BIOINFORMATICS ORIGINAL PAPER

Vol. 26 no. 10 2010, pages 1324–1331
doi:10.1093/bioinformatics/btq136

Systems biology

Advance Access Publication March 30, 2010

Streamlining the construction of large-scale dynamic models using generic kinetic equations

Delali A. Adiamah, Julia Handl and Jean-Marc Schwartz*

Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Studying biological systems, not just at an individual component level but at a system-wide level, gives us great potential to understand fundamental functions and essential biological properties. Despite considerable advances in the topological analysis of metabolic networks, inadequate knowledge of the enzyme kinetic rate laws and their associated parameter values still hampers large-scale kinetic modelling. Furthermore, the integration of gene expression and protein levels into kinetic models is not straightforward.

Results: The focus of our research is on streamlining the construction of large-scale kinetic models. A novel software tool was developed, which enables the generation of generic rate equations for all reactions in a model. It encompasses an algorithm for estimating the concentration of proteins for a reaction to reach a particular steady state when kinetic parameters are unknown, and two robust methods for parameter estimation. It also allows for the seamless integration of gene expression or protein levels into a reaction and can generate equations for both transcription and translation. We applied this methodology to model the yeast glycolysis pathway; our results show that the behaviour of the system can be accurately described using generic kinetic equations.

Availability and implementation: The software tool, together with its source code in Java, is available from our project web site at http://www.bioinf.manchester.ac.uk/schwartz/grape

Contact: jean-marc.schwartz@manchester.ac.uk

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

Received on January 20, 2010; revised on March 9, 2010; accepted on March 25, 2010

1 INTRODUCTION

The current availability of annotated genomes and detailed '-omics' data makes it possible to construct stoichiometric genome-scale metabolic networks that include all reactions, metabolites and proteins. Systems biology aims to examine the properties and dynamics of cellular processes as a whole rather than in isolated parts of a cell or an organism (Kitano, 2002). Integrating cellular components is essential for our understanding of how interactions between these components influence cellular functions. One aspect of this integration is the need to integrate proteins and other cellular components into metabolic networks. For example, Förster and Palsson (2003) manually reconstructed a genome-scale integrative model of gene expression and metabolism of *Saccharomyces cerevisiae* (1175 metabolic reactions, 584 metabolites and 708 open reading frames).

Stoichiometric models, which describe the topology of a metabolic network, provide limited insights into the functioning of cellular processes. To understand the detailed dynamics of cellular functions and their regulation, it is necessary to advance toward kinetic models where the behaviour of a system can be perturbed. The construction of a genome-scale kinetic model of a biological cell requires the integration of genomic, proteomic, metabolomic and fluxomic data along with thermodynamic information (Jamshidi and Palsson, 2008). Attempts for building such large-scale kinetic models are now starting to emerge. Ao *et al.* (2008) provided a systematic method for constructing large-scale kinetic metabolic models and addressed the problem of estimating kinetic parameters. Jamshidi and Palsson (2008) described a framework for building and analysing large-scale kinetic models and presented the mathematical challenges associated with the construction of such models. In a cell-scale model, the number of reactions, metabolites and proteins can reach several thousands, making it time-consuming and costly, if not impossible, to accurately measure individual concentrations of metabolites, fluxes and kinetic parameters.

There is often inadequate knowledge of enzymatic–kinetic laws and their associated parameter values, and usually parameters obtained from the literature are dependent on specific *in vitro* or *in vivo* experimental conditions. Yet, there is growing awareness that exact rate equations and parameters are often not crucial in determining the dynamic properties of large systems. This principle has been illustrated by the development of methods for 'bridging the gap' between structural and kinetic modelling. Steuer *et al.* (2006) proposed a method that aimed to give account of the dynamical capabilities of metabolic systems without requiring explicit information about the rate equations, and they showed that it was possible to acquire a detailed quantitative representation of metabolic systems without explicitly referring to a set of differential equations. Smallbone *et al.* (2007) presented a method for building kinetic models solely based on reaction stoichiometries of a network using linlog kinetics. Their results showed good agreement between the real dynamics and their approximation in a yeast glycolysis model. Ao *et al.* (2008) also acknowledged that the scarcity of experimental data for rate equations and parameters is a major

*To whom correspondence should be addressed.

difficulty in the construction of large kinetic models, and to alleviate this difficulty, they used a generic form of rate equations with a minimum set of parameters to construct a metabolic model of *Methylobacterium extorquens* AM1. Their results showed that attaining the correct dynamical behaviour of a system is possible without the use of extensive and accurately measured rate equations and kinetic parameters. Furthermore, through an analysis of several systems biology models, Gutenkunst *et al*. (2007) suggested that parameter fitting to experimental data still leaves parameters poorly constrained and that biological systems are often robust to large parameter variations. The success of a model is therefore more dependent on an accurate prediction of the main behaviours of a system rather than on a thorough determination of large numbers of parameter values.

In order to streamline the construction of large-scale dynamic models, the difficulties related to the manual assembly of large networks and the generation of customized rate equations and parameters need to be addressed. For this reason, we developed a software tool named GRaPe (Gene-Reaction-Protein integration). GRaPe uses generic reversible Michaelis–Menten rate equations based on the number of substrates and products for all reactions in the network. The Michaelis–Menten relation offers a reliable approximation of the kinetics obeyed by most enzyme-catalytic reactions. Furthermore, most reactions of importance in biochemistry are reversible in the practical sense (Cornish-Bowden, 2004). We make two distinctive assumptions, namely that compartmentalization of the cell and metabolite–enzyme interactions play a negligible role in determining the behaviour of a system. GRaPe then creates a kinetic model of the metabolic system using ordinary differential equations (ODEs) that are automatically generated based on the stoichiometric matrix of the network.

While many tools exist for the modelling and simulation of complex biological dynamic systems, e.g. CellDesigner (Funahashi *et al*., 2003), COPASI (Mendes *et al*., 2006), Biological-Networks (Baitaluk *et al*., 2006), E-Cell (Tomita *et al*., 1999), CADLIVE (Kurata *et al*., 2003) and Cellware (Dhar *et al*., 2004), none of these tools allows for the generation of rate equations from the stoichiometry and for the seamless integration of gene or protein levels into a metabolic network without time-consuming and error-prone manual intervention. Our aim is not to duplicate these tools by creating another simulation software, but to introduce an upstream solution for the rapid generation of large-scale dynamic models, which can be exported for simulation by existing software applications. Being consistent with the standards of systems biology, GRaPe supports the exchange of Systems Biology Markup Language (SBML; Hucka *et al*., 2003) level 2 version 1 and 2 documents.

In this article, we provide an overview of the main features of GRaPe and present a proof-of-principle of the applicability of our approach to the construction of large-scale kinetic models. In particular, we compare the features of a model of the yeast glycolysis pathway based on generic equations automatically generated by GRaPe with a model of the same pathway constructed by Teusink *et al.* (2000) that was based on an experimental determination of rate equations and parameters. Our results show an excellent agreement between both models, supporting the hypothesis that kinetic models using generic equations could successfully reproduce the global behaviour of large metabolic systems without requiring detailed knowledge of the *in vivo* kinetics of each individual reaction.

2 FEATURES OF GRAPE

2.1 General features

GRaPe provides a user-friendly graphical user interface (GUI) for importing, creating, editing and exporting biological models in SBML. GRaPe automatically integrates every metabolic reaction with either an enzyme species or with a gene expression process (Fig. 1). When only proteomic data is defined in the SBML document, GRaPe adopts the Reaction-Protein (RP) representation. When transcriptomic data is given, GRaPe then adopts the Gene Reaction-Protein (GRP) representation. Transcription, translation and degradation of both mRNA and proteins are then expressed mathematically.

GRaPe also provides functionality to manually construct GRP and/or RP network models. In both model-building processes, GRaPe automatically generates a Michaelis–Menten reversible rate equation for each reaction based on its stoichiometry and enzyme mechanism by iterating through the metabolic network. Each reaction in the network can have up to two substrates and products. See Supplementary Material 1 for the detailed list of rate equations used by GRaPe.

GRaPe implements two robust methods, the Levenberg–Marquardt method (LMA) and a genetic algorithm (GA), for estimating kinetic parameters, in addition to the Steady-State Enzyme Estimator (SSEE) method. See Section 3.3 for more details about the estimation methods and their application. The two parameter estimation methods attempt to find the values of missing kinetic parameters, given the experimental time series data. Both methods work interchangeably so that when one method fails to find a suitable solution set (a low objective function) the other is employed. The parameter set from the method that returns the best objective function is then taken.

The time series data used as an input for all estimation methods must be in a tab-delimited plain text file. GRaPe matches the identifiers (ids) in the model to the ids in the data file during parameter estimation and throws an error if any of the ids in the model are not found in the input data. The data file for all parameter estimation methods is the same; which means GRaPe treats the last row of the data file as the steady-state data for the SSEE method.

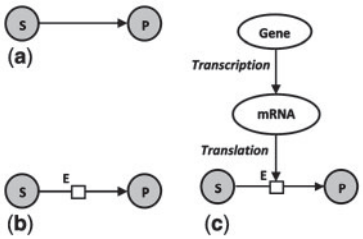

Fig. 1. (**a**) A traditional representation of a metabolic reaction where a substrate, S, gets converted into a product, P. The protein concentration is expressed as being fixed, usually in the V_{max}, a constant in the rate equation. (**b**) An RP representation, where reaction (a) is integrated with only its protein concentration, E. In the rate equation, E is expressed as an independent variable, which can be varied. (**c**) A GRP representation, where an RP reaction is fully integrated with its gene expression module. E is now expressed as a function of transcription, translation and degradation of both mRNA and E.

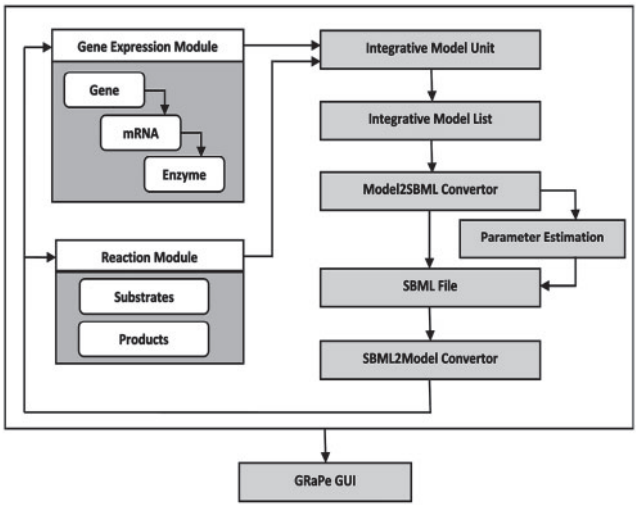

Fig. 2. System architecture of GRaPe. The 'Gene Expression' module takes in the gene(s), mRNA and enzyme species of a reaction. Transcription, translation and degradation of mRNA and enzyme are expressed in this module. The 'Reaction' module constructs a reaction based on the number of substrates, products and reaction mechanism. An 'Integrative Model Unit' module is then created from either 'Gene Expression' or just the enzyme species only and 'Reaction' modules. These units are stored in a list ('Integrative Model List') before being converted into a SBML file by the Model2SBMLConvertor. The 'Parameter Estimation' module provides methods for estimating kinetic parameters in a model. SBML files imported by GRaPe are disintegrated into separated units by the SBML2ModelConvertor. The GUI serves as a platform for creating and editing the model.

The time series data can correspond to experimental or simulated, continuous dynamic or steady-state data.

The estimation procedure constrains parameters to experimental data and therefore assignment of initial concentrations of species must be the same in the model as in the experimental data. It is also recommended that the precision of experimental values be limited to two decimal places for faster estimation. The time taken to estimate parameters for all reactions in a model is dependent on the amount of input data. For dynamic time series data, the total estimation time tends to take longer compared to steady-state time series data. Although there is no limit to the number of data points in the input data, large datasets increase estimation time.

2.2 System architecture

In order to make GRaPe platform independent and easy to use by the biological community, Java was chosen as the programming language. Figure 2 shows the architecture of GRaPe and the interactions between the main components of the system. GRaPe uses the JigCell Parser (Vass *et al*., 2004) for importing and exporting SBML level 2 version 2 documents. Each reaction in the model consists of its substrates, products and either the enzyme alone or the full gene expression process of that reaction. The reactions are stored in a list that later gets converted by the JigCell parser into a SBML file. The inverse of this process can be achieved, i.e. by parsing a SBML file that is then decomposed into reactions, species/metabolites and enzymes.

3 METHODS

3.1 Modelling gene expression

Many methods have been used to model gene expression, which include Boolean networks (Kim *et al*., 2007; Li, 2007), differential equations (Chen *et al*., 1999), dynamic Bayesian networks (Kim *et al*., 2003; Li, 2007) and neural networks (Gagneur and Klamt, 2004). Boolean networks are inexpensive with respect to computational complexity (D'haeseleer, 1999). An advantage of Bayesian networks, like Boolean networks, is that they do not require the explicit determination of kinetic parameters. However, both methods are poor in capturing some important aspects of network dynamics (Schlitt and Brazma, 2007).

We model gene expression using ordinary differential equations; this enables details of the dynamics of the network to be captured by explicitly modelling changes in concentrations of mRNA and proteins over time (Chen *et al*., 1999). Also, gene expression levels tend to be continuous rather than discrete; discretization can lead to loss of information (D'haeseleer, 1999). Smolen *et al*. (2003) and Garcia-Martinez *et al*. (2007) studied the relationship among variables that characterize gene expression at the genome-level in living organisms. Their studies concluded that the amount of both mRNA and proteins primarily depends on their transcription rate (K_{Tr}), translation rate (K_{Tl}), mRNA concentration ([*mRNA*]) and protein concentration ([*Protein*]).

We model the change in mRNA concentration over time as:

$$\frac{\mathrm{d}([mRNA])}{\mathrm{d}t} = v_{\mathrm{TR}}\mathrm{Gene}(t) - k_{\mathrm{mRNADeg}}[mRNA] \quad (1)$$

where [*mRNA*] is the mRNA concentration, v_{TR} is the transcription or mRNA synthesis rate, and k_{mRNADeg} is the mRNA degradation rate. Gene(t) is an expression function which may take any real value, enabling external regulators of gene expression to be incorporated into models. However, in the examples presented here, no such regulation was included and these values were taken to be Boolean, where a value of 1 indicated that the gene was expressed and 0 otherwise.

The change in protein concentration over time is modelled as:

$$\frac{\mathrm{d}([P])}{\mathrm{d}t} = k_{\mathrm{Tl}}[mRNA] - k_{\mathrm{ProteinDeg}}[P] \quad (2)$$

where [P] is the protein concentration, k_{Tl}, the protein synthesis or translation rate, $k_{\mathrm{ProteinDeg}}$ is the protein degradation rate and [*mRNA*], the concentration of mRNA. Both the equations (1) and (2) are based on first order mass action kinetics.

Methods have been developed to determine or estimate the synthesis and degradation rates from microarray data (D'haeseleer, 1999; Wu *et al*., 2004). When these parameters are unknown, we present an alternative modelling approach where only the protein level is integrated into the reaction instead of the full gene expression process (Fig. 1b). The concentration of enzyme can be set as fixed over time or varied during simulation. This can be achieved by the use of SBML 'Events' that represent time-dependent changes within the system. An event can be triggered if a certain condition is reached; for example, set the concentration of enzyme 'A' to 0.8 mM if time is >10 min. Furthermore, the modelling framework enables the incorporation of isoenzymes that can be modelled as individual enzyme species with their own gene expression processes.

3.2 Enzyme kinetics and rate equations

GRaPe automatically generates a rate equation for a reaction in a network based on the assumed enzyme mechanism governing that reaction, and its number of products and substrates. The enzyme mechanism of a reaction can either be of random order or compulsory order. If the binding order of substrates and releasing order of products are unknown then the random-order mechanism is recommended. The compulsory-order mechanism requires the knowledge of the correct order of binding of substrates to the

proteins and releasing of products to be known. This mechanism proceeds in an ordered series of steps, i.e. the substrate must bind in particular order and the product is released in a specified order.

The automatic generation of generic rate equations is a key advantage of GRaPe. This is time-efficient and less error-prone for a relatively large model compared to the manual definition of each rate equation by the user. COPASI provides predefined rate equations for reactions in a system, but these equations must be manually assigned to reactions by the user and protein levels cannot be explicitly assigned in rate equations. SBMLSqueezer is a plug-in for CellDesigner, which generates rate equations for a biochemical reaction (Dräger, 2008); but the user has to manually select the type of rate equation for each reaction in the network, and protein levels are not integrated in rate equations. On a large scale, these solutions are impractical and leave the model with unknown parameters.

GRaPe uses the King & Altman method (King and Altman, 1956) to derive rate equations based on a reaction's stoichiometry and the enzyme mechanism under the steady-state assumptions. Supplementary Material 1 describes the generic Michaelis–Menten rate equations used by GRaPe for the different reaction types. It generates kinetic rate equations for reactions of up to two substrates or products, i.e. reactions can be of type uni–uni, uni–bi, bi–uni or bi–bi. A reaction of more than three substrates or products needs to be decomposed into these reaction types based on its biochemistry.

3.3 Parameter estimation

Kinetic models are shown to produce accurate and testable results. However, the number of large-scale kinetic models has been very low due to the enormous number of kinetic parameters needed to define the system. Furthermore, as observed in Teusink *et al.* (2000), *in vitro* measurements of kinetic constants may not necessarily be representative of their numerical values *in vivo* (Jamshidi and Palsson, 2008). Various software tools can now perform parameter estimation: COPASI has a list of methods for estimation including a GA; SBML-PET (Zhike and Klipp, 2006) uses a stochastic ranking evolution strategy method for parameter estimation; however, it excludes constraints on the flux of a reaction implying that a zero flux may be obtained even in non-equilibrium conditions. COPASI requires that columns specified in the experimental data file must be associated with model elements. Having a flux for a reaction in the experimental data file throws an error with COPASI, as fluxes are not explicitly expressed in a model.

While the exact values of kinetic parameters are not necessarily crucial to determine the behaviour of a biochemical system, it is nevertheless necessary to estimate the values of missing parameters in order to run simulations. In GRaPe, we have introduced a simple but effective algorithm, the SSEE that estimates the concentration of enzyme needed for a reaction to reach a steady state, v. The SSEE algorithm focuses on solving for e_0,the concentration of the enzyme, by assigning all kinetic parameters in the models to a constant value of 1. For a uni–uni reaction, e_0 is calculated as follows:

$$e_0 = \frac{v\left(1+\frac{a}{K_{mA}}+\frac{p}{K_{mP}}\right)}{(K_A a - K_P p)} \tag{3}$$

where K_{mA}, K_{mP}, K_A and K_P are the kinetic parameter associated to substrate A and product P, respectively, assigned a value of 1; a is the concentration of substrate and p is the concentration of the product. SSEE allows for the rapid simulation of steady-state behaviour in a system without prior knowledge of kinetic parameters.

In addition, GRaPe implements two methods for parameter estimation from time series of experimental data: the LMA and GA. The LMA is an upgrade from Nocedal and Wright (1999), which was constrained to work with our integrative models. GA is the predominant algorithm for estimating kinetic parameters in GRaPe. However, when GA does not return a good solution based on an objective function, then the LMA is used. GRaPe returns the solution of the algorithm with the better objective function.

4 RESULTS

In this section, we present a generic model of yeast glycolysis generated by GRaPe and compare its properties to a model developed by Teusink *et al.* (2000), which was based on an experimental determination of kinetic parameters and rate equations. The comparison is aimed at determining whether the behaviour of such a metabolic system can be predicted without accurately measuring the rate equations and detailed kinetics of every enzyme.

4.1 Case study: *S.cerevisiae* glycolytic pathway

We modelled the glycolysis pathway in *S.cerevisiae* using generic rate equations and kinetic parameters estimated by GRaPe and compared it to a detailed model by Teusink *et al.* (2000). Since the main objective of this work is to show that a generic kinetic model can provide results of similar quality to a detailed model, we compare our results with simulations of the Teusink model rather than experimental values.

The yeast glycolysis pathway has been extensively studied (Bakker *et al.*, 1997; Hynne *et al.*, 2001; Lambeth and Kushmerick, 2002; Pritchard and Kell, 2002; Teusink *et al.*, 2000) and a vast amount of genomic and enzymatic data is therefore available. In Teusink *et al.* (2000), the authors examined whether *in vivo* kinetics behaviour can be understood in terms of *in vitro* kinetics of enzymes in yeast glycolysis. They produced two models, one where branched reactions were ignored and a comprehensive model that included all branched reactions. Their results suggested that half of the enzymes matched their *in vivo* fluxes within a factor of 2, and the calculated deviation between *in vivo* and *in vitro* kinetic characteristics of the other enzymes could explain discrepancy between *in vivo* and *in vitro* kinetics. Fluxes and metabolites concentrations were experimentally determined. Fluxes of trehalose and glycogen were expressed in units of glucose, and kinetic parameters were also determined under the same experimental condition. The unbranched model used experimentally determined metabolite concentrations and calculated conserved moieties but no steady state was reached. The branched model, however, reached a steady state with the original parameter set that had been determined *in vitro*. Both models used a set of ordinary differential equations to describe the time dependence of metabolite concentration.

We modelled the glycolysis pathway of *S.cerevisiae* using GRaPe based on the branched topology used by Teusink *et al.* (2000). The initial concentration of metabolites was the same as in the Teusink model. The model includes all enzymes involved in the pathway from glucose uptake to the production of pyruvate and ethanol. All reactions were assumed to be of a random-order mechanism. GRaPe then generated generic reversible rate equations for the reactions in our glycolysis model, which were combined with the stoichiometry of the network to produce ordinary differential equations. We made three distinctive changes in our model. First, in the Teusink model, a metabolic pool represented by an independent variable, P, was defined to represent the sum of high-energy phosphate in adenine nucleotides. In our model, an equation is used for the conservation of adenine nucleotides moiety as:

$$\frac{\mathrm{d[ADP]}}{\mathrm{d}t} = -\frac{\mathrm{d[ATP]}}{\mathrm{d}t} \tag{4}$$

where [ADP] and [ATP] are the concentration of adenosine diphosphate (ADP) and adenosine triphosphate (ATP), respectively and t is the time.

Secondly, since adenosine monophosphate (AMP) does not partake in any reaction, we excluded the adenylate kinase reaction. Thirdly, in the Teusink model the triosephosphate isomerase (TPI) reaction was modelled using an equilibrium equation such that the ratio of glyceraldehyde 3-phosphate (GAP) to glycerone phosphate (DHAP) was at equilibrium. An independent variable, *Trio2-P*, was introduced, which was the sum of the concentration of GAP and DHAP. In our model, we included the TPI reaction and modelled the change in DHAP and GAP concentrations using uni–uni reversible rate equation. These changes make it possible to study the effects of varying the concentration of ATP, ADP, DHAP and GAP on the system. The initial concentration of all metabolites was then assigned using data given in the Teusink model. The concentrations of ATP and ADP were calculated based on conserved moiety equations given in Teusink *et al.* (2000). As cofactors play an important part in the global regulation of glycolysis, their concentrations were treated as free metabolic variables. The kinetic equation for each reaction was generated automatically by GRaPe based on the number of substrates and products and the enzyme mechanism of the reaction. Our glycolysis model in Figure 3 contains 23 metabolites (22; here, the bracketed data corresponds to respective entity in the Teusink model), 15 enzyme species (0), 116 kinetic parameters (88) and 18 fluxes (17). A complete list of the flux conversation equations and abbreviations used in our model is provided in Supplementary Material 2.

4.2 Data acquisition and parameter estimation

Using JWS online (Olivier and Snoep, 2004), a web tool for simulating kinetic models, we collected steady-state data for the Teusink model with glucose uptake concentrations of 10 and 50 mM. These values were then merged to create our input dataset for parameter estimation. The dataset contains values of all metabolite concentrations and reaction fluxes in the glycolysis pathway at every 10 min; the glucose concentration was at 50 mM from time 0 to 30 min and at 10 mM from 30 to 100 min; the precision of values was limited to two decimals points for faster estimation. Next, GRaPe was used to estimate the kinetic parameters for each reaction in our model so that the distance between the input dataset and the values calculated by the model was minimized. Due to the absence of gene expression and enzyme amount data in this example, the concentration of enzymes was set to a default value of 1 in both the model and dataset for parameter estimation. GA was used to estimate the kinetic parameters for each reaction after just one run of estimation. The calculated error over our input data ranged from 4.5e-13 to 2.13e-10 for a reaction's kinetic parameters set.

4.3 Experiment 1: model validation on training data

After parameter estimation was completed, the model was simulated in CellDesigner using the SBML ODE Solver (SOSlib). The results obtained from our simulations were then compared with results from the Teusink model. The first experiments were to verify whether the model correctly reproduced the behaviour of the system at steady state, without any perturbation, when glucose uptake is at 10 and 50 mM. Our results show a near-perfect agreement between our model and the Teusink model (Tables 1 and 2). These results

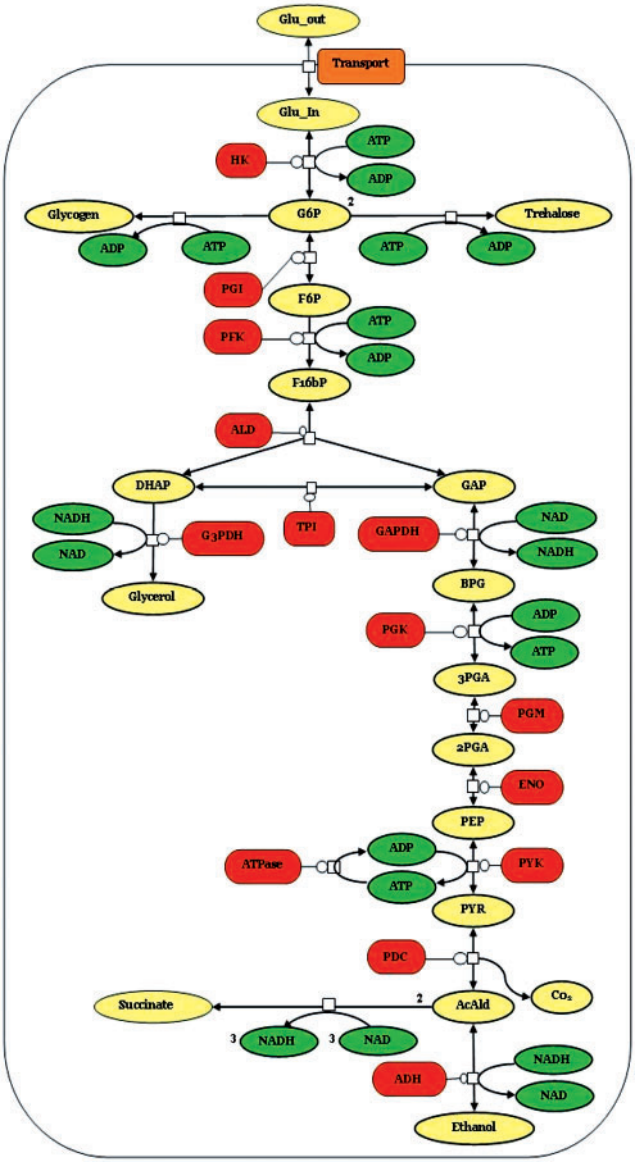

Fig. 3. Topology of our yeast glycolysis model.

confirmed that the training and parameter estimation algorithms successfully identified a solution, where the model reproduces the correct concentration and flux values used in the training data. The dataset used for parameter estimation is provided in Supplementary Material 3.

Using 'events' in SBML, we moreover replicated the effect of a dynamically reduced uptake of glucose: after 30 min, the concentration of glucose was reduced from the original 50 to 10 mM. Results from this experiment (Fig. 4a–b) again show an excellent agreement between our model and the Teusink model when the same reduction in glucose uptake is applied. These results confirm that the integrity of our estimated parameters is conserved in a dynamic experiment. Our model of glycolysis, with events for changing the level of glucose uptake, has been provided in Supplementary Material 4.

Table 1. Metabolite concentrations (in mM) and fluxes (in mmol·min^{-1}·L-cytosol^{-1}) at steady state for the model generated by GRaPe with glucose uptake levels at 1, 10, 50, 100 and 200 mM

	Concentration of glucose (mM)										
Metabolite concentrations (mM)	1	10	50	100	200	Fluxes (mmol·min^{-1} ·L-cytosol^{-1})	1	10	50	100	200
Glucose (In)	0.002	0.01	0.1	0.16	0.37	Glucose transport	88.92	88.85	88.15	87.7	89.81
ADP	2.35	1.34	1.29	1.2	1.2	HK	38.75	80.13	88.15	89.48	89.97
ATP	1.45	2.46	2.51	2.6	2.6	Glycogen	4.95	5.96	6.0	6.06	6.06
G6P	0.19	0.73	1.03	1.07	1.1	Trehalose	1.98	2.39	2.4	2.42	2.43
F6P	0.014	0.07	0.11	0.11	0.12	PGI	29.84	69.39	77.35	78.58	79.06
NAD	1.45	1.54	1.55	1.55	1.55	PFK	29.84	69.39	77.35	78.58	79.06
NADH	0.14	0.05	0.04	0.04	0.04	ALD	29.84	69.39	77.35	78.58	79.06
F16bP	0.1	0.44	0.59	0.63	0.64	G3PDH	7.6	17.2	18.2	18.67	18.73
DHAP	0.032	0.72	0.74	0.78	0.79	TPI	22.24	52.19	59.15	59.91	60.33
GAP	0.026	0.03	0.03	0.03	0.03	GAPDH	52.08	121.59	136.5	138.49	139.39
BPG	8.01e-06	1.91e-04	3.30e-04	3.88e-04	4.03e-04	PGK	52.08	121.59	136.5	138.49	139.39
3PGA	0.068	0.27	0.36	0.37	0.38	PGM	52.08	121.59	136.5	138.49	139.39
2PGA	0.008	0.03	0.04	0.04	0.04	ENO	52.08	121.59	136.5	138.49	139.39
PEP	0.01	0.05	0.07	0.08	0.08	PYK	52.08	121.59	136.5	138.49	139.39
PYR	1.88	6.73	8.52	8.8	8.92	ATPase	28.64	85.30	99.1	100.44	101.26
AcAld	0.067	0.16	0.17	0.18	0.18	PDC	52.08	121.59	136.5	138.49	139.39
						ADH	49.04	114.71	129.22	131.03	139.90
						Succinate	1.52	3.44	3.64	3.73	3.75

The kinetic parameters were trained on data with glucose levels at 10 and 50 mM. After estimation, the model was simulated in CellDesigner using the SBML ODE Solver. G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F16bP, fructose 1,6-bisphosphate; BPG, 1,3-bisphosphoglycerate; 3PGA, 3-phosphoglycerate; 2PGA, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; AcAld, acetaldehyde; HK, hexokinase; PGI, glucose 6-phosphate isomerase; ALD, fructose-bisphosphate aldolase; G3PDH, glycerol 3-phosphate dehydrogenase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; PDC, pyruvate decarboxylase; ADH, alcohol dehydrogenase; NAD/NDAH, nicotinamide adenine dinucleotide.

4.4 Experiment 2: model validation outside the training range

The second experiment was carried out to verify how well the model would predict a new state of the glycolysis pathway outside the range of training data and without re-estimating the kinetic parameters. We carried out simulations by changing the level of glucose to 1, 100 and 200 mM. The results, also shown in Tables 1 and 2, show an excellent agreement between our model and the Teusink model with glucose increased to 100 and 200 mM.

Our model still produced results with very low concentrations of glucose (1 mM), while the Teusink model reported an error during simulation when the glucose input was <2 mM. The generic model thus appears to be more robust than the detailed model. We repeated a dynamic experiment, changing the glucose concentration from 50 to 100 mM after 30 min (Fig. 4c–d), which was again successful. The results obtained from these experiments demonstrate the ability of our generic model to predict new steady-state behaviours that were not used for training. Overall, our results demonstrate that it is possible to predict system behaviour using generic reversible rate equations, without addressing detailed mechanisms at the level of each component.

5 DISCUSSION AND CONCLUSION

We have introduced GRaPe, a platform-independent software tool aimed at streamlining the construction of large-scale dynamic models. GRaPe enables the automated construction of reaction-protein or gene-reaction-protein networks. A novel feature of GRaPe is its ability to generate generic rate equation for models of relatively large sizes. Another important feature is its capability to explicitly integrate gene expression processes or enzyme species into reactions, making it a convenient tool for the construction of integrative protein-reaction networks.

The current availability of high-throughput fluxomic, metabolomic, proteomic and genomic data makes it possible to envisage building integrative genome-scale metabolic models, but convenient tools for assembling such heterogeneous data on a large scale are still lacking. An aim of systems biology is to understand cellular processes as a whole rather than in isolation. Integrating cellular components is essential for our understanding of how interactions between these components influence cellular functions. A few manually constructed integrative metabolic models have now been created (Ao *et al*., 2008; Förster and Palsson, 2003; Jamshidi and Palsson, 2008); however, no computational tool for integrating protein levels into metabolic models exists. The integration of proteomics data into metabolic models could increase our understanding of the role of enzymes on metabolism. Another important feature of GRaPe is its ability to convert existing metabolic models in SBML format into either gene-reaction or gene-protein-reaction networks. This will enable, for example, the import of high-throughput quantitative proteomics data into metabolic models.

Table 2. Metabolite concentrations (in mM) and fluxes (in mmol·min^{-1}·L-cytosol^{-1}) at steady state for the Teusink model of glycolysis with glucose uptake levels at 10, 50, 100 and 200 mM

	Concentration of glucose (mM)								
Metabolite concentrations (mM)	10	50	100	200	Fluxes (mmol·min^{-1} ·L-cytosol^{-1})	10	50	100	200
Glucose (In)	0.01	0.1	0.1	0.1	Glucose Transport	80.16	88.15	88.12	88.1
ADP	n/a	1.29	n/a	n/a	HK	80.16	88.15	89.25	89.81
ATP	n/a	2.51	n/a	n/a	Glycogen	6.0	6.0	6.0	6.0
G6P	0.72	1.03	1.09	1.13	Trehalose	2.4	2.4	2.4	2.4
F6P	0.07	0.11	0.12	0.13	PGI	69.36	77.35	78.45	79.01
NAD	1.55	1.55	1.55	1.55	PFK	69.36	77.35	78.45	79.01
NADH	0.05	0.04	0.04	0.04	ALD	69.36	77.35	78.45	79.01
F16bP	0.44	0.59	0.63	0.64	G3PDH	17.24	18.2	18.34	18.41
DHAP	n/a	n/a	n/a	n/a	TPI				
GAP	n/a	n/a	n/a	n/a	GAPDH	121.48	136.5	138.57	139.62
BPG	2.00e-04	3.30e-04	3.56e-04	3.71e-04	PGK	121.48	136.5	138.57	139.62
3PGA	0.27	0.36	0.37	0.38	PGM	121.48	136.5	138.57	139.62
2PGA	0.03	0.04	0.05	0.05	ENO	121.48	136.5	138.57	139.62
PEP	0.05	0.07	0.08	0.08	PYK	121.48	136.5	138.57	139.62
PYR	6.72	8.52	8.85	9.03	ATPase	85.04	99.1	101.03	102.01
AcAld	0.16	0.17	0.17	0.17	PDC	121.48	136.5	138.57	139.62
					ADH	114.59	129.21	131.23	132.25
					Succinate	3.45	3.64	3.67	3.68

The data were obtained using the JWS Online web simulation tool. No simulation with glucose level at 1 mM was obtainable using the Teusink model. G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F16bP, fructose 1,6-bisphosphate; BPG, 1,3-bisphosphoglycerate; 3PGA, 3-phosphoglycerate; 2PGA, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; AcAld, acetaldehyde; HK, hexokinase; PGI, glucose 6-phosphate isomerase; ALD, fructose-bisphosphate aldolase; G3PDH, glycerol 3-phosphate dehydrogenase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; PDC, pyruvate decarboxylase; ADH, alcohol dehydrogenase; NAD/NDAH, nicotinamide adenine dinucleotide.

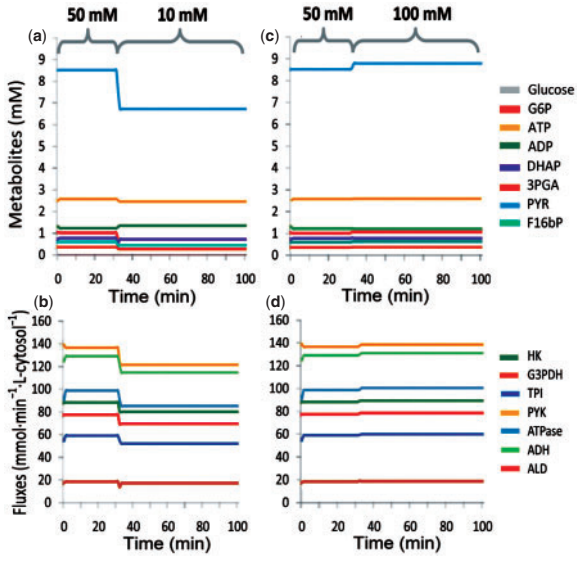

Fig. 4. Dynamic experiments using the generic glycolysis model with the level of glucose uptake being changed after 30 min. (**a**–**b**) A decreased glucose uptake from 50 to 10 mM. (**c**–**d**) An increased glucose uptake from 50 to 100 mM.

Our research ties in with previous investigations indicating that the dynamic behaviour of metabolic systems can be predicted without accurately measuring all rate equations and detailed kinetic parameters. Ao *et al*. (2008) have already used generic rate equations to construct a metabolic model of *M.extorquens* AM1. Their results showed that it is possible to attain the dynamical behaviour of a system without the use of extensive and accurately measured rate equations and kinetic parameters. GRaPe follows this principle to enable the building of large models. It generates generic rate equations for all reactions in a metabolic network, and thus assumes that the global behaviour of a system should be relatively independent of precise kinetic properties and parameter values. It is worth noting that metabolic systems have long been known to be robust to perturbations and maintain relatively stable intracellular metabolite and flux levels in response to changing external conditions. This property was reflected by the Teusink model, as we have shown in a previous study (Schwartz and Kanehisa, 2006), and it is conserved in our generic model.

Parameter estimation (optimization) has become an area of significant importance in kinetic modelling due to the fact that it is often prohibitively expensive and time-consuming to measure vast numbers of kinetic parameters experimentally. Some repositories such as Sabio-RK (Rojas *et al.*, 2007) and BRENDA (Schomburg *et al.*, 2002) store kinetic parameters and enzymatic information for various pathways in different organisms.

However, it is difficult to compare parameters of the same pathway in different models due to different assumptions and

experimental conditions. Gutenkunst *et al.* (2007) suggested that modellers should focus on predicting the behaviour of the system rather than parameters due to parameter 'sloppiness', meaning that parameters are often poorly constrained. GRaPe introduces a simple but effective method for estimating the amount of enzyme concentration required to give a particular steady state. This method enables analysis of the steady-state behaviour without detailed knowledge of kinetic parameters. GRaPe also has two methods for parameter estimation: the LMA algorithm and GA. Both methods are robust and work interchangeably in estimating kinetic parameters.

The capability of GRaPe to convert reactions into ODEs based on their stoichiometric matrix for small or large-scale networks, its main innovation, makes it complementary to other existing software tools. GRaPe is not designed to compete with well-developed simulation softwares, but to complement existing applications by providing an upstream solution for the efficient design of large-scale dynamic models. Models created using GRaPe can be run using existing simulation tools such as CellDesigner, COPASI and any other tools that support SBML. In the future, we aim to interface GRaPe with existing databases of metabolic reactions and kinetic parameters in order to make it capable of rapidly constructing large- or genome-scale integrative kinetic models.

Funding: Biotechnology and Biological Sciences Research Council, UK (to D.A.A.); Special Training Fellowship in Bioinformatics from the Medical Research Council, UK (to J.H.).

Conflict of Interest: none declared.

REFERENCES

Ao,P. *et al*. (2008) Towards kinetic modeling of global metabolic networks: Methylobacterium extorquens AM1 growth as validation. *Chin. J. Biotechnol.*, **24**, 980–994.

Baitaluk,M. *et al*. (2006) BiologicalNetworks: visualization and analysis tool for systems biology. *Nucleic Acids Res.*, **34**, W466–W471.

Bakker,B.M. *et al*. (1997) Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. *J. Biol. Chem.*, **272**, 3207–3215.

Chen,T. *et al*. (1999) Modeling gene expression with differential equations. *Pac. Symp. Biocomput.*, **4**, 29–40.

Cornish-Bowden,A. (2004) *Fundamentals of Enzyme Kinetics*. 3rd edn. Portland Press, London, pp. 30–51.

D'haeseleer,P. *et al*. (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pac. Symp. Biocomput.*, **4**, 41–52.

Dhar,P. *et al*. (2004) Grid Cellware: The first Grid-enabled tool for modeling and simulating cellular processes. *Bioinformatics*, **20**, 1319–1321.

Dräger,A. *et al*. (2008) SBMLsqueezer: a CellDesigner plug-in to generate kinetic rate equations for biochemical networks. *BMC Syst. Biol.*, **2**, 39

Förster,J. *et al*. (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res.*, **13**, 244–253.

Funahashi,A. *et al*. (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO*, **1**, 159–162.

Gagneur,J. and Klamt,S. (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, **5**, 175.

Garcia-Martinez,J. *et al*. (2007) Common gene expression strategies revealed by genome-wide analysis in yeast. *Genome Biol.*, **8**, R222.

Gutenkunst,R.N. *et al*. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, **3**, 189.

Hynne,F. *et al*. (2001) Full scale model of glycolysis in Saccharomyces cerevisiae. *Biophys. Chem.*, **94**, 121–163.

Hucka,M. *et al*. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Jamshidi,N. and Palsson,B.Ø. (2008) Formulating genome-scale kinetic models in the post-genome era. *Mol. Syst. Biol.*, **4**, 171.

Kim,S.Y. *et al*. (2003) Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief. Bioinform.*, **4**, 228–235.

Kim,H. *et al*. (2007) Boolean networks using the chi-square test for inferring large-scale gene regulatory networks. *BMC Bioinformatics*, **8**, 37.

King,E.L. and Altman,C. (1956) A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J. Phys. Chem.*, **60**, 1375–1378.

Kitano,H. (2002) Systems biology: a brief overview. *Science*, **295**, 1662–1664.

Kurata,H. *et al*. (2003) CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle. *Nucleic Acids Res.*, **31**, 4071–4084.

Lambeth,M.J. and Kushmerick,M.J. (2002) A computational model for glycogenolysis in skeletal muscle. *Ann. Biomed. Eng.*, **30**, 808–827.

Li,P. (2007) Comparison of probabilistic Boolean network and dynamic Bayesian network approaches for inferring gene regulatory networks. *BMC Bioinformatics*, **8** (Suppl. 7), S13.

Mendes,P. *et al*. (2006) COPASI: a COmplex PAthway SImulator. *Bioinformatics*, **22**, 3067–3074.

Nocedal,J. and Wright,S.J. (1999) Numerical Optimization. Springer, New York, pp. 262–266.

Olivier,B.G. and Snoep,J.L. (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics*, **20**, 2143–2144.

Pritchard,L. and Kell,D. (2002) Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. *Eur. J. Biochem.*, **269**, 3894–3904.

Rojas,I. *et al*. (2007) Storing and annotating of kinetic data. *In Silico Biol.*, **7**, S37–S44.

Schomburg,D. *et al.* (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem. Sci.*, **27**, 54–56.

Schwartz,J.M. and Kanehisa,M. (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinformatics*, **7**, 186.

Smallbone,K. *et al*. (2007) Something from nothing: bridging the gap between constraint-based and kinetic modelling. *FEBS J.*, **274**, 5576–5585.

Smolen,P. *et al*. (2003) Mathematical modeling of gene networks. *Neuron*, **26**, 567–580.

Steuer,R. *et al.* (2006) Structural kinetic modelling of metabolic networks. *Proc. Natl Acad. Sci. USA*, **103**, 11868–11873.

Teusink,B. *et al*. (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.*, **267**, 5313–5329.

Tomita,M. *et al*. (1999) E-CELL: software environment for whole cell simulation. *Bioinformatics*, **15**, 72–84.

Vass,M. *et al*. (2004) The JigCell model builder and run manager. *Bioinformatics*, **18**, 3680–3681.

Wu,C.C. *et al*. (2004) GeneNetwork: an interactive tool for reconstruction of genetic networks using microarray data. *Bioinformatics*, **20**, 3691–3693.

Zi,Z. and Klipp,E. (2006) SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics*, **22**, 2704–2705.