Application of Lumping Analysis in Modelling the Living Systems – A Trade-off Between Simplicity and Model Quality

G. Maria
Laboratory of Chemical & Biochemical Reaction Engineering,
University Politehnica of Bucharest, P.O. 35–107 Bucharest, Romania
Email: gmaria99m@hotmail.com

Original scientific paper
Received: February 2, 2006
Accepted: September 15, 2006

General chemical engineering modelling principles are valuable tools to represent both stationary and dynamic characteristics of complex cell processes. Elaboration of reduced (lumped) dynamic models uses all types of information 'translated' from the 'language' of molecular biology to that of mechanistical chemistry, by preserving cell structural hierarchy and component functions. A combination of non-/conventional estimation methods reported significant model quality improvements by accounting for qualitative/quantitative data and global properties of the living system. Derivation of a satisfactory model is closely related to the ability of selecting the suitable lumping rules, key-parameters, and influential terms that better realize a trade-off between model simplicity and its predictive quality. Several examples, on the modular modelling of protein synthesis regulation, genetic regulatory networks, and on the successive drug-ligand release in human plasma from a complex multivalent support, illustrate the advantages but also the over-simplifications introduced by various lumping rules.

Key words:
Lumping analysis, genetic regulatory networks, modules, dynamic models

Introduction

Living cells are organized, self-replicating, self-adjustable, evolvable and responsive structures to environmental stimuli, able to convert 'raw materials' (nutrients) from environment into additional copies of themselves. Due to the highly complex and partly unknown aspects of the metabolic processes, the detailed mathematical modelling at a molecular level remains still an unsettled issue, even if remarkable progresses and developments of extended cell simulation platforms have been reported, by using large amounts of information and data (see review of *Maria*[1]). Given these developments, as well as tremendous advances in computing power, it is tempting to believe that reliable whole-cell models with predictive power will be forthcoming once complete sets of 'bio-omic' information become available. However, a better theoretical understanding of detailed cell metabolic processes is necessary before to understand how the process of life emerges out of complex networks of molecular-level interactions between cellular components.

Reliable and sufficiently accurate mechanistic models are very effective tools in representing a chemical/biological process, its influential variables, and physical meaning of parameters and reaction steps. Classical modelling rules use experimental information to build-up a mathematical structure based on conservation and thermodynamic relationships, hypothetical reaction mechanism, kinetic expressions, and known stoichiometry. Then, conventional identification steps, derived from the statistical estimation theory, lead to adapt the model structure/size according to the available data and utilization scope. The costly estimation rules are linked to the statistical methods because the observed data are always subjected to experimental errors, and multiple constraints are usually imposed to the kinetic parameters.[2]

These general modelling rules, based on physico-chemical-biological and chemical engineering principles, are more difficult to be applied to living systems when the analysis is expanded to complex metabolic networks. That is because cell processes at a molecular level present a low observability and data reproducibility vs. very large number of species (states), reactions and transport parameters (many of them poorly understood), while mechanistic details and standard kinetic data are difficult to be obtained. To represent the cell process complexity, by including all key-species, reactions and large number of interactions, modern modelling and lumping techniques have been developed in two complementary approaches: structure-oriented topological analysis and kinetic (dynamic) models.[1] Kinetic models, using various types of variables (continuous, Boolean/discrete, stochastic, or hybrid/mixed) are suited to represent various

processes such as species-interconnectivity, transcriptional regulation, gene expression, internal autocatalysis, system homeostasis and regulation hierarchy, variable volume and isotonic osmolarity, continuous perturbations, cell signalling, gene mutation, molecular diffusion, etc.[34]

To overcome structural low identifiability, current trend in dynamic modelling is to apply advanced numerical techniques that use all types of information 'translated' from the 'language' of molecular biology to that of mechanistic chemistry by preserving the cell structural hierarchy and component functions together with a combination of unconventional – conventional estimation and lumping procedures that account for qualitative/(poor) quantitative information and global properties of the system (e.g. regulatory properties, structural and functional cell hierarchy, key-species homeostatic properties, equilibrated growth stability, cell process periodicity and succession of events, etc.).[1,35]

One key aspect during elaboration of reduced dynamic models in molecular biology is the adequate application of *lumping rules* that group species of similar functions or metabolic reactions in *lumps*. This approach must keep a satisfactory predictability on key-species and steps, species inter-connectivities, structural and functional hierarchy, multi-cascade control with adjustable intermediate levels and multiple effectors in feedback loops. Due to the cell process large complexity, i.e. 10^4–10^6 species, 10^6–10^7 reactions and parameters, and multiple cell sub-structures and functions, application of lumping techniques present important advantages in modelling, such as:

I) increase in the model identifiability and estimability, compensating the data low-observability, low-reproducibility and process variability by reducing the accounted number of reactions and variables and by keeping the most influential terms in the kinetic and transport relationships;

ii) computational tractability allowing application of engineering rules and dynamic systems theory to better characterize the cell system in terms of stationary state (homeostasis) multiplicity, stability, flexibility, robustness to environmental perturbations, etc.;

iii) easier application of effective numerical rules to develop integrated databank – modularised platforms for process simulation, similarity analysis, generation of model algebraic-differential equations from reaction biochemistry, storage of model equations and parameters.

Among the model reduction costs are to be mentioned: I) loss of information on certain species and reactions; ii) biased estimate of model parameters; iii) loss in model generality, prediction capabilities and physical meaning of the derived apparent (overall) rate constants; iv) increased number of reduced structures (rival models) to be discriminated; v) alteration of systemic/holistic predicted properties (such as stability, sensitivity/robustness to perturbations, regulatory effectiveness, response rate to external stimuli, system flexibility, tight control, etc.). For instance, reduced dynamic models indicate too idealized cell regulatory functions, altered response to stationary or dynamic perturbations (by re-allocating species functions to a fewer number of components), reduced possibility to include intermediates that increase the system robustness and to account for the cell-content "inertial" and "ballast" effects. Besides, if elementary serial or parallel steps are lumped together, the identified overall rate constants are always smaller or larger than those of the original steps, leading to apparent reaction parameters of lower significance.

In spite of such drawbacks, lumping in cell dynamic modelling is indispensable, being a widely used technique to adequately represent the cell complexity. Model quality tests, parameter and species sensitivity analysis, principal component and algorithms to find invariant subspaces are common rules to reduce the extended model structures.

The scope of this paper is to illustrate, with some examples, how elaboration of reduced kinetic models of satisfactory quality is closely related to the ability of selecting the suitable heuristic/rigorous lumping rules, key-parameters, and influential terms, and to apply unconventional identification strategies that better realize a trade-off between model simplicity and its predictive quality.

Examples on the gradual lumping analysis include elaboration of genetic regulatory network (GRN) models for simulating the mechanism by which genes and proteins interact to regulate the gene expression. Various semi-autonomous lumped kinetic modules are elaborated, based on experimental observations, in order to represent the protein synthesis regulation. Individual modules are separately investigated in terms of structure and regulatory efficiency, and then linked in regulatory chains accordingly to certain rules that ensure the overall network efficiency in conditions that mimic the stationary and perturbed cell growth, system homeostasis, variable volume and isotonic osmolarity. Advantages and limitations of such a lumping strategy are also discussed.

Another example refers to elaboration of lumping relations between extended (intrinsic, low-estimable) reaction pathway structures of moderate size and apparent (reduced) biokinetic models, with exemplification for the successive drug-ligand release in human plasma. Such links can reveal kinetic

characteristics of the elementary steps to be accounted for the design of a drug-support of desirable properties and release characteristics.

Lumping rules applied in (bio)chemical kinetic modelling

The low-observability and low identifiability of very complex cell biochemical systems lead to the impossibility to identify all parameters and terms of extended mechanistic models. Because in the cellular processes there is a large excess of degrees of freedom in adjustable parameters than in observed and manipulated variables, adequate modelling can lead to multiple solutions, even if reduced model structures are checked. Insensitive and uncertain parameters to the input data changes can cause intrinsic poor-conditioning of the estimator, biased and poor-quality solution, ambiguous estimate, and estimation algorithm failure. If no supplementary information is available, a considerable increase in estimate quality is obtained by reducing the extended models, thus decreasing the model over-parameterization or internal degeneracy due to additive and/or multiplicative parametric terms of unobservable separate influences. The obtained reduced models of higher identifiability can be effective tools to investigate the cell sub-systems and certain cell properties.

Kinetic model reduction is realised by means of special experimental methods (not discussed here)[2] and computational algorithms. As reviewed by *Maria*[1] and *Gorban* et al.,[3] the 'reduced description of a chemical system' means: (i) to shorten the list of species, by eliminating inessential components and/or lumping some species; (ii) to shorten the list of reactions, by eliminating inessential side-reactions and/or assuming quasi-equilibrium for some reaction steps; (iii) to decompose the kinetics into fast and slow 'parts' allowing a separate study and application of the quasi-steady-state-approximation (QSSA) to reduce its dimensionality.

Among used methods, are to be mentioned: parameter and species sensitivity analysis, principal component, ridge parameter selection, and algorithms to find invariant subspaces and invariant manifolds of the original dynamic system, with accounting for physico-chemical-biological restrictions. All these methods can suggest the key-parameters and lumped species subsets that can compensate the model uncertainty.[4,5]

From the mathematical point of view, lumping of continuous variables in (bio)chemical kinetic models implies reduction of the dimensionality of the state (concentrations $\boldsymbol{c}$) and parameter (rate constants $\boldsymbol{k}$) vectors, that is:

$$\frac{\mathrm{d}\boldsymbol{c}}{\mathrm{d}t}=\boldsymbol{r}=\boldsymbol{f}(\boldsymbol{c},\boldsymbol{k}) \Rightarrow \frac{\mathrm{d}\hat{\boldsymbol{c}}}{\mathrm{d}t}=\hat{\boldsymbol{f}}(\hat{\boldsymbol{c}},\hat{\boldsymbol{k}});\ [\hat{\boldsymbol{c}},\hat{\boldsymbol{k}}]=\boldsymbol{h}(\boldsymbol{c},\boldsymbol{k});$$
$$\hat{n}_\mathrm{s}=\mathrm{size}(\hat{\boldsymbol{c}})<n_\mathrm{s}=\mathrm{size}(\boldsymbol{c}); \quad (1)$$
$$\hat{n}_\mathrm{p}=\mathrm{size}(\hat{\boldsymbol{k}})<n_\mathrm{p}=\mathrm{size}(\boldsymbol{k}).$$

The lumped species $\hat{\boldsymbol{c}}$ are related to the original ones $\boldsymbol{c}$ by some lumping functions $\boldsymbol{h}$, which can be linear or non-linear. The link between reduced and extended models depends on the way in which lumping has been realised. Moreover, various constraints must also be accounted during application of lumping rules, representing thermodynamic, physico-chemical-biological limitations, reaction hypotheses, QSSA for intermediates, or some thermodynamic conditions in reversible cyclic reactions.[2] Among the large variety of lumping rules, some techniques can be relevant for biochemical system representation.

Reaction lumping. Elimination, or reaction lumping, lead to simplify a reaction schema, but inherently lead to apparent reaction rates, rate constants and reaction orders.[6] One method to identify low-significant reactions is based on sensitivity measures of model predictions (in terms of species concentrations) vs. rate constants:[7]

$$\sum_{i=1}^{n_\mathrm{s}} |\partial \ln c_i / \partial \ln k_j|_{\boldsymbol{x}_u};$$
$$\sum_{i=1}^{n_\mathrm{s}} |\partial \ln c_i / \partial \ln k_j|^2_{\boldsymbol{x}_u}; \quad (2)$$
$$j=1,\ldots,n_\mathrm{p},$$

(where $\boldsymbol{x}_u$ = independent variable vector for the run $u = 1,\ldots n;$; n_s = number of species). As these sensitivities vary with the reaction time, a threshold of k_j-unimportance may be dependent on species and reaction progress. A uniform threshold of 5 % can be adopted to identify redundant reactions to be left from the scheme, even if the risk to obtain an oversimplified model is not diminished (see discussion of *Tomlin* et al.[4]). Otherwise, elimination of k_j can be decided based on small values of the overall sensitivity measure $\sum_{u=1}^{n}\sum_{i=1}^{n_\mathrm{s}}(\partial \ln c_{iu}/\partial \ln k_j)^2$. A similar integral sensitivity measure leads to the parameter rejection test:[8]

$$\Delta_{p,j}=\frac{1}{n}\sum_{u=1}^{n}\sum_{i=1}^{n_\mathrm{s}}\frac{k_j}{c_{iu}}\left|\frac{\partial c_j}{\partial k_j}\right|_{\boldsymbol{x}_u}<1-3. \quad (3)$$

Similarly, relative sensitivities of species i vs. parameter k_j in every run u, lead to the rejection test:[9]

$$s^*_{ij,u} < 0.01; \quad s_{g,j} = \sum_{u=1}^{n}\sum_{i=1}^{n_s} s^*_{ij,u} < 0.01(n_s n);$$
$$s^*_{ij,u} = |c_i(t_u,\boldsymbol{k}) - c_i(t_u,\boldsymbol{k})_{k_j=0}| \big/ c_i(t_u,\boldsymbol{k}). \tag{4}$$

Another lumping method is based on a simultaneous analysis of reaction sensitivity and reaction significance in the model, i.e. the so-called principal component analysis (PCA[10]). By evaluating the species sensitivities vs. parameters, $\boldsymbol{S}_u = [\partial \ln c_{iu} / \partial \ln k_m]$, the "information" matrix $\boldsymbol{S}^T\boldsymbol{S}$ can be computed for all runs, $\boldsymbol{S}^T = [\boldsymbol{S}_1 \ \boldsymbol{S}_2 \ \ldots \ \boldsymbol{S}_n]$. As revealed by *Vajda* et al.,[5,9] very small $\boldsymbol{S}^T\boldsymbol{S}$ eigenvalues ($\lambda_{\min} = 0$) correspond to linear dependencies among sensitivities and indicate a structurally unidentifiable model. Thus, a parameter k_j rejection test is proposed of the form:[9]

$$\lambda_j \le 10^{-4}(n_s n), \tag{5}$$

(where λ_j = eigenvalues of $\boldsymbol{S}^T\boldsymbol{S}$). A way to lump kinetic parameters uses identified approximate linear dependencies, by inspecting the eigenvectors $\boldsymbol{X}_i$ corresponding to $\boldsymbol{S}^T\boldsymbol{S}$ eigenvalues λ_j smaller than the noise variance, $\lambda_i < \sigma^2$. The proportional components of such eigenvectors are then used to construct rate constant lumps, of the form $k_l \big/ k_j^{\cos\tan t}$.

Side reactions, of less importance in the model, can be identified by means of the ridge selection analysis (RSA[11–13]). The method is based on the system (1) matrix, $\boldsymbol{S}_u = [\partial f_{iu} / \partial k_m]$, and on the eigenvectors λ_j of the modified information matrix $\left(\sum_{u=1}^{n} \boldsymbol{S}_u^T \boldsymbol{S}_u + \alpha \boldsymbol{I}\right)$ (where α is the Hoerl's factor, and $\boldsymbol{I}$ the identity matrix). The k_j rejection test, of the form $\lambda_j / \sigma^2_{\min} < 1(3)$ (with $\alpha \approx \sigma^2_{\min}$), is related to the observation that the biased estimate of redundant/insensitive parameters strongly depends on the Hoerl factor.[13,14] As α increases as the information that comes from experimental data is more and more dumped, and the k_j estimate is more and more independent on such information. In fact, this model intrinsic poor-conditioning reveals the incapacity to assert k_j and to extract information in the direction of $\boldsymbol{X}_j$ from the available data.

When elimination of low-sensitive/low-estimable parameters k_j is not desirable, the alternative is to fix these parameters to some apriori values (i.e. the so-called principal component regression PCR[8,13,15,16]). PCR identifies low-significant parameters corresponding to small eigenvalues of the $\boldsymbol{S}^T\boldsymbol{S}$ matrix ($\lambda_j / \lambda_{\max}$ < threshold). By reducing the estimation vector-size, the estimate dispersion is diminished, but it also becomes biased and more and more independent on the data.

Species lumping. Elimination, or species lumping in a kinetic model is performed when there is insufficient information to characterize the dynamics of all compounds, or when by-products and intermediate separate prediction is not crucial for the process analysis. In other terms, the reduced reaction scheme must not contain redundant species, while the information is condensed in a smaller set including groups of species represented as single variables. Some techniques perform simple manipulations of differential model equations by reducing the set-size and by replacing some individual species with convenient lumps.[6] When chain reactions include at least one equilibrium step, a pre-equilibrium assumption can reduce the parameter vector size.[17] One of the most used methods to reduce the species vector-size is application of the quasi-steady-state approximation QSSA.[4] Assuming quasi-negligible reaction rates for some (low observable) intermediates, their elimination from the model is realised based on small values for the product of target species lifetime ($LT_i = -1/J_{ii}$) and their production rate r_i (where the Jacobian elements are $J_{ik} = \partial f_i(\boldsymbol{c},\boldsymbol{k}) / \partial c_k$).[18] The model simplification cost is the introduction of a corresponding prediction error Δc_i (which includes an instantaneous error $\Delta c_i^s(t) = r_i / \boldsymbol{J}_{ii}$ and contributions from other QSSA in the same model).

Another variant is based on the sensitivity analysis of reaction rates vs. individual species. Redundant species i are detected based on a small global sensitivity index, $B_i = \sum_{m=1}^{n_r} s_{im}^2(t)$, defined over all reactions, $s_{im}(t) = \partial \ln(r_m) / \partial \ln(c_i)$, by including all direct and indirect effects.[4,7]

If lumping functions $\boldsymbol{h}$ (linear or non-linear) can be established in a systematic and coherent way, links between the species and parameters of the extended and reduced models can be used to interpret low-observable/low-estimable extended models from using apparent/identifiable kinetics. Necessary and sufficient conditions for an *exact linear* lumping of linearizable kinetic models have been establish:[19–22]

$$\frac{\mathrm{d}\boldsymbol{c}}{\mathrm{d}t} = \boldsymbol{J}^T\boldsymbol{c} \ \Rightarrow \ \frac{\mathrm{d}\hat{\boldsymbol{c}}}{\mathrm{d}t} = \boldsymbol{M}\boldsymbol{J}^T\boldsymbol{c} = \boldsymbol{M}\boldsymbol{J}^T\boldsymbol{M}^+\hat{\boldsymbol{c}} = \hat{\boldsymbol{J}}^T\hat{\boldsymbol{c}};$$
$$\hat{\boldsymbol{c}} = \boldsymbol{M}\boldsymbol{c}; \qquad \hat{\boldsymbol{f}}(\hat{\boldsymbol{c}}) = \boldsymbol{M}\boldsymbol{f}(\boldsymbol{c}). \tag{6}$$

These conditions require that $\boldsymbol{M}\boldsymbol{f}(\boldsymbol{c})$ is a function of $\hat{\boldsymbol{c}}$, and that an inverse (or a generalised inverse) of $\boldsymbol{M}$ exists, because $\boldsymbol{c} = \boldsymbol{M}^+\hat{\boldsymbol{c}}$. The lumping

rule consists of finding a suitable lumping matrix $\boldsymbol{M}$, of a chosen dimension, and its inverse.

A systematic approach to determine the lumping matrix $\boldsymbol{M}$ is based on a suitable decomposition of $\boldsymbol{J}^T$ and on the invariant subspaces of the original system.[20] The *exact* linear lumping matrix $\boldsymbol{M}$ can be constructed from the eigenvectors of the Jacobian $\boldsymbol{J}^T$ (i.e. $\boldsymbol{X} = [\boldsymbol{x}_j]$ from $\boldsymbol{J}^T\boldsymbol{X} = \lambda\boldsymbol{X}$), because any subspace spanned by a subset of the eigenvectors is a $\boldsymbol{J}^T$ invariant. As a consequence, [span{0}, span{$\boldsymbol{x}_1$},…] gives a 1-dimensional lumping matrix $\boldsymbol{M}$; [span{$\boldsymbol{x}_1$,$\boldsymbol{x}_2$}, span{$\boldsymbol{x}_2$,$\boldsymbol{x}_3$},…] gives a 2-dimensional lumping matrix $\boldsymbol{M}$ with rows formed with the $\boldsymbol{X}$-columns, etc. Then, the lumping matrices of different dimensions can be simply formed by taking the columns of $\boldsymbol{X}$, or *any* linear combination of them. In the exact lumping, the eigenvalues of the reduced system $\boldsymbol{J}^T$ represent a subset of the eigenvalues of the full $\boldsymbol{J}^T$. In the *approximate* lumping, there are several accepted transformation and model prediction errors.

Concerning the *nonlinear* lumping of species, there is any general rule to derive the link functions $\boldsymbol{h}$ because $\boldsymbol{J}^T(\boldsymbol{c})$ is not a constant matrix in general. As a consequence, $\boldsymbol{J}^T$ conversion into a canonical form to find invariant subspaces is not a simple task, even if necessary and sufficient conditions for nonlinear lumping have been already assigned.[4,23]

Other lumping rules, based on Markov chain theory, stochastic/entropy measures of the kinetic process, and/or fuzzy information on species similarities have been reported.[24]

A special strategy to apply the lumping rules refers to the analysis of the reaction *time-scales*.[4,17] The method of 'time-scale separation' starts from the observation that lacks in chemical system observability and in an efficiently modelling comes from the large range of time-scales of complex mechanisms. Intermediate species of small concentrations may have relaxation times up to ten orders of magnitude less than stable species, leading to stiff systems. This is also the case of cell biochemical systems where small amount of intermediates are responsible for quick adjustment of the regulatory and synthesis path efficiency. Moreover, the osmotic equilibrium with the extra-cellular medium and the cell energetic 'effort' to produce a large number of metabolic intermediates impose low concentration levels for enzymes and intermediates comparatively to substrates, metabolites, or key-products. A common measure of the relaxation process is the characteristic time of intermediate species, defined from the original system (1) Jacobian matrix:[17]

$$\tau_i = \frac{1}{|Re(\lambda_i)|};$$

$$\lambda_i = \text{eigenvalues of } \boldsymbol{J} = \partial \boldsymbol{f}(\boldsymbol{c},\boldsymbol{k})/\partial c; \qquad (7)$$

$$i = 1,\ldots,n_s.$$

The very fast time-scales are usually associated with the local equilibrium processes (in the reaction coordinates). Often, by assuming local equilibriums fulfilled, it is possible to decouple these processes from the whole mechanism and to reduce the model size and stiffness. This time-scale separation allows identification of a 'slow manifold' from the large gap in the eigenvalues of the liniarized system Jacobian $\boldsymbol{J}$. The slow manifold describes the long-term system behaviour once the fast *modes* (i.e. eigenvectors of $\boldsymbol{J}$ corresponding to large λ_i) have died away. A $\tau_{\min}$ threshold (of ca. 1s in cell systems[17]) is usually used to separate species subjected to QSSA. Other model reduction variants, such as singular perturbation method, slow manifold approach or approximate lumping in systems with time-scale separation can be applied to decouple fast processes, and to investigate the low-dimensional, quasi-invariant, inertial manifold.[4] A better and slow/fast manifold separation and detection of possible inversions during the system evolution towards the equilibrium points/stable limit cycle have been reported recently.[51–53] The variable time-scale hierarchy is proved to depend not only on the model structure and its parameters, but also on the process operating conditions (system environment) and operating time. The system global properties are determinant, determining the structure of slow manifolds and prevailing over the local system behaviour.

Application of lumping methods to stiff cellular kinetic systems, such as QSSA, pre-equilibrium assumptions, model quasi-linearizations/modal analysis[17] must be made with caution, especially when modelling fast dynamic metabolic steps, regulatory functions, system response to dynamic perturbations or cell signalling processes for which the lumped model structure is partly known. Importance of individual fast equilibriums and intermediates has to be separately checked, and approximate lumping in system variables has to be based on slow sub-spaces and lumping of some intermediates that lead to an acceptable loss of information about the system dynamics.

The wide-separation of time constants in cell systems is called *time hierarchy*. Hierarchic organisation (structural, functional, and temporal) is a characteristic of the living matter in general. Only fast and slow modes are of interest, while the very slow processes are neglected or treated as parame-

ters (such as the external nutrient or metabolite evolution). Aggregate pools (combining fast reactions) are used in dynamic models in a way that they are produced and consumed only by irreversible reactions.[17] Besides, application of lumping rules to metabolic processes must also account for physical significance, species interactions, and *systemic* properties of the metabolic pathway. The only separation of components and reactions based on the time-constant scale (as in the modal analysis of the $\boldsymbol{J}$ matrix case) has been proved to be insufficient.

Lumped GRN models

A large variety of lumped dynamic models have been developed to represent various complex cell processes.[1,34] A worthy route to develop reduced models is to base the analysis on the concepts of '*reverse engineering*' and '*integrative understanding'* of the cell system.[34,35] Such a rule allows disassembling the whole system in parts (modules) and then, by performing tests and suitable numerical/sensitivity analysis, to define rules that allow recreation of the whole and its characteristics, reproducing the real system. Such an approach, combined with derivation of lumped modules, allows reducing the model complexity by relating the cell response to certain perturbations to the response of few inner regulatory loops instead of the response of thousands of gene expression and metabolic circuits.

One remarkable application of the reverse engineering and lumping techniques is the modular construction of GRN models. Such dynamic models (Boolean, continuous, or stochastic), of adjustable size in accordance to the available experimental information, are used 'to divide' the complex gene circuits in sub-systems (modules) of a more tractable size. By representing the transcriptional mechanism and gene interactions, the architecture of the cell regulatory network is related to the physiological characteristics of the organism.[34,38] Semi-autonomous lumped modules are elaborated for representing various regulatory units used in protein synthesis, and then linked to efficiently cope with cell perturbations, and to ensure an equilibrated growth during the cell cycle, with an optimised resource consumption (substrate, metabolic energy). Besides, the gene expression multi-cascade control presents a monotonous response that implies an intrinsic system modularity.[38] This approach allows reduction of the analysis complexity by investigating individual modules, and then their relation to the holistic cell properties.

Due to the large size of the identification problem and the time-related data type, co-regulated genes are clustered together, and (un-)structured model is derived. In the *structured* alternative,[35] a three-level control is defined for subcellular structure, nuclear connectivity and dynamic interactions by the standard kinetic model:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{g}(\boldsymbol{u}(t),\boldsymbol{x}(t),p_{\mathrm{g}}),\ \text{(in nucleus)}$$
$$\boldsymbol{u}(t) = \boldsymbol{h}(\boldsymbol{x}(t),\boldsymbol{v}(t),p_{\mathrm{h}}),\ \text{(in sub-cellular units)} \tag{8}$$

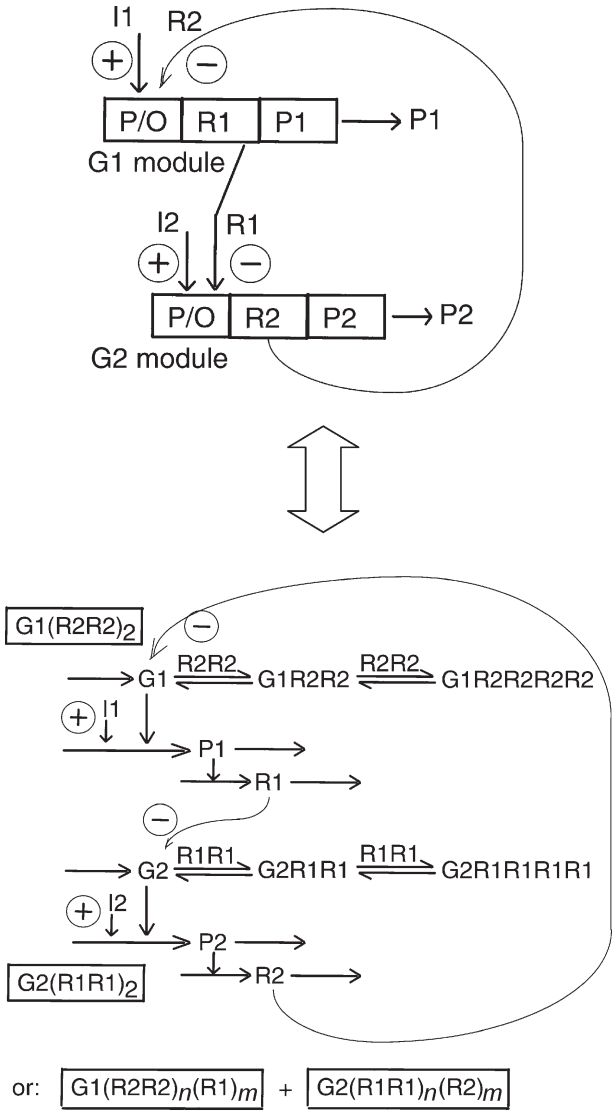
[where $\boldsymbol{v}$ = vector of external variables (stimuli); $\boldsymbol{x}$ = state vector (usually species concentrations, mRNA-levels); $\boldsymbol{u}$ = gene transcription factors (TFs); $\boldsymbol{p}$ = model parameters; t = time]. Due to the GRN system complexity [$p_{\mathrm{g}} \sim O(10^3–10^4)$, number of states and $p_{\mathrm{g}} \cdot p_{\mathrm{t}}$ parameters; p_{t} = number of TFs; $p_{\mathrm{t}} \sim O(10^3)$] and due to few available data, modular approach is used as a valuable modelling technique in various lumping alternatives: gene clustering, minimum number of gene interactions, structure reduction based on various system constraints (stability, sensitivity, multiplicity).[34,35]

A potential application of lumped modular GRN models is the so-called 'genetic circuit engineering', by which simulation of gene expression is used to *in-silico* design organisms that possess specific and desired functions.[36,37] By inserting the new GRNs into organisms, one can create a large variety of mini-functions/tasks (or desired 'motifs') in response to external stimuli. The induced functions in gene circuits are diverse, such as: switches (decision-making branch points between on/off states according to the presence of inducers), oscillators (cell systems evolving among two or several quasi-steady-states), signal / external stimuli amplifiers, amplitude filters, genetic 'memory' storage. The genetic components may be considered as 'building blocks' because they can be extracted, replicated, altered, and spliced into new biological organisms.

By combining the induced motifs in modified cells one can create potent applications in industrial and medical fields, e.g. the production of biosensors used in medicine or for environmental engineering applications. The design of modular GRN construction must present some characteristics:[37] a tight control of gene expression (i.e. low-expression in the absence of inducers and accelerated expression in the presence of specific external signals); a quick dynamic response and high sensitivity to specific inducers; gene circuit robustness (low sensitivity) vs. undesired inducers (external noise).

As an example, *Kaznessis*[36,37] designed a bistable switch genetic circuit, by using two gene modules extracted from the *lac operon* of *E. coli*. The transcriptional regulation is modelled by using

a stochastic approach accounting for 40 reactions and 27 species (reduced model) or 70 reactions and 50 species (extended model). Such a regulatory scheme (Fig. 1), including dimeric repressors (RR) and mutual repression following the presence in excess of one of the activating inducers (I), can be illustrated by means of the conventional representation of *Yang* et al.[26] and *Maria*[1] (see below paragraph). Several lumped alternatives are possible for the two linked modules: simple mutual repression of the gene expression, $G1(R2R2)_2 + G2(R1R1)_2$ (G = gene; R = repressor); combined self- and mutual-repression, $G1(R2R2)_n(R1)_m + G2(R1R1)_n(R2)_m$ (with n = 1–4; m = 1–2). The advantage of such a modular approach is the possibility to adapt the model size to the available information.



F i g . 1 – *The bistable switch circuit of two gene regulatory lumped modules designed by Salis & Kaznessis[36,37] (up) in the conventional representation of Yang et al.[26] and Maria[1] (down). Notations: P = promoter or protein, G = gene; O = operators; I = inducer, R = repressor.*

When linking modules, various system properties have to be ensured in terms of metabolic efficiency (minimum energy and substrate consumption), individual or associative component functions of components, hierarchic organization, system homeostasis, equilibrated cell growth, minimum intermediate levels, etc. Several linking rules have been advanced such as:[1,27] linking reactions between modules must be set slower comparatively to the module core reactions; cooperative/mutual catalysis scheme must avoid internal competition among components displaying similar functions (referring to the same substrate or synthesis reactions); individualized functions must be set to the components into the cell; intermediate species levels and allosteric regulation loops must be adjusted accordingly to the GRN size; use variable cell-volume environment must be considered for an adequate representation of the secondary connectivity effects (cell 'ballast' and 'inertia'). According to the intermediate role and specific synthesis scheme, temporal hierarchy and event succession into the cell can also be accounted for. In such large GRN models, unconventional lumping rules are based on a modular representation combined with a sensitivity analysis to relate the GRN holistic properties to the structure, function and efficiency of certain individual modules to cope with perturbations.[1,39]

The difficulty to precise the very large number of parameters in complex GRNs leads to inclusion of lumped unstructured representations of rate expressions of power-law (S-system)[40,41] or hyperbolic type,[42] explicitly accounting for the activator/repressors influence on the individual operon activity. Even if resulted fractional orders of reactions produce a biased representation of the real process, promising practical implementations are reported, being able to simulate cell system multi-stability, bifurcations, oscillatory behaviour, and hysteresis.[38,41] Various criteria to define the modular system functional effectiveness have been defined (in terms of stability, responsiveness, selectivity, robustness, efficiency)[43] while multi-objective criteria allow identification and optimization of GRNs (in terms of gene connectivity, stability, redundancy, robustness/low sensitivity vs. external noise and high regulatory performance/response rate and overshoot).[1,44,45] Alternative lumped modular GRN structures are discriminated based on the system constraints, experimental observation, physical meaning of lumped components and reactions.[45]

Lumped modular approach is also a common approach in developing topological models for representing GRN and metabolic networks. Thus, metabolic reaction modules are identified based on the topological measures characterising the complex reaction pathway graph: geodesic distance, graph ec-

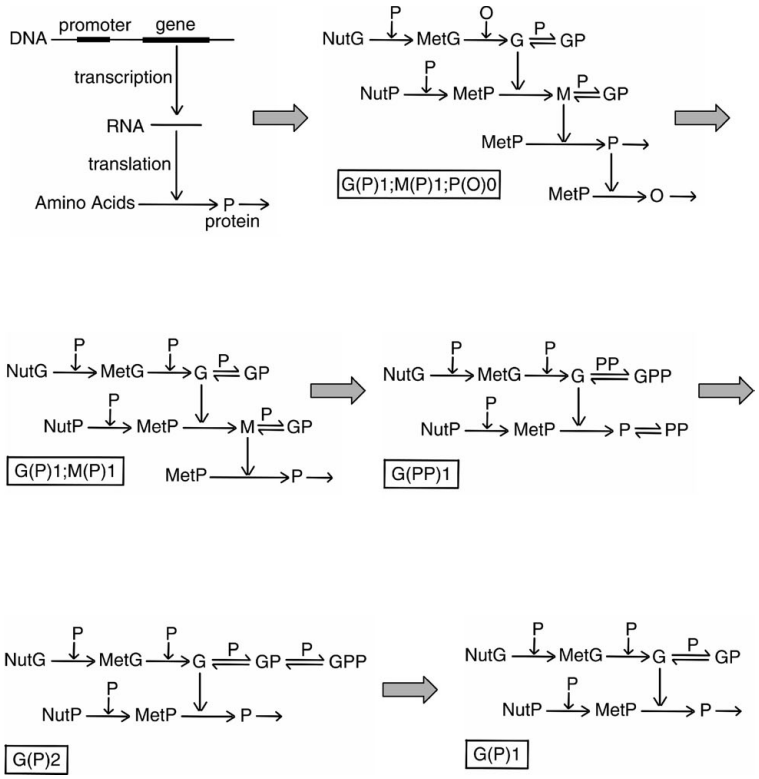
centricity, graph diameter (larger distance between any two vertices), node centrality, robustness capacity, fragility vs. removal of certain nodes, and modularity coefficient (i.e. the quality of the path decomposition).[46] In such a manner, efficient rules to decompose metabolic network into functional modules, based on the global connectivity indices, have been reported.[47] The same rules have been successfully applied to analyse the hierarchical structure of the GRN, for instance in *E. coli*[48] or yeast,[49] pointing-out that most of the gene expressions are regulated by two or more interacting feed forwards loops or by even more complicated GRN motifs together with TFs. The GRN presents a multi-layer hierarchical structure inferring global regulators with network motifs.

Case study 1: Linked regulatory modules used to simulate protein synthesis

Protein synthesis regulation in living cells is a very complex process ensuring a balanced and flexible cell growth under a wide range of environmental conditions. The process is only partially understood, but a multi-cascade control with negative feedback loops seem to be the key elements of gene expression (review of Maria[1]). The activity of enzymes that catalyse the synthesis is allosterically regulated by means of positive or negative 'effector' molecules, via fast reversible 'buffering' reactions, while a cooperative binding of regulation factors in a structured cascade regulation scheme amplify the effect of a change/signal from inner cell or from environment. Inter-connected proteins and proteic complexes act as 'nodes' of a regulatory network that provides a balanced response to perturbations, promotes a catalytically efficient sequence of reactions, 'channels intermediate metabolites', and ensures vital basic functions of the cell (as permease, polymerase, metabolase). Hierarchically structured regulatory networks adjust the metabolic synthesis to maintain homeostasis, i.e. the quasi-invariance of key species concentrations (enzymes, proteins, metabolites), despite external perturbations (in nutrients and metabolites) or internal cell changes. Feedback in gene transcription, metabolic pathways, signal transduction and other species interactions complete the cell regulatory mechanisms.

Among various alternatives used to model the complex protein synthesis and GRNs, the modular approach mentioned in the previous paragraph allows reducing the model complexity by connecting the GRN response to perturbations with the regulatory loops inside the sub-units (modules) and linking elements (reaction, intermediates). Such modules are linked in chains, thus forming hierarchized regulatory networks able to efficiently cope with cell perturbations and to ensure the equilibrated growth during the cell cycle.[25,26] This approach presents several advantages, allowing:[1] i) the analysis of individual modules (of limited types) and check for their regulatory efficiency in conditions that mimic the stationary and perturbed cell growth; ii) investigation of module linking rules used to build-up GRNs of optimised efficiency that ensure system homeostasis and holistic properties (cell functions) with a minimum consumption of metabolic energy; iii) elaboration of whole-cell models, accounting for separate cell sub-units and metabolic functional modules, organized in structured simulation platforms.

Cell regulatory modules are based on semi-autonomous groups of reactions and species, i.e. functional units generating the identifiable cell function. Inherently, any model representation involves a large number of simplifications and lumps in species and reactions. In this category, several types of mechanistic modules have been advanced for representing regulation of the gene expression process.[1,26–28] At a generic level, a regulatory module must include the target protein (P) and its encoding gene (G), the metabolite lumps involved in their synthesis (MetP, MetG), regulatory effectors (R), and other intermediates. Different degrees of simplification are assumed, but the species and reaction lumps must ensure the transcriptional and translational control of the P synthesis, a certain regulatory effectiveness vs. (internal & external) stationary and dynamic perturbations, minimum intermediate levels, system homeostasis, and proteic functions. Optimization of such a regulator modules is limited by the model structure. To manipulate easier various types of modules, *Yang* et al.[26] propose a nomenclature of type $[[L_1(R_1)n_1;\ldots;L_i(R_i)n_i]$, including assembled regulatory units $L_i(R_i)n_i$. One unit i is formed by the component L_i (e.g. enzymes or even genes G, P, M = mRNA, etc.) at which regulatory element acts, and $n_i = 0,1,2,\ldots$ number of 'effector' species R_i (i.e. P, PP, PPPP, etc.) binding the 'catalyst' L_i of the regulated reaction. For instance, in Fig. 2 the simplest G(P)1 module, describing regulation of the P/G pair synthesis, includes one binding step of 'catalyst' G with the synthesis product P (which here play also the role of 'effector'). The intermediate species GP, produced by the rapid buffering reaction G + P ⇔ GP, is inactive catalytically, while the mass conservation law is all time fulfilled, i.e. $\sum_{i=0}^{1} G(P_i)$ = constant. As proved, the maximum regulatory efficiency of the P-synthesis at steady-state (QSS, in-



F i g . 2 – *Examples of lumped regulatory modules for protein P synthesis (horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; G = gene encoding P; M = mRNA; NutP, NutG = nutrients; MetP, MetG = metabolites; GP, GPP, MP = catalytically inactive species; O = DNA polymerase).*

dex 's') corresponds to $[G]_s/[G]_{total} = 1/2$, when the maximum regulation sensitivity vs. perturbations in $[P]_s$ is reached.[27] Subsequent allosterically control of G activity leading to inactive species [GP*n*], with usually $n \geq 5$, amplifies the regulatory efficiency and performance index (P.I.) of the module, usually with a 1.3–2 multiplicative factor for every new added unit/buffering reaction. Such a representation must also account for the protein concentration diminishment due to the cell-growth dilution effect, but can also include the protein degradation by proteolysis.

Among homeostatic regulatory modules, the G(P)*n* units are less realistic but computationally attractive when simulation of large systems are required. The G(PP)*n* units represent better the 'real' regulatory loops in which multiple copies of effectors bind to promoter site of the DNA. Moreover, the low-level of dimer PP, produced by rapid reversible reactions, increases the system efficiency at the expense of a low metabolic "effort". Modules of type [G(P)*n*;M(P)*n′*] (e.g. G(P)1;M(P)1 in Fig. 2), account for the two-step transcription/translation cascade reactions, each of them controlled allosterically. The control in cascade amplifies the response/stimuli ratio. Other regulatory modules, of higher complexity, can also be approached (e.g. G(PP)*n*;M(PP)*n′*, G(RR)*n*;M(RR)*n′*, etc.) but the identification of parameters for every module from the regulatory chain leads to a power-law increase in the computational effort.

When evaluating the regulatory efficiency, beside the module complexity (expressed as the number of involved species n_s and involved reactions n_r), various quantitative measures have been proposed in the literature:[1] stationary regulation effectiveness vs. stationary perturbations of QSS; dynamic regulation effectiveness vs. dynamic perturbations of QSS; regulatory robustness; species interconnectivity; QSS stability strength. It is proved that regulatory performance indices P.I. increase with the regulatory scheme complexity (in the order G(P)*n* < G(PP)*n* < G(P)*n*;M(P)*n*' < …), and linearly increase with the number of effectors in the allosteric control of the catalyst activity at every cascade level.[1]

In order to exemplify application of lumping rules to derive regulatory kinetic modules and to investigate the lumping effect on the module effectiveness, a variable cell-volume model with continuous variables has been adopted.[1] The model mass balance equations, conservation relationships, and the isotonic osmolarity hypothesis are presented in Table 1. The model parameters, i.e. rate constants $\boldsymbol{k}$ and unobservable stationary concentrations $\tilde{\boldsymbol{c}}_s$, can be estimated from QSS mass balance equations and by using non-conventional estimators that optimise global cell regulatory properties, such as:[1,44,45] maximum recovering rate after a dynamic perturbation,[26] smallest QSS sensitivity to inner/external perturbations,[28] stability strength of QSS,[28,50] oscillatory properties and system flexibility.[29,30] *Maria*[1] proposes optimization of a module P.I.-index to adjust $\tilde{\boldsymbol{c}}_s$ of some intermediates under restrictions of QSS mass balance ($\boldsymbol{g}(\boldsymbol{c}_s, \boldsymbol{k}) = 0$), catalyst conservation (active/inactive forms), isotonic osmolarity and maximum regulator sensitivity vs. perturbations (i.e. $[L]_{active}/[L]_{total} = ½$).[27] Dissociation rate constants of catalytically inactive forms in reversible buffering reactions have been adopted at values much higher than the dilution rate ($k_{diss} \approx 10^7\ D_s$).[1,26]

A hypothetical cell, similar to *Escherichia coli*, has been considered in an equilibrated growth of constant Log-growing rate D_s. Cell characteristics and QSS species concentrations are presented in Table 2. Environmental conditions are considered constant over a cell cycle, with lumped nutrient

Table 1 – *Variable cell-volume dynamic model and its basic hypotheses*[1]

Mass balance and state equations	Remarks
$\frac{dc_j}{dt} = \frac{1}{V}\frac{dn_j}{dt} - Dc_j = g_j(\boldsymbol{c},\boldsymbol{k});$ $\frac{1}{V}\frac{dn_j}{dt} = r_j(\boldsymbol{c},\boldsymbol{k}); \quad j = 1,\ldots,n_s;$	continuous variable dynamic model representing the cell growing phase (ca. 80% of the cell cycle)
$V(t) = \frac{RT}{\pi}\sum_{j=1}^{n_s} n_j(t);$	Pfeffer's law in diluted solutions[33]
$D = \frac{1}{V}\frac{dV}{dt} = \left(\frac{RT}{\pi}\right)\sum_{j}^{n_s}\left(\frac{1}{V}\frac{dn}{dt}\right);$	D = cell content dilution rate = cell volume logarithmic growing rate
$\frac{RT}{\pi} = \frac{V}{\sum_{j=1}^{n_s} n_j} = \frac{1}{\sum_{j=1}^{n_s} c_j} = \frac{1}{\sum_{j=1}^{n_s} c_{jo}} = \text{constant.}$	constant osmotic pressure constraint
$\left(\sum_{j}^{all} c_j\right)_{cyt} = \left(\sum_{j}^{all} c_j\right)_{env}$	isotonic osmolarity constraint

Hypotheses:
– negligible inner-cell gradients;
– open cell system of uniform content;
– semi-permeable membrane, of negligible volume and resistance to nutrient diffusion, following the cell growing dynamics;
– constant osmotic pressure, ensuring the membrane integrity ($\pi_{cyt} = \pi_{env}$ = constant);
– nutrient and overall environment concentration remain unchanged over a cell cycle;
– logarithmic growing rate of average $D_s = \ln(2)/t_c$; volume growth of $V = V_o e^{D_s t}$;
– homeostatic stationary growth of $(dc_j/d/t)_s = g_j(c_s,\boldsymbol{k}) = 0$;
– perturbations in cell volume are induced by variations in species copynumbers under the isotonic osmolarity constraint: $V_{perturb}/V = (\Sigma n_j)_{perturb}/(\Sigma n_j)$.

Table 2 – *Nominal stationary conditions (index 's') for the considered E. coli cell*

Symbol	Significance	Value
$v_{cyt,o}$	initial cell (cytoplasma) volume	$1.66 \cdot 10^{-15}$ L
t_c	cell cycle period	100 min
D_s	cell-volume stationary logarithmic growing rate	ln(2)/100 min^{-1}
$c_{NutG,s}$	nutrient for G-production (concentration in the environment)	$3 \cdot 10^6$ nmol L^{-1}
$c_{NutP,s}$	nutrient for P production (concentration in the environment)	$3 \cdot 10^8$ nmol L^{-1}
$\Sigma_j c_{MetP_j,s}$	overall metabolite concentration for P-production; $(c_{MetP_{j,s}})_{per\ module} = (\Sigma_j c_{MetP_j,s})/(\text{no. of modules})$	$3 \cdot 10^8$ nmol L^{-1}
$\Sigma_j c_{MetG_j,s}$	overall metabolite concentration for G production (derived from the restriction: $\sum_{j}^{all} c_{j,cyt} = \sum_{j}^{all} c_{j,env}$); $(c_{MetG_{j,s}})_{per\ module} = (\Sigma_j c_{MetG_j,s})/(\text{no. of modules})$	~10^6 nmol L^{-1}
$c_{P1,s}$, $c_{01,s}$	protein P1 or lump O1 concentration (in module 1)	10^3 nmol L^{-1}
$c_{P2,s}$	protein P2 concentration (in module 2)	10^2 nmol L^{-1}
c_{G1s}, c_{G2s}	gene G1 (in module 1) or G2 (in module 2) concentration	½ nmol L^{-1}
c_{M1s}, c_{M2s}	intermediate M1 (in module 1) or M2 (in module 2) concentration	½ nmol L^{-1}
$c_{GjPj,s}$, $c_{GjPjPj,s}$	concentration of catalytically inactive forms of G_j (in module j)	½ or ¼ nmol L^{-1}
$c_{M1P1,s}$, $c_{M2P2,s}$	concentration of catalytically inactive forms of M1 (in module 1) or M2 (in module 2)	½ nmol L^{-1}
$c_{P1P1,s}$, $c_{P2P2,s}$	concentration of active parts of P1 or P2 dimers in buffering reactions	0.01 nmol L^{-1}

concentrations of $c_{NutG,s} = 3 \cdot 10^6$ nmol L^{-1}, $c_{NutP,s} = 3 \cdot 10^8$ nmol L^{-1}. Because only a few individual species are accounted in the model, the cell 'ballast' and 'inertial' effect are mimicked by adopting high levels for metabolite concentrations, i.e. $\sum_j c_{MetP_{j,s}} = 3 \cdot 10^8$ nmol L^{-1}; $\sum_j c_{MetG_{j,s}} \approx 10^6$ nmol L^{-1}.

In the regulatory modules, fast reversible buffering and oligomerization reactions are responsible for fast adjustment of the catalyst activity, intermediate levels, system optimization and its flexibility. However, application of rigorous lumping rules combined with model quasi-linearizations and modal analysis for determining the system invariant sub-spaces, QSSA applied to species with small relaxation times, assumptions of fast equilibrium or pre-equilibrium reactions, all these conventional model reduction strategies are not easy to be implemented in the regulatory module cases due to the systems high complexity. Besides, some intermediate species, of quickly adjustable low-concentrations, cannot be eliminated from the mechanism based on QSSA only when the module lumped structure is not well defined, their production-consumption rate being directly responsible for the dynamic and stationary regulatory characteristics, synthesis path efficiency, but also for some metabolic system global properties (such as an optimised energetic balance when recovering from perturbed QSS). When lumped models of adjustable complexity are checked, the intermediate optimal levels result not from simple application of QSSA but rather from conferring optimal characteristics to the regulatory chain of the metabolic path. For instance, one of the key elements in quick control of the catalyst L activity is fulfilment at QSS of the condition $[L]_{active}/[L]_{total} = ½$ and of a large sensitivity of this ratio vs. perturbations.[27] Inactive forms of the catalyst, of quickly adjustable levels, are connected to the oligomeric intermediates and effector species, of levels adjusted according to the system global properties and to the proposed lumping level.

Lumping in modelling the gene expression must preserve the essential regulatory characteristics even if an important loss of information on certain species and side-reactions dynamics is inherent. Because the elementary reaction scheme and stoichiometry are only partly known in cell regulatory schemes, application of species and reaction lumping rules are rather based on the sensitivity of the system P.I.s vs. module structure (i.e type and number of effectors and cascade control levels) and module interactions. Even if unknown (or unidentifiable) parts of the mechanism are ignored and/or included in the lumps, model reduction must preserve an acceptable predictability for key-species homeostatic levels, functions and cell systemic properties (structural, functional, and temporal hierarchy).

The gene expression can be represented by regulatory modules of variate complexity according to the accepted trade-off between the model simplicity, its estimability (vs. available information), computing tractability and predictive quality. Simplified models are desired when simulating complex regulatory networks including hundreds or thousands of protein synthesis modules, but the increase of bias from the real process characteristics can lead to an unsatisfactory representation of the regulatory properties of the network. To exemplify how the lumping is related to the individual module P.I.s, the analysis starts with a simplified representation of the gene expression, of the form [G(P)1;M(P)1;P(O)0] (see Fig. 2), i.e. including one buffering reaction for every control level. Because only one module is analysed in the hypothetical cell, the protein P (i.e. the product) is assumed to perform several functions: the effector that dynamically adjust the catalysts activity (G,M); the catalyst for synthesis of a lump species O (polymerase) responsible for encoding gene synthesis; the permease that catalyses the import of NutG and NutP from environment; the metabolase that converts nutrients into metabolites MetG and MetP. The result is that G and P syntheses are mutually autocatalytic. When coupling several modules in a regulatory chain, individual proteins exhibit different or similar functions contributing to gene expression optimization in an interconnected reaction schema. Otherwise, uncooperative and non-interacting module linking may lead to a degenerated (not-viable) cell system.[1]

Even if the chosen regulatory module is of very simplified form, successive application of a lumping rule can lead to even more simplified modular structures, such as (Figure 2): [G(P)1;M(P)1] (with P becoming polymerase); [G(PP)1] (with no cascade control, but with a dimeric effector); [G(P)2] (with P effector and two buffering reactions); [G(P)1] (with P effector and one buffering reaction). To determine the lumping effect on module P.I., several effectiveness indices have been evaluated under a specified QSS of Table 2, for a hypothetical cell similar to *E. coli* (see also the review on P.I.s of *Maria*[1]):

– dynamic efficiency vs. individual species, i.e. the recovering time-constant τ_j necessary to species j to return to its stationary c_{js} after a "standard" dynamic (impulse-like) perturbation; evaluation is made by simulating the $c_j(t)$ recovering path until a 1% c_{js} tolerance is reached after a ±10% $c_{P1,s}$ impulse perturbation;

– overall dynamic efficiency of the module, i.e. the average of species recovering times, AVG(τ_j);

– the standard deviation of species recovering times, i.e. STD(τ_j), related to the overall species interconnectivity and mutual assistance/synchronisation during the QSS recover;

– stationary efficiency vs. individual species, i.e. the species c_{js} absolute sensitivity to $c_{\mathrm{NutP,s}}$ changes in the environment $[S^{j}_{\mathrm{NutP}} = (\partial c_j / \partial c_{\mathrm{NutP}})_{\mathrm{s}}]$; this P.I., related to the module efficiency to cope with stationary perturbations, is evaluated by solving the stationary set of sensitivity equations derived from differentiation of QSS mass balances:[1]

$$[\partial \boldsymbol{g}/\partial \boldsymbol{c}]_{\mathrm{s}}[\partial \boldsymbol{c}/\partial c_{\mathrm{NutP}}]_{\mathrm{s}} + [\partial \boldsymbol{g}/\partial c_{\mathrm{NutP}}]_{\mathrm{s}} = 0. \quad (9)$$

By comparing the dynamic and stationary P.I.s of each module of Fig. 2, it appears that an increase in the module complexity leads to a diminishment of τ_{P1} (Fig. 3 and 6), revealing better regulatory efficiency of the cascade control for the P-synthesis. When only an individual module is analysed, is more difficult to ensure species connectivity and their quick recovering in a non-scattered (mutual-assisted) way. Thus, by keeping only two buffering reactions, the overall dynamic-P.I. declines with the increase of n_{s} due to the increase in the AVG(τ_j) and STD(τ_j). Oppositely, the stationary-P.I. seems to be less affected by the increase in module complexity, as revealed by the quasi-uniform s^{j}_{NutP} in Fig. 6.

In fact, the superiority of complex vs. lumped regulatory schemes concerning realized P.I.s can be proved when linking P-synthesis modules in regulatory chains, placed in a variable cell volume to which all species contribute. The link strategy ensures the whole system homeostasis, optimum regulatory properties of the whole chain, and distinct functions for each protein in the cell (see the linking rules presented in the 'Lumped GRN models' paragraph). The cooperative link assumes that each protein performs a certain function in the profit of one or several modules, avoiding internal competition among components displaying similar functions. For instance, in a two-module system P1 is a permease and metabolase, while P2 is a polymerase. In a hypothetic three-module chain,[1] P1 and P3 are permeases and metabolases for respectively NutG and NutP import, while P2 is kept as a polymerase. The same rule can be applied to every added module, leading to a large modular chain to be included in the cell regulatory network. The superiority of complex regulatory modules, that includes efficient regulatory units and a cascade con-

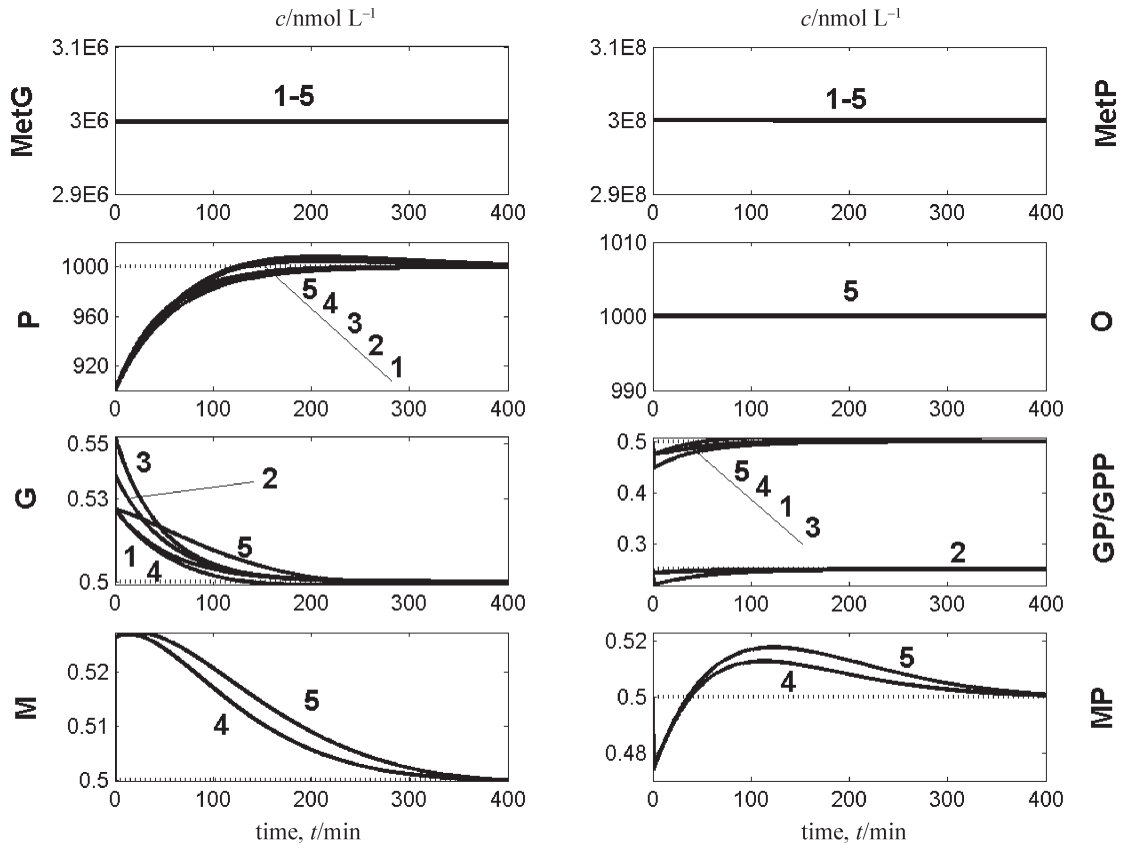


Fig. 3 – *Comparative cell components recover after a dynamic (impulse) perturbation in $[P]_s$ for modules G(P)1 (curves 1), G(P)2 (curves 2), G(PP)1 (curves 3), G(P)1;M(P)1 (curves 4), G(P)1;M(P)1;P(O)0 (curves 5). The –10 % impulse perturbation in [P]s = 1000 nmol L^{-1} was applied at t = 0.*

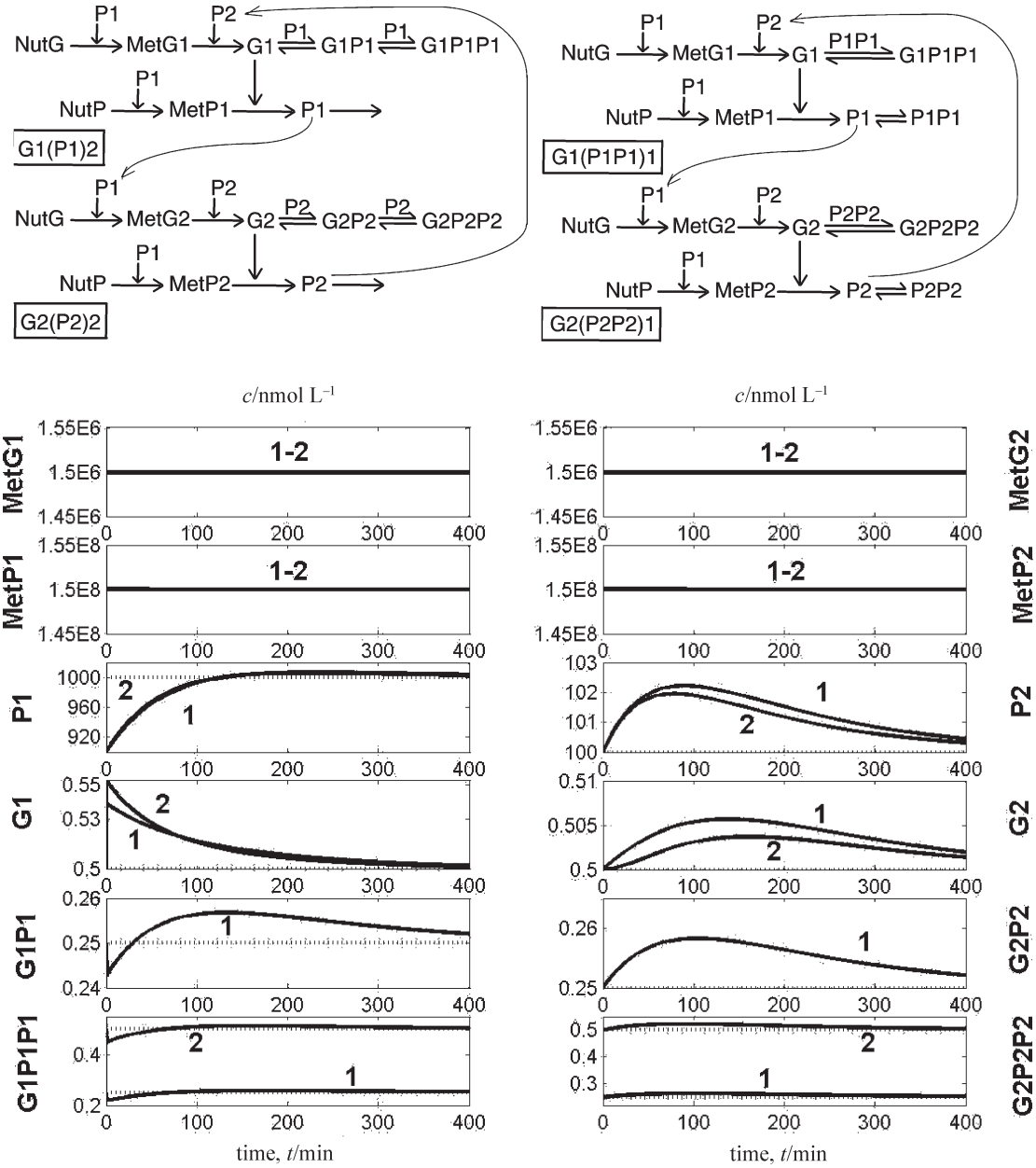


Fig. 4 – *(Up) Examples of two-coupled regulatory modules. (Down) Cell components recover after a –10 % impulse perturbation in [P1]s = 1000 nmol L^{-1} applied at t = 0 for coupled [G1(P1)2] + [G2(P2)2] (curves 1) comparatively to the coupled [G1(P1P1)1] + [G2(P2P2)1] (curves 2).*

trol, can be simply exemplified by comparing two linked modules [G1(P1P1)1 + G2(P2P2)1] vs. the [G1(P1)1 + G2(P2)1] system. As proved by *Maria*[1], the use of dimeric PP as transcription factors instead of simple P is closer to real situation and improves P.I.s due to their adjustable low-levels in buffering reactions. The result is improvement in τ_{P1} and STD(τ_j) with 15%, and in AVG(τ_j) with 26 %.[1] Similarly, the cooperative link of two modules [G1(P1)2 + G2(P2)2], compared to the link [G1(P1P1)1 + G2(P2P2)1] proves that, in spite of two buffering reactions in the G(P)2 unit, the use of PP dimmers in one buffering reaction unit G(PP)1 leads to better dynamic-P.I., i.e. lower τ_{P1} AVG(τ_j) and STD(τ_j)(see results from Fig. 4 and 7). The stationary-P.I. (S^{j}_{NutP}) are practically unchanged, while the module complexity is comparable (n_s = 12 and n_r = 8 for both systems). The low-concentrations of the oligomeric effectors (of type PP, PPP, …) are determined not by a QSSA but from optimising the global properties of the overall modular regulatory chain.

When the cooperative link is realized by using more complex structures, as for instance [G1(P1)1;M1(P1)1 + G2(P2)1;M2(P2)1] accounting for a cascade control (n_s = 14), an improvement in P.I.s is expected. Indeed, by comparing this as-

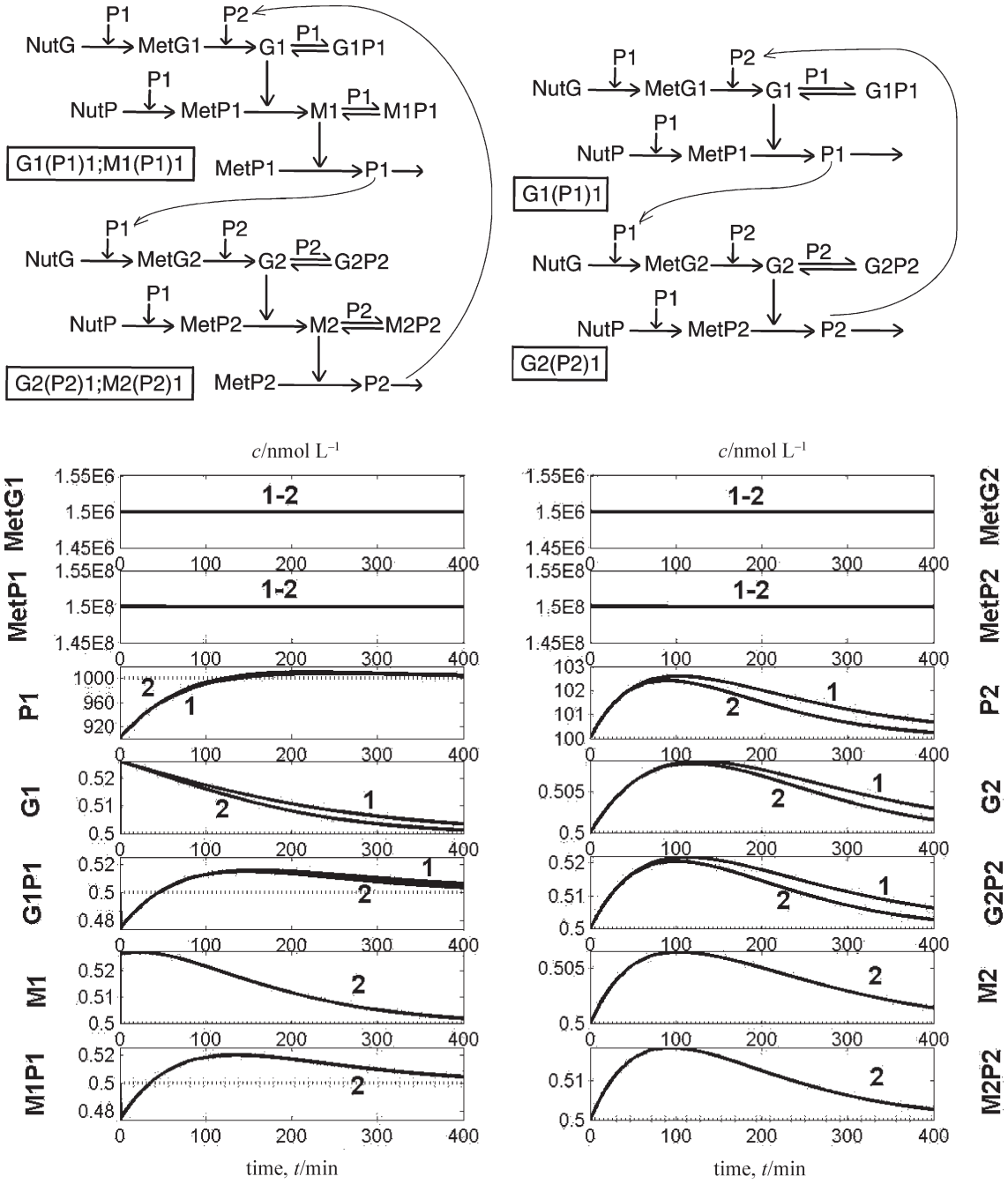
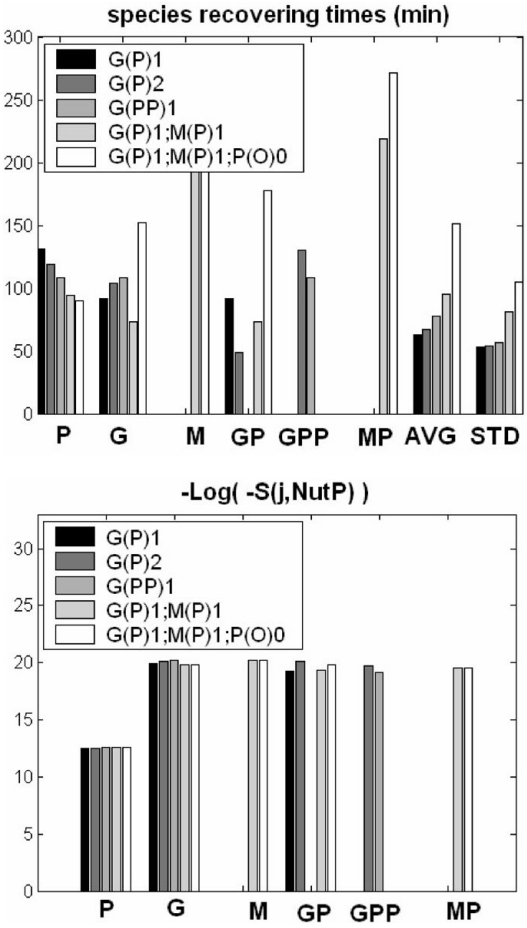
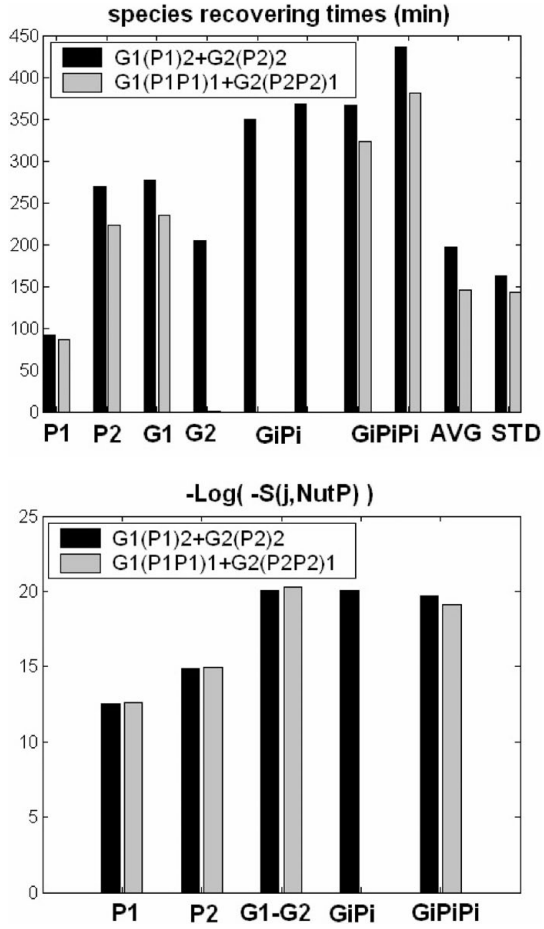


Fig. 5 – *(Up) Examples of two-coupled regulatory modules. (Down) Cell components recover after a –10 % impulse perturbation in [P1]s = 1000 nmol l^{-1} applied at t = 0 for coupled [G1(P1)1] + [G2(P2)1] (curves 1) comparatively to the coupled [G1(P1)1;M1(P1)1] + [G2(P2)1;M2(P2)1] (curves 2).*

sembly to the [G1(P1)1 + G2(P2)1] system (n_s = 10) in Figures 5 and 8, it appears that, in spite of a considerable increase in model complexity, τ_j is lower for all species (except P1), the species interconnectivity index STD(τ_j) is better (i.e. a low value), while QSS-resistance to external perturbations are practically unchanged (i.e. the S^j_{NutP} sensitivities). The identified rate constants of the lumped models are larger than of the extended ones due to the lumped parallel reactions. For instance, AVG($\boldsymbol{k}$) = 3.47 · 10^{-6} and STD($\boldsymbol{k}$) = 3.45 · 10^{-6} for module G(P)1, are larger comparatively to AVG($\boldsymbol{k}$) = 2.31 · 10^{-6} and STD($\boldsymbol{k}$) = 3.26 · 10^{-6} for module G(P)1;M(P)1;P(O)0 (not accounting for fast buffering reactions). Complex regulatory schemes demonstrate superior P.I.s in spite of involvement of more species and reactions. The increased difficulty in synchronizing species response during recovering after a perturbation in QSS is compensated by an increased mutual assistance, and increased system flexibility. These are offered by a larger number of intermediates of quickly adjustable levels, by



F i g . 6 – *Comparative species recovering times (τ_j, min) and stationary sensitivities (s^j_{NutP}) for individual modules: G(P)1 (n_s = 5), G(P)2 (n_s = 6), G(PP)1 (n_s = 6), G(P)1;M(P)1 (n_s = 7), G(P)1;M(P)1;P(O)0 (n_s = 8) [$s^j_{NutP} = (\partial c_j / \partial c_{NutP})_s$ = species j stationary level sensitivity to NutP; τ_j = species j recovering time to QSS, with a 1% tolerance, after a ±10% $c_{PI,s}$ impulse perturbation; AVG= average of τ_j (min); STD = standard deviation of τ_j (min); $s^j_{NutP} < 0$ for all represented species; $\tau_{MetG} = \tau_{MetP} \approx 0$; species PP and O are not included in the bar-plots].*



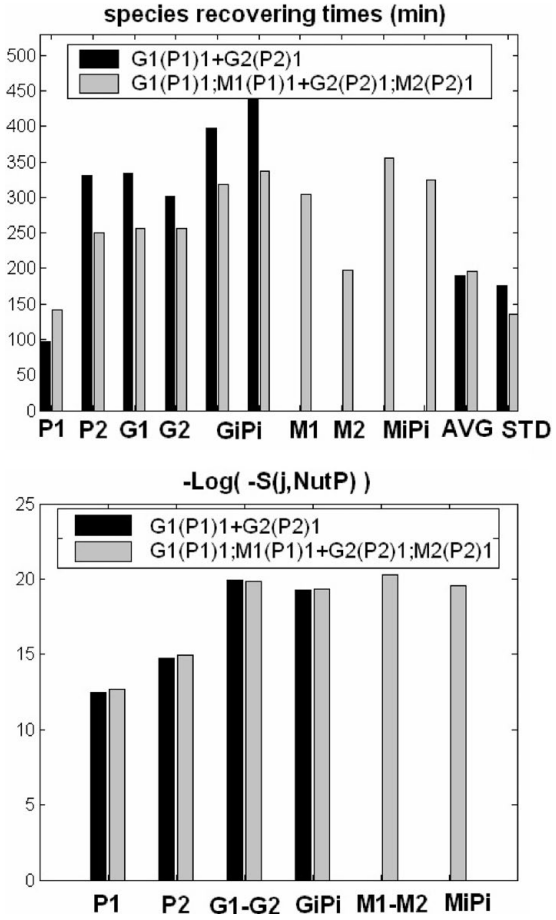
F i g . 7 – *Comparative species recovering times (τ_j, min) and stationary sensitivities (s^j_{NutP}) for coupled modules G1(P1)2+G2(P2)2 (n_s = 12) and G1(P1P1)1+G2(P2P2)1 (n_s = 12) [$s^j_{NutP} = (\partial c_j / \partial c_{NutP})_s$ = species j stationary level sensitivity to NutP; τ_j = species j recovering time to QSS, with a 1% tolerance, after a ±10% $c_{PI,s}$ impulse perturbation; AVG = average of τ_j (min); STD = standard deviation of τ_j (min); $s^j_{NutP} < 0$ for all represented species; $\tau_{MetG} = \tau_{MetP} \approx 0$; species P1P1 and P2P2 are not included in the bar-plots].*

a larger number of effectors at several control levels disposed in cascade, and by an increased number of buffering reactions in the allosteric control of the catalyst activity.[1,26,28]

The model reduction cost in representing cell regulatory modules is a loss of information on certain intermediates and side (buffering) reactions, but also an altered capability to represent the 'real' properties of the modular chain under stationary or dynamic perturbations. By re-allocating species functions to a fewer number of components in the module, a reduced possibility to include some intermediates exhibits a lower system flexibility, robustness, and possibility to represent the cell-content inertial effect in "smoothing" perturbations from the environment (see discussion of *Maria*[1]). Consequently, a combination of sensitivity analysis of the complex GRN properties vs. the characteristics and size of lumped modules can successfully replace the classical theoretical lumping rule, difficult to be applied to large cell dynamic models.

Case study 2: Lumped dynamic models for drug release in human plasma

Modern drugs are designed to realize a "programmed" release of active principles (ligands) in human plasma from a macromolecular support structure, leading to a long-lasting active-drug concentration to the receptor. The successive drug-ligand release process involves chain reactions but also parallel (isomerisation) processes, all making the whole choreography to be complex and difficult to model. However, an increased knowledge of the intrinsic reactions and establishment of direct relations between the elementary kinetics and the drug-ligand/macromolecular-support structure are



F i g . 8 – *Comparative species recovering times (τ_j, min) and stationary sensitivities (s^j_{NutP}) for coupled modules G1(P1)$_1$+G2(P2)$_1$ (n_s = 10) and G1(P1)$_1$;M1(P1)$_1$+G2(P2)$_1$;M2(P2)$_1$ (n_s = 14) [$s^j_{NutP} = (\partial c_j / \partial c_{NutP})_s$ = species j stationary level sensitivity to NutP; τ_j = species j recovering time to QSS, with a 1% tolerance, after a ± 10% $c_{PI,s}$ impulse perturbation; AVG= average of τ_j, (min); STD= standard deviation of τ_j, (min); $s^j_{NutP} < 0$ for all represented species; $\tau_{MetG} = \tau_{MetP} \approx 0$].

essential aspects to better "design" the release kinetics according to a desired scenario. On the other hand, an apparent (observable and identifiable) kinetics, accounting for only few observed species lumps is ineffective to reveal intrinsic effects. By applying a suitable lumping analysis to an extended reaction scheme, the links between apparent and intrinsic rate and equilibrium constants can be established, revealing unobservable intrinsic characteristics of the process.

To exemplify the lumping analysis, the drug-release process studied by *Zhang* et al.[31] and analysed by *Maria*[32] has been approached. The four hydrophobic dansyl-ligand (L) groups, attached to the support by disulphide bonds (-S-S-), are successively released from a macromolecular dendritic support (based on melamine) in a reducing synthetic medium (including a DTTr agent) that mimics the human plasma conditions. The DTTr mediates the thiol-disulphide exchange, leading to a free thiol on the dendrimeric-support, a free drug-ligand LSH, and the oxidized DTTo (Fig. 9). After each ligand release, a more reduced dendrimer and more hydrophobic environment will alter the release rate of other ligands from the support. Experiments have been repeated under variate environmental conditions, by using various dendrimeric supports and number of ligands, thus allowing evaluation of the drug release comparative effectiveness.[32]

The ligand-release process depends on a large number of variables, including a drug structure, ligand and support nature, environmental characteristics and release conditions (pH, reducing agent properties). In order to experimentally identify the release intermediates for various drug-supports, a low-reducing medium has been checked, leading to a longer release time (ca. 6000–7200 min). Starting from a four-ligands dendritic structure (denoted by A in Fig. 9), a large number of stereoisomers have been identified during the release process, including 3, 2 or 1 SL-groups attached to the support-molecule. Even if cyclic isomers have not been observed, it is quite impossible to record consistent kinetic data on each isomer dynamics during the process. As a consequence, a reduced model accounting for only four reversible reactions and five lumps ($\hat{A},\hat{B},\hat{C},\hat{D},\hat{E}$, Fig. 9) has been proposed and the apparent forward and reverse reaction rates (k_{ij},k_{ji}; $i,j = \hat{A} \div \hat{E}$) together with the apparent equilibrium constants (K_{ij}, $i,j = \hat{A} \div \hat{E}$) have been estimated (Fig. 10). It is to observe that apparent equilibrium constants present quasi-uniform values, with a slight tendency to increase as the dendrimeric support contains less coupled ligands. Such an unexpected result is in contradiction with independent observations that release rate is faster as less ligands are attached to the support. Such a result indicates that the apparent model (not accounting for the support free surface) is imperfect and can not be used to develop quantitative relationships between support-ligand structure and intrinsic release kinetics.

To decipher the real mechanism, the lumping analysis starts from a proposed extended (intrinsic) reaction path involving 16 conformational isomeric intermediates: A, B1, B2, B3, B4, C1, C2, C3, C4, C5, C6, D1, D2, D3, D4, E (Fig. 9). Assuming only pseudo-first order reactions (with quasi-constant [DTTr], [DTTo], [LSH]), the extended schema stoichiometric matrix accounts for 64 reactions and 16 species, resulted from including all reversible release reactions among all dendritic isomers. The stoichiometric analysis identifies 15 independent reactions, 15 independent species, and two steady-state invariant relationships of molecular species conservation.[32]

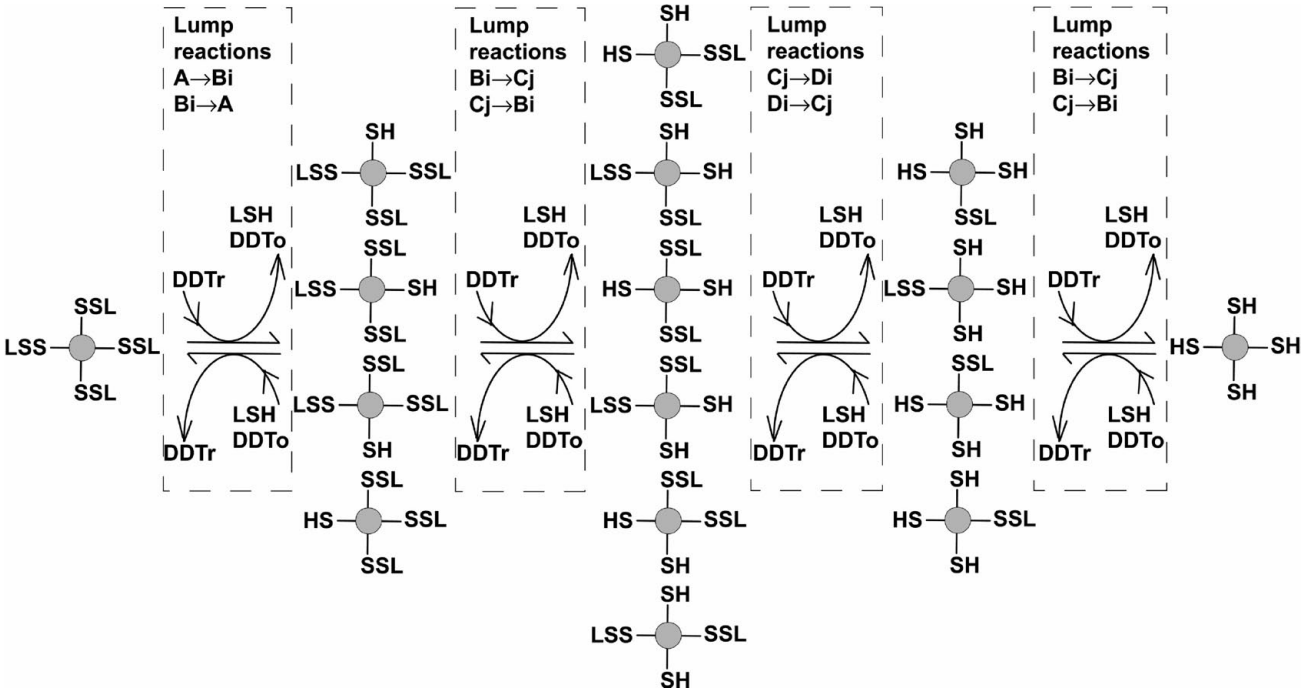


Fig. 9 – *Chain reactions for successive ligand (L) release from a dendrimeric drug support structure in a reducing environment. Kinetic schema accounts for all reversible release reactions for each dendritic stereoisomer (i.e. a total of 64 reactions involving 16 species).*[32]

By applying the exact model lumping methodology, described below eq. (6), the invariant sub-spaces of the system are checked among $\boldsymbol{J}^T$ eigenvectors corresponding to the smallest $\boldsymbol{J}^T$ eigenvalues. Species lumps are also selected based on their physical significance and observability. For simplification, one considers quasi-equal rate constants for synthesis of stereoisomers with the same number of ligands (i.e. $k_{AB_i} = k_{AB}$, $k_{B_iA} = k_{BA}$, $k_{B_iC_j} = k_{BC}$, $k_{C_jB_i} = k_{CB}$, $k_{C_jD_1} = k_{CD}$, $k_{D_1C_j} = k_{DC}$, $k_{D_1E} = k_{DE}$, $k_{ED_1} = k_{ED}$), and all rate constants in $\boldsymbol{J}^T$ are set to 1 (because the "intrinsic" rate constants are àpriori unknown). Such a strategy leads to build-up only five observable lumps: $\hat{A} = A$; $\hat{B} = B1 + B2 + B3 + B4$; $\hat{C} = C1 + C2 + C3 + C4 + C5 + C6$; $\hat{D} = D1 + D2 + D3 + D4$; $\hat{E} = E$. The lumping matrix $\boldsymbol{M}_{5\cdot16}$ allows to reduce the reaction system Jacobian matrix $\boldsymbol{J}^T$ from a [16·16] size to $\hat{\boldsymbol{J}}^T = \boldsymbol{M}\boldsymbol{J}^T\boldsymbol{M}^+$ of dimension [5·5]. The exact lumping is check by the equality of $\boldsymbol{J}^T$ and $\hat{\boldsymbol{J}}^T$ eigenvalues.

The next analysis step consists of determining quantitative link-relations among kinetic terms of the extended and reduced models, by comparing the analytical expressions of the $\hat{\boldsymbol{J}}^T$ elements with the identified reduced Jacobian of the apparent rate model. The identified links are:

$$k_{\hat{A}\hat{B}} = 4k_{AB}; \quad k_{\hat{B}\hat{A}} = k_{BA};$$

$$K_{\hat{A}\hat{B}} = k_{\hat{A}\hat{B}} / k_{\hat{B}\hat{A}} = 4K_{AB};$$

$$\begin{aligned}
k_{\hat{B}\hat{C}} &= 3k_{BC}; \quad k_{\hat{C}\hat{B}} = 2k_{CB}; \\
K_{\hat{B}\hat{C}} &= k_{\hat{B}\hat{C}} / k_{\hat{C}\hat{B}} = 3/2\, K_{BC}; \\
k_{\hat{C}\hat{D}} &= 2k_{CD}; \quad k_{\hat{D}\hat{C}} = 3k_{DC}; \\
K_{\hat{C}\hat{D}} &= k_{\hat{C}\hat{D}} / k_{\hat{D}\hat{C}} = 2/3\, K_{CD}; \\
k_{\hat{D}\hat{E}} &= k_{DE}; \quad k_{\hat{E}\hat{D}} = 4k_{ED}; \\
K_{\hat{D}\hat{E}} &= k_{\hat{D}\hat{E}} / k_{\hat{E}\hat{D}} = 1/4\, K_{DE}.
\end{aligned} \tag{10}$$

By comparing the obtained rate constants in Fig. 10, it is to observe that apparent ones are larger, by including all parallel reversible reactions of each release step and the loss in system diversity (and entropy). On the other hand, the intrinsic equilibrium constants for every release step are much smaller than the lumped equilibrium constants at the beginning of the process (the first and second drug-ligand release; Fig. 10) and much larger at the end of the process (the third and the fourth ligand release from the dendrimeric support). Such a result cannot be deduced from the experimental data directly, and only the numerical lumping analysis can reveal how the successive release equilibriums are favoured by the release progress. Such a conclusion is important for the correct interpretation of the drug ligand-macromolecule bond liability, correctly suggesting that dansyl groups exchange faster as the macromolecular architecture of the drug-support becomes less sterically hindered.[31]

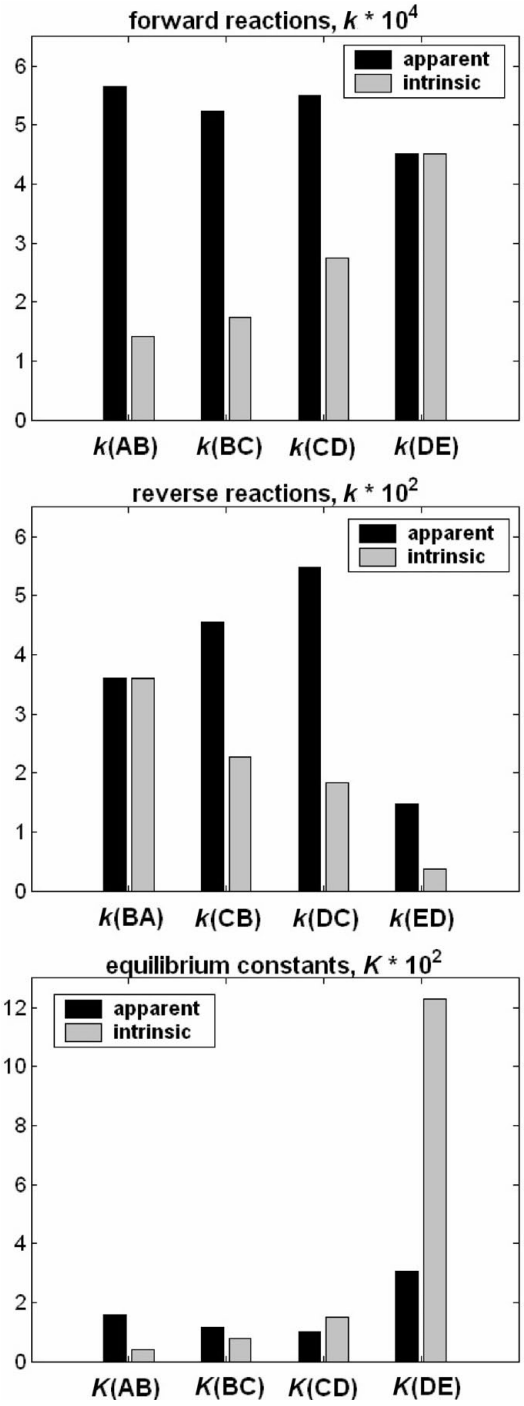


Fig. 10 – *Estimated apparent and intrinsic rate (k) and equilibrium (K) constants for the successive drug release process (units are in min and mmol L^{-1}).*

As a conclusion, application of the lumping analysis to reduce an extended kinetic model presents the advantage of derivation of lumping functions, which connect the apparent (observable) characteristics of the reduced model to the 'intrinsic' (unobservable) properties. The loss of information on steps and intermediates can be compensated if an exact or approximate lumping matrix can be formulated from the liniarized system Jacobian invariants.

Conclusions

Lumping rules are valuable tools to model complex (bio)chemical reacting systems when process low observability, identifiability and estimability are present. Exemplification in the case of modelling the protein synthesis regulatory dynamics reveals how various lumping degrees can be investigated, with the aim of making the modular representation more computationally tractable. However, a continuous model simplification, by replacing species and reactions with few lumps, not only leads to the loss of information on reaction intermediates, species functions, and side-reaction characteristics, but also to diminishment of model prediction capability on the modular chain regulatory characteristics under stationary and dynamic perturbations. An overadvanced model lumping for high sensitive and fast cell processes is proved to have a negative effect, leading to an unrealistic representation of regulatory P.I. and species inter-connectivities, to an altered/unsynchronized dynamic response to perturbations, and to a less flexible / adaptive cell system.

Application of rigorous lumping rules, such as QSSA, pre-equilibrium assumptions, model quasi-linearizations, invariant system sub-space identification to model living cell systems is more difficult, due to the analysed system high complexity, lack of knowledge of the regulatory network, process elementary steps, species functions and inter-connectivities. The lumping analysis is rather focused on principal component and sensitivity analysis of individual modules and on the whole-model construction predictive properties (e.g. network regulatory efficiency, flexibility, degree of inter-connectivity, etc.). Application of lumping rules to metabolic processes must account for physical significance, species interactions and *systemic* properties of the metabolic pathway, than to only separation of components and reactions based on the time-constant scale.

Importance of individual fast equilibriums and intermediates has to be separately checked, and approximate lumping in system variables has to be based on slow sub-spaces presenting the acceptable loss of information about the system dynamics. However, some intermediate species, of quickly adjustable low-concentrations, cannot be eliminated by simple QSSA applied to lumped model of adjustable complexity, their optimal levels resulting from dynamic and stationary optimal regulatory characteristics, synthesis path efficiency, and some global properties of the metabolic system. Model reduction by including in lumps the unknown or unidentifiable parts of the metabolic mechanism, must preserve an acceptable predictability for

key-species homeostatic levels, functions, and cell systemic properties (structural, functional, and temporal hierarchy).

Application of a successive lumping strategy starting from an extended model presents the advantage of better understanding of the intrinsic process dynamics under continuous environmental changes, of the importance of intermediate species and steps, species inter-connectivities, structural and functional hierarchy, multi-cascade control with adjustable intermediate levels and multiple effectors in feedback loops. In such a manner, a trade-off between reduced model simplicity and its predictive quality can be better realized.

When the lumped model components are fully observable, derivation of a reduced and identifiable kinetic model from an extended reaction path, by using a stoichiometric analysis, a lumping strategy, and identifiable species lumps, is of high interest both from theoretical and practical reasons. This is exemplified for the case of modelling the "programmed" release of active drug-principles (ligands) in human plasma from a macromolecular support. The systematic approach leads to evaluate "intrinsic" constants from the estimated apparent rate constants, allowing a deep interpretation of reaction steps, thermodynamic limitations, and relations between drug-support structure and release characteristics.

In general, derivation of relations between apparent and extended model structures is of high interest for a process correct interpretation and to characterise relative importance of various reaction steps. The apparent steps tend to compensate the loss in system diversity introduced by the lumping rule and cannot fully describe the real interactions among reaction intermediates. Because the apparent parameters (identified from experimental data) present values smaller or larger than those corresponding to elementary steps, physical meaning of lumps can also play important role in choosing the most suitable lumping route from the large number offered by the theoretical analysis.

In conclusion, when large cell dynamic models are developed, application of unconventional lumping strategies are recommended by combining suitable system modularisation (in functional sub-units) with application of the sensitivity analysis to relate metabolic network holistic properties (hierarchic organization and regulatory efficiency) to the individual module properties. When reduced reaction pathways are analysed, application of theoretical (exact) lumping rules lead to the possibility to directly relate the apparent (observable) model parameters to the intrinsic process characteristics, thus exhibiting a detailed and correct process interpretation.

ACKNOWLEDGEMENT

The author is grateful to the (late) Prof. Wolf-Dieter Deckwer group of Biochemical Engineering (GBF Braunschweig) and to the Prof. An-Ping Zeng group of Bioinformatics (TU Hamburg-Harburg) for interesting and constructive discussions occasioned by the author's seminars on this subject, on the Summer 2006.

Nomenclature

B_i – species i global sensitivity measure
$\boldsymbol{c}$ – species concentration vector
D – cell-content dilution rate (i.e. cell-volume logarithmic growing rate)
$\boldsymbol{f},\boldsymbol{g},\boldsymbol{h}$ – model function vectors
$\boldsymbol{h}$ – lumping function vector
$\boldsymbol{I}$ – identity matrix
$\boldsymbol{J} = \partial \boldsymbol{f} / \partial \boldsymbol{c}$ – system Jacobian matrix
$\boldsymbol{k}$ – rate coefficient vector
K – equilibrium constant
$\boldsymbol{M}$ – linear lumping matrix
n,n_i – number of runs, or number of species Oi (effectors)
n_r – number of reactions
n_s – number of species
n_p – number of parameters
n_j – species j, amount of substance (moles)
$\boldsymbol{p}$ – dynamic model parameters
$\boldsymbol{r}$ – species reaction rate vector
R – universal gas constant
$\boldsymbol{s}$ – vector of species, or reaction sensitivities
$\boldsymbol{S}$ – species or reaction sensitivity matrix
$S_{\mathrm{g},j}$ – species j global sensitivity
t – time
t_c – cell cycle period
T – absolute temperature
$\boldsymbol{u}$ – gene transcription factors
$\boldsymbol{v}$ – vector of cell external variables (inputs, stimuli)
V – cell volume
$\boldsymbol{x}$ – state variable vector, or independent variable vector
$\boldsymbol{X} = \lfloor \boldsymbol{x}_j \rfloor$ – eigenvectors

Greeks

α – Hoerl factor
Δ – difference operator
λ_i – i-th eigenvalue
π – osmotic pressure
σ^2 – experimental error variance
τ_j – species j recovering time, or characteristic-time

Index

cyt – cytoplasma
diss – dissociation
env – environment
max – maximum
min – minimum
0 – initial
perturb – perturbed
s – (quasi-)steady-state

Superscipts

\+ – psudo-inverse
^ – lumped
~ – unobservable

Abbreviations

AVG – average
DTTr, DTTo – reduced or oxidized forms of the DTT (dithiothreitol)
G – gene (DNA)
GRN – genetic regulatory network
L – ligand, or species at which regulatory element acts
M – mRNA
Met – metabolite
Nut – nutrient
nmol L^{-1} – nano-molar (i.e. 10^{-9} mol L^{-1} concentration)
P – protein
O – effector species, or polymerase
O – order of magnitude
PCA – principal component analysis
PCR – principal component regression
P.I. – regulatory performance index
R – repressor
RSA – ridge selection analysis
QSS – quasi-steady-state
QSSA – quasi-steady-state approximation
Re – real part
STD – standard deviation
TF – transcription factor
size(.) – vector dimension
[.] – concentration

References

1. *Maria, G.,* Chemical and Biochemical Engineering Quarterly **19** (2005a) 213.
2. *Maria, G.,* Chemical and Biochemical Engineering Quarterly **18** (2004) 195.
3. *Gorban, A. N., Karlin, I. V., Zinovyev, A. Y.,* Physica A **333** (2004) 106.
4. *Tomlin, A. S., Turanyi, T., Pilling, M. J.,* Mathematical tools for the construction, investigation and reduction of combustion mechanisms, in: *Pilling, M. J., Hancock, G.* (Eds.), Low temperature combustion and autoignition, Elsevier, Amsterdam, 1997, p. 293–437.
5. *Vajda, S., Rabitz, H., Walter, E., Lecourtier, Y.,* Chem. Eng. Commun. **83** (1989) 191.
6. *Connors, K. A.,* Chemical kinetics: The study of reaction rates in solution, VCH Publ., New York, 1990.
7. *Turanyi, T.,* J. Mathematical Chemistry **5** (1990) 203.
8. *Lei, F., Jorgensen, S. B.,* J. Biotech. **88** (2001) 223.
9. *Vajda, S., Valko, P., Turanyi, T.,* Int. J. Chem. Kinetics **17** (1985) 55.
10. *Malinowski, E. R.,* Factor analysis in chemistry, Wiley, New York, 1991.
11. *Maria, G.,* Canadian J. Chem. Eng. **67** (1989) 825.
12. *Maria, G., Rippin, D. W. T.,* (1993) Chem. Eng. Sci. **48** (1993) 3855.
13. *Le Roux, C., Antonio, G., Joulia, X.,* Récents Progrès en Génie des Procedés (Grenoble) **7** (1993) 43.
14. *Maria, G., Muntean, O.,* Chem. Eng. Sci. **42** (1987) 1451.
15. *Marquardt, D. W.,* Technom. **12** (1970) 591.
16. *Bard, Y.,* Nonlinear parameter estimation, Academic Press, New York, 1974.
17. *Heinrich, R., Schuster, S.,* The regulation of cellular systems, Chapman & Hall, New York, 1996.
18. *Turanyi, T., Tomlin, A. S., Pilling, M. J.,* J. Phys. Chem. **97** (1993) 163.
19. *Wei, J., Kuo, J. C. W.,* Ind. & Eng. Chem. Fundam. **8** (1969) 114.
20. *Li, G., Rabitz, H.,* Chem. Eng. Sci. **45** (1990) 977.
21. *Li, G., Rabitz, H.,* Chem. Eng. Sci. **46** (1991a) 583.
22. *Li, G., Rabitz, H.,* Chem. Eng. Sci. **46** (1991b) 95.
23. *Li, G., Rabitz, H., Toth, J.,* Chem. Eng. Sci. **49** (1994) 343.
24. *Iordache, O., Maria, G., Pop, G.,* Ind. & Eng. Chem. Research **27** (1988) 2218.
25. *Gabaldon, T., Huynen, M. A.*, Cell Mol Life Sci **61** (2004) 930.
26. *Yang, Q., Lindahl, P., Morgan, J.*, J. theor. Biol. **222** (2003) 407.
27. *Sewell, C., Morgan, J., Lindahl, P.*, J. theor. Biol. **215** (2002) 151.
28. *Maria, G.*, Chemical and Biochemical Engineering Quarterly **17** (2003) 99.
29. *Tsuchiya, M., Ross, J.,* J. Phys. Chem. A **105** (2001) 4052.
30. *Morgan, J. J., Surovtsev, I. V., Lindahl, P. A.*, Theor. Biology **231** (2004) 581.
31. *Zhang, W., Tichy, S., Perez, L., Maria, G., Lindahl, P., Simanek, E.,* Journal of American Chemical Society **125** (2003) 5086.
32. *Maria, G.,* Chemical Engineering Science **60** (2005b) 1709.
33. *Wallwork, S. C., Grant, D. J. W.*, Physical Chemistry, Longman, London, 1977.
34. *Styczynski, M. P., Stephanopoulos, G.,* Computers & Chemical Engineering **29** (2005) 519.
35. *Zak, D. E., Vadigepalli, R., Gonye, G. E., Doyle III, F. J., Schwaber, J. S., Ogunnaike, B. A.,* Computers & Chemical Engineering **29** (2005) 547.
36. *Salis, H., Kaznessis, Y.,* Computers & Chemical Engineering **29** (2005) 577.
37. *Kaznessis, Y. N.,* Chemical Engineering Science **61** (2006) 940.

38. *Angeli, D., Ferrell, J. E. Jr., Sontag, E. D.*, Proceedings of the National Academy of Sciences of the USA **101** (2004) 1822.
39. *Kholodenko, B. N., Kiyatkin, A., Bruggeman, F. J., Sontag, E., Westerhoff, H. V., Hoek, J. B.,* Proceedings of the National Academy of Sciences of the USA **99** (2002) 12841.
40. *Savageau, M. A.,* Math. Biosciences **180** (2002) 237.
41. *Voit, E. O.*, IEE Proc. Syst. Biol. **152** (2005) 207.
42. *Ofiteru, D. I.,* Optimal control strategies for bioprocesses, PhD Thessis, University Politehnica of Bucharest (Romania), 2006.
43. *Hlavacek, W. S., Savageau, M. A.,* J. Mol. Biol. **266** (1997) 538.
44. *Van Someren, E. P., Wessels, L. F. A., Backer, E., Reinders, M. J. T.,* Signal Processing **83** (2003) 763.
45. *Schwacke, J. H., Voit, E. O.,* Theoretical Biology and Medical Modelling **1** (2004) 1.
46. *Silva, M. R.,* Bioinformatics tools for the visualization and structural analysis of metabolic networks, PhD Thesis, TU Braunschweig (Germany), 2006.
47. *Ma, H. W., Zhao, X. M., Yuan, Y. J., Zeng, A. P.,* Bioinformatics **20** (2004) 1870.
48. *Ma, H. W., Kumar, B., Ditges, U., Gunzer, F., Buer, J., Zeng, A. P.,* Nucleic Acids Research **32** (2004) 6643.
49. *He, F., Zeng, A. P.,* BMC Bioinformatics **69** (2006) 1.
50. *Chang, Y. J., Sahinidis, N. V.,* Computers & Chemical Engineering **29** (2005) 467.
51. *Adrover, A., Cerbelli, S., Giona, M.,* Int. J. Bifurcation and Chaos **12** (2002) 353.
52. *Kaper, H. G., Kaper, T. J.,* Physica D **165** (2002) 66.
53. *Kutznetsov, Y. A.,* Elements of applied bifurcation theory, Springer-Verlag, Berlin, 1998.