Reconstruction of dynamic metabolic networks: challenges, limitations and alternative solutions

Rafael S. Costa, Daniel Machado, Isabel Rocha, Eugénio C. Ferreira
IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, P-4710-057 Braga, Portugal

1. MOTIVATION:

Dynamic modelling of metabolic networks is a powerful tool to guide experimentation and to explain properties of complex biological systems. The large-scale kinetic models at the network reaction level are usually constructed using mechanistic enzymatic rate equations and a large number of kinetic parameters. However, one of the biggest obstacles to construct accurate dynamic models is the lack of detailed knowledge of the rate laws and the difficulty in the identification of their associated kinetic parameters. In this work, we provide a critical overview of the most important limitations found during the reconstruction of the central carbon metabolism dynamic mechanistic model from *E. coli* (based on kinetic data available) developed in our group [1]. We suggest briefly some strategies that will hopefully allow the systems community to improve the traditional construction of large-scale metabolic dynamic kinetic models. Furthermore, while a huge amount of standard kinetic information is not available, we test a suitable alternative modelling approach with a relatively few number of kinetic parameters composed on approximated linlog kinetics and constraint-based method. The approach is illustrated to reconstruct the large-scale central carbon metabolism model of *E. coli*.

2. BACKGROUND:

The great challenge in the post-genomic era is to understand the dynamic behaviour of microbial cells. During the last years, the tremendous increase in the availability of biological data due to novel high-throughput analytical techniques allowed an unprecedented insight on intracellular dynamics [2]. However, due to the intrinsic complexity of biological systems, more detailed mathematical models are necessary to integrate these experimental data in the interest of understanding cellular metabolism under a quantitative aspect. The most suitable large-scale cell models are constraint-based models and models that account for dynamics at the enzyme level [3,4]. Dynamic modeling of large scale metabolic systems predominantly use non-linear ordinary differential equations (ODE's) and require *a priori* knowledge on the network structure and a large amount of experimental information, such as, initial concentration of metabolites, kinetic parameters and detailed rate laws. A major challenge with such models, however, is that they often possess many kinetic parameters [5]. While network information has been compiled in public databases [6], there are currently limited methods for measuring kinetic parameters [7]. In addition, for a large number of enzymes, the kinetic parameters are usually unknown or are available in the literature or in databases only as general values obtained by *in vitro* experiments by enzymologists [8]. These parameters should be used with care by modelers, since enzymologists in general work under optimal conditions for the enzyme and do not perform the enzyme characterization under physiological conditions, restricting their *in silico* applicability (see limitations in section 3). A common approach to address this issue has been the use of time course *in vivo* data in response to a stimulus [9] for kinetic parameter estimation by minimizing an objective function [10]. On the other hand, the true mechanistic enzyme-kinetic rate law for a specific reaction is frequently not known for most of the enzymes. For these reasons, the applicability of this traditional dynamic approach to kinetic models requires a large amount of experimental data to represent the physiological kinetic behaviour and has been limited to biochemical networks of small size [11], with the exception of the human red blood cell model [12].

Alternatively, the constraint-based modeling approach, which is used usually to predict the effect of gene knockouts in metabolic phenotypes in microorganisms requires only stoichiometric information and physicochemical constraints [13]. Although these models can be used to predict steady-state behaviour using flux analysis, they fail to capture the transient behaviour. These facts reveal the need to develop alternative approaches for large scale dynamic models. Recently, a great effort has been carried out by researchers in developing alternative approaches for modeling large-scale metabolic networks, like statistical frameworks, approximate kinetic formats, and hybrid modelling approaches[14-18]. The most important advantage of the approximated rate equations like linear-logarithmic (linlog) or power law kinetics is the lower number of parameters and a universal applicability when the catalytic mechanism is unknown. Parameters of such approximated kinetics can be estimated from time course and/or steady-state experimental data [19,20] and also inferred from the stoichiometry of the reactions [15]. Another

approach was developed by Yugi *et al*. [17]. The proposed method aims to reduce the number of enzyme kinetic assays necessary to build a dynamic model, by considering a dynamic and a static part. The static module is calculated by metabolic flux analysis (MFA) constrained by the dynamic reactions. There are some limitations to be considered for accurate simulations, such as the need to obtain elasticity coefficients at boundary reactions between modules and inconsistencies in the static module caused by the inclusion of irreversible reactions. More recently, Smallbone and co-workers [15] proposed a method combining two modelling approaches (linlog kinetics and constraint-based modelling), in which the parameters (elasticities) are given by the stoichiometric coefficient for the respective metabolites and the steady state fluxes by the flux balance analysis (FBA) approach.
This paper is organized as follows: some problems found during the reconstruction of dynamic metabolic networks are briefly given in section 3; in section 4, the methods for the alternative modelling approach are presented; afterwards, in section 5 the results are presented and discussed; and finally, section 6 provides the conclusions.

3. Some limitations of available kinetic data to set-up large-scale dynamic models:

One of the steps of a typical dynamic model building cycle is the collection of various kinetic/thermodynamic parameter values with the respective kinetic rate laws from public databases, the literature and/or their estimation from time course experimental data. However, while information on pathways has been compiled in several available publicly databases [6], there are currently few databases collecting kinetic data. In addition, the available kinetic data has several inconsistencies for constructing dynamic models of metabolism: (1) Kinetic parameters are usually available in databases like BRENDA and SABIO-DK but lack the kinetic equations describing the associated rate law [21]; (2) In those data sources, for reversible reactions, it is frequent to find only the parameters for the forward or reverse reaction and rarely for both [22]; (3) Experimental conditions under which the kinetic parameters had been determined and the methodology of the assay are rarely available. On the other hand, even in the cases where some of this information is available, it refers generally only to temperature and pH. Therefore, data standardization is necessary to reach comparability of enzyme kinetic data and to ensure data quality. A good starting point for the experimentalists to achieve standardization of the kinetic data in the future is to follow the recent recommendations from the STRENDA (Standard for Reporting Enzymological Data) commission (www.strenda.org); (4) The maximum velocity (v_{max}) depends on the amount of enzyme present and is often measured in test tubes. However, in the dynamic modeling we are interested in the *in vivo* data. Furthermore, the values usually reported in the literature are of specific activities and come as μmol min^{-1} $mg^{-1}_{protein}$. However, since information on the conditions in which the assay was performed (pure enzymes or cellular extracts, for example) is scarce, it is impossible to convert these units to maximum velocities; (5) Sometimes values regarding an inhibition constant are given in literature or databases, but information on the type of inhibition they refer to (competitive, noncompetitive, uncompetitive, etc.) is not available.

4. METHODS:

Dynamic *E. coli* central carbon metabolic network and parameter estimation

The reconstructed model is based on the full mechanistic model of the central carbon metabolism of *E. coli* formulated by Chassagnole *et al*. [3] available in SBML format from Biomodels online database [23]. The original model integrates the reactions of glycolysis, pentose phosphate pathway and the phosphotransferase system (PTS) and was extended to represent also the TCA cycle, glyoxylate bypass and the acetate metabolism [1]. Only reactions confirmed in literature that are active in *E. coli* are included. The reconstructed *E. coli* kinetic model consists on 30 metabolites, 7 cometabolites and 111 kinetic parameters (elasticities) in the linlog formulation. The simulations were performing by solving the differential equations (ODE's) using the numerical algorithm available in the Complex Pathway Simulator (Copasi) software tool *v.4.4* [24]. The parameter estimation for the whole reconstructed linlog model was performed with all the 18 observable pseudo-experimental time series data sets generated by simulation of the full mechanistic *E. coli* model and were used as noise-free pseudo metabolome data. The metabolite concentration time series data sets were obtained at sampling interval of 1.2 seconds. The parameter estimation was conducted using the evolutionary programming (EP) algorithms available in Copasi and the number of population size was set to 100. To make sure that this algorithm does not "stop" at sub-optimal local minima five different estimation runs were performed. We have considered the EP method for parameter estimation because evolutionary algorithms have proven to have key advantages in large inverse problems of quantitative mathematical models [25].

Flux Balance Analysis (FBA)

The FBA approach applied to constraint-based models uses linear programming (LP) optimization to maximize or minimize an objective function under different constraints and is based on a steady state approximation to the internal metabolites concentrations [26]. In mathematical terms, FBA use of LP to predict the metabolic flux distribution vector *v* is formalized as follows:

$$\begin{aligned} &\text{maximize} \quad Z = c^T v_j \\ &\text{subject to} \quad \sum_{j=1}^{N} S_{ij} v_j = 0, \qquad i = 1,...M \\ &\qquad\qquad v_j^{min} \le v_j \le v_j^{max}, \qquad j = 1,...N \end{aligned} \tag{1}$$

For metabolic engineering applications, the linear objective function (*Z*) to be maximized can correspond to different objectives ranging from a particular design objective (e.g. optimization of a desired metabolite) to the maximization of biomass growth. The mass balance constrains are imposed by a system of linear equations, where S_{ij} is an $i \times j$ stoichiometric matrix, in which *i* is the number of metabolites and *j* is the number of reactions, and v_j represents the flux of reaction *j*. Some other constraints based on physicochemical or physiological aspects may be applied, such as thermodynamic considerations that restrict the capacity and flow direction by setting v^{min} and v^{max} as lower and upper bounds on flux values.

In our genome-scale metabolic model, we use the *E. coli* metabolic network of Reed *et al*. [27] containing 1075 reactions catalyzed by 904 enzymes. The LP problem is performed for a steady-state flux distribution that maximizes biomass growth rate and only the glucose consumption rate was always set as constraint to calculate the FBA solution. The FBA computations were performed employing an in-house developed software tool OptFlux (www.optflux.org).

LinLog Kinetics

The non-mechanistic linlog representation [28] is based on the notion that the relation between the rate of reaction and the thermodynamic driving force is proportional. In this type of kinetics the parameters are the elasticities (ϵ^0_S, ϵ^0_P, ϵ^0_I and ϵ^0_A) and the steady-state fluxes (J^0). All the reactions have the same mathematical structure with linearity in the elasticities and the effect of metabolites levels on the flux is described as a linear sum of logarithmic concentration term given by:

$$r = J^0 \frac{e}{e^0}\left(1 + \sum_i \varepsilon^0_{Si} \ln(\frac{S_i}{S_i^0}) + \sum_i \varepsilon^0_{Pi} \ln(\frac{P_i}{P_i^0}) + \sum_i \varepsilon^0_{Ii} \ln(\frac{I_i}{I_i^0}) + \sum_i \varepsilon^0_{Ai} \ln(\frac{A_i}{A_i^0})\right) \tag{2}$$

where, e/e^0 represents the relative enzyme activities. S_i/S_i^0, P_i/P_i^0, I_i/I_i^0 and A_i/A_i^0 are the relative concentrations of the substrates, products, inhibitors and activators metabolites, respectively. Thus, one single elasticity per metabolite is involved in each reaction. The superscripts (^{0}) denote the reference state (e.g., steady-state from wild-type). In this work, the initial metabolite concentrations of the reconstructed model were taken as our reference state and the enzyme ratio level (e/e^0) is set to be 1 assuming that the enzyme level remains constant during the simulation.

5. RESULTS AND DISCUSSION:

The motivation for our dynamic modelling strategy is to provide a suitable approach to reconstruct large scale dynamic networks containing a small number of kinetic parameters that can be simultaneously estimated. For this purpose, we apply the special variant of linlog kinetics approximation, to the mechanistic reconstructed *E. coli* model [1]. The crucial point in our approach is to identify the steady state fluxes. Experimentally, to obtain this steady state fluxes in a network with known stoichiometry, isotopic tracer experiments with ^{13}C-labeled molecules are in general required [29]. Here, the steady state fluxes were estimated by applying FBA to the *E. coli* network [27], while defining as input cellular growth as the objective function to maximize and constraining only the glucose consumption rate to 0.200043 mM s^{-1}. This rate value was obtained from the original mechanistic model which is in agreement with the literature experimental value (0.210006 mM s^{-1}) at a dilution rate of 0.1h^{-1} [2]. The results from the real steady-state fluxes from Chassagnole *et al*. model or *in vivo* data and the corresponding fluxes computed by FBA are given in Table 1.

Table 1: Comparison between fluxes from Chassagnole *et al*. model [3] or *in vivo* data [2] and estimated FBA fluxes (J^0), from *E. coli* at $0.1h^{-1}$ dilution rate.

Reaction (Chassagnole / FBA)	Flux (mM s^{-1})		Reaction (Chassagnole / FBA)	Flux (mM s^{-1})		
	Chassagnole	FBA		Chassagnole	*in vivo*	FBA
PTS / GLCpts	0.200	0.198	**DAHPS / DDPA**	0.00787		0.00788
PGI / PGI	0.0766	0.0826	**PDH / PDH**	0.188		0.147
PGM / PGMT	0.00268	0.00333	**MethSynth / TRPAS2**	0.00226		0.00104
G6PDH / G6PDH2r	0.121	0.114	**PGDH / GND**	0.121		0.114
PFK / PFK	0.147	0.135	**R5PI / RPI**	0.0498		0.0594
TA / TALA	0.0395	0.0308	**Ru5P / RPE**	0.0712		0.0543
TKa / TKT1	0.0395	0.0313	**PPK / PRPPS**	0.0103		0.0180
TKb / TKT2	0.0317	0.0230	**G1PAT / GLGC**	0.00266		0.00295
MurSynth / -	0.000437	-	**PTA / PTAr**		0.0	0.0110
ALDO / FBA	0.147	0.135	**ACKA / ACKr**		0.0	0.0110
GAPDH / GAPD	0.325	0.301	**CS / CS**		0.173	0.0867
TIS / TPI	0.145	0.132	**ACN / ACONT**		0.173	0.0867
TrpSymth / -	0.00104	-	**ICD / ICDHyr**		0.119	0.0867
G3PDH / G3PD2	0.00182	0.00317	**KDH / TEST_AKGDH**		0.101	0.0654
PGK / PGK	0.325	0.3013	**SYN / PHETA1**		0.0174	0.00034
SerSynth / PGCD	0.0178	0.0409	**ScAS / SUCOAS**		0.101	0.0654
PgluMu / PGM	0.307	0.2604	**SDH / SUCD1i**		0.155	0.0664
ENO / ENO	0.307	0.2604	**FUM / FUM**		0.155	0.0855
PK / PYK	0.0380	0.0	**ICL / ICL**		0.054	0.00096
PepCxylase / PPC	0.0459	0.0448	**MS / MALS**		0.054	0.00096
Synth1 / -	0.0144	-	**ACOOAsynth / G1PACT**		0.059	0.00138
Synth2 / ACLS	0.0536	0.0159				

For the new reactions in the reconstructed network the estimated FBA fluxes are compared with *in vivo* fluxes."-" denotes "no data".

Generally, we observe that the discrepancies in the flux values were relatively small, with a low relative error for most of the reactions. Moreover, the accuracy of such results could be improved through incorporation additional constraints to FBA solution.

Using the steady state fluxes computed by FBA in Table 1, and the elasticities estimated from parameter fitting as described in the Methods section for the reconstructed *in silico* network under study, we used equation (2) to define the ODE system and simulate the model. The simulated time course data for some metabolites after glucose impulse at time zero are shown in Figure 1 resulting from 5 estimation runs. The trajectories represent simulations for the mechanistic model and the reconstructed linlog model. Although a rather simple variant of linlog kinetics for each single reactions has been applied, the whole network describes satisfactory well the experimentally metabolic change and only some of the parameters obtained large variance among the results of the estimation runs. A limitation in applying the proposed method is that we assume that the concentrations of the 18 metabolites are measurable. However, the new high-throughput approaches for detecting a huge number of metabolites, can be used for such measurements [30,31].

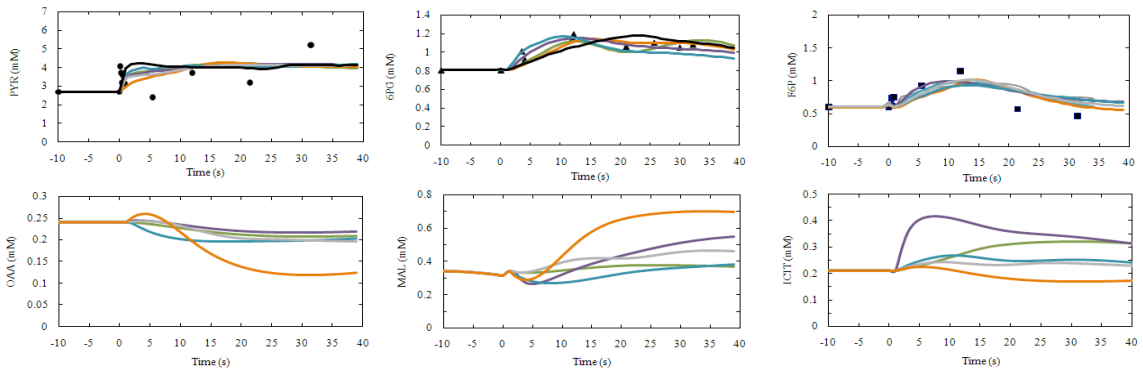

Figure 1: Comparison of simulated metabolite concentrations over time course produced by special variant of the linlog model (color lines) and the mechanistic model (black solid lines) for tree metabolites (PYR, pyruvate; 6PG, 6-phosphogluconate and F6P, fructose-6-phoshate) of the reconstructed *E. coli* dynamic metabolic network with kinetic parameters resulting from five different estimations sets. For some of the new metabolites OAA, oxaloacetate; MAL, malate and ICIT, isocitrate, only simulation results for the reconstructed model are shown. The experimental time series data are represented by symbols.

6. CONCLUSIONS:

In this work, we have described some limitations and inconsistencies detected during the reconstruction of the dynamic model of *E. coli* based on the literature and public databases, such as missing information on the experimental conditions in which the parameters had been determined. In addition, while a huge amount of standard kinetic data is not available, the set-up of large scale dynamic networks based in constraint based approach and linlog kinetics, as described in this work, appears to be a promising method to overcome the limitation of the high number of unknown kinetic parameters and rate laws.

Acknowledgments: Rafael S. Costa would like to thank Fundação para a Ciência e Tecnologia for providing the grant SFRH/BD/25506/2005 and the MIT-Portugal project "Bridging Systems and Synthetic Biology for the development of improved microbial cell factories" MIT-Pt/BS-BB/0082/2008.

REFERENCES:

1. Costa RS, Machado D, Rocha I, Ferreira EC: **Large Scale Dynamic Model Reconstruction for the Central Carbon Metabolism of *Escherichia coli* .** *Lecture Notes in Computer Science* 2009, **5518:** 1079-1083.
2. Ishii N, et.al.: **Multiple high-throughput analyses monitor the response of *E. coli* to perturbations.** *Science* 2007, **316:** 593-597.
3. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M: **Dynamic modeling of the central carbon metabolism of *Escherichia coli*.** *Biotechnology and Bioengineering* 2002, **79:** 53-73.
4. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO: **Reconstruction of biochemical networks in microorganisms.** *Nature Reviews Microbiology* 2009, **7:** 129-143.
5. Ingraham PJ, Stumpf MPH, Stark J: **Network motifs: structure does not determine function.** *BMC Genomics* 2006, **7:** 108-120.
6. Karp P, Riley M, Saier M, Paulsen I.T., Paley SM, Pellegrini-Toole A: **The EcoCyc and MetaCyc databases.** *Nucleic Acids Research* 2000, **28:** 56-59.
7. Maerkl SJ, Quake SR: **A systems approach to measuring the binding energy landscapes of transcription factors.** *Science* 2007, **315:** 233-237.
8. Schomburg I, Chang A, Schomburg D: **BRENDA, enzyme data and metabolic information.** *Nucleic Acids Research* 2002, **30:** 47-49.
9. Vaseghi S, Baumeister A, Rizzi M, Reuss M: **In vivo dynamics of the pentose phosphate pathway in *Saccharomyces cerevisae*.** *Metab Eng* 1999, **1:** 128-140.
10. Mendes P, Kell DB: **Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation.** *Bioinformatics* 1998, **14:** 869-883.
11. Ishii N, Suga Y, Hagiya A, Watanabe H, Mori H, Yoshino M *et al.*: **Dynamic simulation of an in vitro multi-enzyme system.** *Febs Letters* 2007, **581:** 413-420.
12. Jamshidi N, Edwards JS, Fahland T, Church GM, Palsson BO: **Dynamic simulation of the human red blood cell metabolic network.** *Bioinformatics* 2001, **17:** 286-287.
13. Shlomi.T., Berkman O, Ruppin E: **Regulatory on/off minimization of metabolic flux changes after genetic perturbations.** *Proc Natl Acad Sci U S A* 2005, **102:** 7695-7700.
14. Jamshidi N, Palsson BO: **Formulating genome-scale kinetic models in the post-genome era.** *Molecular Systems Biology* 2008, **4:** 171-180.
15. Smallbone K, Simeonidis E, Broomhead DS, Kell DB: **Something from nothing - bridging the gap between constraint-based and kinetic modelling.** *Febs Journal* 2007, **274:** 5576-5585.
16. Resendis-Antonio O: **Filling kinetics gaps: Dynamic Modeling of metabolism where detailed kinetic information is lacking.** *Plos ONE* 2009, **4:** 1-11.
17. Yugi K, Nakayama Y, Kinoshita A, Tomita M: **Hybrid dynamic/static method for large-scale simulation of metabolism.** *Theoretical Biology and Medical Modelling* 2005, **2:** 42-53.
18. Heijnen JJ: **Approximative kinetic formats used in metabolic network modeling.** *Biotechnology and Bioengineering* 2005, **91:** 534-545.
19. Nikerel IE, van Winden WA, van Gulik WM, Heijnen JJ: **A method for estimation of elasticities in metabolic networks using steady-state and dynamic metabolomics data and linlog kinetics.** *BMC Bioinformatics* 2006, **7:** 540-563.
20. Vilela M, Chou I-C, Vinga S, Vasconcelos ATR, Voit E, Almeida SJ: **Parameter estimation in S-systems models.** *Bmc Systems Biology* 2008, **2:** 35-48.
21. Ruijter GJG, Visser J: **Characterization of Aspergillus niger phosphoglucose isomerase. Use for quantitativa determination of erythrose 4-phosphate.** *Biochimi* 1999, **81:** 267-272.
22. Fifif T, Scopes RK: **Purification of 3-Phosphoglycerate from diverse sources by affinity elution chromatography.** *Biochemical Journal* 1978, **175:** 311-319.
23. Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H *et al.*: **BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.** *Nucleic Acids Research* 2006, **34:** D689-D691.
24. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N *et al.*: **COPASI — a COmplex PAthway SImulator.** *Bioinformatics* 2006, **22:** 3067-3074.
25. Moles CG, Mendes P, Banga JR: **Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods.** *Genome Research* 2003, **13:** 2467-2474.
26. Kauffman KJ, Prakash P, Edwards JS: **Advances in flux balance analysis.** *Current Opinion in Biotechnology* 2003, **14:** 491-496.
27. Reed JL, Vo TD, Schilling CH, Palsson BO: **An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR).** *Genome Biology* 2003, **4:** R54.
28. Visser D, Heijnen JJ: **Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics.** *Metabolic Engineering* 2003, **5:** 164-176.
29. Wiechert W: **13C Metabolic Flux Analysis.** *Metabolic Engineering* 2001, **3:** 195-206.
30. Kell DB: **Metabolomics and systems biology: making sense of the soup.** *Current Opinion in Microbiology* 2004, **7:** 296-307.
31. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T: **Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry.** *Journal of Proteome Research* 2003, **2:** 488-494.