Pruning, pooling and limiting steps in metabolic networks

Ovidiu Radulescu[1,2], Alexander N. Gorban[3], Andrei Zinovyev[4]

[1] DIMNP - UMR 5235 CNRS/UM1/UM2, Pl. E. Bataillon, Univ. of Montpellier 2, Bat. 24 CP 107, 34095 Montpellier Cedex 5, France
[2] Team Symbiose, INRIA-IRISA, Campus de Beaulieu, 35042 Rennes, France
[3] University of Leicester, LE1 7RH Leicester, UK
[4] Institut Curie, U900 INSERM/Curie/Mines ParisTech, 26 rue d'Ulm, F75248 Paris, France

Abstract

Dynamics of metabolic systems can be modelled by systems of differential equations. Realistic models of metabolism allowing to integrate genome scale data should have very large size and thus face problems related to incompleteness of the information on their structure and parameters. We discuss how model reduction techniques that use qualitative information on the order of magnitude of parameters can be applied to simplify large models of differential equations.

1 Introduction

In spite of steady advance in the omic sciences (metabolomics, transcriptomics, proteomics), modelling of large biochemical networks, based on standard mathematical approaches, faces obstacles such as incompleteness of network description (structural and parametric) and lack of exact knowledge of kinetic parameters. In the particular case of the modelling of metabolic pathways, although genome scale reaction models are available for certain well studied organisms[2], complete and reliable information on the kinetic parameters of enzymatic reactions (Vm, Km) are not available for such very large models.

Constraint based approaches (such as flux balance analysis FBA [11]) circumvent these obstacles by using optimality principles and replacing the network by a set of stoichiometric constraints. FBA is well suited for global studies of perturbations of metabolism. Thermodynamics imposes constraints that can be dealt with within the same approach. FBA has been successfully applied to the global study of the metabolism of various organisms to identify the effects of gene knock-outs in various media, as well as for defining the concept of minimal supporting growth media. In spite of these successes FBA has two major drawbacks. It can not deal with time dynamics. Moreover it can not predict concentrations of metabolites (the predicted variables are the fluxes), that is a major defect when dealing with metabolomic studies. Dynamic effects are particularly important in parmacokinetik when, depending on the dose time-scenario, the application of a drug could trigger or not compensatory mechanisms.

Dynamic modelling using differential equations was successfully used to study small or medium models (usually from ten to several tens of variables) for which complete sets of kinetic parameters can be measured or reverse-engineered. In order to study much larger models we could use model reduction techniques, which is a different way to tame complexity.

Model reduction is now common practice in combustion modelling where systems of thousands or tens of thousands of chemical reactions are reduced to much simpler sets of equations (see methods such as CSP, ILDM, invariant manifold [14, 12, 6, 5, 7, 1]). The feasibility of the reduction is guaranteed by generic properties of dissipative systems, that after a quick transition converge to a dynamics with a few degrees of freedom (invariant manifold [6, 5, 7, 1]). Applied to metabolic pathways such methods give a reduced description of dynamics in terms of synthetic variables that most of the time are difficult to interpret. This could be enough for numerical purposes such as stiffness elimination, but is not always appropriate for metabolic modelling. In this case we are looking for reduced variables that are easy to interpret and we need reduction methods that can cope with incomplete kinetic information. A recent extension of limitation theory [9, 15, 10] satisfies both these requirements. It can deal with both precise (exact values of constants) and qualitative (order of constants, such as much slower or much quicker) information. Also, the reduced models can be deduced from the initial model by simple constructions such as pruning and pooling.

2 Differential equations models

Continuous dynamical models of metabolism are conveniently represented as systems of differential equations. In such models, the state of the system is a vector $\boldsymbol{x} \in \mathbb{R}^n$ containing the concentrations of all metabolites. Each reaction (elementary step) in the model (indexed by an integer $i \in 1, \ldots, r$) is described by a stoichiometric vector $\boldsymbol{\nu}^i = \boldsymbol{\beta}^i - \boldsymbol{\alpha}^i$ ($\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ corresponds to the stoichiometries of products and reactants) and a rate R_i.

The rates (or fluxes) R_i, which are expressed in units of transformed mass per unit time, are functions of the concentrations. For reactions with no intermediate steps, occurring by random collisions of molecules, one can obtain the mass action law:

$$R(\boldsymbol{x}) = k_+ \prod_j x_j^{\alpha_j} - k_- \prod_j x_j^{\beta_j} \tag{1}$$

The set of reactions representing the step by step transformations of the metabolites is also called reaction mechanism or reaction network. The dynamics of a reaction network is given by the following system of differential equations:

$$\frac{d\boldsymbol{x}}{dt} = \sum_{i=1}^{r} R_i(\boldsymbol{x})\boldsymbol{\nu}^i \tag{2}$$

For instance, consider the Michaelis-Menten mechanism for enzymatic reactions $S + E \rightleftharpoons ES \longrightarrow E + P$.

For this model we have $\boldsymbol{x} = \begin{pmatrix} [S] \\ [E] \\ [ES] \\ [P] \end{pmatrix}$, $\boldsymbol{\nu}_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$, $R_1 = k_1^+[S][E] - k_1^-[ES]$, $\boldsymbol{\nu}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$, $R_2 = k_2[ES]$. The system of differential equations reads:

$$\frac{d[S]}{dt} = -k_1^+[S][E] + k_1^-[ES] \tag{3}$$

$$\frac{d[E]}{dt} = -k_1^+[S][E] + k_1^-[ES] + k_2[ES] \tag{4}$$

$$\frac{d[ES]}{dt} = k_1^+[S][E] - k_1^-[ES] - k_2ES \tag{5}$$

$$\frac{d[P]}{dt} = k_2[ES] \tag{6}$$

3 Traditional rate limiting step theory

In the IUPAC Compendium of Chemical Terminology (2007) one can find the following definition of limiting steps: "A rate-controlling (rate-determining or rate-limiting) step in a reaction occurring by a composite reaction sequence is an elementary reaction the rate constant for which exerts a strong effect - stronger than that of any other rate constant - on the overall rate." Hans Krebs coined the term "pacemaker" for rate-limiting enzymes, that could play important role in targeting metabolism with drugs.

Although it is obvious from this definition that a rate-limiting step does not always exist (among the control functions generically there is a biggest one, but this is not necessarily much bigger than all the others), biochemists tend to believe that each metabolic pathway has a unique limiting step even if most often do not agree on which one this is. On the other extreme, metabolic control theorists suggest that pathways rates depend to various degrees on rate constants of all the reactions, and thus limitation theory has limited utility. This theoretical prediction (relying on summation theorems for control coefficients [3]) seem to be verified experimentally. As cited by David Fell, a 3.5-fold increase of the amount of limiting enzyme phosphofructokinase in yeast have no significant effect on the anaerobic glycolytic flux [3]. True pacemaker enzymes allowing flux re-distributions are difficult to find.

However, for the notion of limiting step that is used in practice, there are important dynamical differences between systems without limiting step and

systems with limiting step. The behavior of the later in terms of dynamics of intermediates and distribution of fluxes can be understood even if kinetic information is only partially quantitative. Finally, metabolic control and limitation theory can be unified in a common methodology. Limitation based model reduction can provide simpler models whose control coefficients can be more easily studied. Removing dominated (inessential) reactions allows to solve the problem of "sloppy sensitivities" identified in the context of gene networks, but also valid for regulation of metabolism.

4 Reducing linear networks with separated constants: pruning, glueing, and restoring

4.1 Linear networks

Linear reaction mechanisms include monomolecular networks or more generally first order networks.

The structure of monomolecular reaction networks can be completely defined by a simple digraph, in which vertices correspond to chemical species A_i, edges correspond to reactions $A_i \to A_j$ with rate constants $k_{ji} > 0$. In this case, the stoichiometric vector for the reaction (i, j) has -1, 1 in positions i, j, respectively and zeros elsewhere. The rate function is proportional to the concentration of the substrate $R_{ji}(\boldsymbol{x}) = k_{ji}x_i$.

The system of kinetic differential equations is

$$\frac{dx_i}{dt} = \sum_j k_{ij}x_j - (\sum_j k_{ji})x_i, \text{ or in matrix form } \frac{d\boldsymbol{x}}{dt} = \boldsymbol{K}\boldsymbol{x} \qquad (7)$$

where $\boldsymbol{K}$ is the kinetic matrix.

Monomolecular mechanisms are conservative, ie the total number of molecules is a constant of the dynamics $\sum_i x_i = const.$. This means that if a single molecule is transformed by the mechanism, at each time there will be a single molecule somewhere in the mechanism.

The methods that we discuss here can be applied more generally to pseudo-conservative first order mechanisms.

First order reaction networks can contain reactions that are not monomolecular, such as $A \to A + B$, or $A \to B + C$. There is always a unique substrate and the rates proportional to the concentration of the substrate. These mechanisms are not conservative because they allow overall molecule production. Pseudo-conservative first order mechanism conserve total number of molecules of some species of interest that we call internal but can consume or produce external species. In pseudo-conservative first order mechanism $A \to A + B$ reactions are allowed, provided that B is external; similarly $A \to B + C$ reactions are allowed, provided that either B or C is external. Degradation reactions can be studied in this framework by considering a special component (sink), that collects degraded molecules.

Metabolic networks (or subnetworks) are rarely monomolecular or first order. However, when all substrates and cofactors are in excess, except for one, metabolic reactions can be also considered to be first order because in this case the rate is proportional to the concentration of the substrate that is not in excess. Though the general applicability of this method should not be taken for granted, linear formalism can provide new insights into metabolic network design.

4.2 Reduction algorithm for monomolecular networks

In [9, 15] we propose an algorithm to simplify monomolecular networks with total separation of the rate constants. Total separation of the constants means that either $k_I << k'_I$ or $k'_I << k_I$ for all $I = ij$, $I' = i'j'$. The algorithm, justified by estimates for the eigenvalues and eigenvectors (inspired, but not fully covered by Gershgorin theorem) of the kinetic matrix [8], consists of three stages:

I. Constructing of an auxiliary reaction network: pruning.

For each A_i branching node (substrate of several reactions) let us define κ_i as the maximal kinetic constant for reactions $A_i \to A_j$: $\kappa_i = \max_j\{k_{ji}\}$. For correspondent j we use the notation $\phi(i)$: $\phi(i) = \arg\max_j\{k_{ji}\}$.

An auxiliary reaction network $\mathcal{V}$ is the set of reactions obtained by keeping only $A_i \to A_{\phi(i)}$ with kinetic constants κ_i and discarding the other, slower reactions. Auxiliary networks have no branching, but they can have cycles and confluences. The correspondent kinetic equation is

$$\dot{c}_i = -\kappa_i c_i + \sum_{\phi(j)=i} \kappa_j c_j, \tag{8}$$

If the auxiliary network contains no cycles, the algorithm stops here.

II gluing cycles and restoring cycle exit reactions

In general, the auxiliary network $\mathcal{V}$ has several cycles $C_1, C_2, ...$ with periods $\tau_1, \tau_2, ... > 1$.

These cycles will be "glued" into points and all nodes in the cycle C_i, will be replaced by a single vertex A^i. Also, some of the reactions that were pruned in the first part of the algorithm are restored with renormalized rate constants. Indeed, reaction exiting a cycle are needed to render the correct dynamics: without them, the total mass accumulates in the cycle, with them the mass can also slowly leave the cycle. Reactions $A \to B$ exiting from cycles ($A \in C_i$, $B \notin C_i$) are changed into $A^i \to B$ with the rate constant renormalization: let the cycle C^i be the following sequence of reactions $A_1 \to A_2 \to ...A_{\tau_i} \to A_1$, and the reaction rate constant for $A_i \to A_{i+1}$ is k_i (k_{τ_i} for $A_{\tau_i} \to A_1$). For the limiting (slowest) reaction of the cycle C_i we use notation $k_{\lim\, i}$. If $A = A_j$ and k is the rate reaction for $A \to B$, then the new reaction $A^i \to B$ has the rate constant $kk_{\lim\, i}/k_j$. This rate is obtained using quasi-stationary distribution for the cycle.

The new auxiliary network $\mathcal{V}^1$ is computed for the network of glued cycles. Then we decompose it into cycles, glue them, iterate until a acyclic network is obtained $\mathcal{V}^n$.

III Restoring cycles

The dynamics of species inside glued cycles is lost after the second part. A full multi-scale approximation (including relaxation inside cycles) can be obtained by restoration of cycles. This is done starting from the acyclic auxiliary network $\mathcal{V}^n$ back to $\mathcal{V}^1$ through the hierarchy of cycles. Each cycle is restored according to the following procedure:

For each glued cycle node A_i^m, node of $\mathcal{V}^m$,

- Recall its nodes $A_{i1}^{m-1} \rightarrow A_{i2}^{m-1} \rightarrow ... A_{i\tau_i}^{m-1} \rightarrow A_{i1}^{m-1}$; they form a cycle of length τ_i.
- Let us assume that the limiting step in A_i^m is $A_{i\tau_i}^{m-1} \rightarrow A_{i1}^{m-1}$
- Remove A_i^m from $\mathcal{V}^m$
- Add τ_i vertices $A_{i1}^{m-1}, A_{i2}^{m-1}, ... A_{i\tau_i}^{m-1}$ to $\mathcal{V}^m$
- Add to $\mathcal{V}^m$ reactions $A_{i1}^{m-1} \rightarrow A_{i2}^{m-1} \rightarrow ... A_{i\tau_i}^{m-1}$ (that are the cycle reactions without the limiting step) with correspondent constants from $\mathcal{V}^{m-1}$
- If there exists an outgoing reaction $A_i^m \rightarrow B$ in $\mathcal{V}^m$ then we substitute it by the reaction $A_{i\tau_i}^{m-1} \rightarrow B$ with the same constant, i.e. outgoing reactions $A_i^m \rightarrow ...$ are reattached to the beginning of the limiting steps
- If there exists an incoming reaction in the form $B \rightarrow A_i^m$, find its prototype in $\mathcal{V}^{m-1}$ and restore it in $\mathcal{V}^m$
- If in the initial $\mathcal{V}^m$ there existed a "between-cycles" reaction $A_i^m \rightarrow A_j^m$ then we find the prototype in $\mathcal{V}^{m-1}$, $A \rightarrow B$, and substitute the reaction by $A_{i\tau_i}^{m-1} \rightarrow B$ with the same constant, as for $A_i^m \rightarrow A_j^m$ (again, the beginning of the arrow is reattached to the head of the limiting step in A_i^m)

The result of the algorithm is a reduced network that has no cycles and no branchings. Some reactions necessarily disappear from the initial model in order to break cycles and eliminate branchings, so the global operation can be called pruning. Pruning expresses the domination relations between pathways. Simple in acyclic networks (the quicker branch dominates, the much slower branches are pruned), these relations can be quite intricate in the presence of cycles. Rate constants of some of the remaining reactions are changed into monomial functions of the initial constants.

For the reduced network the calculation of the dynamics is straightforward.

Solution of the homogeneous linear dynamic equations (7) are:

$$x(t) = \sum_{k=1}^{n} r^k (l^k, x(0)) \exp(\lambda_k t) \tag{9}$$

where λ_k, l^k, r^k are the eigenvalues, left and right eigenvectors of the kinetic matrix $\boldsymbol{K}$, respectively: $l^k \boldsymbol{K} = \lambda_k l^k$, $\boldsymbol{K} r^k = \lambda_k r^k$.

Computing eigenvalues and eigenvectors is straightforward for acyclic networks with no branching.

The eigenvalues are $\lambda_i = -\kappa_i$, one for each node in the network. If a node i is a sink (it has no successor) we consider that $\lambda_i = 0$.

Right eigenvectors r^i are obtained by recursion, in the forward direction along the reaction graph. One has $r^i_j = 0$ for $j < i$. Starting with the normalised value $r^i_i = 1$, the coordinates $r^i_{\phi^k(i)}$ ($k = 1, 2, ..$) are obtained by:

$$\begin{aligned} r^i_{\phi^{k+1}(i)} = \frac{\kappa_{\phi^k(i)}}{\kappa_{\phi^{k+1}(i)} - \kappa_i} r^i_{\phi^k(i)} &= \prod_{j=0}^{k} \frac{\kappa_{\phi^j(i)}}{\kappa_{\phi^{j+1}(i)} - \kappa_i} \\ &= \frac{\kappa_i}{\kappa_{\phi^{k+1}(i)} - \kappa_i} \prod_{j=1}^{k} \frac{\kappa_{\phi^j(i)}}{\kappa_{\phi^j(i)} - \kappa_i}. \end{aligned} \tag{10}$$

Left eigenvectors are also obtained by recursion, but in the reverse direction. Thus, $l^i_j = 0$ for $j > i$. Starting with the normalised value $l^i_i = 1$, the coordinates l^i_j are obtained as:

$$l^i_j = \frac{\kappa_j}{\kappa_j - \kappa_i} l^i_{\phi(j)} = \prod_{k=0}^{q-1} \frac{\kappa_{\phi^k(j)}}{\kappa_{\phi^k(j)} - \kappa_i}. \tag{11}$$

In the case of fully separated systems, these expressions are significantly simplified and do not require knowledge of the exact values of κ_i. Thus, for the left eigenvectors $l^i_i = 1$ and, for $i \neq j$,

$$l^i_j = \begin{cases} 1, \text{ if } \phi^q(j) = i \text{ for some } q > 0 \text{ and } \kappa_{\phi^d(i)} > \kappa_i \text{ for all } d = 0, \dots q-1 \\ 0, \text{ else} \end{cases} \tag{12}$$

For the right eigenvectors we suppose that $\kappa_f = 0$ for a sink vertex A_f. Then $r^i_i = 1$ and

$$r^i_{\phi^k(j)} = \begin{cases} -1, \text{ if } \kappa_{\phi^k(i)} < \kappa_i \text{ and } \kappa_{\phi^m(i)} > \kappa_i \text{ for all } m = 1, \dots k-1 \\ 0, \text{ else} \end{cases} \tag{13}$$

A monomolecular network with totally separated constants have rate-limiting step. Supposing that the reduced network is a chain, the rate-limiting step is

the slowest reaction in the chain. However, this is not always the slowest reaction of the initial network.

Broken cycle The simplest example illustrating this counterintuitive possibility is a cycle of reactions. Consider an isolated cycle with total separation. The reduced acyclic model is the chain obtained from the the cycle by removing its slowest constant. The rate limiting step in the chain is the second slowest constant of the cycle. A cycle with total separation behaves like a chain ensuring transport of the mass to the beginning of the slowest step.

Interrupted pathway The effect of pruning can also lead to pathway interruption.

For instance let us consider the example in Fig.1. The network preprocessing consists in pruning reaction $A_4 \to A_2$ because this is dominated by the much faster reaction $A_4 \to A_5$. The resulting auxiliary network has the cycle (A_3, A_4, A_5), that is glued to A_3 in the step 3). The reaction $A_4 \to A_2$ exiting the cycle is restored with renormalized $k'_{24} = k_{24}k_{35}/k_{54}$ constant in step 3). This produces a new cycle that is glued to A_2. Depending on the order relation between the renormalized constant k'_{24} of the exit reaction and the constant k_{32}, the limiting step of the glued cycle A_2 can change. After elimination of cycle limiting step and cycle restoration 3.1.3-4 or 3.2.3-4 there are two possible reduced neworks, both of them chains. In the case 3.1.4 the limiting step for the transformation of A_1 into A_5 is the reaction $A_2 \to A_3$, the slowest reaction in the initial mechanism, but in 3.2.4 the reduced mechanism no longer contains this transformation.

Futile cycles and switching Metabolic networks contain many cyclic structures. As discussed in [3] a futile cycle converts a metabolite into another and back. It produces no net change but dissipates energy. Among various potential roles of futile cycles (heat production, increased control coefficients) there is the possibility of switching the direction of the flux (see Fig.2).

5 Quasi-stationarity and quasi-equilibrium: pooling species and reactions

Quasi-stationarity and quasi-equilibrium are useful concepts that can be used for model reduction of rather general reaction mechanisms.

Quasi-equilibrium reactions are reversible reactions at thermodynamic equilibrium. Species involved in quasi-equilibrium reactions bear well defined algebraic relations between their concentrations. A mass action quasi-equilibrium reaction would imply:

$$\frac{\prod_j x_j^{\beta_j}}{\prod_j x_j^{\alpha_j}} = K_{eq} = \exp\left(-\frac{\Delta G}{RT}\right) \tag{14}$$

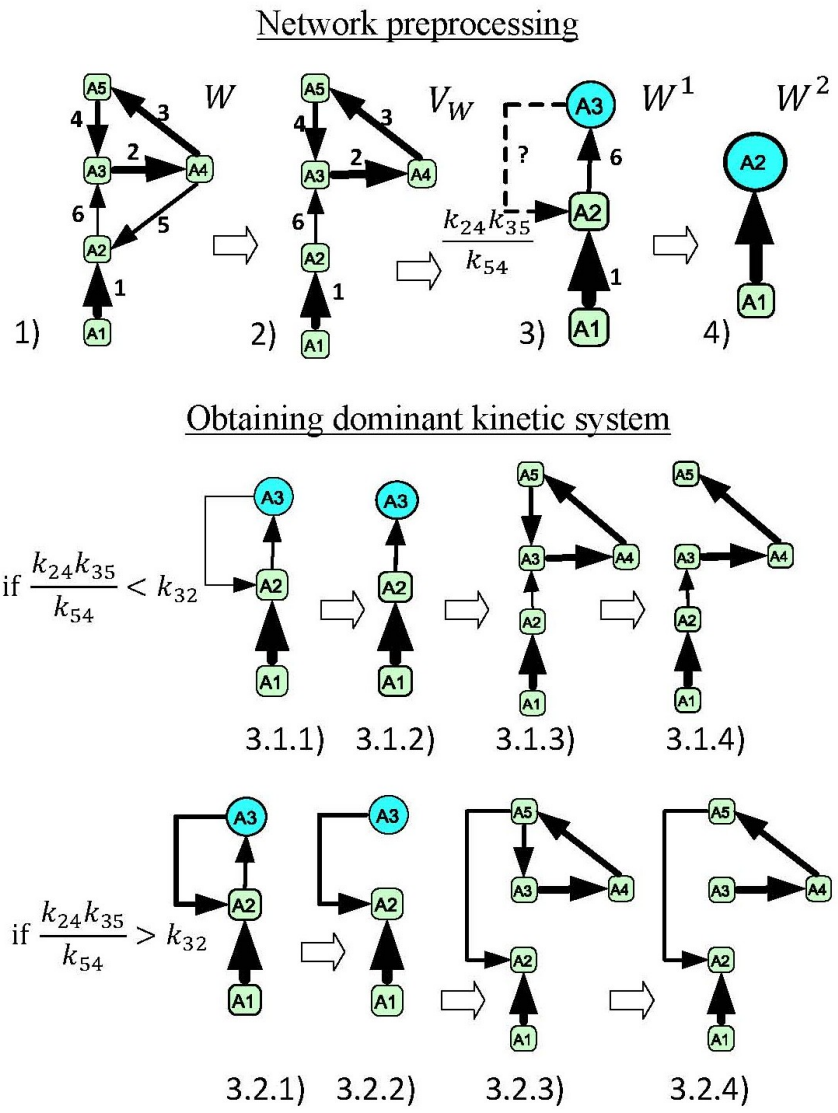

Figure 1: Example of calculation of the dominant approximations for a monomolecular reaction network with total separation of the constants (from [15]). The order of kinetics parameters is shown both by integer numbers (ranks) and the thickness of arrows (faster reactions are thicker).

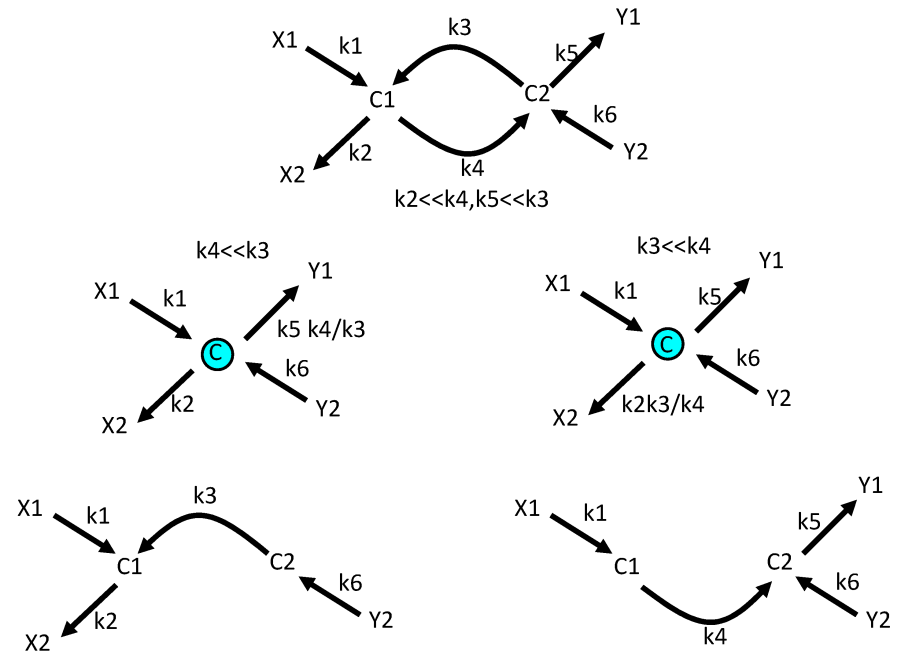

Figure 2: Futile cycle used as a metabolic switch. The cycling condition reads $k_2 << k_4, k_5 << k_3$.

where R is the universal gas constant, T is the temperature, ΔG is the Helmholtz free energy change, $K_{eq} = k^+/k^-$.

Slightly more generally, quasiequilibrium approximation uses the assumption that a group of reactions is much faster than other and goes fast to its equilibrium. This can be studied by using singular perturbations[18, 19], by introducing a small positive parameter ϵ representing the ratio of timescales of slow and fast reactions. Then the dynamics reads:

$$\frac{d\boldsymbol{x}}{dt} = \sum_{s,slow} R_s \boldsymbol{\gamma}^s + \frac{1}{\epsilon} \sum_{f,fast} R_f \boldsymbol{\gamma}^f \tag{15}$$

To separate slow/fast variables, we have to study the spaces of linear conservation law of the initial system and of the fast subsystem:

$$\frac{d\boldsymbol{x}}{dt} = \frac{1}{\epsilon} \sum_{f,fast} R_f \boldsymbol{\gamma}^f \tag{16}$$

In general the system (15) can have several conservation laws. These are linear functions $b^1(\boldsymbol{x}), \ldots, b^m(\boldsymbol{x})$ of the concentrations that are constant in time. The conservation laws of the system (16) provide variables that are constant on the fast timescale. If they are also conserved by the full dynamics, the system has no slow variables (variables are either fast or constant). In this case, the dynamics of the fast variables is simply given by

Eq.(16). Suppose now that the system (16) has some more conservation laws $b^{m+1}(\boldsymbol{x}), \ldots, b^{m+l}(\boldsymbol{x})$ that are not conserved by the full system (15). Then, these provide the slow variables of the system. The QE equation $\sum_{f,slow} R_f \boldsymbol{\gamma}^f = 0$ serves to compute fast variables as functions of the slow ones [10].

The quasisteady-state (QSS) assumption was invented in chemistry for description of systems with radicals or catalysts [20]. In the most usual version [16], the species are split in two groups with concentration vectors $\boldsymbol{c}^s$ ("slow" or basic components) and $\boldsymbol{c}^f$ ("fast intermediates" or QSS species).

The small parameter ϵ used in singular perturbation theory is now the ratio of small concentrations of fast intermediates to the concentration of other species. After rescaling $\boldsymbol{c}^s$ and $\boldsymbol{c}^f$ to order one, the set of kinetic equations reads:

$$\frac{d\boldsymbol{c}^s}{dt} = \boldsymbol{W}^s(\boldsymbol{c}^s, \boldsymbol{c}^f) \tag{17}$$

$$\frac{d\boldsymbol{c}^f}{dt} = (1/\epsilon)\boldsymbol{W}^f(\boldsymbol{c}^s, \boldsymbol{c}^f) \tag{18}$$

where the functions $\boldsymbol{W}^s$, $\boldsymbol{W}^f$ and their derivatives are of order one ($0 < \epsilon << 1$).

The standard singular perturbation theory[18, 19] provides the QSS algebraic condition $\boldsymbol{W}^f(\boldsymbol{c}^s, \boldsymbol{c}^f) = 0$. These equations, together with additional balances for c^f (conservation laws) are enough to deduce the fast variables c^f as functions of the slow variables c^s and to eliminate them [20, 13, 15]. The slow dynamics is given by Eq.(17).

However, not all fast species are small concentration intermediates. The simplest such example is a fast irreversible cycle with a slow exit reaction. This example does not correspond to the traditional definition of quasi-equilibrium because it lacks reversibility and can not fulfill detailed balance. A singular perturbation analysis similar to the one for QE shows that the total mass of the cycle (this can be arbitrarily big) represents a slow variable, while each one of the concentrations of species inside the cycle are fast variables. The algebraic relations for fast species are those for QSS because they express the steady-state condition for the fast cycle.

The simplest illustration of these two approximation is provided by the Michaelis-Menten model for enzymatic reaction (see section 2).

One can have quasi-equilibrium if the first equation is fast: $k_1^{\pm} = \kappa^{\pm}/\epsilon$, where $\epsilon > 0$ is small. Then, the quantities conserved by the rapid reaction form two slow pool global variables, namely $C^s = [S] + [ES]$ and $b_E = [E] + [ES]$. Actually, b_E is conserved by all the reactions of the mechanism, so it is a kinetics constant $b_E = const.$. The algebraic quasi-equilibrium condition reads $k_1^+(C^s - [ES])(b_E - [ES]) = k_1^-[ES]$. This gives a dependence of the pool variable $[ES]$ on the pool global variables, namely $[ES] = c_{ES}(C^s, b_E)$. The final slow dynamics, obtained from Eq.6 and by adding term by term the Eqs.3,5, reads:

$$\dot{C^s} = -k_2 c_{ES}(C^s, b_E) \quad (19)$$
$$\dot{P} = k_2 c_{ES}(C^s, b_E) \quad (20)$$

The reduced QE mechanism is a single reaction transforming the pool C^s into the product P, with a rate law $R(C^s) = k_2 c_{ES}(C^s, b_E)$ (the constant b_E is the total quantity of enzyme).

The QSS approximation is obtained from the full mechanism when the enzyme is in much less quantity than the substrate $C^s >> b_E$. Under this condition, E and ES are fast intermediates (QSS species). b_E is conserved and constant like before. The slow variables are the concentrations of non-intermediate species, namely $[S], [P]$.

The QSS algebraic condition reads $k_1^+[S][E] = (k_1^- + k_2)[ES]$ which gives a dependence of the fast variables on the concentrations of the other species:

$$[ES] = c_{ES}([S], b_E) = k_1^+[S]b_E/(k_1^+[S] + k_1^- + k_2) \quad (21)$$

The dynamics of the external (non-intermediate) species reads:

$$[\dot{S}] = -k_1^+[S](b_E - c_{ES}([S])) + k_1^- c_{ES}([S]) = -k_2 c_{ES}([S], b_E) \quad (22)$$
$$[\dot{P}] = k_2 c_{ES}([S], b_E) \quad (23)$$

The reduced QSS mechanism is a single reaction transforming the substrate S into the product P, with a rate law $R_{MM}(C^s, b_E) = k_2 c_{ES}(C^s, b_E) = V_m[S]/([S] + K_m)$, where R_{MM} is the well known Michaelis-Menten rate law, $V_m = k_2 b_E$, $K_m = (k_1^- + k_2)/k_1^+$.

The single step of the reduced M-M mechanism can be seen as resulting from merging (considering them to be simultaneous) two steps $S + E \longrightarrow ES$ and $ES \longrightarrow E + P$ of the initial full mechanism. This merged step, or "pool of reactions", is a combination of reactions involving the two rapid species E and ES such that the resultant reaction does not change the concentrations of any of the fast species (these combinations conserve the fast species). Given a reaction mechanism and a set of fast species, there may be several reaction pools that preserve fast species (notice that for pool definition, reversible reactions are considered as two steps, one for each direction). There is only one such pool for the M-M mechanism.

The difference between the QSS and the QE in this example is obvious.

QE corresponds to pooling of species. For QE we first identify pools of species that are rapidly transformed one into the other by rapid QE reactions. The number of pools is given by the number of conservation laws of the set of QE reactions that are not conservation laws of the full mechanism. Then, QE conditions allow to express the rates of reactions exiting the pools as functions of global variables of the pools (conservation laws of QE reactions).

The reduced mechanism is made out of the pools, the remaining species and reactions with rate laws thus computed.

QSS corresponds to a pooling of reactions. A reaction pool (also called reaction route [20, 17]) is a linear combination with positive integer coefficients of reactions in the mechanism (reversible reactions counted twice, one reaction for each direction). We are interested in those pools (routes) that transform slow species into other slow species and conserve the intermediate fast species. In [15] we have also imposed a simplicity criterion for the pools, by choosing only simple sub-mechanisms. Simple sub-mechanisms are pools (routes) with a minimal number of reactions, transforming slow species without producing accumulation or depletion of the intermediate fast species. According to this definition, simple sub-mechanisms are elementary modes [4] of the set of reactions involving fast species. The QSS conditions and the internal balances are used to express the concentration of intermediate species and the rate laws of pooled reactions as functions of the concentrations of slow species [15].

Coming back to the previous section we would like to relate the reduction algorithm for monomolecular networks to the general concepts of QE and QSS.

Monomolecular networks with completely separated constants can not be considered to be at quasi-equilibrium, because they do not include reversible reactions (if both forward and reverse fluxes are allowed, then one of them dominates the other). Although quasi-equilibrium ideas have been used as an intermediate step of the reduction algorithm (gluing cycles), the reduced model is a acyclic graph with no pooling of species.

Monomolecular networks with completely separated constants can contain QSS species. These can be easily identified in the reduced model which is a chain or a set of chains with confluences. For instance, if the reduced model is a chain, the QSS species are those species that are consumed by fast reactions. In totally separated chains, QSS species concentrations can be set to zero (they are consumed by fast reactions). For instance, in the example shown in Fig.1, case 3.1.4 for timescales of the order of the inverse of the rate limiting step, three species A_1, A_3, A_4 are QSS. The reaction pool $(A_2 \longrightarrow A_3) + (A_3 \longrightarrow A_4) + (A_4 \longrightarrow A_5)$ gives the reduced mechanism $A_2 \longrightarrow A_5$ of constant rate k_{32}.

As suggested above, identification of QSS species, QE reactions and limiting steps is not easy in general. The QSS and QE nature of species and reactions as well as the limiting steps are global properties of the reaction mechanism that can not be easily obtained by comparing rate constants of individual reactions. Furthermore, the idea of "rapid reactions" can lead to very complex kinetic situation and should be used with care in the reduction of models.

Last but not least, we must emphasize an important difference between QSS and QE. Contrary to QE, QSS is a purely kinetic concept and has no relation to equilibrium thermodynamics (it does not have to obey detailed

balance for instance). Thus, in the QSS situation, rate constants can not be related to thermodynamic potentials. This makes QE a simpler situation from the point of view of parameter identification.

6 Conclusion

Dynamics of metabolic networks can be studied by systems of differential equations. Large models with incomplete information are not suited for immediate analysis by traditional approaches and have to be simplified.

We have presented several model reduction techniques allowing to transform large reaction networks into simpler networks, whose dynamics can be readily studied. These techniques exploit the separation of the timescales of the complex networks. In the process of simplification, non-critical elements are removed from the models, and only essential elements are kept.

For monomolecular networks with total separation of the rate constants, we propose a reduction algorithm allowing to transform any such network into an acyclic network without branching, whose dynamics is computed analytically. The global transformation leading to simpler monomolecular networks can be defined as pruning. This transformation eliminates dominated reactions and computes a dominant subnetwork. The limiting step, easily identified on the reduced network, can be different from the slowest reaction of the full mechanism. Monomolecular models, though not always realistic, can teach us about design principles of large networks.

More general concepts such as quasi-equilibrium and quasi-steady state approximation can be applied to simplify non-linear as well as linear networks. We showed how these approximations can be related to pooling of species and of reactions.

Pooling of species and of reactions can also result from decompositions of the Jacobian (matrix defining the linearised dynamics) of nonlinear systems of differential equations (2), once all rate constants are known. This method has been applied in [11] to analyse pooling of a fully parametrized glycolysis model. However, one would like to obtain the pools without knowing numerical values of all the parameters, using only the order relations between time scales and/or rate constants.

There is still much to do in this direction to propose simple general rules allowing for correct identification of limiting steps, QSS species and QE reactions. The next case to study will be the linear networks with partial separation, that could be approached by a combination of pooling and pruning. QSS and QE, combined with techniques for dominant solutions of algebraic equations represent a promising approach to the reduction of non-linear models (see [15, 13]).

References

[1] E. Chiavazzo, A.N. Gorban, and I.V. Karlin. Comparison of invariant manifolds for model reduction in chemical kinetics. *Communications in Computational Physics*, 2(5):964–992, 2007.

[2] 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.

[3] D. Fell. *Understanding the control of metabolism*. Portland Press London, 1997.

[4] J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC bioinformatics*, 5(1):175, 2004.

[5] Alexander N. Gorban, Iliya V. Karlin, and Andrei Yu. Zinovyev. Constructive methods of invariant manifolds for kinetic problems. *Physics Reports*, 396(4-6):197 – 403, 2004.

[6] A.N. Gorban and I.V. Karlin. Method of invariant manifold for chemical kinetics. *Chemical Engineering Science*, 58(21):4751–4768, 2003.

[7] A.N. Gorban and I.V. Karlin. *Invariant manifolds for physical and chemical kinetics, Lect. Notes Phys. 660*. Springer Verlag, Berlin, Heidelberg, 2005.

[8] A.N. Gorban and O. Radulescu. Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Syst. Biol*, 1(4):238–246, 2007.

[9] AN Gorban and O. Radulescu. Dynamic and static limitation in reaction networks, revisited . In David West Guy B. Marin and Gregory S. Yablonsky, editors, *Advances in Chemical Engineering - Mathematics in Chemical Kinetics and Engineering*, volume 34 of *Advances in Chemical Engineering*, pages 103–173. Elsevier, 2008.

[10] AN Gorban, O. Radulescu, and AY Zinovyev. Asymptotology of chemical reaction networks. *Chemical Engineering Science*, 65:2310–2324, 2010.

[11] N. Jamshidi and B.Ø. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Molecular Systems Biology*, 4:171, 2008.

[12] SH Lam and DA Goussis. The CSP method for simplifying kinetics. *International Journal of Chemical Kinetics*, 26(4):461–486, 1994.

[13] Mark Z. Lazman and Gregory S. Yablonsky. Chapter 2 overall reaction rate equation of single-route complex catalytic reaction in terms of hypergeometric series. In David West Guy B. Marin and Gregory S. Yablonsky, editors, *Advances in Chemical Engineering - Mathematics in Chemical Kinetics and Engineering*, volume 34 of *Advances in Chemical Engineering*, pages 47 – 102. Elsevier, 2008.

[14] U. Maas and S.B. Pope. Simplifying chemical kinetics- Intrinsic low-dimensional manifolds in composition space. *Combustion and Flame*, 88(3-4):239–264, 1992.

[15] O. Radulescu, A.N. Gorban, A. Zinovyev, and A. Lilienbaum. Robust simplifications of multiscale biochemical networks. *BMC systems biology*, 2(1):86, 2008.

[16] L.A. Segel and M. Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review*, 31(3):446–477, 1989.

[17] ON Temkin, A.V. Zeigarnik, and D. Bonchev. *Chemical Reaction Networks: a graph-theoretical approach*. CRC, Boca Raton, 1996.

[18] A. N. Tikhonov, A. B. Vasileva, and A. G. Sveshnikov. *Differential equations*. Springer, Berlin, 1985.

[19] W. Wasow. *Asymptotic Expansions for Ordinary Differential Equations*. Wiley, New York, 1965.

[20] GS Yablonskii, VI Bykov, AN Gorban, and VI Elokhin. *Kinetic models of catalytic reactions. Comprehensive Chemical Kinetics*, volume 32. Elsevier, Amsterdam, 1991.

$$\frac{dX}{dt} = F1 - F2(X)$$

$$F2(X) = KX$$