lost or gained in the observed data and smoothes the data if necessary (e.g. [28–35]). Apparent losses may be due to experimental noise or the failure to measure all relevant metabolites.

*Step 2:* This replaces each ordinary differential equation (ODE) with a set of algebraic equations. This replacement is based on the recognition that the differential on the left-hand side of an ODE can be interpreted as the slope of the time course of a variable at a given point in time. Thus, Step 2 consists of estimating slopes of the time courses at $N$ time points and substituting these for the differentials on the left-hand side of each differential equation. This substitution is computationally very advantageous, because it circumvents the need to integrate differential equations and decouples the system of $n$ differential equations into $n$ sets of $N$ algebraic equations each. These sets may be evaluated simultaneously, sequentially or in parallel. The slope substitution strategy [16, 36–38] has become a de facto standard for most estimation methods from time-series data [26].

*Step 3:* This considers each flux in the system as a time-dependent variable. Because pathway models consist of fluxes entering and leaving metabolite pools, the dynamics in each equation is naturally given as sums and differences of fluxes, which thus form a linear system. This system is easily solved at each time point if the system has full rank, or through linear regression if the system is over determined. If this system is under determined, DFE has problems, which are discussed and ameliorated in the next section.

The first, model-free phase of DFE results in a representation of each flux as a numerically characterised function of time and as a function of all contributing metabolites and other variables in the system. This representation is not explicit, but purely numerical and consists of points in flux–time or flux–metabolite plots. The second phase of DFE converts these plots into mathematical representations. As in other modelling approaches, this step requires the choice of a model and the estimation of its parameters. However, these tasks are easier now, because they are addressed one flux at a time and because many candidate models, including Michaelis–Menten rate laws and power-law functions, are available.

DFE offers significant advantages. It reveals inconsistencies within the data, and between data and the alleged system topology, and permits quantitative diagnostic tools of whether – or to what degree – the assumed mathematical formulations are appropriate or in need of improvement. These important features are novel and not available in other standard estimation methods. Moreover, because DFE identifies parameters based on fluxes – as opposed to entire differential equations – issues of faulty extrapolation are reduced to a minimum.

Under ideal conditions, DFE appears to be as close to perfect as it is currently possible. However, its two very significant limitations are the requirement of comprehensive data, which are seldom available, and the fact that the flux system needs to have full rank. We discuss in the following how the issues may be overcome by resorting to information from additional sources.

## 3 Complementation of DFE with additional information

A direct, unique solution of the flux equations in DFE is only possible if the flux system is of full rank. The most frequent case in practical applications, however, is an under-determined system, because most actual pathway systems contain more fluxes than metabolites. As a consequence, the best purely algebraic solution possible is the expression of some fluxes as functions of other fluxes, which is not very useful per se. However, in most practical cases, other information about the system is known, and this information may be used to complement DFE. This complementation does not come for free and either requires assumptions about functional forms of fluxes, mechanistic details or inferences regarding missing time series.

In this section, we discuss prominent issues potentially afflicting DFE, provide theoretically feasible solutions and demonstrate some of the solutions with specific examples. As an illustration throughout, we use the glycolytic pathway in the bacterium *Lactococcus lactis* (Fig. 1), which has become one of the de facto test cases in the field (e.g. [23, 39–42]). Experimental data (Fig. 2) were measured in the laboratory of Drs. Helena Santos and Ana Rute Neves with methods of in vivo nuclear magnetic resonance (NMR) under anaerobic conditions following a 40 M glucose bolus [43]. While glucose 6-phosphate (G6P) was not measured for this specific bolus, it was adapted from a corresponding NMR experiment with a 20 M glucose bolus. Thus, data on the key metabolites were available, but data on less important metabolites were not. Many kinetic details of the pathway are known, and at least some information about all enzymatic steps can be found in the literature. Altogether, this test case is about as well documented as one may currently hope for.

We showed elsewhere that a simplified model of the glycolytic pathway under aerobic conditions was amenable to

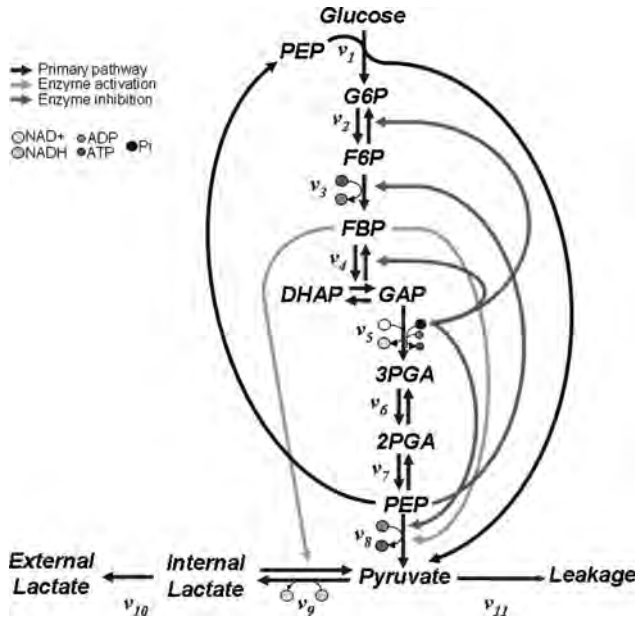


**Figure 1** *Schematic representation of the glycolytic pathway in L. lactis*

Black arrows indicate material flux, light grey arrows activation and dark grey arrows inhibition. Abbreviations: G6P: glucose 6-phosphate; F6P: fructose 6-phoshpate; FBP: fructose 1,6-bisphosphate; DHAP: dihydroxyacetone phosphate; GAP glyceraldehyde 3-phosphate; 3-PGA: 3-phosphoglycerate; 2-PGA: 2-phosphoglycerate; PEP: phosphoenolpyruvate; ATP: adenosine triphosphate; ADP: adenosine diphosphate; Pi: inorganic phosphate; $NAD^+$: nicotinamide adenine dinucleotide (oxidised); NADH: nicotinamide adenine dinucleotide (reduced)

various analyses, including DFE [27, 41, 42]. For illustration purposes, we consider here the more detailed pathway in Fig. 1, which evades a direct DFE analysis, because profiles of some metabolites [e.g. fructose 6-phosphate (F6P)] are missing.

In generic terms, non-ideal situations that require complementation of DFE arise from a combination of the following issues.

Issue 1: The connectivity of the system is not fully known.

Issue 2: Some time series were not measured, although it is known that the corresponding metabolites are involved in the pathway. A typical example for this situation is a metabolite that is very quickly converted into another product, thereby precluding accurate measurements.

Issue 3: Some unknown or not measured metabolites are in fact important. The exclusion of these metabolites is a potential reason for mass imbalances in the system.

Issue 4: All relevant metabolites have been measured as time series, but the flux system is underdetermined. This situation is the rule rather than the exception.

Resolving these issues seems only possible if additional information is available and/or if assumptions are made regarding the functional forms of some of the fluxes in the system.

## 3.1 Solution strategies for issue 1

Distinctly different methods have been developed for computationally inferring the unknown or ill-characterised connectivity of biological pathways (for a recent review see [26]). They include a wide spectrum of techniques, ranging from causality models [44–46] to perturbation methods [47], correlation-based approaches [48] and probabilistic graph models for deducing causality [49]. Some methods (e.g. [36, 40, 50–54]) used time-series data as the basis for their analysis. Specifically for metabolic pathways, methods like alternating regression (AR) [55] and eigenvector optimisation (EO) [56] were proposed as structure identification methods that do not necessarily require knowledge of the connectivity or regulation of the pathway system.

If information is scarce or if the data are noisy, purely computational estimations are not always reliable, and

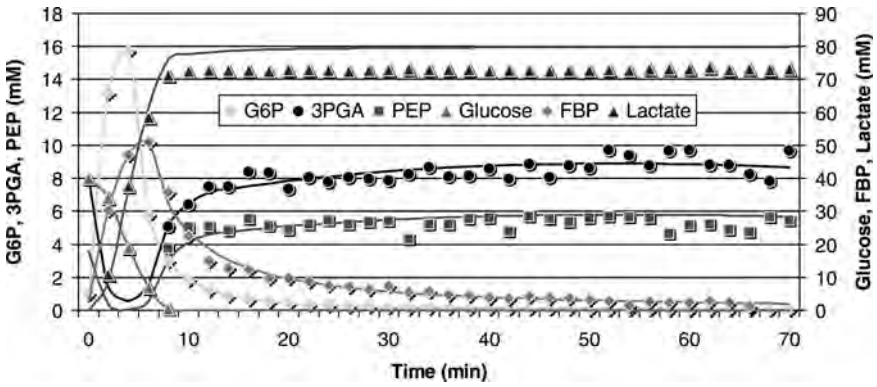


**Figure 2** *In vivo NMR measurements of metabolites of the glycolytic pathway in L. lactis (adopted from [43])*

Symbols show data points, while lines are the results of the DFE analysis. The time course of lactate seems overestimated. This systematic deviation is due to the fact that the raw data exhibited 10% loss in mass over time ('leakage'), which was not included in the estimation. Accounting for leakage makes data and model consistent



doi: 10.1049/iet-syb.2008.0180

within term, within equation and between equation error compensation may become a significant issue [27]. Instead of relying on structure identification algorithms alone, it may be useful to employ simpler algorithms that merely attempt to establish the connectivity pattern within the pathway. An example is a linearisation procedure that generates probabilities for a given equation to be affected by combinations of system variables [54]. A different approach consists of an algorithm that reconstructs equations from the bottom up, testing first the data fit with the most parsimonious parameter set and gradually increasing the complexity of the equation [53]. It is also possible to optimise parameters for a predefined set of biochemically feasible candidate models [40].

## 3.2 Solution strategies for issue 2

The lack of time-series data for certain metabolites may or may not be serious. An important determinant is the mass of the missing metabolite pools during the experiment. If this mass is small, methods of compensatory mass balancing [27] may provide a solution that is not overly damaging. However, significant amounts of missing mass cannot be ignored. If enzymatic information is available for fluxes producing and degrading a metabolite in question, it is sometimes possible to reconstruct its unknown time profile from neighbouring time series. Obviously, this use of kinetic information is subject to some degree of bias and uncertainty, especially if the information was obtained in vitro and/or from a different organism. If large portions of mass are unaccounted for, and if no pertinent kinetic information is available, it might not be possible to continue the parameterisation of the model with reliability.

## 3.3 Example

Consider the reversible isomerisation of G6P to F6P, which is catalysed by phosphoglucose isomerase (PGI). The kinetics of PGI has been characterised for both directions, and if one assumes a reversible Michaelis–Menten rate law for the net flux (1), pertinent parameters are readily obtained from the literature [57–59]. By combining this kinetic in vitro information with the time series data on G6P and the in vivo G6P degradation flux estimates for $v_2$ at the measured time points, which we obtained with DFE, we can deduce the time series for the unknown metabolite F6P. This is accomplished by expressing (1) with F6P as the dependent variable and solving it for all measured time points

$$v_2 = \frac{v_{\max}^{\text{for}}([\text{G6P}]/K_{m\text{G6P}}) - v_{\max}^{\text{rev}}([\text{F6P}]/K_{m\text{F6P}})}{1 + ([\text{G6P}]/K_{m\text{G6P}}) + ([\text{F6P}]/K_{m\text{F6P}}) + ([P_i]/K_{mP_i})} \quad (1)$$

The reconstructed F6P profile is similar to the G6P profile (Fig. 3), but at a scale of about 1:10. Both the shape and scaling factor are consistent with the common understanding of a fast equilibrium between the two.

## 3.4 Solution strategies for issue 3

The consequences of unknown or not measured metabolite pools may range from irrelevant to utterly detrimental for any estimation effort, depending on the extent of lacking information. A diagnostic aid for this situation is the checking of mass balance in the entire system throughout the experimental time period. If significant changes in balance are observed, because non-negligible amounts of mass are gained or lost, additional biological insight will be needed to remedy the situation. If the masses are more or less balanced, it is still possible that important fluxes or metabolites are missing. There is currently no obvious defense in this situation.

A slightly different situation occurs if relevant cofactors or modulators were not measured. For instance, $NAD^+$ and

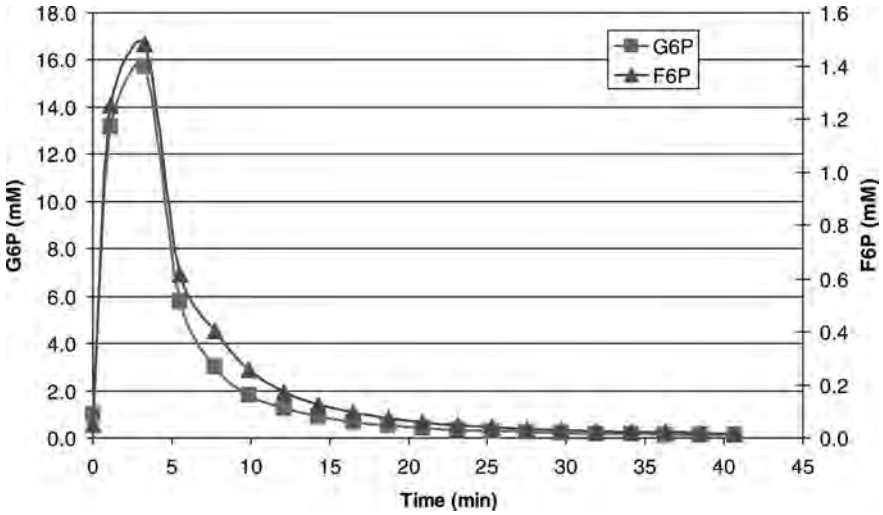


**Figure 3** *In vivo NMR measurements of G6P in L. lactis (adapted from [43]) and reconstructed time series of F6P derived from a combination of DFE and kinetic literature information (parameters from [57–59])*

NADH may affect the speed of a reaction, but because of moiety conservation, no change in (carbon) mass is observable, so that the (carbon) mass in the system is perfectly balanced. Nonetheless, factors influencing the $NAD^+$/NADH ratio may significantly affect the dynamics of the pathway. Again, this situation requires a case-by-case treatment.

### 3.5 Example

Recently, we discussed mass balancing in the context of raw experimental data in *L. lactis* [27]. The detected imbalance was too severe to be attributable to acceptable measurement noise, and smoothing efforts still left 10% of the supplied glucose unaccounted. It turned out that several secondary metabolites and fluxes had not been included, and accounting for these enabled the balancing of the system.

### 3.6 Solution strategies for issue 4

In the ideal situation, the stoichiometric matrix is of full rank and the first phase of DFE is free of assumptions. The direct and unambiguous result then consists of numerical relationships between variables and fluxes, which are further analysed in the second phase of DFE. However, if the flux system is under determined, it is necessary to complement the information from the flux system with information from other sources. This complementation always involves assumptions, but ultimately leads to the characterisation of some fluxes, for instance, with a kinetic bottom-up approach. Distinct options are available for supplementing information embedded in the flux system, at least in principle. First, it may be possible to obtain fluxes directly from experiments. In a few cases, flux–substrate relationships were measured (e.g. see parameter estimation in [25] from flux data in [60]), but such data are rare. Much more prevalent is the information on the kinetic properties of enzymes and the reactions they catalyse. This information is closely linked to an alleged functional form for each flux. For instance, if a Michaelis–Menten rate function is deemed appropriate and if applicable $K_M$ and $V_{max}$ values can be found, the parameters and the time-series data may be entered into the rate function to compute the appropriate flux value at each time point. Of course, this use of literature information is potentially compromised by the fact that pertinent parameters were often obtained in other labs, under different conditions, and sometimes even from other species.

As an alternative, or if relevant kinetic information is unavailable, it has been shown that regression methods, genetic algorithms (GA), as well as specialised methods like AR and EO [55, 56], have the potential of determining parameter values in pathway models from metabolic time-series data [26]. This feature renders it possible in principle to determine the necessary number of missing fluxes and to use them in the first phase of DFE. A considerable drawback of this strategy is that GA and the various regression methods must a priori assume specific mathematical representations of the fluxes that are to be estimated. However, the most appropriate representations are often unknown. This situation becomes less of a hindrance if some of the variables and fluxes operate within relatively small ranges, because one might expect that the typical canonical approximations, such as products of power-law functions or lin-log expressions, would be sufficiently accurate throughout these limited ranges. Thus, while many combinations of fluxes could theoretically be chosen to supplement DFE in an under-determined estimation task, it is advisable to choose variables and fluxes that remain relatively close to some normal operating values. At the same time, variables that do not vary much at all contain relatively weak information, which may lead to misestimation, so that the choice of fluxes requires a compromise. In addition to the fact that estimation algorithms must assume specific functions, they are also susceptible to error compensation between the terms of an equation.

### 3.7 Example

As an illustration for the use of kinetic information, we pretend that the glycolytic system under investigation was underdetermined. To obtain an additional flux outside DFE, we study the phosphofructokinase (PFK) step ($v_3$ in Fig. 1), in which a phosphoryl group is transferred from adenosine triphosphate (ATP) to F6P, yielding fructose 1,6-bisphosphate (FBP) and adenosine diphosphate (ADP). Since F6P is not observed under the given experimental conditions, it is not possible to estimate the PFK flux directly from our time-series data. However, it is well established that G6P and F6P are in rapid equilibrium, and because F6P is below the detection limit (2.5 mM), we assume that its accumulation pattern is one-tenth that of the G6P at all time points (see discussion above and Fig. 3). It is furthermore known that the PFK reaction is essentially irreversible under physiological conditions and that the enzyme is allosterically inhibited by ATP, FBP and PEP, while being activated by ADP. Several rate laws have been proposed for the PFK reaction (e.g. [61, 62] and references therein). We choose the model of Hoefnagel *et al.* (2); [58]), because it was developed specifically for *L. lactis* under comparable conditions.

Using this model with the published parameter values [58] and the time series of G6P (divided by 10 for the expected profile of F6P), we obtained a parameterised, mechanistic PFK model. This deduced model rather closely reflects the process in vivo, as it was obtained with DFE (Fig. 4). This result is quite remarkable, first, because it confirms that kinetic information can indeed be used under opportune conditions to supplement DFE and, second, because it confirms that the entirely model-free




doi: 10.1049/iet-syb.2008.0180

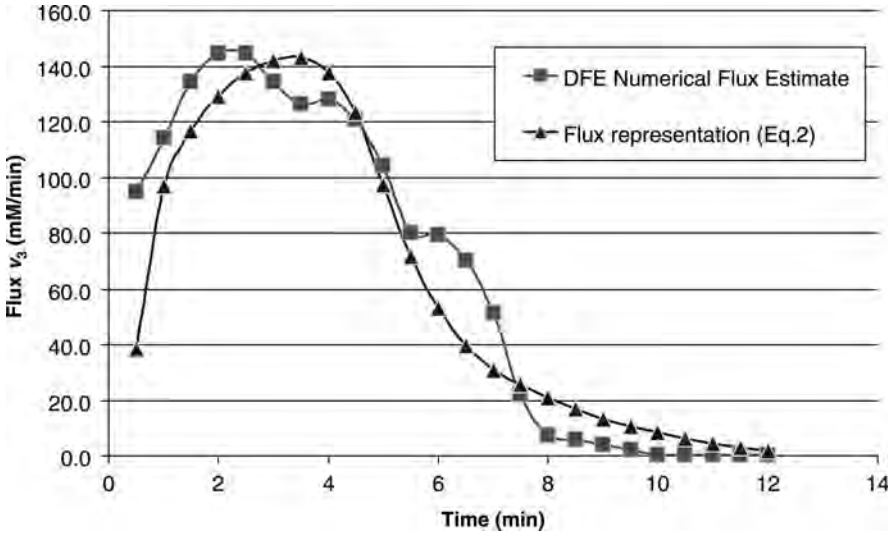


**Figure 4** *Flux $v_3$, obtained with DFE as a numerical estimate (square symbols), and formulated as a published rate function with parameter values directly taken from the literature ((2); [58]; triangular symbols*)

Symbols indicate where data points were used for estimation. Interestingly, the 'bumps' in the numerical DFE estimate reflect different phases of glucose uptake, which are visible in slight yet distinct changes in slopes and which may be due to a differential affinity of the cellular transporters to the $\alpha$ and $\beta$ forms of glucose. The mechanistic flux model does not distinguish these phases

phase of DFE yields very reasonable, numerical flux representations. (See (2))

Under opportune conditions, we could use GA, AR, EO or some other regression algorithm to supply additional flux information. However, in the present case all variables vary within wide ranges, especially, if they are logarithmically transformed, thus raising doubt about any type of approximation. Indeed, we used AR, which uses power-law approximations, and obtained good fits for entire equations, but not for individual fluxes, which demonstrates the hidden problem of error compensation among mathematical terms in a systems model (results not shown). As an alternative, if we had good functional candidates for all production and degradation processes of a given variable, for instance, in the form of Michaelis–Menten functions or other, more complicated rate laws [6], we could estimate the entire decoupled equation for this variable and retrieve from the optimal fit the representations of the individual fluxes. Thus, in contrast to the solution discussed above, we would not need parameter values like $K_M$, $K_I$, $K_{eq}$ or $V_{max}$, but only adequate symbolic formulations. These formulations, together with the metabolic time series, would be used to optimise the kinetic parameter per regression or GA.

## 4 Discussion

Parameter estimation continues to be the bottleneck of many studies in computational systems biology. This complicated task has so far mandated an early, fundamental decision, namely whether to pursue a bottom-up or a top-down approach. In the past this had been a simple decision, because time-series data were very rare. However, dynamic profile data have become much more frequent in recent years, and if efficacious methods of analysis are made available, the appeal of time-series measurements will certainly increase.

A flurry of recent articles on model estimation from time profiles (see [26]) reveals three facts. First, there seem to be no silver bullet solutions that are able to tackle the majority of practical applications. Second, almost all methods have been focusing on good fits and the criterion of algorithmic speed, but not on issues of error compensation within the model, issues of extrapolation and diagnostics of assumptions. Third, except for artificial examples, hardly any estimation studies had the luxury of 'complete' data.

Recently we proposed the DFE method, which addresses some of the above issues and works very well, but only if rather comprehensive data are available. Even then, DFE has the severe limitation that the fluxes in the pathway have to form a system of full rank. For more or less linear pathways, this assumption may be true, but as soon as pathway systems with cycles are under investigation, DFE cannot be applied directly, because the fluxes outnumber the metabolites. Other complicating factors are missing time series and uncertainties with regard to structure and regulation of the metabolic pathway system.

Faced with the situation that ideal scenarios allowing direct application of DFE are rare, we investigated here to what

$$v_3 = \frac{v_{max}(1 - (PEP^{n3PEP}/PEP^{n3PEP} + K_{mPEP}^{n3PEP}))([F6P]/K_{mF6P})^{n3}([ATP]/K_{mATP})}{(1 + ([F6P]/K_{mF6P})^{n3} + ([FBP]/K_{mFBP}))(1 + ([ATP]/K_{mATP}) + ([ADP]/K_{mADP}))} \quad (2)$$

degree DFE may be supplemented with other information. Indeed, our study addressed the question of whether distinctly different approaches to parameter estimation may be successfully combined. The short answer is that this is possible in different ways, at least in principle. The options for DFE supplementation span a range of methods. If all significant metabolic time series are available, and if some of the enzymes in the system are well characterised under pertinent conditions, it may be possible to construct flux–time and flux–variable relationships and use these as substitutes for unknown fluxes in DFE. Sufficient kinetic information may even allow the construction of time-series profiles of metabolites that were not measured. Alternatively, or in addition, if one can reasonably assume functional forms for a few of the fluxes within the system, then a GA or more specialised methods like AR or EO can be employed to estimate a sufficiently large subset of fluxes to execute DFE on the rest of the flux system.

The combination of methods presented here serves primarily as a proof of concept, and it is to be expected that targeted work on combined estimation methods will lead to refined and possibly even entirely novel estimation strategies. Such strategies will become increasingly important, because one should expect a rapidly growing number of time-series data of high quality, which however will very seldom be comprehensive enough for a unidirectional estimation approach.

## 5 Acknowledgments

The authors are grateful to Drs. Santos and Neves at ITQB, Portugal for sharing their in vivo NMR data on *Lactococcus*. This work was supported in part by a 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. Voit, PI) from the National Science Foundation, a BioEnergy Science Center grant from the Department of Energy (M. Keller, PI), and an endowment from the Georgia Research Alliance. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring institutions.

## 6 References

[1] VOIT E.O., BRIGHAM K.L.: 'The role of systems biology in predictive health and personalized medicine', *Open Pathol. J.*, 2008, **2**, pp. 68–70

[2] VOIT E.O.: 'A systems-theoretical framework for health and disease: inflammation and preconditioning from an abstract modeling point of view', *Math. Biosci.*, 2009, **217**, pp. 11–18

[3] VEFLINGSTAD S.R., DAM P., XU Y., VOIT E.O.: 'Microbial pathway models', in XU Y., GOGARTEN J.P. (EDS): 'Computational methods for understanding bacterial and archaeal genomes' (Imperial College Press, London, 2008), pp. 315–344

[4] VOIT E.O.: 'Modeling networks using power-laws and S-systems' in WOLKENHAUER O. (ED.): 'Systems biology' (Portland Press, Portland, 2008), pp. 29–40

[5] HODGKIN A.L., HUXLEY A.F.: 'A quantitative description of membrane current and its application to conduction and excitation in nerve', *J. Physiol.*, 1952, **117**, (4), pp. 500–544

[6] SCHULZ A.R.: 'Enzyme kinetics: from diastase to multi-enzyme systems' (Cambridge University Press, Cambridge, New York, 1994)

[7] HERNÁNDEZ-BERMEJO B., FAIRÉN V.: 'Lotka–Volterra representation of general nonlinear systems', *Math. Biosci.*, 1997, **140**, (1), pp. 1–32

[8] LOTKA A.: 'Elements of physical biology' (Williams and Wilkins, Baltimore, 1925)

[9] MAY R.E.: 'Theoretical ecology: principles and applications' (Blackwell, Oxford, 1976)

[10] VOLTERRA V.: 'Variazioni e fluttuazioni del numero d'individui in specie animali conviventi', *Mem. R. Accad. dei Lincei.*, 1926, **2**, pp. 31–113

[11] SAVAGEAU M.A.: 'Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation', *J. Theor. Biol.*, 1969, **25**, (3), pp. 370–379

[12] SAVAGEAU M.A.: 'Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions', *J. Theor. Biol.*, 1969, **25**, (3), pp. 365–369

[13] SAVAGEAU M.A.: 'Biochemical systems analysis: a study of function and design in molecular biology' (Addison-Wesley Pub. Co. Advanced Book Program, Reading, MA, 1976, vol. xvii), p. 379

[14] TORRES N.V., VOIT E.O.: 'Pathway analysis and optimization in metabolic engineering' (Cambridge University Press, Cambridge, UK, 2002), p. 305

[15] VOIT E.O. (ED.): 'Canonical nonlinear modeling. S-system approach to understanding complexity' (Van Nostrand Reinhold, NY, 1991), p. xi + 365

[16] VOIT E.O.: 'Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists' (Cambridge University Press, Cambridge, UK, 2000), p. xii + 530

[17] HATZIMANIKATIS V., BAILEY J.E.: 'MCA has more to say', *J. Theor. Biol.*, 1996, **182**, (3), pp. 233–242



doi: 10.1049/iet-syb.2008.0180

[18] VISSER D., HEIJNEN J.J.: 'The mathematics of metabolic control analysis revisited', *Metab. Eng.*, 2002, **4**, (2), pp. 114–123

[19] FELL D.A.: 'Understanding the control of metabolism' (Portland Press, London, 1997)

[20] HEINRICH R., RAPOPORT T.A.: 'A linear steady-state treatment of enzymatic chains. General properties, control and effector strength', *Eur. J. Biochem.*, 1974, **42**, (1), pp. 89–95

[21] KACSER H., BURNS J.A.: 'The control of flux', *Symp. Soc. Exp. Biol.*, 1973, **27**, pp. 65–104

[22] 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**, (5), pp. 1259–1277

[23] GOEL G.: 'Reconstructing biochemical systems: systems modeling and analysis tools for decoding biological designs' (VDM Verlag Dr. Müller, Saarbrücken, Germany, 2008)

[24] GOEL G., CHOU I-C., VOIT E.O.: 'Biological systems modeling and analysis: a biomolecular technique of the twenty-first century', *J. Biomol. Tech.*, 2006, **17**, (4), pp. 252–269

[25] ALVAREZ-VASQUEZ F., SIMS K.J., HANNUN Y.A., VOIT E.O.: 'Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models', *J. Theor. Biol.*, 2004, **226**, (3), pp. 265–291

[26] CHOU I-C., VOIT E.O.: 'Recent developments in parameter estimation and structure identification of biochemical and genomic systems', *Math. Biosc.*, 2009, **219**, pp. 57–83

[27] GOEL G., CHOU I-C., VOIT E.O.: 'System estimation from metabolic time-series data', *Bioinformatics*, 2008, **24**, (21), pp. 2505–2511

[28] BURDEN R.L., FAIRES J.D.: 'Numerical analysis' (PWS Publishing Co., Boston, MA, 1993, 5th edn.)

[29] DE BOOR C.: 'A practical guide to splines. Applied mathematical sciences' (Springer-Verlag, New York, 1978, vol. 27), vol. xxiv, p. 392

[30] DE BOOR C., HÖLLIG K., RIEMENSCHNEIDER S.D.: 'Box splines. Applied mathematical sciences' (Springer-Verlag, New York, Hong Kong, 1993, vol. 98), vol. xvii, p. 200

[31] GREEN P.J., SILVERMAN B.W.: 'Nonparametric regression and generalized linear models: a roughness penalty approach. (*Monographs on statistics and applied probability*)' (Chapman & Hall, London, New York, 1994, vol. 58, 1st edn.), vol. ix, p. 182

[32] SEATZU C.: 'A fitting based method for parameter estimation in S-systems', *Dynam. Syst. Appl.*, 2000, **9**, (1), pp. 77–98

[33] ALMEIDA J.S.: 'Predictive non-linear modeling of complex data by artificial neural networks', *Curr. Opin. Biotechnol.*, 2002, **13**, (1), pp. 72–76

[34] EILERS P.H.C.: 'A perfect smoother', *Anal. Chem.*, 2003, **75**, (14), pp. 3631–3636

[35] VILELA M., BORGES C.C., VINGA S., *ET AL.*: 'Automated smoother for the numerical decoupling of dynamics models', *BMC Bioinf.*, 2007, **8**, p. 305

[36] VOIT E.O., ALMEIDA J.: 'Decoupling dynamical systems for pathway identification from metabolic profiles', *Bioinformatics*, 2004, **20**, (11), pp. 1670–1681

[37] VOIT E.O., SAVAGEAU M.A.: 'Power-law approach to modeling biological systems; II. Application to ethanol production', *J. Ferment. Technol.*, 1982, **60**, (3), pp. 229–232

[38] VOIT E.O., SAVAGEAU M.A.: 'Power-law approach to modeling biological systems; III. Methods of analysis', *J. Ferment. Technol.*, 1982, **60**, (3), pp. 233–241

[39] DEL ROSARIO R.C., MENDOZA E., VOIT E.O.: 'Challenges in lin-log modelling of glycolysis in *Lactococcus lactis*', *IET Syst. Biol.*, 2008, **2**, (3), p. 136

[40] SRIVIDHYA J., CRAMPIN E.J., MCSHARRY P.E., SCHNELL S.: 'Reconstructing biochemical pathways from time course data', *Proteomics*, 2007, **7**, (6), pp. 828–838

[41] VOIT E., NEVES A.R., SANTOS H.: 'The intricate side of systems biology', *Proc. Natl. Acad. Sci. USA*, 2006, **103**, (25), pp. 9452–9457

[42] VOIT E.O., ALMEIDA J., MARINO S., *ET AL.*: 'Regulation of glycolysis in *Lactococcus lactis*: an unfinished systems biological case study', *IEE Proc. Syst. Biol.*, 2006, **153**, (4), pp. 286–298

[43] NEVES A.R., RAMOS A., NUNES M.C., *ET AL.*: 'In vivo nuclear magnetic resonance studies of glycolytic kinetics in *Lactococcus lactis*', *Biotechnol. Bioeng.*, 1999, **64**, (2), pp. 200–212

[44] ARKIN A., ROSS J.: 'Statistical construction of chemical-reaction mechanisms from measured time-series', *J. Phys. Chem.*, 1995, **99**, (3), pp. 970–979

[45] TORRALBA A.S., YU K., SHEN P., OEFNER P.J., ROSS J.: 'Experimental test of a method for determining causal connectivities of species in reactions', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (4), pp. 1494–1498

[46] VANCE W., ARKIN A., ROSS J.: 'Determination of causal connectivities of species in reaction networks', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, (9), pp. 5816–5821

[47] SONTAG E., KIYATKIN A., KHOLODENKO B.N.: 'Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data', *Bioinformatics*, 2004, **20**, (12), pp. 1877–1886

[48] EISEN M.B., SPELLMAN P.T., BROWN P.O., BOTSTEIN D.: 'Cluster analysis and display of genome-wide expression patterns', *Proc. Natl. Acad. Sci. USA*, 1998, **95**, (25), pp. 14863–14868

[49] SACHS K., PEREZ O., PE'ER D., LAUFFENBURGER D.A., NOLAN G.P.: 'Causal protein-signaling networks derived from multiparameter single-cell data', *Science*, 2005, **308**, (5721), pp. 523–529

[50] CRAMPIN E.J., MCSHARRY P.E., SCHNELL S.: 'Extracting biochemical reaction kinetics from time series data' (lecture notes in artificial intelligence Springer-Verlag, 2004), pp. 329–336

[51] KIKUCHI S., TOMINAGA D., ARITA M., TOMITA M.: 'Pathway finding from given time-courses using genetic algorithm', *Genome Inf.*, 2001, **12**, pp. 304–305

[52] MAKI Y., UEDA T., OKAMOTO M., UEMATSU N., INAMURA Y., EGUCHI Y.: 'Inference of genetic network using the expression profile time course data of mouse P19 cells', *Genome Inf.*, 2002, **13**, pp. 382–383

[53] MARINO S., VOIT E.O.: 'An automated procedure for the extraction of metabolic network information from time series data', *J. Bioinf. Comput. Biol.*, 2006, **4**, (3), pp. 665–691

[54] VEFLINGSTAD S.R., ALMEIDA J., VOIT E.O.: 'Priming nonlinear searches for pathway identification', *Theor. Biol. Med. Model*, 2004, **1**, p. 8

[55] CHOU I.-C., MARTENS H., VOIT E.O.: 'Parameter estimation in biochemical systems models with alternating regression'. Theor. Biol. Med. Model., 2006, 3, p. 25

[56] VILELA M., CHOU I-C., VINGA S., VASCONCELOS A.T., VOIT E.O., ALMEIDA J.S.: 'Parameter optimization in S-system models', *BMC Syst. Biol.*, 2008, **2**, p. 35

[57] EVEN S., LINDLEY N.D., COCAIGN-BOUSQUET M.: 'Molecular physiology of sugar catabolism in *Lactococcus lactis* IL1403', *J. Bacteriol.*, 2001, **183**, (13), pp. 3817–3824

[58] HOEFNAGEL M.H., VAN DER BURGT A., MARTENS D.E., HUGENHOLTZ J., SNOEP J.L.: 'Time dependent responses of glycolytic intermediates in a detailed glycolytic model of *Lactococcus lactis* during glucose run-out experiments', *Mol. Biol. Rep.*, 2002, **29**, (1–2), pp. 157–161

[59] TEUSINK B., PASSARGE J., REIJENGA C.A., *ET AL.*: 'Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry', *Eur. J. Biochem.*, 2000, **267**, (17), pp. 5313–5329

[60] WU W.I., MCDONOUGH V.M., NICKELS J.T. JR., *ET AL.*: 'Regulation of lipid biosynthesis in *Saccharomyces cerevisiae* by fumonisin B1', *J. Biol. Chem.*, 1995, **270**, (22), pp. 13171–13178

[61] PESKOV K., GORYANIN I., DEMIN O.: 'Kinetic model of phosphofructokinase-1 from *Escherichia coli*', *J. Bioinf. Comput. Biol.*, 2008, **6**, (4), pp. 843–867

[62] HOFMEYR J., ROHWER J., SNOEP J.L.: 'Concepts in computational systems biology: structural analysis, kinetics, control and regulation of cellular systems' (lecture notes for the mini course BCH714: computational systems biology, 2007) (http://www.jjj.sun.ac.za/minicourse/)



doi: 10.1049/iet-syb.2008.0180