# Reduced Modular Representations Applied to Simulate Some Genetic Regulatory Circuits**

**GHEORGHE MARIA***
University Politehnica of Bucharest,, Laboratory of Chemical & Biochemical Reaction Engineering, 1 Polizu, 011061 , Bucharest, Romania

*General chemical engineering modelling principles are valuable tools to represent the topology and the kinetics of complex cell processes. Elaboration of reduced models is based on a large amount of qualitative-quantitative information 'translated' from the 'language' of molecular biology to that of mechanistic chemistry. In the 'reverse engineering' approach, linked modules of adjustable complexity are used to simulate the genetic expression regulatory networks by adequately lumping of species or metabolic reactions exhibiting similar functions. The resulted model must keep a satisfactory predictability on key-species and steps, an adequate representation of species inter-connectivity, structural, functional, and temporal cell hierarchy, and of the multi-cascade control loops of adjustable intermediate levels. Application of (non-) conventional identification and lumping methods can lead to a satisfactory prediction of local and global properties of the protein synthesis regulatory network. Examples on modelling the genetic expression illustrate the advantages but also the over-simplifications introduced by various reduced modular representations. Potential applications of the lumped simulation in 'genetic circuit engineering' reveal the advantages of using modular constructions for 'in-silico' design of new organisms that possess desired specific functions (e.g. biosensor design).*

*Keywords: lumping analysis, genetic regulatory networks, modular approach*

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 [1].

Reliable and sufficiently accurate mechanistic kinetic models are very effective tools in understanding a chemical/biological process, its influential variables, and physical meaning of parameters. However, complex kinetic modelling requires costly investigations under a wide range of operating conditions and inputs, by using observed variables sampled at various scales. Finally, the model degree of complexity depends on the adopted hypotheses/assumptions, amount of available information and utilisation scope. Classical modelling rules use standard kinetic data (species concentrations vs. reaction time) and a succession of conventional identification steps derived from the statistical estimation theory and from the physico-chemical-biological modelling principles [2]. The estimation rule is linked with the statistical methods because the observed data are always subjected to experimental errors, and multiple constraints are usually imposed to the kinetic parameters.

Modelling very complex (bio)chemical systems, such as metabolic processes at a molecular level, becomes an extremely difficult task because the systems present a low observability vs. the large number of species, reactions, and transport parameters (many of them poorly understood). However, advances in –omics information lead to a continuous expansion of bioinformatic databases, while advanced numerical techniques, mixtures of conventional and non-conventional estimation procedures, accounting for qualitative/quantitative information and global properties of the system, and massive software platforms organized on a modular basis, reported progresses in formulating reliable mechanistic cell models [1].

When modelling complex cell metabolic mechanisms, two main approaches have been developed over decades [1,5,13]:

- the structure-oriented analysis (mainly known as the 'metabolic control analysis' MCA) is focused on characterizing the pathway topology and the stationary cell growth. MCA uses various types of sensitivity coefficients (the so-called 'response coefficients'), which are quantitative measures of how much a perturbation affects the cell-system states (e.g. reaction rates, mass fluxes, component concentrations) around the steady-state. The systemic response of fluxes or concentrations to perturbation parameters (i.e. the 'control coefficients'), or of reaction rates to perturbations (i.e. the 'elasticity coefficients') have to fulfil the 'summation theorems', which reflect the network structural properties, and the 'connectivity theorems' related to the properties of single enzymes vs. the system behaviour.

- the dynamic or kinetic models based on a hypothetical reaction mechanism, kinetic equations, and known stoichiometry. When the analysis is expanded to large-

* Tel.: 0744830308

** Paper presented at 15-th Romanian International Conference on Chemistry and Chemical Engineering (RICCCE-15), Sinaia, 19-22 September 2007

scale metabolic networks, lumped representations of species and reactions are usually used due to the difficulty to obtain consistent experimental kinetic information and complete mechanistic details. Valuable structured dynamic models have been developed for simulating various cell (sub)systems, such as: 'whole-cell' models; single cell growth; oscillatory metabolic paths; regulatory networks for gene expression; cell cycles and oscillatory metabolic processes; cellular communications, intracellular signalling, neuronal transmission, networks of nerve cells etc.

Each theory presents strengths and shortcomings in providing an integrated predictive description of the cellular regulatory network. However, a precondition for a reliable modelling is the correct identification of both topological and kinetic properties of the system. In solving the modelling problem, lumped (grouped) representations of cell metabolic subsystems are used, such as the functional subunits called 'modules' (e.g. amino acid or protein synthesis regulation module, protein degradation module, mitochondria metabolic path, etc. [16]). As few kinetic data are present in a standard form, the current trend in developing dynamic models is to use the so-called unconventional identification algorithms [12]. Such an approach accounts for various types of information on cell subsystem properties and functions, mixtures of estimation methods, and advanced lumping algorithms to increase the (reduced) model statistical quality. Structured models of a modular construction, 'circuit-like' network, or compartmented, including Boolean, continuous or stochastic variables, are currently used to represent various cell processes [1,2].

To overcome the structural low kinetics identifiability of complex (bio)chemical systems, the model reductions by lumping reactions and/or variables, and by keeping the most influential terms in the kinetic and transport relationships, are currently used. Estimate quality tests, parameter sensitivity analysis, principal component, ridge parameter selection, or elaborated algorithms to find invariant subspaces of the extended kinetic model are common rules to reduce an extended model structure [4]. Combined statistical and non-conventional estimators are used, by exploiting the system global properties, such as regulatory effectiveness, process periodicity, flexibility, certain succession of events, etc. The model reduction cost is a loss of information on certain species and reactions, a certain loss in model generality, prediction capabilities, and physical meaning for some rate constants [2-4]).

The use of advanced model reduction algorithms and strategies in modelling (bio)chemical kinetic systems present an important number of advantages, such as:

- increase in the statistical quality of the kinetic model (identifiability, estimability) vs. available data, and an increase in the computing tractability of the complex biosystem.

- establishment of linking relations between extended (low-estimable) kinetic model structures and reduced (estimable, apparent) model structures, allowing an useful interpretation of rate/equilibrium constants of extended models from using reduced model parameters identified from few observations.

- establishment of computationally tractable alternatives to represent complex bio-molecular processes, such as cell regulatory networks, synthesis networks, metabolic cycles etc.

The modular approach in developing structured dynamic models is very attractive due to important offered advantages. Thus, in the so-called 'reverse engineering' and 'integrative understanding' analysis the cell system is disassembled as much as possible, performing tests, and learning the structure of the whole and its parts with the hope of being able to 'recreate' the same system from scratch [6, 17]. Linked modules of adjustable complexity are used to simulate the genetic expression regulatory networks by adequately lumping of species or metabolic reactions exhibiting similar functions. The resulted model must keep a satisfactory predictability on key-species and steps, an adequate representation of species inter-connectivity, structural, functional, and temporal cell hierarchy, and of the multi-cascade control loops of adjustable intermediate levels. Application of unconventional identification and lumping methods can lead to a satisfactory prediction of local and global properties of the cell regulatory network. Potential applications of the lumped simulation in 'genetic circuit engineering' reveal the advantages of using modular constructions for 'in-silico' design of new organisms that possess desired specific functions (e.g. biosensor design [5]).

The scope of this paper is to illustrate, with some examples, how elaboration of reduced kinetic models of satisfactory quality to represent GRC (genetic regulatory circuits) is closely related to the ability of selecting the suitable lumping rules (of species and/or reactions), key-parameters, influential terms, and to apply unconventional identification strategies and linking rules that better realize a trade-off between model simplicity and its predictive quality.

*Lumped GRN models*

The gene expression is a highly regulated process, auto- and mutually catalysed by means of synthesized activation and repressor proteins (transcription factors) implied in negative/positive feedback regulatory loops [18]. The process regulation is realized by means of a hierarchical organized genetic regulatory network (GRN). The modular modelling approach implies the use of a gradual lumping analysis of the GRN for simulating the mechanism by which genes and proteins interact to regulate the gene expression. Thus, various semi-autonomous lumped kinetic modules can be constructed, based on experimental observations. The negative regulatory loops and a cascade control of the gene transcription and translation steps speed-up the response time of the GRC to perturbations and make it more sensitive and effective [13,18]. As an example, in Fig. 1 are presented various modular lumped models describing the expression of a single gene [19], a bistable switch circuit of two gene expression modules [5], or a three module cross-regulated GRC from E. coli (i.e. the repressilator of Ellowitz & Leibler [20]). As an application, the bistable GRC switch of two gene expression modules can be used to design mutant cells functioning as biosensors.

Under such a representation, 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. The emergent field of such efforts is the so-called 'gene circuit engineering' and a large number of examples have been reported with in-silico creation of novel GRC conferring new properties/functions to the

mutant cells (i.e. desired 'motifs' in response to external stimuli), such as: toggle-switches, hysteretic GRC, GRC oscillator, GRC signalling etc. [5,21,22]. A new research topics, called Synthetic Biology, interpreted as the engineering-driven building of complex biological entities, aims at applying engineering principles of systems design to biology with the idea to produce predictable and robust systems with novel functions in a broad area of applications, such as therapy of diseases (gene therapy), design of new biotechnological processes, new devices based on cell-cell communicators, biosensors etc. [23].

Such a lumping strategy presents several advantages but also limitations. Dynamic models (Boolean, continuous, or stochastic), of adjustable size, are used 'to divide' the complex gene circuits in sub-systems (modules) of a more tractable complexity. By representing the transcriptional mechanism and gene interactions, the architecture of the cell regulatory network is thus related to the physiological characteristics of the organism [6,7]. 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). Beside, the gene expression multi-cascade control presents a monotonous response that implies an intrinsic system modularity. This approach allows reducing the analysis complexity by investigating individual modules, and then to relate them to the holistic cell properties.

The difficulty to precise the very large number of parameters in complex GRNs lead to include lumped unstructured representations of rate expressions of power-law [8,9] or hyperbolic type [10] 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 [7,9]. Various criteria to define the modular system functional effectiveness have been defined (in terms of stability, responsiveness, selectivity, robustness, efficiency [11,15]) 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 [1,5]. Alternative lumped modular GRN structures are discriminated based on the system constraints, experimental observation, physical meaning of lumped components and reactions.

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 an acceptable loss of information about the system dynamics. However, some intermediate species, of quickly adjustable low-concentrations, cannot be eliminated by simple QSSA (quasi-steady-state assumption) 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)[12].

*Example of constructing modular GRC models*

In order to exemplify various rules for linking GRC modules, the variable volume cell model [13] have been adopted. Under this approach, the cell volume is explicitly included in the mass-balance equations. The kinetic expressions, evaluated at a certain time during the cell growth, are linked through the osmotic pressure and the state-law (linking the cell volume, osmotic pressure, temperature, and cell content in terms of species number of moles). The main model hypotheses are the followings (table 1): i) negligible inner-cell gradients; ii) open cell system of uniform content; iii) semi-permeable membrane, of negligible volume and resistance to nutrient diffusion, following the cell growing dynamics; iv) constant inner/external osmotic pressure, ensuring the membrane integrity; v) average logarithmic growing rate($D_s = \ln(2) / t_c$; $t_c$ = cell cycle time); vi) volume growth of the approximate exponential law ($V = V_0 \exp(D_s t)$); vii) homeostatic stationary growth ($dc_j / dt_s = 0$ (where $c_j$ = species j concentration); viii) perturbations in cell volume are induced by variations in species copynumbers under the isotonic osmolarity constraint.

Under such a representation, the individual or lumped reaction rates and the internal/external perturbations in species levels are explicitly linked with the evolution of the cell volume and dilution factor, which in turn will directly influence (by means the so-called 'secondary perturbations') the cell component concentrations. The result is a more accurate representation of the cell-growth [1,13] and cell-division phases [24], and a more realistic representation of the local and global regulatory properties of the GRC. The rates of individual reactions are constrained by the periodicity of the cell-cycle and by the requirement that molar amounts of all components and the volume must double in exactly one cell-cycle [25]. To be consistent with the hypotheses, such a 'whole-cell' modelling framework requires that each cell process to be included at some level of detail, i.e. as an individual or lumped species and/or reaction. The model analysis allows the full characterization of the GRC, by explicitly linking the model parameters to the system properties and its effectiveness (P.I.). Among module P.I.-s are to be mentioned the followings [5]:

-system local stability condition, and stability strength: i)stationary regulation, i.e. large margins of stability in the state variable space vs. stationary perturbations; ii) dynamic regulation, i.e. fast $\tau_{rec,j}$ = species j recovering time of the steady-state (QSS) with a tolerance of 5% [11] or 1% [1], after an impulse-like perturbation;

-high responsiveness to (exo/endogeneous) signalling species of repression or de-repression, that is small rise-times (transition times $\tau_j$) and tolerable overshoots in the level of enzymes repressing or de-repressing the gene expression;

-GRC selectivity, the regulator protein being sufficiently insensitive to changes in the level of effector protein [i.e. small sensitivities $S(c_{R1};c_{R2}) = \partial \ln(c_{R1}) / \partial \ln(c_{R2})$ ] (fig.1), or to other species from the GRC;

-GRC robustness, that is small sensitivities of the system performances vs. its kinetic parameters [i.e. small sensitivities $S(c_{i,s};\boldsymbol{k}) = \partial \ln(c_{i,s}) / \partial \ln(\boldsymbol{k})$, or $S(\tau_{rec,j};\boldsymbol{k}) = \partial \ln(\tau_{rec,j}) / \partial \ln(\boldsymbol{k})$ ]

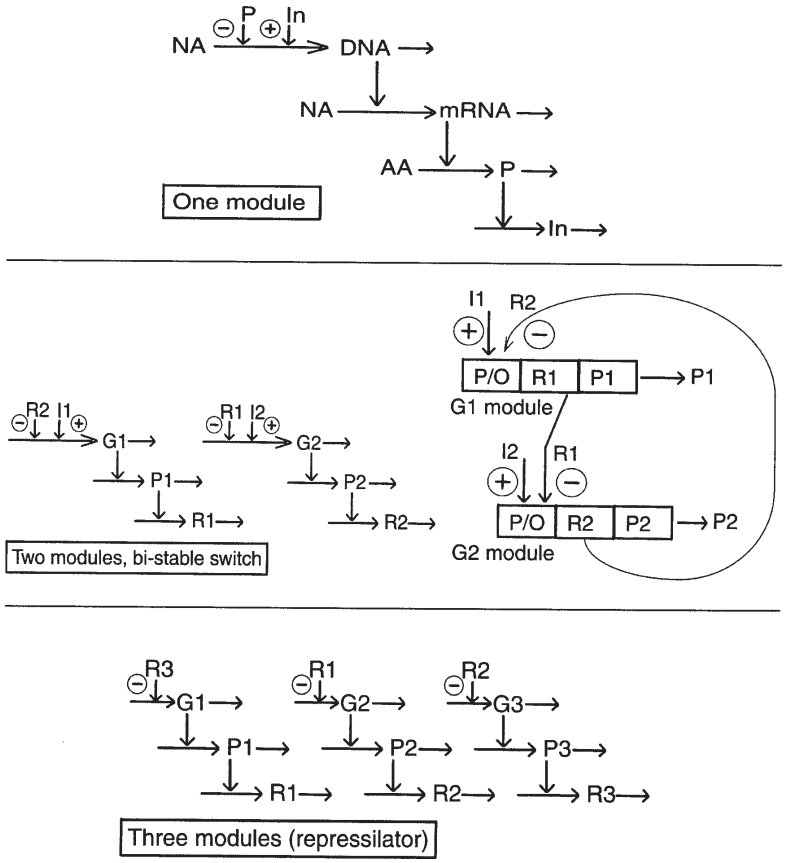


Fig. 1. Example of lumped representations of simple GRC (1-3 regulatory modules).
(Up) One gene expression regulatory module with induction and repression loops [19].
(Middle) The bistable switch circuit of two gene expression regulatory modules, oftype G1(R2R2)n(R1)m + G2(R1R1)n(R2)m, [5,13].
(Down) Three module GRC circuit, i.e. the repressilator from *E. coli* [20]. Notations: I, In= inducer; AA= aminoacids; P= promoter or protein, G= gene; O= operators; R= repressor

-system regulatory efficiency in terms of ensuring an appropriate steady-state stability in response to dynamic perturbations in internal or external species [i.e. small QSS-sensitivities $S(c_{j,s}; c_{perturb.})$];

-species connectivity in terms of synchronized and efficient treatment of a dynamic perturbation for recovering the steady-state [i.e. small STD= standard deviation of $\tau_{rec,j}$ ];

In this paper, the checked regulatory P.I.-s are the followings: the species *j* sensitivity vs. stationary environmental perturbations in nutrient levels $S(c_{j,s}; c_{Nut.})$; the QSS-recovering times of species $\tau_{rec,j}$ after a dynamic inner perturbation; AVG= average of $\tau_{rec,j}$; STD= standard deviation of species $\tau_{rec,j}$ after a dynamic inner perturbation; AVG= average of $\tau_{rec,j}$; STD= standard deviation of $\tau_{rec,j}$.

For exemplification, the *E. coli* cell with the nominal (index 's') growing conditions[13] has been approached: $.V_{cyt,o} = 1.66 . 10^{-15}$ L; t = 100 min; $c_{NutG,s} = 3 . 10^6$ nmol $L^{-1}$; $c_{NutP,s} = \Sigma_j c_{Met}P_{j,s} = 3 . 10^8$ nmol $L^{-1}$; $c_{P1,s} = 10^3$ nmol $L^{-1}$ ; $c_{P2,s} = 10^2$ nmol $L^{-1}$ ; $c_{G1s}$, $c_{G2s}$, $c_{M1s}$, $c_{M2s}$, $c_{GjPj,s}$, $c_{GjPjPj,s}$, $c_{M1P1,s}$, $c_{M2P2,s}$, are 1/2 or 1/4 nmol $L^{-1}$; $c_{P1P1,s}$, $c_{P2P2,s}$, are intermediates of adjustable levels that maximize the P.I.-s of GRC (Notations: 'cyt'= cytoplasma; G= gene; P= protein; 'o'= initial; M= mRNA; NutP, NutG = nutrients; MetP, MetG= metabolites; GP, GPP, MP = catalytically inactive species).

Several linking rules have been establish and checked. Among them, are to be mentioned the followings: linking reactions between modules must be set slower comparatively to the module core reactions; use cooperative and mutual catalysis; individualized functions must be allocated to each component into the cell; intermediate species levels and allosteric regulation loops must be adjusted accordingly to the GRN size; variable cell-volume and isotonic modelling environment must be considered for a more realistic representation of the secondary connectivity effects (cell 'ballast' and 'inertia'). One important rule is to try to use effectors in dimmer form (e.g. PP instead P) to adjust the gene activity (fig. 2). The transcriptional control with multiple operators binding repressor dimers Gi(PjPj)n is highly effective, the quickly adjustable dimer levels $[PjPj]_s$ via fast $Pj+Pj \underset{k_{diss}}{\overset{k_{bind}}{\rightleftarrows}} PjPj$ reactions will confer more flexibility to the gene expression regulatory module in ranging the stability, the dynamic characteristics, and the GRC flexibility vs. environmental changes. For instance, the cooperative link of two modules [G1(P1)2 + G2(P2)2], compared to the link [G1(P1P1)1 + G2(P2P2)1] in figure 2 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 $\tau_{P1}$, AVG($\tau_{rec,j}$) and STD($\tau_{rec,j}$). The stationary $S(c_j;NutP)$ are practically unchanged, while the module complexity is comparable ($n_s$ = 12 species and $n_r$ = 8 reactions for both systems). The low-concentrations of the oligomeric effectors (of type PP, PPPP, ...) are determined not by a QSSA but from optimising the global properties of the overall modular regulatory chain.

When the GRN is extended, by keeping the same gene-expression module type, the whole network P.I.-s

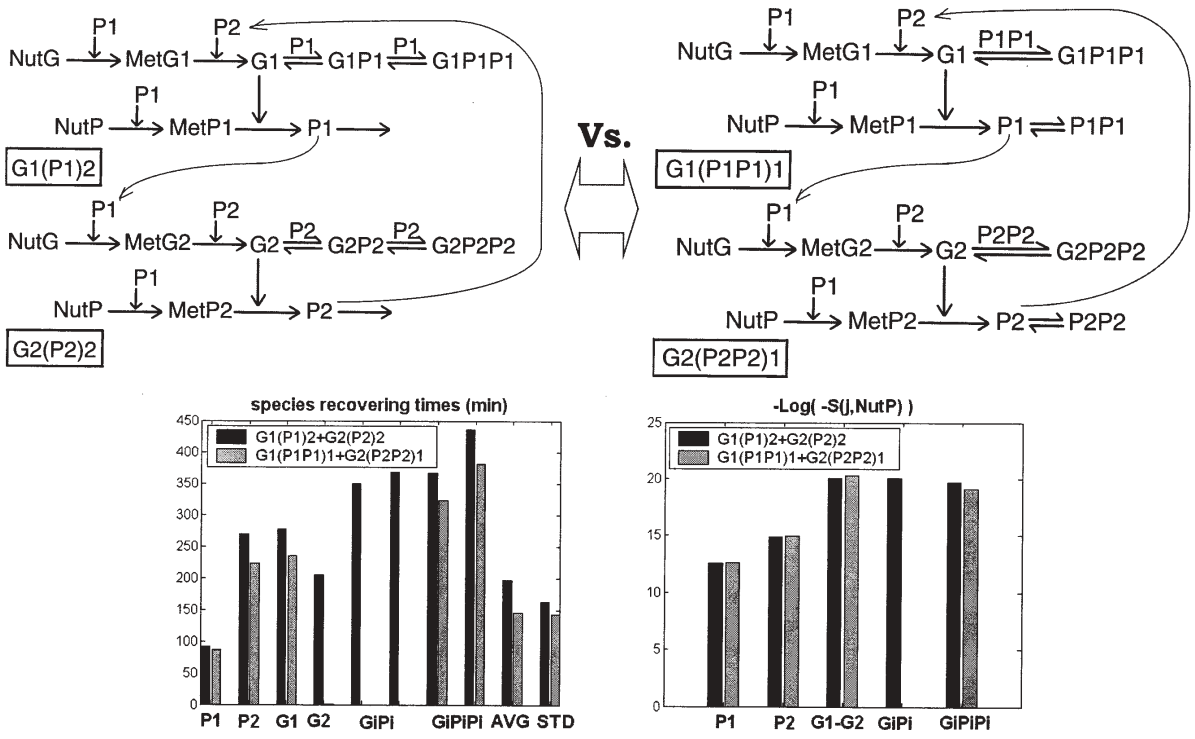


Fig. 2. Example of two linked GRC modules: multiple dimmer effectors (of optimized level) increase the system flexibility. (Up) The $G1(P1)_2$ + $G2(P2)_2$ system compared to the $G1(P1P1)_1$ + $G2(P2P2)_1$ system. (*Down left*) Species recovering times $\tau_{rec,j}$ to steady state after a ±10%$[P1]_s$ impulse perturbation; (*Down right*) Sensitivities vs. nutrients $S(c_j;NutP)$ of the species stationary levels [13].

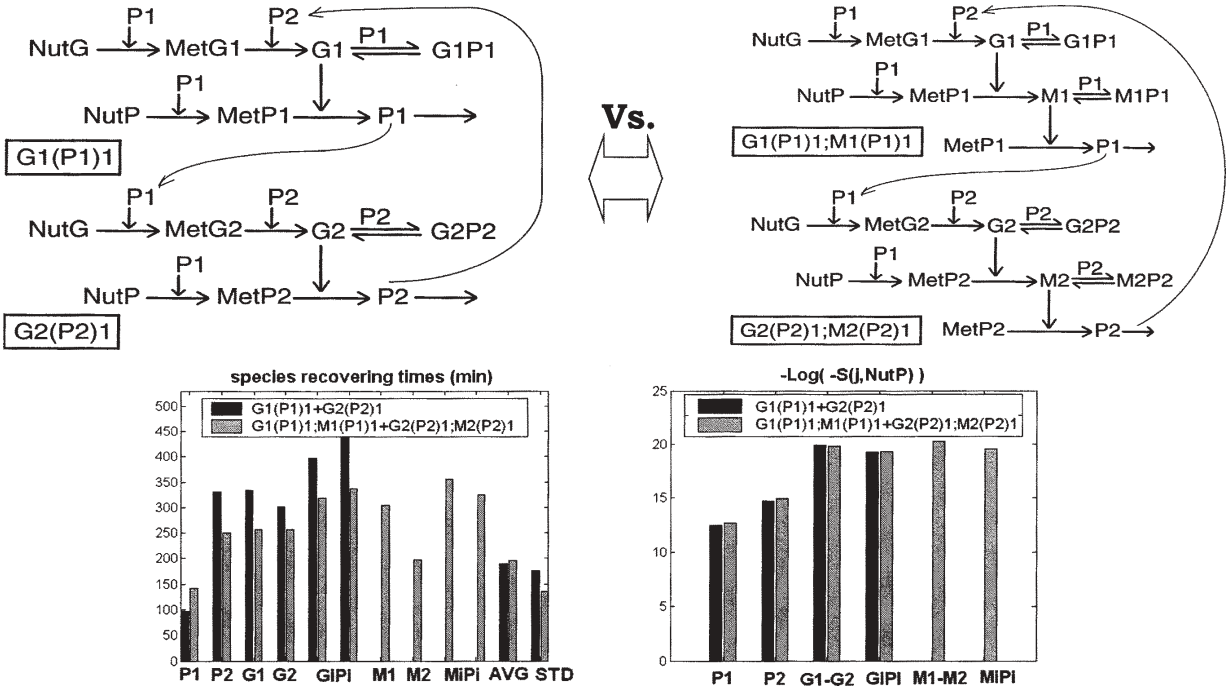


Fig. 3. Example of two linked GRC modules: cope with complexity and response synchronization by using a cascade control of gene expression. (*Up*) The $G1(P1)_1$ + $G2(P2)_1$ system compared to the $G1(P1)_1$;$M1(P1)_1$ + $G2(P2)_1$;$M2(P2)_1$ system. (*Down left*) Species recovering times $\tau_{rec,j}$ to steady state after a ±10%$[P1]_s$ impulse perturbation; (*Down right*) Sensitivities vs. nutrients $S(c_j;NutP)$ of the species stationary levels [13].

tend to decline due to an increased complexity of the system ($n_s$ number), and an increased difficulty to synchronize the efficient response of all components vs. perturbations. Due to such reasons, as the GRC is extended, as more effective modular representations (with cascade control and multiple effectors) should be used. For instance, in figure 3 a two-module GRC is modelled in the variant of a simpler $G1(P1)_1$ + $G2(P2)_1$ representation ($n_s$ =10 species) comparatively to the $G1(P1)_1$;$M1(P1)_1$ + $G2(P2)_1$;$M2(P2)_1$ system ($n_s$ =14 species). In spite of an increased complexity, the use of a more effective regulatory schema leads to adequate regulatory P.I.-s. Indeed, by comparing the two assembly alternatives in figure 3, $\tau_{rec,j}$ is lower for all species (except for P1), the species interconnectivity index $STD(\tau_j)$ is better (i.e. a low value), while QSS-resistance

**Table 1**

VARIABLE CELL-VOLUME DYNAMIC MODEL AND ITS BASIC HYPOTHESES [1]

| *Mass balance and state equations* | *Remarks* |
|---|---|
| $\frac{dc_j}{dt} = \frac{1}{V}\frac{dN_j}{dt} - D c_j = g_j(\boldsymbol{c},\boldsymbol{k})$ | continuous variable dynamic model representing the cell growing phase (ca. 80% of the cell cycle) |
| $\frac{1}{V}\frac{dN_j}{dt} = r_j(\boldsymbol{c},\boldsymbol{k})\ ;\ j = 1,...,n_s$ | |
| $\frac{RT}{\pi} = \frac{V}{\sum_{j=1}^{n_s} N_j} = \frac{1}{\sum_{j=1}^{n_s} c_j} = \frac{1}{\sum_{j=1}^{n_s} c_{jo}} = \text{constant}$ | constant osmotic pressure constraint (Pfeffer's law in diluted solutions) |
| $\left(\sum_j^{all} c_j\right)_{cyt} = \left(\sum_j^{all} c_j\right)_{env}$ | 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 stationary growing rate of average $D_s = ((dV/dt)/V)_s = ln(2)/t_c$ ; | |
| - homeostatic stationary growth of $\left(dc_j/dt\right)_s = g_j(\boldsymbol{c}_s,\boldsymbol{k}) = 0$ ; | |
| - perturbations in cell volume are induced by variations in species copynumbers under the isotonic osmolarity constraint: $V_{perturb}/V = \left(\sum N_j\right)_{perturb} / \left(\sum N_j\right)$. | |
| *Notations*: $c_j$ = species j concentration; $r_j$ = species j reaction rate; $N_j$ = species j no. of moles; $V$ = cell volume; $T$ = temperature; $R$ = universal gas constant; $t$ = time; index 'o'= initial; $\boldsymbol{k}$ = rate constant vector; $t_c$ = cell cycle period | |

to external perturbations are practically unchanged (i.e. the $S(c_j; NutP)$ sensitivities).

## Conclusions

Modelling synthetic gene circuits for in-silico GRC-design is an important step in advancing the understanding on the regulatory cell network, with important theoretical and practical implications. The modular approach, with accounting for both local and holistic GRC properties and observations from bio-molecular databanks, makes this computational approach effective allowing: similarity analysis of models (structure vs. predictions); lumping analysis; system characterization (QSS-multiplicity, stability, flexibility, robustness, efficiency); system modularisation and development of cell simulation platforms.

Even if a generic GRC from E. coli has been analysed, the variable-volume and whole-cell modelling framework, with explicitly considering the link between the volume-growth and the reaction rates for all species into the cell, appears to be a more promising approach to evaluate the GRC characteristics in a cell, by mimicking the equilibrated or perturbed growth. Such models can avoid over-estimation of some regulatory properties (i.e. responsiveness, efficiency, connectivity), accounting for the role of cell-ballast in smoothing internal/external perturbations, for direct or indirect perturbations of species levels (transmitted via chain reactions and cell-volume variation).

Lumping rules are proved to be effective tools for modelling the cell regulatory process complexity and dynamics, coping with the cell-system low observability, identifiability and estimability. Power-law or Hill-type representation of modular GRC, including apparent rate constants, can reproduce a wide-range of cell functions and dynamic behaviour. However, the model predictability is strongly dependent on the lumping degree, on the key-species selection and ability to realise the suitable trade-off between model simplicity, its predictive power and physical-meaning of terms. A sensitivity analysis applied to model terms can help in relating the GRC holistic properties to the individual regulatory module structure.

Derivation of relations between apparent and extended model structures is of high interest for a process correct interpretation, and to characterise the 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, the physical meaning of lumps can also play an important role in choosing the most suitable lumping route from the large number offered by the theoretical analysis.

When large cell dynamic models are developed, application of unconventional model reduction 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.

## References


1. MARIA, G., Chem. & Biochem. Eng. Quarterly **19**, 2005b, p. 213
2. MARIA, G., Chem. & Biochem. Eng. Quarterly **18** , 2004, p. 195
3. MARIA, G., Chem. & Biochem. Eng. Quarterly **17**, 2003. p. 99
4. MARIA, G.., Chemical Engineering Science **60**, 2005a, p. 1709
5. MARIA, G.., Chem. & Biochem. Eng. Quarterly **21**, 2007, p. 417

6. STYCZYNSKI, M.P., STEPHANOPOULOS, G., Comp. & Chemical Engineering **29**, 2005, p. 519
7. ANGELI, D., FERRELL, J.E. Jr., SONTAG, E.D., Proceedings of the National Academy of Sciences of the USA **101**, 2004, p. 1822
8. SAVAGEAU, M.A., Math. Biosciences **180** , 2002, p. 237
9. VOIT, E.O., IEE Proc. Syst. Biol. **152** , 2005) 207
10. OFITERU, D.I., PhD Thessis, University Politehnica of Bucharest (Romania), 2006
11. HLAVACEK, W.S., SAVAGEAU, M.A., J. Mol. Biol. **266**, 1997, p. 538
12. MARIA, G.., Invited SFB-578-53 Kolloquium, Integration gen- und verfahrenstechnischer Methoden zur Entwicklung biotechnologischer Prozesse: Von Gen zum Produkt, TU Braunschweig (Germany), July 2006b
13. MARIA, G.., Chem. & Biochem. Eng. Quarterly **20**, 2006a, p. 353
14. YANG, Q., LINDAHL, P., MORGAN, J., J. theor. Biol. **222**, 2003, p. 407
15. MA, H.W., ZHAO, X.M., YUAN, Y.J., ZENG, A.P., Bioinformatics **20** , 2004, p. 1870
16.KHOLODENKO, B.N., SCHUSTER, S., GARCIA, S., WESTERHOFF, H.V., CASCANTE, M., Biochimica Biophysica Acta 1379, 1998, p. 337
17. ZAK, D.E., VADIGEPALLI, R., GONYE, G.E., DOYLE III, F.J., SCHWABER, J.S., & OGUNNAIKE, B.A., Comp. & Chem. Eng. 29,2005, p. 547
18. ALON, U., An introduction to systems biology, Chapman & Hall/ CRC, Boca Raton, 2007
19. SCHWACKE, J.H., VOIT, E.O., Theoretical Biology and Medical Modelling 1,2004, p. 1
20. ELOWITZ, M.B., LEIBLER, S., Nature 403, 2000, p. 335
21. KAZNESSIS, Y.N., Chemical Engineering Science 61, 2006, p. 940
22. ATKINSON, M.R., SAVAGEAU, M.A., MYERS, J.T., NINFA, A.J., Cell 113, 2003, 597
23. HEINEMANN, M., PANKE, S., Bioinformatics 22, 2006, p. 2790
24. MORGAN, J.J., SUROVTSEV, I.V., LINDAHL, P.A., Jl. theor. Biology 231, 2004, p. 581
25. TYSON, J.J., NOVAK, B., J. theor. Biol. 210, 2001, p. 249

Manuscript received: 18.12.2007