

Setting Some Milestones when Modelling Cell Gene Expression Regulatory Circuits Under Variable-volume Whole-cell Modelling Framework II

GHEORGHE MARIA^{1*}, ANDREEA GEORGIANA SCOBAN

University Politehnica of Bucharest, Department of Chemical & Biochemical Engineering, 1-7 Polizu Str., 011061, Romania

While in the first part of the study, general concepts of the novel whole-cell simulation of metabolic processes in living cells are presented, by considering a variable-volume modelling framework, in the present paper exemplification is made for approaching several case studies when building-up modular model structures, for instance by developing modular kinetic representations of the homeostatic gene expression regulatory modules (GERM) that control the protein synthesis and homeostasis of metabolic processes. Past and current experience with GERM linking rules is presented in order to point-out how optimized globally efficient kinetic models for the genetic regulatory circuits (GRC) can be obtained to reproduce experimental observations. Based on quantitative regulatory indices evaluated vs. simulated dynamic and stationary environmental perturbations, the paper exemplifies with GERM-s from *E. coli*, at a generic level, how this methodology can be extended to characterize the GERM module efficiency, species connectivity, and system stability.

Keywords: kinetic modelling of cell protein synthesis; homeostatic regulation; gene expression regulatory modules (GERM); linking GERM-s

Define performance indices (PI) of a GERM to homeostatically regulate a gene expression under a deterministic WCVV modelling approach

To evaluate and compare the regulatory efficiency of various GERM structures when maintaining cell homeostasis, some quantitative performance indices P.I. have to be defined [1]. These P.I.-s fall in two categories of indices, defined under stationary ('step' like) or dynamic ('impulse' like) continuous perturbations of key-species stationary concentrations. Random perturbations, due to interactions of P-synthesis GERM with other metabolic processes, or due to environmental changes, lead to a GERM response that tends to maintain the key-component functions and homeostasis. Module efficiency depends on the GERM regulatory structure, species inter-connectivity, quasi-steady-state (QSS) characteristics, cell size and perturbation magnitude. The definitions introduced by Maria [1] are the followings:

Stationary perturbations refers to permanent modifications of nutrient / metabolite levels, leading to new stationary concentrations inside cell. Referring to the target protein P, the regulatory module tends to diminish the

deviation $[P]_s - [P]_{ns}$ between the *nominal* QSS (unperturbed set-point, of index *n*) and the new QSS reached after perturbation. Equivalently, the P-synthesis regulatory module will tend to maintain $[P]_{ns}$ within certain limits, $[P]_{min} \leq [P]_{ns} \leq [P]_{max}$ (a relative $R_{ss} = \pm 10\%$ maximum deviation has been proposed by Sewell et al. [2]). A measure of species *i* steady-state concentration (C_{is}) *resistance* to various perturbations (in rate constants, k_j , or in nutrient concentrations, C_{Nut_j}) is given by the magnitude of relative sensitivity coefficients at QSS, i.e. $S_{kj}^{C_i}$ and $S_{Nut_j}^{C_i}$ respectively, where $S_{perturb}^{state} = \partial(\text{State}) / \partial(\text{Perturbation})$ are the state sensitivities vs. perturbations; [3]. A regulatory index, $A_{unsync} = k_{syn} \times k_{decline}$, has been introduced to illustrate the maximum levels of (unsynchronized) stationary perturbations in synthesis or consumption rates of a key-species tolerated by the cell within defined limits [2]. The sensitivities $S_{Nut_j}^{C_i}$ are computed from solving a nonlinear algebraic set (1) obtained by assuming QSS conditions of the ordinary differential ODE kinetic model, and known nominal species stationary concentrations C_s :

$$\left(\frac{dC_j}{dt}\right)_s = \left(\frac{1}{V} \frac{dn_j}{dt}\right)_s - D_s C_{js} = h_{js}(C_s, k) = 0; j = 1, \dots, n_s; D_s = \left(\frac{RT}{\pi}\right) \sum_j \left(\frac{1}{V} \frac{dn_j}{dt}\right)_s;$$

$$\frac{RT}{\pi} = \frac{V(t)}{\sum_{j=1}^{n_s} n_j(t)} = \frac{1}{\sum_{j=1}^{n_s} C_{j0}} = \text{constant} \quad (1)$$

where: V = cell system volume; n_j = number of moles of *j* species; t = time; D = cell-content dilution rate (i.e. cell-volume logarithmic growing rate); Nut = nutrients; t = time; T = absolute temperature; R = universal gas constant; π = osmotic pressure.

Then, differentiation of the steady-state conditions eqn. (1) leads to evaluation of the state sensitivity vs. nutrient levels, i.e. $S_{Nut_j}^i = \left(\frac{\partial C_i}{\partial C_{Nut_j}}\right)_s$ by using (*s* index denotes stationary condition) :

* email: gmaria99m@hotmail.com ; Phone: (+40)744830308

Paper dedicated to the memory of Prof. Octavian Smigelschi (Univ. Politehnica of Bucharest, and Lummus Crest, GmbH, Weisbaden, Germany)

$$\left[\frac{\partial h_i(C, C_{Nut}, k)}{\partial C_i} \right]_S \left[\frac{\partial C_i}{\partial C_{Nut_j}} \right]_S + \left[\frac{\partial h_i(C, C_{Nut}, k)}{\partial C_{Nut_j}} \right]_S = 0 \quad (2)$$

In the previous relationship, the ODE model Jacobian $J_C = [\partial h_i / \partial C_j]_S$ is numerically evaluated for the cell-system stationary-state (1).

Dynamic perturbations are instantaneous changes in the concentration of one or more components that arise from a process lasting an infinitesimal time (impulse-like perturbation). After perturbation, the system recovers and returns to their stable nominal QSS. The *recovering time* of the key-species P (τ_j) and the recovering rate (denoted with R_D by Yang et al. [4]) can be approximated from the solution of the linearized system model [1, 5], or by simple simulation of the GERM system dynamics after such an impulse/dynamic perturbation:

$$\begin{aligned} dC / dt &= h(C, k) ; C(t=0) = C_s ; \\ C(t) &= C_s + \sum_{i=1}^{n_s} d_i b_i \exp(\lambda_i t) \end{aligned} \quad (3)$$

where: C = concentration vector; λ_i = eigenvalues of the system Jacobian matrix at QSS, $J_C = (\partial h(C, k) / \partial C)_S$; b_i, d_i = constants depending on the system characteristics at stationary conditions; t = time. If the real parts of eigenvalues are all negative, then the stationary state C_s is stable. The recovering rate R_D reflects the recovering properties of the regulated P-synthesis by the GERM system. The species j recovering times $\tau_j \sim 1/R_D$ are evaluated by simulating the system behavior, and by determining the times necessary to a certain species concentration to return to its stationary C_s concentration, with a certain tolerance and for a defined perturbation magnitude (Maria [1] proposed a 1% recovery tolerance for a *standard* $\pm 10\%$ C_s impulse perturbation).

Steady-state C_s stability strength is related to the GERM system characteristics. As $\text{Max}(\text{Re}(\lambda_i)) < 0$ is smaller (with the λ_i of the Jacobian evaluated at a certain QSS), as this QSS is more stable. When analyzing the predicted QSS and the regulatory characteristics of a P-synthesis GERM, the stability strength can also be associated to an index against periodic oscillations. This can be evaluated from

Table 1

THE REGULATORY EFFICIENCY PERFORMANCE INDICES PI-S PROPOSED TO EVALUATE THE BEHAVIOUR VS. PERTURBATIONS OF A GERM (AFTER [1]). MIN = TO BE MINIMIZED; MAX = TO BE MAXIMIZED

Index	Suitable objective	Expression
stationary regulation	Min	$R_{SS} = ([P]_S - [P]_{ns}) / [P]_{ns}$
stationary regulation	Max	$A_{unsync} = k_{syn} \times k_{decline}$
stationary regulation	Min	$S_{NutP_j}^i = \left[(\partial C_i / C_{is}) / (\partial C_{Nut_j} / C_{Nut_{js}}) \right]_S$
stationary regulation	Min	$S_{k_j}^i = \left[(\partial C_i / C_{is}) / (\partial k_j / k_j) \right]_S$
dynamic regulation	Min	$R_D = \text{Max}(\text{Re}(\lambda_i))$, $\text{Re}(\lambda_i) < 0$
dynamic regulation	Min	τ_j ; τ_P
regulatory robustness	Min	$(\partial R_D / \partial k)$
species interconnectivity	Min	$AVG(\tau_j) = \text{average}(\tau_j)$
species interconnectivity	Min	$STD(\tau_j) = \text{st.dev.}(\tau_j)$
QSS stability strength	Min	$\text{Max}(\text{Re}(\lambda_i))$, $\text{Re}(\lambda_i) < 0$
QSS stability strength*	Min	$ \lambda_A < 1$

Footnote: n = nominal value; s = stationary value; (*) see [3], and eqn. (4) for the monodromy matrix A matrix calculation; λ_i = i -th eigenvalues; A = monodromy matrix, defined in eqn. (4), and by Maria, [3]; τ_j = species j recovering time; Nut = nutrient; Re = real part; AVG = average; STD = standard deviation; C_j = species j concentration; R_D = dynamic regulatory (recovering) index.

the linearized form of the system model, by calculating the monodromy matrix $A(T)$ after a checked period T of time [3]:

$$\begin{aligned} dC / dt &= h(C, k) ; C(0) = C_s ; \\ dA / dt &= J_C A ; A(0) = I \end{aligned} \quad (4)$$

For a stable QSS, i.e. $|\lambda_{A_i}| < 1$, as $|\lambda_{A_i}|$ are smaller, as the system QSS stability strength is higher [λ_{A_i} are the eigenvalues of the $A(T)$ matrix; I = identity matrix].

Species interconnectivity in a GERM (a modular gene expression regulatory schema of reactions) can be viewed as a degree of which the involved species assist each other during the system synchronous recovering. Cell species connections appear due to common reactions, or common intermediates participating to chain reactions, or from the common cell volume to which all species contribute (under constant osmotic pressure, see WCVV model hypotheses in the Appendix-part 1 of the paper). Vance et al. [10] reviewed and proposed several quick experimental - computational rules to check a reaction schema via species inter-connectivities. By inducing experimental perturbations to a (bio)chemical system, by means of tracers, or by fluctuating the inputs of the system, one can measure the perturbation propagation through the consecutive / parallel reaction path. Then, various techniques can determine the *distance* among observed species, and rules to include this information in elaborating a reaction schema. In the present study, one proposes a similar approximate measure of species interconnectivity related to the species recovering-times after a dynamic perturbation, that is: $AVG(\tau_j)$ and $STD(\tau_j)$, i.e. the average and standard deviation of τ_j . As AVG and STD are larger, as the cell dynamic regulatory effectiveness is lower, species being less interconnected, with components that recover more disparately. As the number of effectors and buffering reactions is higher, as these dynamic regulatory indices of the module are better [6,1-3,7-9].

By summarizing, the regulatory efficiency performance indices P.I.-s proposed to evaluate the behaviour vs. perturbations of a GERM or of a chain of GERM-s, are given in table 1.

Efficiently linking GERM-s in a WCVV modelling framework

To exemplify the GERM linking analysis in a simple way, one considers a hypothetical cell, similar to *E. coli*, in a balanced growth under isothermal and isotonic conditions, with a cell cycle period of $t_c = 100$ min, and a quasi-constant logarithmic growing rate of $D_s = \ln(2)/t_c$. The nominal concentrations of the individual and lumped cell species corresponds to a cell of high ballast, and are given in table 2, being similar to those of Maria [1], i.e. $C_{NutG,s} = 3 \times 10^6$ nM, $C_{NutP,s} = 3 \times 10^8$ nM (nM=nano-molar, i.e. 10^{-9} mol/L concentration). As only a few number of individual species are accounted in the model, the cell *ballast* is mimicked by adopting high levels for me^9 nM (nM=nano-molar, i.e. 10^{-9} mol/L concentration).

$$\sum_j C_{MetP_{j,s}} = 3 \times 10^8 \text{ nM};$$

$$\sum_j C_{MetG_{j,s}} = C_{NutG,s} + C_{NutP,s} - \sum_{j \neq MetG_j}^{cell} C_{j,s} \approx 3 \times 10^6 \text{ nM}.$$

For the genes, proteins, and other intermediates, the nominal stationary concentrations are displayed in the table 2. The nutrient concentrations in the environment are assumed to be constant during the cell cycle.

a) Ranging the number of transcription factors TF and buffering reactions. For selecting the suitable GERM structure matching the available data, the first problem to be solved is related to the number of buffering reactions of type $G + P \rightleftharpoons GP$ or $M + P \rightleftharpoons MP$ (fig. 1 of [13]) necessary to be included. Calculation of various P.I.-s for a large number of GERM structures, indicates that the dynamic regulatory efficiency of $[G(P)_n]$ modules is nearly linearly increasing with the number (n) of buffering reactions (fig. 1). Moreover, the plots reveal that this increase is more pronounced in the case of using dimeric TF (that is PP), and for $[G(P)1; M(P)_n]$ modules that use a expression control scheme in cascade.

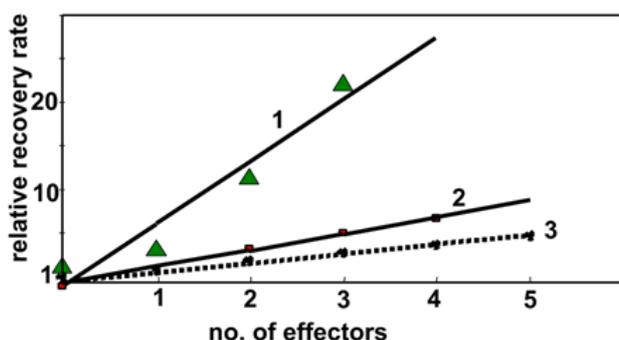


Fig. 1. The dependence of the GERM dynamic regulatory efficiency (the P species relative recovering rate after a $\pm 10\%$ impulse perturbation) for various GERM types, as function of the number n of buffering reactions (i.e. effectors (adapted following the results of Maria, 2003[3], 2005[1], and of Yang et al.[4]. Curve 1 refers to $[G(P)1; M(P)_n]$ modules; Curve 2 refers to $[G(PP)n]$; Curve 3 refers to $[G(P)_n]$ modules (after [1,3,4,7])

Such a module efficiency ranking is valid not only in the case of species relative recovering rate after a $\pm 10\%$ impulse perturbation in the key-protein P, but also for other P.I.-s, such as the stationary regulatory effectiveness; low sensitivity to stationary perturbations; stability strength of the homeostatic QSS [3]. As underlined by Maria [3], the recovering trajectories in the G/P phase plane is more linear for the efficient GERM-s, presenting a lower amplitude, thus not disturbing other cell processes.

As tested by Maria [1,3,7] with WCVV models of $[G(P)n]$, $[G(PP)n]$, and for $[G(P)1; M(P)_n]$ GERM-s, it is to underline that:

-Modules reporting high stationary-regulation P.I.-s also report high dynamic-regulation P.I.-s.

-The catalyst activity control at a single enzyme level (that is lacking of buffering reactions able to modulate the gene G catalytic activity) appears to be of lowest efficiency.

-Multiple copies of effector molecules (i.e. O, R, P in figure 1 of [13]), which reversibly and sequentially (allosterically) bind the catalyst (G, M) in negative feedbacks, improve the regulation effectiveness.

-A structured cascade control of several enzyme activities, with negative feedback loops at each level, improves regulation and amplifies the effect of a change in a stimuli (inducer). The rate of the ultimate reaction is amplified, depending on the number of cascade levels and catalysis rates. As an example in figure 1 of [13], by placing regulatory elements (O, R) at the level of mRNA (i.e. species M), and at the level of DNA (i.e. species G) in the module $[G(P)_n; M(P)_n]$ is highly effective.

-The nearly linear increase of GERM P.I.-s with the number of effectors (P, PP, O, R) acting in the i -th allosteric unit $L_i(O)_n$, of buffering reactions applied at various level of control of the gene expression, is valid for both dynamic and stationary P.I.-S of table 1 $[R_D, AVG(\tau_j), STD(\tau_j)]$.

-P.I. improves ca. 1.3-2 times (or even more) for every added regulatory unit to the same GERM type. Multiple regulatory units lead to average recovering times $AVG(\tau_j)$ much lower than the cell cycle period t_c , under constant logarithmic volume growing rate, $D = \ln(2) / t_c$.

-Combinations of regulatory schemes and units (with different effectors) might improve the regulatory P.I.-s (to be proved).

-Certain regulatory modules reported an increased flexibility, due to adjustable intermediate transcription factors TF species levels. This is the case, for instance, of adjusting $[M]_s$ in module $[G(P)_n; M(P)_n]$ and of $[PP]_s$ in modules $[G(PP)n]$. Optimal levels of these species can be set accordingly to various optimization criteria, rendering complex regulatory modules to be more flexible in reproducing certain desired cell-synthesis regulatory properties.

b) The effect of the mutual G/P synthesis catalysis.

One essential aspect of the $[G(P)n]$, $[G(PP)n]$, and $[G(P)_n; M(P)_n]$ kinetic models of GERM is the mutual catalysis of G and its encoding protein P synthesis. If one adds the WCVV modeling constraints eq. (1) and eqn. (10) of [13], this direct and indirect link ensures a quick G, P recovering after any perturbation. To prove this, one considers a GERM of $[G(P)1]$ type and the nominal homeostasis of a high ballast cell of table 2 but, in both alternatives; without mutual catalysis (see the left scheme in fig. 2), or with mutual catalysis (see the right scheme in fig. 2).

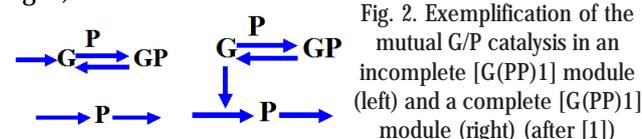


Fig. 2. Exemplification of the mutual G/P catalysis in an incomplete $[G(PP)1]$ module (left) and a complete $[G(PP)1]$ module (right) (after [1])

By applying a-10% impulse perturbation in the $[P]_s = 1000$ nM at $t=0$, the incomplete $[G(P)1]$, reports perturbations in all GERM species (fig. 3) with any or very slow recovering tendencies. On the contrary, as revealed by simulation of the complete $[G(P)1]$ behaviour plotted in figure 3, the G, and P species present relatively short

Species	Low ballast cell (nM)		High ballast cell (nM)	
	QSS conc. (nM)	Recovery time (min.)	QSS conc. (nM)	Recovery time (min.)
C_{NutP}	3000	NG	3×10^8	NG
C_{NutG}	3000	NG	3×10^6	NG
Lump $\sum_j C_{MetG_{j,s}}$	~2000	NG	$\sim 3 \times 10^6$	NG
Lump $\sum_j C_{MetP_{j,s}}$	3000	NG	3×10^8	NG
$C_{P,s} = [P]_s$	1000	103	1000	133
$C_{G,s} = [G]_s$	0.5	223	0.5	93
$C_{GP,s}$	0.5	246	0.5	93
SUM $\sum_j C_j$	12001		$\sim 6.06 \times 10^8$	

Table 2
THE NOMINAL (HOMEOSTATIC QSS) E. COLI CELL CONDITIONS, AND THE RECOVERING RATES OF A [G(P)1] GENE EXPRESSION MODULE AFTER a -10% IMPULSE PERTURBATION IN THE [P]_s OF 1000 nM AT AN ARBITRARY t=0. CELL INITIAL VOLUME OF THE CONSIDERED E. COLI CELL, IS OF $V_{cyt,0} = 1.66 \cdot 10^{-15}$ L (ADAPTED FROM MARIA [1])

Footnotes: The lump $\sum_j C_{MetG_j}$ results from the isotonic constraint: $\sum_j C_{MetG_j} = C_{NutG} + C_{NutP} - \sum_{j=MetG}^{cell} C_j - \sum_{j=MetG_j}^{cell} C_j$; The considered cell life cycle is of $t_c = 100$ min; the cell-volume logarithmic growing rate is $D = \ln 2 / t_c$. The $\text{Max}(\text{Re}(\lambda_i)) < 0$ indicates a stable QSS homeostasis of the cell, where λ_j are the Jacobian (2) eigenvalues of the ODE kinetic model of the WCVV cell system (1). The rate constants of the G(P)1 model (1) results from solving the stationary model (that is for $\frac{dC_j}{dt} = 0$) with known stationary species concentrations displayed in the above table. The only $G + P \rightleftharpoons GP$ buffer reaction was considered with the reverse reaction rate constant of 10^5 1/min [7]. Notations: NutP and NutG are substrates used in the synthesis of metabolites MetP and MetG used for P and G synthesis; G = a gene (DNA); P = a protein; M = RNA; GP = the inactive complex of G with P; C_j = species j concentration; cyt = cytoplasm; o = initial; 's' index refers to the QSS; NG = negligible.

recovering rates, and negligibles for the other species (table 3). Also, the species connectivity increases in the complete [G(P)1] is better compared to an incomplete [G(P)1], or a [G(P)0], (that is smaller STD(τ)). Consequently, removal of P as a catalyst for G synthesis will:

- decrease the species inter-connectivity
- increase the recovering time of species
- decrease the standard deviation of recovering times.

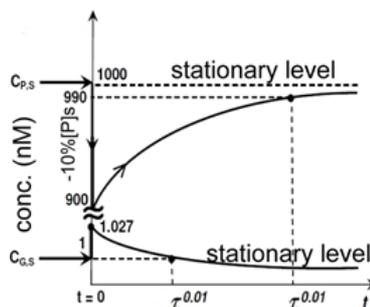


Fig. 3. Exemplification of the auto-mutual G/P catalysis after a -10% impulse perturbation in the [P]_s = 1000 nM at t=0 for a GERM of [G(PP)1] type. The cell nominal conditions are those of table 2 (a high ballast cell, with [G]_s = 1 nM)

One aspect of the WCVV kinetic models in the way by which the variable cell-volume plays an important role is the realistic representation of species inter-connectivity (direct or indirect) in the same GERM regulatory module or among linked modules. Even if connectivity of species can be expressed in several ways [10], it is directly dependent on the manner in which species in a GERM regulatory module or from several linked modules recover more or less independently after a perturbation. When the species connectivity increases, they recover with a more comparable rate (or equivalently, over the same time). When the species are more disconnected, they recover in a more disparate way, 'assisting' each other less to cope with a perturbation.

Consequently, the mutual autocatalytic synthesis reactions increase species inter-connectivity in a GERM, as proved by the table 3. Also, the recovery time (τ_{rec_j}) was smaller for the key-species. Similar differences were observed for steady-state regulation, state sensitivities to external nutrients being smaller. Thus, mutual autocatalysis appears to *interconnect* member components such that they are regulated more as a unit than would otherwise be the case. Interconnectivities (the degree to which a perturbation in one component influences others) may arise from a direct connection between components (e.g. when they are involved in the same chain of reactions), or from an indirect connection (via volume changes). Our analysis indicates that mutual autocatalysis is a particularly strong type of interaction that unifies the regulatory response, and they serve to *smooth* the effects of perturbations. It also suggests a way to quantitatively evaluate inter-connectivities between all cellular components: each component could be perturbed one at a time, and recovery rates or some other measure of regulatory effectiveness could be calculated for all components. The resulting relationships would thus reflect the holistic properties of the GRC-s.

c) The effect of system isotonicity.

The effect of the isotonicity constraint eq. (1) of a WCVV model can be easily proved in the same way as done in the previous chapter (b-the effect of the mutual G/P synthesis catalysis). By simulating the effect of applying a -10% impulse perturbation in the [P]_s = 1000 nM at t=0, on a [G(P)1] model, perturbations in only key-species (G, P, GP) are observed (fig. 3). However, when the isotonicity constraint is missing from the model, the key-species do not recover. On the contrary, as revealed by simulation of the complete [G(P)1] behaviour plotted in figure 3, the

GERM type	Species C_j	$\partial C_j / \partial \text{NutG},s$	$\partial \ln C_j / \partial \ln \text{NutG},s$	Recovery time τ_{rec} (min)
G(P)0	P	-4.53	-0.452	156.5
	G	4.76×10^{-4}	0.047	NG
	MetG	52.43	0.524	NG
	MetP	-47.89	-0.478	NG
	AVG			39.12
	STD			78.25
G(P)1	P	-3.7	-0.365	127.1
	G	1.2×10^{-3}	0.229	118.1
	GP	-6.8×10^{-4}	-0.125	69.5
	MetG	52.43	0.524	NG
	MetP	-48.76	-0.487	NG
	AVG			62.94
	STD			61.49

Table 3
COMPARISON OF THE SPECIES RECOVERING RATES FOR A [G(P)1] COMPARATIVELY TO A [G(P)0] (NOMINAL CONDITIONS OF HIGH BALLAST CELL OF TABLE 2 BUT WITH [G]_s = 1 nM), NG= NEGLIGIBLE.

Notations G= a gene (DNA); P= a protein; M= RNA; GP= the inactive complex of G with P; NutP and NutG are substrates used in the synthesis of metabolites MetP and MetG used for P and G synthesis; C_j = species j concentration.

key-species present relatively short recovering rates, and negligible for the other GERM species (table 3). As proved by the examples of this chapter, the WCVV construction, with including the cell ballast effect, and the G/P mutual autocatalysis, are more flexible and adaptable to environment constructions, being able to smooth the influence of the environmental changes on the cell homeostasis.

d) The effect of a regulatory element involving TF-s in GERM

The effect of regulatory elements on GERM PI-s was proved by Maria [1,3] by using several simple GERM -s. As revealed by the obtained results:

-the dynamic regulatory efficiency increases in the order: {G(P)0} (no buffering reaction) < [G(P)1] (one buffering reaction) < [G(P)1; M(P)1] (cascade control also at the M level) < [G(PP)2] (two buffering reactions, with dimeric TF= PP). Some GERM modules reported an increased PI flexibility, due to adjustable intermediate TF species levels. This is the case, for instance, of adjusting [M]_s in the module and of [G(P)n; M(P)n] and of [PP]_s in modules [G(PP)n]. Optimal levels of these species can be set accordingly to various optimization criteria, rendering complex regulatory modules to be more flexible in reproducing certain desired cell-synthesis regulatory properties.

-the dynamic regulatory efficiency (table 1) decreases in the order:

-- $\text{Min}(\tau_{rec})_P: G(PP)_2 > [G(P)_1; M(P)_1] > G(P)_1 > G(P)_0$

-- $\text{Min}(STD): G(PP)_2 > G(P)_1 > \{G(P)_0; [G(P)_1; M(P)_1]\}$

- the stationary regulatory efficiency (table 1) decreases in the order;

- $\text{Min}\left(\frac{C_P/C_{PS}}{C_{NutG}/C_{NutG,S}}\right): G(PP)_2 > \{G(P)_1; [G(P)_1; M(P)_1]\} > G(P)_0$

e) The effect of the cell ballast on the GERM efficiency.

When constructing WCVV more or less simplified, it is important to know what is the minimum level of simplification to not essentially affect the holistic properties of the cell. This paragraph proves why it is essential to include in a WCVV model the so-called *cell ballast*, that is the sum of concentrations of all species, which are not accounted in the ODE mass balance of the analysed species.

In other words, the PI-s of a GERM are the same in a *rich* cell of high cell content (ballast), compared to those from a *poor* cell of low cell content (ballast)? The answer is no. To simply prove that, one considers a [G(P)1] module placed in an *E. coli* cell with two different nominal conditions given in table 2: a high-ballast cell, and a low-ballast cell. To not complicate these models, the Lumped $\sum_j C_{MetG_{j,s}}$ and the Lumped $\sum_j C_{MetP_{j,s}}$ play also the role of ballast, being set to large levels compared to other cell species. Species trajectories after a -10% impulse perturbation in the [P]_s of 1000 nM at an arbitrary t=0 are presented in the figure 3 of [13].

Selecting appropriate Lumped $\sum_j C_{MetG_{j,s}}$ and $\sum_j C_{MetP_{j,s}}$ Lumped required understanding their effect on cell properties. Low concentrations relative to the total number of other molecules in the cell afforded shorter $(\tau_{rec})_P$ for the key-protein. For instance, in [G(P)1], with

$\sum_j C_{MetP_{j,s}} = 3000 \text{ nM}; \sum_j C_{MetG_{j,s}} = 2000 \text{ nM}$ and $\sum_j C_j = 12001 \text{ nM}$, the resulted $(\tau_{rec})_P$ was 103 min, and $(\tau_{rec})_G$ was 223 min after a -10% impulse perturbation in the [P]_s of 1000 nM at an arbitrary t = 0. Whereas for $\sum_j C_{MetP_{j,s}} = 3 \times 10^8 \text{ nM}; \sum_j C_{MetG_{j,s}} = 3 \times 10^6 \text{ nM}$ and $\sum_j C_j = 6.06 \times 10^8 \text{ nM}$, the resulted $(\tau_{rec})_P$ was 133 min, and $(\tau_{rec})_G$ was 93 min after a -10% impulse perturbation in the [P]_s of 1000 nM at an arbitrary t=0.

We refer to this as the *Inertial Effect*. It arises because the invariance relationships described above require that larger rate constants for P and G synthesis be used to counterbalance lower [Met_p] and [Met_G], and these constants are determinants for key-species recovering rates after a perturbation $(\tau_{rec})_j$. On the other hand, when metabolite concentrations were low, perturbation of cell volume was greater than when they were high (volume increase plots not presented here). The attenuation of perturbation-induced volume changes by large metabolite concentrations is called the *Ballast Effect*. Ballast diminishes the indirect perturbations otherwise seen in concentrations of all cellular components. Thus, [G] was perturbed far less, as a result of an impulse perturbation in [P], for the cell containing higher metabolite concentrations than for that containing lower metabolite concentrations (fig. 3 of [13]). Thus, increasing metabolite concentrations

attenuates the impact of perturbations on all cellular components but negatively influences recovery times.

In fact, the Ballast Effect shows how all components of the cell are interconnected via volume changes. It represents another holistic property of cells, and it is only evident with only variable-volume WCVV modelling. Its importance is related to the magnitude of perturbations and the total number of species in a cell. For a single perturbation in real cells, the *Ballast effect* will be insignificant due to the large number of total intracellular species. However, the sum of all perturbations experienced during a cell cycle might be significant.

f) The effect of GERM complexity on the resulted GRC efficiency, when linking GERM-s

When developing a suitable WCVV kinetic model of a GERM, especially when chains of GERM-s are modelled, it is important to adopt a suitable reduced model structure by means of an acceptable trade-off model simplification- vs.-model quality (i.e. model adequacy [1]).

Adoption of too complex reaction pathways is not desirable when developing cell simulators, these structures being difficult to be modelled and difficult to be estimated by using ODE kinetic models, due to the very large number of parameters. Beside, cell model constructions with too complex cell modules lead to inoperable huge models impossible to be identified and used for cell design purposes. The alternative is to use reduced ODE models with a number of lumped species and reactions enough to fairly reproduce the experimental data, but enough simple to make possible a quick dynamic analysis of the metabolic process and of its regulation [1].

To exemplify how a suitable tradeoff between model simplicity and its capabilities can be obtained, one consider the problem of adequate and efficient linking of two GERM-s (expression of G1/P1 and G2/P2 pairs) such that the resulted GRC to present optimal PI-s. To solve this problem, Maria [1] compared two linking alternatives:

Variant A; [G1(P1)]+[G2(P2)1] (10 individual and lumped components)

Variant B; [G1(P1)1; M1(P1)]+[G2(P2)1; M2(P2)1] (14 individual and lumped components)

Tests have been made by using the nominal conditions of table 2, the high ballast cell case, with [P1]s = 1000 nM. [P2]s=100nM. [G1]s = [G2]s = 0.5 nM, $\sum_j C_{M_{Pj}} = 3 \times 10^8 \text{ nM}$; $\sum_j C_{MetG_{j,s}} = 10^6 \text{ nM}$. By evaluating various PI-s of the GRC including the two linked GERM-s, the following conclusion is derived: in spite of a slightly more complex structure (14 vs. 10 individual and lumped components, and two more buffering reactions), the GRC **variant B** presents much better PI-s, that is: the key-species recovering times after an impulse perturbation of a low AVG and STD indices, species QSS concentrations low sensitivity vs. environmental perturbations. Thus, the right choice of the GERM structures in a GRC is an important modelling step. This example proves how, with the expense of a little increase in the model complexity (4 additional species and 2 buffering reactions), the cascade control of [G(P)n; M(P)n] gene expression modules offer superior regulatory properties of the design GRC, with properties easily adjustable via model parameters, including a better species synchronization when coping with perturbations (i.e. low AVG, STD indices).

g) Cooperative vs. concurrent linking of GERM-s in GRC.

When coupling two GERM modules into the same cell, such as the nutrients, and metabolites in the G/P syntheses

are roughly the same. The modelling problem is what alternative should be chosen? A competitive scheme (due to the common substrate), or a cooperative scheme, the GERM assisting each other? For instance, in the figure 2 of [13], there are tested three alternatives (by using the nominal high-ballast cell condition of Table 2 [1]):

Variant A: Competitive (on common metabolites) linking of [G1(P1)0]+[G2(P2)0];

Variant B: Simple cooperative linking of [G1(P1)0]+[G2(P2)0] modules. P1 is permease and metabolise for both GERM-s; P2 is polymerase for replication of both G1 and G2 genes.

Variant C: cooperative linking (identical to variant B), but adding buffer reversible regulatory reactions to modulate the G1, G2 catalytic activity in the modules [G1(P1)1]+[G2(P2)1].

The tests performed by Maria [1] led to very interesting conclusions:

-In the Variant A, one links two modules [G1(P1)0] + [G2(P2)0], both ensuring regulation of two protein synthesis (P1 and P2), in an uncooperative disconnected way (fig. 2 of [13]). For this hypothetical system, synthesis of P1/G1 and P2/G2 from metabolites is realized with any interference between modules ($C_{P1s} = 1000 \text{ nM}$, $C_{P2s} = 100 \text{ nM}$, $C_{G1s} = 1 \text{ nM}$, $C_{G2s} = 1 \text{ nM}$). The only connection is due to the common cell volume to which both protein syntheses contribute. If one checks this system for stability, by applying a $\pm 10\%$ impulse perturbation in C_{P1s} , it results an *unstable system*, evolving toward the decline and *disappearance of one of the proteins* (i.e. those presenting the lowest synthesis rate [1]). Consequently, the homeostasis condition is not fulfilled, the cell functions cannot be maintained, and the disconnected protein synthesis results as an unfeasible and less plausible GERM linking alternative.

-In the Variant B, the simple cooperative linking of [G1(P1)0]+[G2(P2)0] modules in figure 2 of [13] ensures specific individual functions of each protein, i.e. P1 lumps both the permeases and metabolases, while P2 is a polymerase.

-In the Variant C, the simple cooperative linking of [G1(P1)0] + [G2(P2)0] system of the Variant B has been improved by adding simple effectors for gene activity control. In the cooperatively linked system, thus resulting the system [G1(P1)1]+[G2(P2)1], (fig. 2 of [13]), the effectors P1 and P2 act in two buffering reactions, $G1+P1 \Leftrightarrow G1P1$, and $G2+P2 \Leftrightarrow G2P2$, with the stationary states $C_{G1s} = C_{G1P1s} = 1/2 \text{ nM}$, and $C_{G2s} = C_{G2P2s} = 1/2 \text{ nM}$ ensuring maximum dynamic PI-.

-The same rule of linking GERM-s can continue in the same way, for instance [1], also involving [G(PP)n] modules, the effectors being the dimers PP, acting in n buffering reactions, $G+PP \Leftrightarrow GPP$, etc., with the stationary states $C_{Gs} = C_{GPPs} = 1/2 \text{ nM}$, the rate constants being estimated from the stationary concentrations ($C_{j,s}$ in table 2, adopted from the *E. coli* cell) and by imposing regulatory optimal characteristics given by the criterion eqn. (10) of [13]. The $k_{dis} \gg D$ in eq. (11) of [13] has been adopted, as being ca. $10^7 \cdot D$, while system optimization with criterion eq. (10) of [13] leads to small values for C_{PPs} (i.e. the active parts of dimmers; [1]).

The all three systems' stability and dynamic regulatory characteristics have been determined by studying the QSS-recover after a $\pm 10\%$ C_{P1s} impulse perturbation. The results, presented by Maria [1], reveal the following aspects concerning the systems A, B, C:

i) All three systems are stable [$\max(\text{Re}(\lambda_i)) = -D < 0$] (where λ_i are the eigenvalues of the ODE model Jacobian. Each system recovers after a dynamic perturbation in C_{P1s} ,

It results that the cooperative module linking is superior as PI, by preserving specific functions of each protein inside the cell, is a viable solution ensuring system homeostasis.

ii) The system is as better regulated as the effector is more effective (the use of multiple buffering reactions, with dimeric TF, and a cascade control of the expression (not presented here)).

iii) The use of efficient effectors and multiple regulation units can improve very much the dynamic index, in the following ranking: $G(P)n < G(PP)n < [G(P)n; M(P)n']$.

iv) Dynamic perturbations affect rather species present in small amounts inside the cell, while recovering times for other species (e.g. metabolites MetP, MetG) are negligible.

v) By adding to the Variant B, i.e. to the simple cooperative linking of $[G1(P1)0] + [G2(P2)0]$ regulation effectors of type $G(P)I$ (i.e. Variant C), or even more effective of type $G(PP)I$, the whole system recovers faster after an impulse perturbation (i.e. the resulted AVG is lower), but also the species interconnectivity increases (i.e. STD becomes lower), with a positive effect on the target proteins P1, P2 synthesis regulation.

vi) The stationary regulatory indices (sensitivities of states vs. nutrients) follow the same trend.

In the same way, the regulatory network GRC design procedure can be continued, by accounting for new proteins (and their corresponding GERM-s). For instance, in the simplified representation of Maria [1], a 3-rd GERM for P3 synthesis can be added to the **Variant C**, by allocating specific functions to the P1, P2, P3 proteins, as follows: P1 and P3 lumps permease and metabolase enzymes, which ensure nutrient import inside the cell, and their transformation in gene-metabolites (MetG1-MetG3) and protein-metabolites (MetP1-MetP3) respectively. Protein P2 lumps polymerases able to catalyze the genes G1, G2, G3 production. If one considers the simplest effector case, the resulted cell GRC includes three modules $[G1(P1)1] + [G2(P2)1] + [G3(P3)1]$, which regulate the synthesis of P1, P2 and P3, in a cooperative interconnected way by preserving protein functions.

h) The optimal value of TF

It is self-understood that, in a realistic WCVV model, the holistic properties of the cell should be preserved, and modulated via model structure and parameters. One of the cell modelling principles postulates that the concentration of intermediates used in the GRC-s should be maintained at a minimum level to not exhaust the cell resources, but at the same time at an optimal value to maximize the GERM PI-s. Such a GRC property was simply proved by Maria [9] in the case of a genetic switch of *E. coli* cell, modelled in a WCVV approach. The two considered self- and cross-repressing gene expression modules are of type $[G2(P2P2)1(P3P3)1] + [G3(P3P3)1G3(P2P2)1]$. These well chosen GERM-s allow superior regulatory properties of the genetic switch (i.e. adjustable switch certainty, good responsivity, good dynamic and stationary efficiency). Besides, Maria [9] in-silico proves that it exist an optimal levels of the TF-s of type $[P2P2]_s$, $[P3P3]_s$ that are associated to the optimal holistic regulatory properties of the GRC, and that these TF-s are rather dimmers than monomeric molecules. These in-silico obtained results have been confirmed by the literature data.

i) Some rules to be followed when linking GERM-s.

Cell regulatory networks, and in particular protein synthesis regulation, are poorly understood. The modular approach of studying the regulation path, accounting for its structural and functional organization, seems to be a

promising route to be followed. Because a limited number of GERM types exists, individual regulatory modules GERM-s can be separately analyzed and checked for efficiency in conditions that mimic the stationary and perturbed cell growing conditions. A GERM is as efficient as the dynamic and stationary performance indices (table 2) are more favourable and sensitivity to perturbations is lower. Then, they are linked accordingly to certain rules to mimic the real metabolic process, by ensuring the overall network efficiency, system homeostasis, and protein functions. Module linking rules are not fully established. It seems that modular network is hierarchically organised, and includes a large number of compounds with strong interactions inside a module and weaker interactions among modules, so that the whole cell system efficiency can be adjusted. [1]. By testing several ways to link GERM-s, Maria [1] advanced some rules:

-When the GRC is constructed, linking reactions between GERM modules are set to be relatively slow comparatively with the module core reactions. In such a manner, individual modules remain fully regulated, while the assembly efficiency is adjusted by means of linking reaction and intermediate species, TF levels. To preserve the individual regulatory capacity, the magnitude of linking reactions would have to decline as the number of linked modules increases.

-When linking regulatory modules, the main questions arise on the connectivity mechanism and on the cooperative vs. uncooperative way of which proteins interact over the parallel/consecutive metabolic path. In spite of an apparent 'competition' for nutrient consumption, protein synthesis is a closely cooperative process, due to the specific role and function of each protein inside the cell. In a cooperative linking, common species (or reactions) are used for a cross-control (or cross-catalysis) of the synthesis reactions. Thus, the system stability is strengthened, while species inter-connectivity is increased leading to a better treatment of perturbations.

-Protein interactions are very complex, being part of the cell metabolism and distributed over regulatory network nodes. There are many nodes with few connections among proteins and a small, but still significant, number of nodes with many proteic interactions. These highly connected nodes tend to be essential to an organism and to evolve relatively slowly. At a higher level, protein interactions can be organized in 'functional modules', which reflect sets of highly interconnected proteins ensuring certain cell functions. Specific proteins are involved in nutrient permeation (permeases), in metabolite synthesis (metabolases), or in gene production (polymerases). In general, experimental techniques can point-out molecular functions of a large number of proteins, and can identify functional partners over the metabolic pathways. Moreover, protein associations can ensure supplementary cell functions. For instance, enzyme associations lead to the well-known 'metabolic channelling' (or tunnelling) process, that ensures an efficient intermediate transfer and metabolite consecutive transformation without any release into the cell bulk phase [1].

-It results that an effective module linking strategy has to ensure the cell-functions of individual proteins and of protein associations over the metabolic synthesis network. As a general observation, even not presenting common reactions, the modules are anyway linked through the cell volume (to which all cell species contribute) and due to some intermediates controlling the GRC. The WCVV model is able to account for such cell regulatory characteristics.

-A popular strategy for building complex and realistic cell models is to analyze independent modules or groups of closely interacting cellular components, and then link

them. The WCVV may facilitate this strategy. Each module could be modelled as a separate *entity* growing at the actual rate of the target cell. The volume of newborn cell and the environment characteristics could match those of the target. To allow this, and to reproduce the cell ballast effect, lumped molecular species could be defined into each cell where a GERM is tested, in amounts equal to those of the target cell minus those due to the components of the module. Thus, each tested cell carrying a certain defined GERM would grow at the same observed rate. As a result, linking GERM-s would be a seamless process requiring only that the ballast level to be kept at its experimental level.

-The WCVV modeling approach demonstrates that each and every component of a cell affects, and is affected by, all other cellular components. Indirect interconnectivities arise because all components in a cell contribute to cell volume, and cell volume influences component concentrations. Thus, perturbations in one component reverberate throughout the cell. The importance of these indirect relationships will vary with the diversity and complexity of cellular components. Increasing numbers add ballast to the cell, minimizing these indirect relationships, while increasing diversity allows individual metabolites to be present at lower concentrations, improving dynamic responses of GERM-s and GRC to perturbations. Another issue, thus far unexamined, is how specific types of interconnectivities affect the regulatory behavior of cells. This could be probed using methods developed by Ross, Arkin and coworkers to deduce connectivities in biochemical pathways from the effects of impulse perturbations [10].

-When modeling complex operon structures, simple GERM structures should be adopted to not complicate too much the WCVV model. The default GERM is the [G(P)1]. But according to the experimental data and operon structure of interactions among genes and proteins, more complicated GERM construction can be tested (see the examples of Maria [8,9,11]).

j) The effect of cascade control on the GERM efficiency.

Among GERM-s reviewed and tested by Maria [1, 3, 7], the most significant are the [G(P)n] of effectiveness nearly linearly increasing with the number (n) of buffering reactions, then the [G(PP)] of a more pronounced regulatory efficiency due to the used dimeric TF (that is PP). The most effective are the GERM-s of a cascade control of expression, by means of buffering reactions applied at the G, and M catalyst level, that is [G(P)n; M(P)n'] (fig. 1 of [13]). The superiority of the [G(P)n; M(P)n'] gene expression structure has been in-silico proved by Maria [1, 3, 7]. The conclusions are the followings:

- the very rapid buffering reactions have been proved to be very effective regulatory elements, by quickly adjusting the active/inactive G/GP or M/MP ratios thus coping with the perturbations;

- numerical tests revealed that the P.I.-s of the compared GERM-s increase in the approximate order:

MC[G0] (0 regulatory element) < MC[G(P)1] (1 regulatory element) < MC[G(P)1;M(P)1] (2 regulatory elements) < MC[G(PP)2] (3 regulatory elements)

Roughly, the obtained improvement of P.I. per regulatory element is of ca. 1.3 (under WCVV modelling), while the obtained improvement of P.I. per regulatory element is of ca. 2.5 (under constant volume modelling hypothesis). It clearly appears that the WCVV modelling framework is more realistic, the default constant volume approach tending to over-estimate the P.I.-s of GERM-s.

Conclusions

The reviewed case studies of WCVV modular kinetic models of GERM-s proved that the chemical and biochemical engineering principles, together with the control theory of the nonlinear systems are fully applicable to modelling complex metabolic cell processes, including the sophisticated gene expression regulatory circuits controlling the cell enzymes syntheses and all metabolic fluxes. The ODE kinetic models with continuous variables are fully feasible alternatives to well describe the cell response to stationary or dynamic continuous perturbations from the environment.

In fact, cell process modelling has to 'translate' from the 'language' of molecular biology to that of mechanistic chemistry and mathematics/computing languages, by preserving the structural, functional, and timing hierarchy of the cell components and functions. To avoid very extended ODE cell kinetic models, difficult to be identified, and to be used, the model structure should ensure a satisfactory trade-off between model simplicity and its predictive quality.

The study also proves the importance of using a WCVV modelling environment to get more realistic simulation results. By contrary, the use of the classical default constant-volume cell system leads to distorted results, tending to overestimate the P.I.-s of GERM-s [3] compared to [1, 7, 9].

The cell simulator become more and more valuable tools in designing GMO with desirable characteristics, or for obtaining micro-organisms cloned with desirable plasmids with important applications in industry (new biotechnological processes, optimization of bioreactors [11, 12], production of vaccines), or in medicine (such as therapy of diseases, gene therapy, new devices based on cell-cell communicators, biosensors), etc.

Acknowledgments. Authors are grateful to Prof. Dr. Stefan Szedlacsek from the Department of Enzymology of the Institute of Biochemistry of the Romanian Academy Bucharest for the very constructive suggestions and discussions.

References

1. MARIA, G., Chemical and Biochemical Engineering Quarterly, **19**, 2005, p. 213-233.
2. SEWELL, C., MORGAN, J., LINDAHL, P., J. theor. Biol., **215**, 2002, p. 151-167.
3. MARIA, G., Chemical and Biochemical Engineering Quarterly **17**, 2003, p.99-117.
4. YANG, Q., LINDAHL, P., MORGAN, J., J. theor. Biol., **222**, 2003, p. 407-423.
5. HEINRICH, R., SCHUSTER, S., The regulation of cellular systems, New York, Chapman & Hall, 1996.
6. MARIA, G., Chemical and Biochemical Engineering Quarterly, **21**, 2007, p. 417-434.
7. MARIA, G., Chemical and Biochemical Engineering Quarterly, **20**, 2006, p. 353-373.
8. MARIA, G., Chemical & Biochemical Engineering Quarterly, **28**, 2014, p. 83-99.
9. MARIA, G., Asia-Pacific Journal of Chemical Engineering, **4**, 2009, p. 916-928.
10. VANCE, W., ARKIN, A., ROSS, J., Proceedings of the National Academy of Science USA, **99**, 2002, p. 5816-5821.
11. MARIA, G., LUTA, I., Computers & Chemical Engineering, **58**, 2013, p. 98-115.
12. MARIA, G., LUTA, I., MARIA, C., Chemical Papers, **67**, 2013, p. 1364-1375.
13. MARIA, G., SCOBAN, A.G., Rev. Chim. (Bucharest), **67**, no.12, 2017, p. 3027

Manuscript received: 15.02.2016