Model transformation of metabolic networks using a Petri net based framework

Daniel Machado[1], Rafael S. Costa[1], Miguel Rocha[2], Isabel Rocha[1], Bruce Tidor[3], and Eugénio C. Ferreira[1]

[1] IBB-Institute for Biotechnology and Bioengineering/Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
{dmachado,rafacosta,irocha,ecferreira}@deb.uminho.pt
[2] Department of Informatics/CCTC, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
mrocha@di.uminho.pt
[3] Department of Biological Engineering/Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
tidor@mit.edu

Abstract. The different modeling approaches in Systems Biology create models with different levels of detail. The transformation techniques in Petri net theory can provide a solid framework for zooming between these different levels of abstraction and refinement. This work presents a Petri net based approach to Metabolic Engineering that implements model reduction methods to reduce the complexity of large-scale metabolic networks. These methods can be complemented with kinetics inference to build dynamic models with a smaller number of parameters. The central carbon metabolism model of *E. coli* is used as a test-case to illustrate the application of these concepts. Model transformation is a promising mechanism to facilitate pathway analysis and dynamic modeling at the genome-scale level.

1 Introduction

Systems Biology provides a new perspective in the study of living systems and embraces the complexity emerging of interactions among all biological components. Combining theory and experiments, scientists build models to explain and predict the behavior of the systems under study. Metabolic Engineering is one of the fields where this perspective has proven useful through the optimization of metabolic processes for industrial applications [28, 2].

Modeling in Systems Biology is an iterative process as the life-cycle of a model is comprised of successive refinements using experimental data. Different approaches, such as top-down, bottom-up or middle-out [18] are used depending on the purpose of the model and the type of data available for its construction. In Metabolic Engineering there are macroscopic kinetic models that consider the cell as a black-box converting substrates into biomass and products, which are typically used for bioprocess control. On the other hand, there are reaction-network-level models, either medium-scale dynamic models with detailed kinetic

Recent Advances in Petri Nets and Concurrency, S. Donatelli, J. Kleijn, R.J. Machado, J.M. Fernandes (eds.), CEUR Workshop Proceedings, ISSN 1613-0073, Jan/2012, pp. 103–117.

information derived from literature and experimental data [3], or genome-scale stoichiometric reconstructions derived from genome annotation complemented with literature review [5].

Although the ultimate goal of Systems Biology is a complete understanding of the cell as a whole, not only it is extremely difficult to collect all the kinetic information necessary to build a fully detailed whole-cell model due to the lack of experimental data and model identifiability concerns, but also the computational cost of simulating the dynamics of a system with such detail would be tremendous. Therefore, there is a need to fit the level of detail of a model to the specific problem at hand. For instance, Metabolic Pathway Analysis (MPA) has been useful in the analysis of metabolism as a way to determine, classify and optimize the possible pathways throughout a metabolic network. However, due to the combinatorial explosion of pathways with increasing number of reactions, it is still infeasible to apply these methods in genome-scale metabolic reconstructions without decomposing the network into connected modules [23, 24]. This zooming in and out between different levels of abstraction and connecting parts with different levels of detail is a feature where formal methods and particularly Petri nets may play an important role. Concepts such as subnetwork abstraction, transition refinement or node fusion, among others, have been explored in Petri net theory [8] and may provide the theoretical background for method development.

In previous work, we reviewed different modeling formalisms used in Systems Biology from a Metabolic Engineering perspective and concluded that Petri nets are a promising formalism for the creation of a common framework of methods for modeling, analysis and simulation of biological networks [15]. They are a mathematical and graphical formalism, therefore intuitive and amenable to analysis. The different extensions available (*e.g.:* stochastic, continuous, hybrid) provide the flexibility required to model and integrate the diversity of phenomena occurring in the main types of biological networks (metabolic, regulatory and signaling). Moreover, one may find biological meaning in several concepts in Petri net theory; for instance, the incidence matrix of a Petri net is the equivalent of the stoichiometric matrix, and the minimal *t-invariants* correspond to the elementary flux modes (EFMs).

In this work, we explore strategies of model reduction for Petri net representations of metabolic networks, and the integration of this methodology with recent approaches such as genome-scale dynamic modeling. This paper is organized as follows. Section 2 explores the motivation for the work. Section 3 presents the model reduction and kinetics inference methods, Section 4 discusses their application to *E. coli* and Section 5 elaborates on conclusions and future work.

2 Background

There are different examples of model reduction in the literature. One such method was developed in [17], based on timescale analysis for classification of metabolite turnover time using experimental data. The fast metabolites are

removed from the differential equations and their surrounding reactions are lumped. In [20] the EFMs of a reaction network are calculated in order to create a macroscopic pathway network, where each EFM is a macro-reaction connecting extracellular substrates and products. A simple Michaelis–Menten rate law is assumed for each macro-reaction and the parameters are inferred from experimental data. The method is applied in a network with 18 reactions and a total of 7 EFMs. However it does not scale well to larger networks because, in the worst case, the number of EFMs grows exponentially with the network size.

The combinatorial pathway explosion problem is well known; there are methods for network decomposition in the literature that address this issue. In [23] the authors perform a genome-scale pathway analysis on a network with 461 reactions. After estimating the number of extreme pathways (EPs) to be over a million, the network is decomposed into 6 subsystems according to biological criteria and the set of EPs is computed separately for each subsystem. A similar idea in [24] consists on automatic decomposition based on topological analysis. The metabolites with higher connectivity are considered as external and connect the formed subnetworks. An automatic decomposition approach based on Petri nets is the so-called *maximal common transition sets* (MCT-sets) [22], and consists on decomposing a network into modules by grouping reactions by participation in the minimal *t-invariants* (equivalent to EFMs). A related approach relies on clustering of *t-invariants* for network modularization [9]. A very recent network coarsening method based on so-called *abstract dependent transition sets* (ADT-sets) is formulated without the requirement of pre-computation of the *t-invariants* and thus may be a promising tool for larger networks [12].

Another problem in genome-scale metabolic modeling is the study of dynamic behavior. Genome-scale metabolic reconstructions are stoichiometric and usually analyzed under steady-state assumption using constraint-based methods [1]. Dynamic flux balance analysis (dFBA) allows variation of external metabolite concentrations, and simulates the network dynamics assuming an internal pseudo steady-state at each time step [16]. It is used in [19] to build a genome-scale dynamic model of *L. lactis* that simulates fermentation profiles. However, this approach gives no insight into intracellular dynamics, neither it integrates reaction kinetics. In [26] the authors build a kinetic genome-scale model of *S. cerevisiae* using linlog kinetics, where the reference steady-state is calculated using FBA. Some of the elasticity parameters and metabolite concentrations are derived from available kinetic models, while the majority use default values. Using the stoichiometric coefficients as elasticity values is a rough estimation of the influence of the metabolites on the reaction rates. Moreover, no time-course simulation is performed. Mass action stoichiometric simulation (MASS) models are introduced in [14] as a way to incorporate kinetics into stoichiometric reconstructions. Parameters are estimated from metabolomic data. Regulation can be included by incorporating the mechanistic metabolite/enzyme interactions. A limitation of these models is that mass-action kinetics do not reflect the usual non-linearity of enzymatic reactions and the incorporation of regulation leads to a significant increase in network size.

3 Methods

The idea of this work is closer to the reduction concepts of [17, 20] than the modularization concepts in [23, 24]. In the latter cases a large model is decomposed into subunits to ease its processing by analyzing the parts individually. Instead, our objective is to facilitate the visualization, analysis and simulation of a large-scale model as a whole by abstracting its components. This reduction is to be attained by reaction lumping in a way that maintains biological meaning and valid application of current analysis and simulation tools. The Michaelis–Menten kinetics is a typical example of abstraction, where the small network of mass-action reactions are lumped into one single reaction.

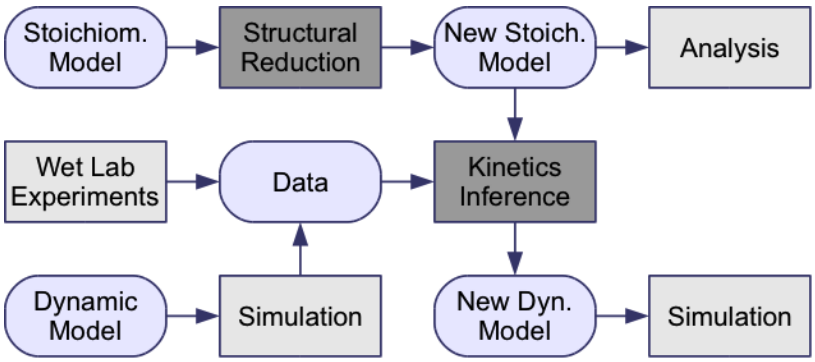

Fig. 1. Overall concept of model reduction and kinetics inference.

The overall idea of the model reduction method is depicted in Fig. 1. A large-scale stoichiometric model can be structurally reduced into a simplified version that can be more easily analyzed by methods such as MPA. Also, one may infer a kinetic structure to build a dynamic version of the reduced model. Due to the smaller size, a lower number of parameters has to be estimated. The data used for estimation may be experimental data found in the literature, or pseudo-experimental data from dynamic simulations if part of the system has been kinetically characterized.

When abstracting a reaction subnetwork into one or more macro-reactions, it is important to consider the assumptions created by such abstraction. As in Michaelis–Menten kinetics, these simplifications result in a pseudo-steady-state assumption for the intermediate species that disappear. While this may not be a problem for flux balance models, it changes the transient behavior of dynamic models because the buffering effect of intermediates in a pathway is neglected. The selection of metabolites to be removed depends on the purpose of the reduction. The network may have different levels of granularity based on the availability of experimental data, topological properties, or simply in order to aggregate pathways according to biological function.

3.1 Basic definitions

The proposed method for model reduction uses several Petri net concepts from the literature. We will use the following definition of an unmarked continuous Petri net (adapted from [4]) for modeling a stoichiometric metabolic network:

$$Pn = < P, T, Pre, Post >$$
$$Pre : P \times T \to \mathbb{R}^+$$
$$Post : P \times T \to \mathbb{R}^+$$

where the set of places P represents the metabolites, the set of transitions T represents the reactions and $Pre, Post$ are, respectively, the substrate and product stoichiometries of the reactions. Note that for the representation of a stoichiometric network, a discrete Petri net usually suffices; however, because some models may contain non-integer stoichiometric coefficients, the continuous version was adopted. Moreover, we will assume that reversible reactions are split into irreversible reaction pairs. We will also use the following definitions:

$$loc(x) = \{x\} \cup {}^\bullet x \cup x^\bullet$$
$$In(p) = \sum_{t \in {}^\bullet p} Post[p,t] \cdot v(t)$$
$$Out(p) = \sum_{t \in p^\bullet} Pre[p,t] \cdot v(t)$$

where ${}^\bullet x, x^\bullet$ are sets representing the input and output nodes of a node x, the set $loc(x) \subseteq P \cup T$ is called the locality of x, function $v : T \to \mathbb{R}_0^+$ is a given flux distribution (or the so-called instantaneous firing rate), and $In, Out : P \to \mathbb{R}_0^+$ are, respectively, the feeding and draining rates of the metabolites.

The method for network reduction consists of eliminating a set of selected metabolites from the network. For each removed metabolite its surrounding reactions are lumped in order to maintain the fluxes through the pathways. This reduction assumes a steady-state condition for the metabolite, *i.e.* $In(p) = Out(p)$.

3.2 Model reduction: Conjunctive fusion

There are two options for lumping the reactions depending on the transformation method applied. The first approach is based on a transformation called *conjunctive transition fusion* [8] and it results in an abstraction that replaces the transition-bordered subnet $loc(p)$ by a single macro-reaction. The drawback of this method is that the flux ratios between the internal reactions are lost. If a known steady-state flux distribution (v) is given, then the stoichiometric coefficients can be adjusted to preserve the ratios for that distribution; however, the space of solutions of the flux balance formulation becomes restricted to a particular solution. In the limiting case, if all the internal metabolites are removed, the cell is represented by one single macro-reaction connecting extracellular substrates and products with the stoichiometric yields inferred from the

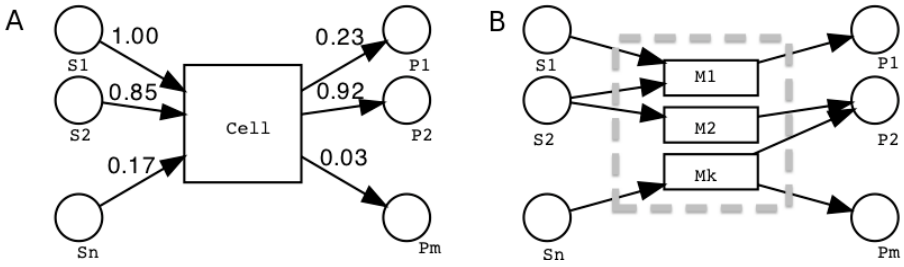

Fig. 2. Exemplification of limit scenarios where all the internal metabolites are removed. (A) In the conjunctive reduction case the result is one single macro-reaction converting substrates into products with the respective yields specified in the stoichiometry. (B) In the disjunctive reduction method, all possible pathways connecting substrates and products are enumerated.

network topology for one particular steady-state (Fig 2A). The transformation method for removing metabolite p in Pn given a flux distribution v is described as follows:

$$
\begin{aligned}
Pn' &= < P', T', Pre', Post' > \\
P' &= P \setminus \{p\} \\
T' &= T \setminus ({}^\bullet p \cup p^\bullet) \cup \{t_p\} \\
Pre' &= \{(p_i, t_j) \mapsto Pre(p_i, t_j) \mid (p_i, t_j) \in dom(Pre) \setminus (P \times ({}^\bullet p \cup p^\bullet))\} \\
&\quad \cup \{(p_i, t_p) \mapsto f_{in}(p_i) \mid p_i \in {}^\bullet({}^\bullet p \cup p^\bullet), p_i \neq p, v'(t_p) \neq 0, f_{in}(p_i) \neq 0\} \\
Post' &= \{(p_i, t_j) \mapsto Post(p_i, t_j) \mid (p_i, t_j) \in dom(Post) \setminus (P \times ({}^\bullet p \cup p^\bullet))\} \\
&\quad \cup \{(p_i, t_p) \mapsto f_{out}(p_i) \mid p_i \in ({}^\bullet p \cup p^\bullet)^\bullet, p_i \neq p, v'(t_p) \neq 0, f_{out}(p_i) \neq 0\} \\
v' &= \{t \mapsto v(t) \mid t \in T \setminus ({}^\bullet p \cup p^\bullet)\} \cup \{t_p \mapsto In(p)\}.
\end{aligned}
$$

where

$$
\begin{aligned}
f_{in}(p_i) &= \frac{\sum_{t \in p_i^\bullet \cap ({}^\bullet p \cup p^\bullet)} Pre(p_i, t) \cdot v(t)}{v'(t_p)} \\
f_{out}(p_i) &= \frac{\sum_{t \in {}^\bullet p_i \cap ({}^\bullet p \cup p^\bullet)} Post(p_i, t) \cdot v(t)}{v'(t_p)}
\end{aligned}
$$

The stoichiometric coefficients of the new reaction may be very high or low, depending on $v'(t_p)$ and so, optionally, one may also normalize them with some scalar λ, such that $Pre''(p_i, t_p) = \frac{1}{\lambda} \cdot Pre'(p_i, t_p)$, $Post''(p_i, t_p) = \frac{1}{\lambda} \cdot Post'(p_i, t_p)$ and $v''(t_p) = \lambda \cdot v'(t_p)$. This will also make the final result independent of the order of the metabolites removed. A good choice for λ is:

$$
\lambda = \max\left(\{Pre(p_i, t_p) \mid p_i \in {}^\bullet t_p\} \cup \{Post(p_i, t_p) \mid p_i \in t_p{}^\bullet\}\right)
$$

3.3 Model reduction: Disjunctive fusion

The second approach is based on a transformation called *disjunctive transition fusion* [8], where every combination of input and output reaction pairs connected by the removed metabolite is replaced by one macro-reaction. Although this approach does not constrain the steady-state solution space of the flux distribution, it has the drawback of increasing the number of transitions, if the metabolite is highly connected, due to the combinatorial procedure. Note that applying this reduction step to metabolite p_i is equivalent to performing one iteration of the *t-invariant* calculation algorithm to remove column i of the transposed incidence matrix. Therefore, in the limiting case where all internal metabolites are removed, the cell is represented by the set of all possible pathways connecting extracellular substrates and products (Fig. 2B), as was done in [20]. The definition, similar to the previous one, is as follows:

$$
\begin{aligned}
Pn' &= < P', T', Pre', Post' > \\
P' &= P \setminus \{p\} \\
T' &= T \setminus ({}^\bullet p \cup p^\bullet) \cup \{t_{xy} \mid (x,y) \in ({}^\bullet p \times p^\bullet)\} \\
Pre' &= \{(p_i,t) \mapsto Pre(p_i,t) \mid (p_i,t) \in dom(Pre) \setminus (P \times ({}^\bullet p \cup p^\bullet)\} \\
&\quad \cup \{(p_i,t_{xy}) \mapsto Pre_0(p_i,x) \cdot Pre(p,y) + Pre_0(p_i,y) \cdot Post(p,x) \\
&\quad \mid (x,y) \in ({}^\bullet p \times p^\bullet), p_i \in {}^\bullet\{x,y\}\} \\
Post' &= \{(p_i,t) \mapsto Post(p_i,t) \mid (p_i,t) \in dom(Post) \setminus (P \times ({}^\bullet p \cup p^\bullet)\} \\
&\quad \cup \{(p_i,t_{xy}) \mapsto Post_0(p_i,x) \cdot Pre(p,y) + Post_0(p_i,y) \cdot Post(p,x) \\
&\quad \mid (x,y) \in ({}^\bullet p \times p^\bullet), p_i \in \{x,y\}^\bullet\}
\end{aligned}
$$

where

$$
Pre_0(p,t) = \begin{cases} Pre(p,t) & \text{if } (p,t) \in dom(Pre) \\ 0 & \text{if } (p,t) \notin dom(Pre) \end{cases}
$$

$$
Post_0(p,t) = \begin{cases} Post(p,t) & \text{if } (p,t) \in dom(Post) \\ 0 & \text{if } (p,t) \notin dom(Post) \end{cases}
$$

Whenever there are pathways with the same net stoichiometry, these can be removed by checking the columns of the incidence (stoichiometric) matrix and eliminating repeats. It should also be noted that in both methods, if a metabolite acts both as substrate and product in a lumped reaction, it will create a redundant cycle that is not reflected in the incidence matrix. If these cycles are not removed, they propagate through the reduction steps; therefore, they should be replaced by a single arc containing the overall stoichiometry. The procedure

works as follows:

$$\begin{aligned} Pre' =& \{(p,t) \mapsto Pre(p,t) \mid (p,t) \in dom(Pre) \setminus dom(Post)\} \\ & \cup \{(p,t) \mapsto Pre(p,t) - Post(p,t) \\ & \quad \mid (p,t) \in dom(Pre) \cap dom(Post), Pre(p,t) > Post(p,t)\} \\ Post' =& \{(p,t) \mapsto Post(p,t) \mid (p,t) \in dom(Post) \setminus dom(Pre)\} \\ & \cup \{(p,t) \mapsto Post(p,t) - Pre(p,t) \\ & \quad \mid (p,t) \in dom(Pre) \cap dom(Post), Post(p,t) > Pre(p,t)\} \end{aligned}$$

The previous arc removing procedure may cause isolation of some nodes when $Pre(p,t) = Post(p,t)$; therefore, the isolated nodes should be removed:

$$\begin{aligned} P' &= \{p \mid p \in P, loc(p) \neq \{p\}\} \\ T' &= \{t \mid t \in T, loc(t) \neq \{t\}\} \end{aligned}$$

3.4 Kinetics inference

Given a stoichiometric model, if metabolomic or fluxomic data are available for parameter estimation, one may try to build a dynamic model by inferring appropriate kinetics for the reactions. In [25] the authors propose that this is performed by assuming linlog kinetics for all reactions using an FBA solution as the reference state and the stoichiometries as elasticity parameters. An integration of Biochemical Systems Theory (BST) with Hybrid Functional Petri Nets (HFPN) is presented in [29], where general mass action (GMA) kinetics is assumed for each transition. The review of kinetic rate formulations is out of the scope of this work and may be found elsewhere [10].

Assuming that all metabolites with unknown concentration were removed, we will extend our definition to a marked continuous Petri net:

$$Pn =< P, T, Pre, Post, m_0 >$$

where $m_0 : P \to \mathbb{R}_0^+$ denotes the initial marking (concentration) of the metabolites. The kinetics inference process consists on defining a firing rate $v : T \to \mathbb{R}_0^+$, which will be dependent on the current marking (m) and the specific kinetic parameters (see [7] for an introduction on marking-dependent firing rates). As we assumed irreversible reactions, each rate will only vary with substrate concentration. The rates can be easily derived from the net topology. In case of GMA kinetics v is given by:

$$v(t) = k_t \prod_{p \in {}^\bullet t} m(p)^{a_{p,t}}$$

where k_t is the kinetic rate of t and $a_{p,t}$ is the kinetic order of metabolite p in reaction t. A usual first approximation for $a_{p,t}$ is $Pre(p,t)$.

Linlog kinetics are formulated based on a reference rate v_0, and defined by:

$$v(t) = v_0(t) \left(1 + \sum_{p \in {}^\bullet t} \varepsilon^0_{p,t} \ln\left(\frac{m(p)}{m_0(p)}\right)\right)$$

where $\varepsilon^0_{p,t}$ is called the elasticity of metabolite p in reaction t, reflecting the influence of the concentration change of the metabolite in the reference reaction rate. As in the previous case, $Pre(p,t)$ can be chosen as an initial approximation for $\varepsilon^0_{p,t}$. The relative enzyme activity term (e/e_0) commonly present in linlog rate laws to account for regulatory effects at larger time scales will not be considered.

4 Results and Discussion

The proposed methods were tested using the dynamic central carbon metabolism model of *E. coli* [3], where the stoichiometric part was used for the application of the reduction methods, and the dynamic profile was used to generate pseudo-experimental data sets for parameter estimation and validation of the kinetics inference method. A Petri net representation of this model (Fig. 3) was built using the Snoopy tool [21]. All reversible reactions were split into irreversible pairs. The net contains a total of 18 places, 44 transitions and is covered by 95 *semipositive t-invariants*, computed with the Integrated Net Analyzer [27].

In the application of the conjunctive method (Fig 4A), the metabolites were classified as in [17] based on their timescale (Table 1), by calculating their turnover time ($\tau : P \to \mathbb{R}^+_0$) using the reference steady-state of the dynamic model, where:

$$\tau(p) = \frac{m_0(p)}{In(p)}$$

Metabolites with small turnover time are considered fast. In this case, all metabolites except the slowest 5 (*glcex, pep, g6p, pyr, g1p*) were removed.

For the application of the disjunctive method (Fig 4B), the metabolites were classified based on their topology (Table 1). We conveniently opted to remove the metabolites with lower connectivity to avoid the combinatorial explosion problem. All metabolites except 5 (*g6p, pyr, f6p, gap, xyl5p*) were removed. This reduction assumes steady-state for the removed metabolites. However, it makes no assumptions on the ratios between the fluxes, therefore preserving the flux-balance solution space.

Because we are assuming that the reference steady-state is known, the conjunctive reduced model was chosen for the application of the kinetics inference method assuming linlog kinetics at the reference state. The elasticity parameters were estimated using COPASI [13]. The pseudo-experimental data was generated from simulation with the original model after a 1 mM extracellular glucose pulse with the addition of Gaussian noise ($std = 0.05$ mM) (Fig. 5A). The fitted model was then validated using pseudo-experimental data from a 2 mM pulse (Fig. 5B). It is possible to observe an instantaneous increase in *pyr* (from 2.67 to 3.93) and an instantaneous decrease *pep* (from 2.69 to 1.26) which the model is unable to reproduce. The poor fitting in some of the intracellular metabolites is expected given the significant reduction to the model. However, the extracellular glucose consumption profile is remarkably good, both in the fitting and validation cases.

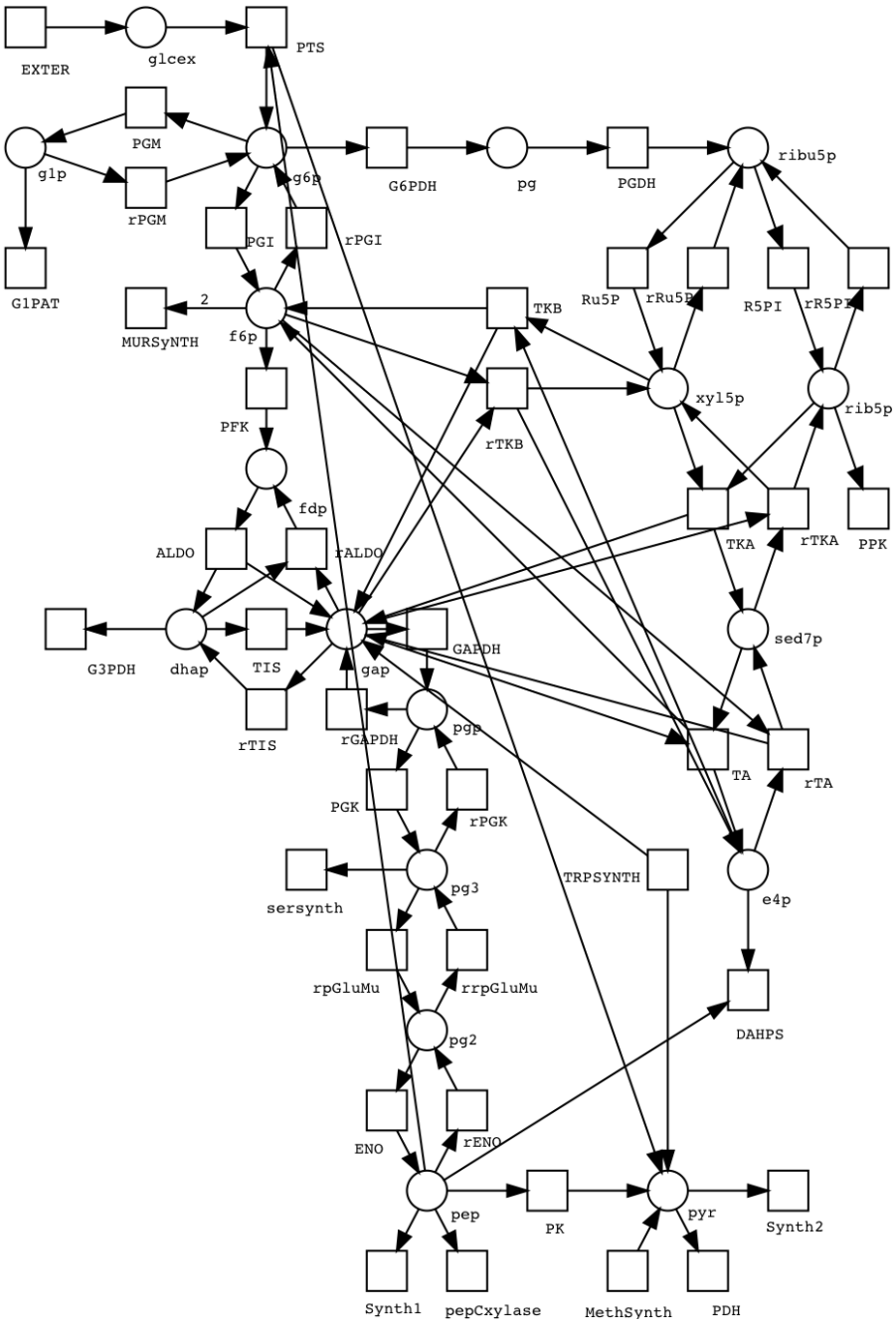

Fig. 3. Petri net model of the dynamic central carbon metabolism model of *E. coli* with reversible reactions split into irreversible pairs.

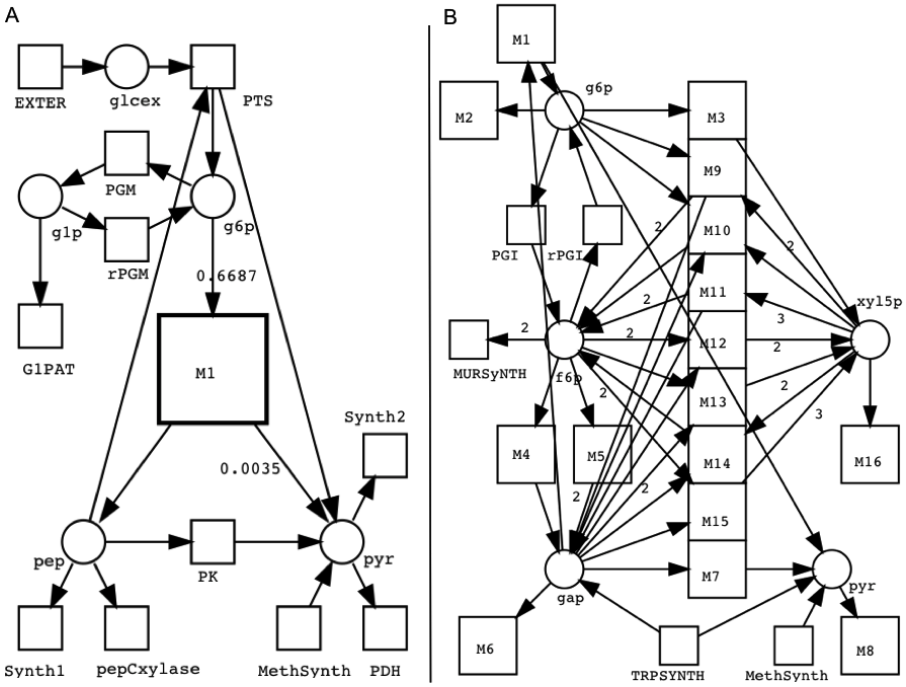

Fig. 4. Reduced versions of the original network. (A) Conjunctive reduction method. (B) Disjunctive reduction method.

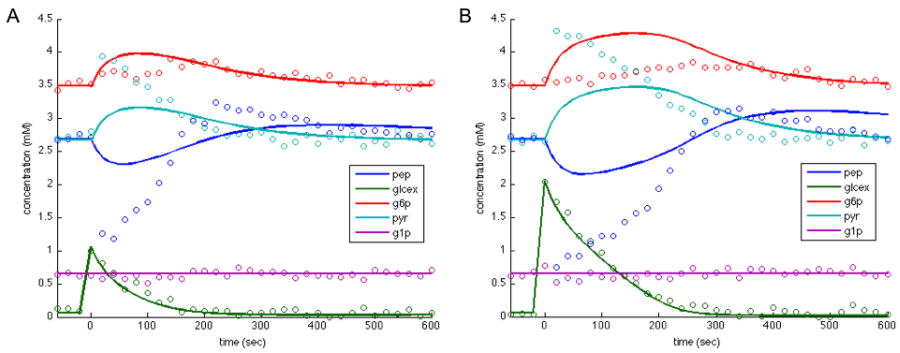

Fig. 5. (A) Results of parameter estimation with pseudo-experimental data with 1 mM extracellular glucose pulse. (B) Validation of the model with a 2 mM extracellular glucose pulse. In both cases, the circles represent the experimental data and the lines represent time-course simulations generated by the reduced model.

Table 1. Metabolite topological properties (input reactions, output reactions, connectivity) and dynamic properties (concentration, flux, turnover time) at the reference steady-state.

Metabolite	$\#({}^\bullet p)$	$\#(p^\bullet)$	$\#({}^\bullet p \times p^\bullet)$	m_0 (mM)	In (mM/s)	τ (s)
glcex	1	1	1	0.0558	0.0031	**18.099**
pep	1	6	6	2.6859	0.3031	**8.8603**
g6p	3	3	**9**	3.4882	0.2004	**17.406**
pyr	4	2	**8**	2.6710	0.2418	**11.044**
f6p	3	5	**15**	0.6014	0.1423	4.2266
g1p	1	2	2	0.6539	0.0023	**278.62**
pg	1	1	1	0.8092	0.1397	5.7929
fdp	2	1	2	0.2757	0.1414	1.9495
sed7p	2	2	4	0.2761	0.0454	6.0757
gap	7	6	**42**	0.2196	0.3661	0.5997
e4p	2	3	6	0.0986	0.0454	2.1684
xyl5p	3	3	**9**	0.1385	0.0839	1.6503
rib5p	2	3	6	0.3994	0.0558	7.1626
dhap	2	3	6	0.1682	0.1414	1.1892
pgp	2	2	4	0.0080	0.3207	0.0251
pg3	2	3	6	2.1437	0.3207	6.6851
pg2	2	2	4	0.4014	0.3031	1.3241
ribu5p	3	2	6	0.1114	0.1397	0.7974

Although both reducing methods can be combined with kinetics inference, the conjunctive version seems more suitable if a steady-state distribution is known, because it generates smaller models, hence less parameters. The disjunctive version is appropriate for analyzing all elementary pathways between a set of metabolites without the burden of calculating the set of EFMs of the whole model. For instance, the macro-reactions *M4* ($ALDO + G3PDH$) and *M5* ($ALDO + TIS$), with net stoichiometries of, respectively, [*fdp* $\rightarrow$ *gap*] and [*fdp* $\rightarrow$ 2 *gap*], are two unique pathways between these two metabolites.

5 Conclusions

This work presents strategies for model reduction of metabolic networks based on a Petri net framework. Two approaches, conjunctive and disjunctive reduction are presented. The conjunctive approach allows the abstraction of a subnetwork into one lumped macro-reaction, however limited to one particular flux distribution of the subnetwork. The disjunctive approach on the other hand, makes no assumptions on the flux distribution by replacing the removed subnetwork with macro-reactions for all possible pathways through the subnetwork, therefore not constraining the steady-state solution space. In both cases, the reduced model may be transformed into a dynamic model using kinetics inference and parameter estimation if experimental data is available. Using the reduced model,

instead of the original, facilitates this process because it significantly decreases the number of parameters to be estimated.

In future work, we intend to build a dynamic genome-scale model of *E. coli* by using the already available central carbon dynamic model [3], complemented with lumped versions of the surrounding pathways in the genome-scale network [5]. Note that some of the reactions on the central carbon model already represent lumped versions of some biosynthetic pathways (*e.g. mursynth, trpsynth, methsynth, sersynth*). However they were not deduced from the genome-scale network and may not be accurate abstractions of these pathways.

Among the extensions available to Petri nets are the addition of different types of arcs, such as read-arcs and inhibitor-arcs, which could be use to represent activation and inhibition in biochemical reactions. They could also be used to integrate metabolic and regulatory networks. Optimization in metabolic processes is usually based on knockout simulations in metabolic networks. However, these simulations do not take into consideration the possible regulatory effects caused by the knockouts. In our transformation methods we removed the arcs with the same stoichiometry in both directions, because these are not reflected in the stoichiometric matrix. In the Michaelis–Menten example this results in removing the enzyme from the network. The proposed methods can be extended to consider read-arcs for these situations, which should be preserved during the reduction steps, therefore establishing connection places to the integration of a regulatory network (Fig 6).

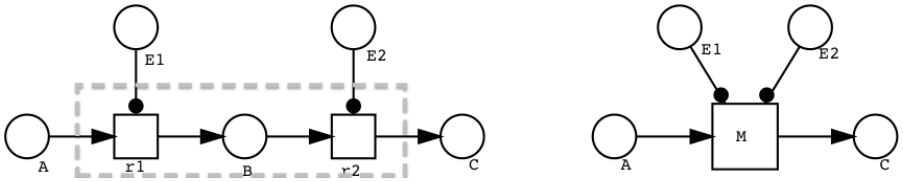

Fig. 6. Reduction step conserving the read-arcs associated with the enzymes of the original reactions.

An alternative to the reduction of the models would be to consider their representation using hierarchical Petri nets. In this case, each macro-reaction would be connected to its detailed subnetwork. Although this would not reduce the number of kinetic parameters in the case of kinetics inference, it would be extremely useful for facilitated modeling and visualization of large-scale networks without compromising detail. It could also be the solution for genome-scale pathway analysis, if it is performed independently at each hierarchical level. The hierarchical model composition proposed for SBML [6] may facilitate the implementation of this alternative. See [11] for an automatic network coarsening algorithm based on hierarchical petri nets applied to different kinds of biological networks.

Acknowledgments. Research supported by PhD grants SFRH/BD/35215/2007 and SFRH/BD/25506/2005 from the Fundação para a Ciência e a Tecnologia (FCT) and the MIT–Portugal Program through the project "Bridging Systems and Synthetic Biology for the development of improved microbial cell factories" (MIT-Pt/BS-BB/0082/2008).

References

1. S.A. Becker, A.M. Feist, M.L. Mo, G. Hannum, B.Ø. Palsson, and M.J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols*, 2(3):727–738, 2007.
2. A.P. Burgard, P. Pharkya, and C.D. Maranas. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.
3. C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.
4. R. David and H. Alla. *Discrete, continuous, and hybrid Petri nets*. Springer Verlag, 2005.
5. A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3(1), 2007.
6. A. Finney. Developing SBML beyond level 2: proposals for development. In *Computational Methods in Systems Biology*, pages 242–247. Springer, 2005.
7. D. Gilbert and M. Heiner. From Petri nets to differential equations-an integrative approach for biochemical network analysis. *Petri Nets and Other Models of Concurrency-ICATPN 2006*, pages 181–200, 2006.
8. C. Girault and R. Valk. *Petri Nets for System Engineering: A Guide to Modeling, Verification, and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.
9. E. Grafahrend-Belau, F. Schreiber, M. Heiner, A. Sackmann, B.H. Junker, S. Grunwald, A. Speer, K. Winder, and I. Koch. Modularization of biochemical networks based on classification of Petri net t-invariants. *BMC bioinformatics*, 9(1):90, 2008.
10. J.J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and bioengineering*, 91(5):534–545, 2005.
11. M. Heiner. Understanding Network Behavior by Structured Representations of Transition Invariants. *Algorithmic Bioprocesses*, page 367, 2009.
12. M. Heiner and K. Sriram. Structural analysis to determine the core of hypoxia response network. *PloS one*, 5(1):e8600, 2010.
13. S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI—a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.
14. N. Jamshidi and B.Ø. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.
15. D. Machado, R.S. Costa, M. Rocha, I. Rocha, B. Tidor, and E.C. Ferreira. A critical review on modelling formalisms and simulation tools in computational biosystems. In *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, pages 1063–1070. Springer, 2009.

16. R. Mahadevan, J.S. Edwards, and F.J. Doyle. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical journal*, 83(3):1331–1340, 2002.
17. I.E. Nikerel, W.A. van Winden, P.J.T. Verheijen, and J.J. Heijnen. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metabolic Engineering*, 11(1):20–30, 2009.
18. D. Noble. The rise of computational biology. *Nature Reviews Molecular Cell Biology*, 3(6):459–463, 2002.
19. G.M. Oddone, D.A. Mills, and D.E. Block. A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. *Metabolic Engineering*, 2009.
20. A. Provost and G. Bastin. Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control*, 14(7):717–728, 2004.
21. C. Rohr, W. Marwan, and M. Heiner. Snoopy-a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics*, 2010.
22. A. Sackmann, M. Heiner, and I. Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC bioinformatics*, 7(1):482, 2006.
23. C.H. Schilling and B.Ø. Palsson. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. *Journal of Theoretical Biology*, 203(3):249–283, 2000.
24. S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae, 2002.
25. K. Smallbone, E. Simeonidis, D.S. Broomhead, and D.B. Kell. Something from nothing-bridging the gap between constraint-based and kinetic modelling. *FEBS Journal*, 274(21):5576–5585, 2007.
26. K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.
27. P.H. Starke. INA: Integrated Net Analyzer. *Reference Manual*, 1992.
28. G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58, 1998.
29. J. Wu and E. Voit. Hybrid modeling in biochemical systems theory by means of functional Petri nets. *Journal of bioinformatics and computational biology*, 7(1):107, 2009.