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

I. Generalities

GHEORGHE MARIA*, ANDREEA GEORGIANA SCOBAN

University Politehnica of Bucharest, Department of Chemical & Biochemical Engineering, Bucharest, Romania

The whole-cell simulation of cell metabolic processes under considering a variable-volume modelling framework has been reviewed to prove their advantages when building-up modular model structures that can reproduce complex protein syntheses inside cells. The more realistic whole-cell-variable-volume (WCVV) approach is exemplified when developing modular kinetic representations of the homeostatic gene expression regulatory modules (GERM) that control the protein synthesis and homeostasis of metabolic processes. In the first part, the general concepts of the WCVV modelling is presented, while in the second part of the paper, 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: i) to characterize the module efficiency, species connectivity and system stability; ii) to build-up modular regulatory chains of various complexity; iii) to prove feasibility of the cooperative vs. concurrent construction that ensures an efficient gene expression, system homeostasis, proteic functions and a balanced cell growth during the cell cycle; iv) to prove the effect of the whole-cell content ballast in smoothing the effect of internal/external perturbations on the system homeostasis.

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

Living cells are self-replicating complex biological structures, able to convert environmental nutrients to replicate the cell content in exactly one cell cycle. Cells present such a highly sophisticated structure, involving O(10^{3-4}) components, O(10^{3-4}) transcription factors (TF-s), activators, and inhibitors, and at least one order of magnitude higher number of (bio)chemical reactions, all ensuring a fast adaptation of the cell to the changing environment. Cell is highly responsive to the environmental stimuli and highly evolvable by self-changing its genome/ proteome and metabolism to get an optimized and balanced growth with using minimum of resources (nutrients/substrates).

Cells have a hierarchic organization (structural, functional and temporal), which is a characteristic of the living matter in general: i) the *structural hierarchy* include all cell components from simple molecules (nutrients, saccharides, fatty acids, aminoacids, simple metabolites), then macromolecules or complex molecules (lipids, proteins, nucleotides, peptidoglycans, coenzymes, fragments of proteins, nucleosides, nucleic acids, intermediates), and continuing with well-organized nano-structures (membranes, ribosomes, genome, operons, energy harnessing apparatus, replisome, partitioning apparatus, Z-ring etc. [1]). To ensure self-replication of such a complex structure through enzymatic metabolic reactions using nutrients (Nut), metabolites (Met), and substrates (glucose/fructose, N-source, dissolved oxygen and micro-elements), all the cell components should be associated with specific functions into the cell, following ii) *functional hierarchy* according to the species structure; e.g. source of energy, intermediates. Sauro and Kholodenko [2] provided examples of biological systems that have evolved in a modular fashion and, in different contexts, perform the same basic functions. Each module, grouping several cell components and reactions, generates an identifiable function (e.g. regulation of a certain reaction, gene expression, etc.). More complex functions, as regulatory networks, synthesis networks, or metabolic cycles can be built-up from basic *building blocks*. Such a building blocks structure is very tractable when developing cell reduced dynamic models by defining various metabolic sub-processes, such as: regulatory functions for the enzymatic reactions, energy balance functions for ATP/ ADP/AMP renewable system, electron donor systems of the NADH, NADPH, FADH, FADH2 renewable components, hydrophobic effects; or functions related to the metabolism regulation (regulatory components / reactions of the metabolic cycles, gene transcription and translation); genome replication / gene expression regulation (protein synthesis, storage of the genetic information, harness cell energy), functions for cell cycle regulation (nucleotide replication and partitioning, cell division); iii) the wide-separation of time constants of the metabolic reactions in the cell systems is called *time hierarchy*. Thus, the reactions are separated in slow and fast according to their time constant (see definition in [3]); in fact, only fast and slow reactions are of interest, while the very slow processes are neglected or treated as parameters (such as the external nutrient or metabolite evolution). Aggregate pools (combining fast reactions) are usually used in building-up cell dynamic models in a way that intermediates are produced in a minimum quantity and consumed only by irreversible reactions. All cell processes obey a certain succession of events, while stationary or dynamic perturbations are treated by maintaining the cell

* *email: gmaria99m@hotmail.com*

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

components homeostasis, the recovering or transition times after perturbations being minimum.

To model such an astronomically complex cell system with a detailed kinetic model is practically impossible even if expandable bio-molecular data are continuously added in –omics databanks (e.g. KEGG, EcoCyc, Prodoric, Brenda, CRGM, NIH [4]). However, as underlined by Tomita et al. [5], and Tomita [6], whole-cell (WC) simulation of metabolic processes with mechanistic kinetic models of continuous variables, represents the grand challenge of the 21st century. Such a huge effort is justified by the very large number of immediate application: design genetically modified micro-organisms (GMO) with desirable characteristics to be used in industry (new bio-technological processes, production of vaccines). A large number of GMO applications are in medicine, such as therapy of diseases (gene therapy), new devices based on cell-cell communicators, biosensors etc.

As underlined by Tomita [6] Computer models and *in silico* experiments are necessary to understand and predict phenotypes of the cell, especially when they are polygenic phenotypes. After all, most biological and pathological phenomena in which the pharmaceutical industry has a great interest, such as cancer and allergy, are polygenic.

Comparatively to chemical systems, and lacking of enough and reproducible experimental data, the cell biochemical processes present the advantage of being approached in a modular way; every subsystem can be separately studied/modelled [6-10], that is: *the central carbon metabolism* phosphotransferase PTS-system for glucose transport into the cell, the pentose- phosphate pathway PPP for nucleotides and aminoacids production, and tricarboxylic acid cycle TCA, [10]); regulation of gene expression, cell cycle, signal transduction, and various metabolic pathways [6]. However, although these models made significant contributions to the development of *in silico* biology, the programs were only able to handle only specific subsystems, and it was difficult to combine different subsystem models into one single-cell model [6]. Encouraging results have been reported for the design of artificial gene networks, for reprogramming signaling pathways, for refactoring of small genomes, or re-design of metabolic fluxes with using switching genes. By assembling functional parts of an existing cell, or by designing new gene regulatory networks on a modular basis, it is possible to reconstitute an existing cell (the so-called integrative understanding) or to produce novel biological entities with new properties. The genetic components may be considered as building blocks because they may be extracted, replicated, altered, and spliced into the new biological organisms [7].

In a whole-cell approach, due to the mentioned insuperable detailed modelling difficulties, various cell modelling alternatives have been developed over decades [7].

The structure-oriented analyses ignore some mechanistic details and the process kinetics, and use the only network topology, the so-called 'Metabolic Control Analysis' (MCA) being focus on using various types of sensitivity coefficients (the so-called 'response coefficients'), which are quantitative measures of how much environmental perturbations (influential variable x_j) affects the cell-system states y_j (e.g. r = reaction rates, J = fluxes, C = concentrations) around the steady-state (QSS, of index 's'), i.e. $\left[(\partial y_i/\partial y_{is})/(\partial x_j/\partial x_{js})\right]_s$. The systemic response of fluxes (i.e. stationary metabolic reaction rates), or of 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 [7].

The Boolean or topological approach can not characterize the dynamics of the metabolic processes.

Even if regulation mechanisms are not fully understood, metabolic regulation at a low-level is generally better clarified, and conventional ordinary differential ODE kinetic models with continuous variables, with a mechanistic description of reactions tacking place among individual species (proteins, mRNA, intermediates etc.) have been proved to be a convenient route to analyse continuous metabolic / regulatory processes and continuous perturbations. When systems are too large or poorly understood, coarser and more phenomenological kinetic models may be postulated (e.g. protein complexes, metabolite channelling etc.). In dynamic models, only essential reactions are retained, the model complexity depending on measurable variables and available information. An important problem to be considered is the distinction between the qualitative and quantitative process knowledge, stability and instability of involved species, the dominant fast and slow modes of process dynamics, reaction time constants, macroscopic and microscopic observable elements of the state vector. Such ODE kinetic models can be useful to analyse the regulatory cell-functions, both for stationary and dynamic perturbations, to model cell cycles and oscillatory metabolic paths, and to reflect the species inter-connectivity or perturbation effects on cell growth [7].

To reduce from the modelling effort, structured reduced WC-kinetic models are used according to the available –omics/experimental data and utilization scope, including lumped species and/or reactions. These modular models present the advantage of being easily extensible. The rate constants are usually estimated from stationary data (see Appendix). As revealed by Visser et al. [12], Traditionally, kinetic metabolic models are based on mechanistic rate equations, which are derived from in-vitro experiments. However, due to large differences between in-vivo and in-vitro conditions, it is unlikely that the in-vitro obtained parameters are valid in-vivo. Thus, the kinetic parameters must be adjusted, using data on in-vivo metabolite levels and fluxes obtained in dynamic experiments. Due to the complexity of mechanistic rate equations, which often contain a considerable amount of parameters, this requires a large experimental and mathematical effort. Such an approach is computational tractable, a large number of chemical engineering, non-linear system control algorithms being available.

Besides, application of lumping rules to metabolic processes must also account for physical significance, species interactions, and for preserving the *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 *Jacobian* matrix J case, see Appendix for J definition) has been proved to be insufficient.

Applications of such cell dynamic simulators, especially of GERM chains (the GRC) controlling the cell metabolism are immediate: design new micro-organisms of desirable characteristics; *in-silico* re-programming the cell metabolism; design of biosensors; drug target release; industrial bioprocess optimization and control using GMO; gene therapy; optimal cell cloning, etc. Consequently, the

cell metabolism can be changed by modifying/designing GRC-s, thus conferring new properties/functions to the mutant cells (i.e. desired 'motifs'), while engineered/ synthetic gene circuits can be designed by using the Synthetic Biology tools [12].

Of course, the use of reduced metabolic kinetic models present a series of disadvantages, such as: loss in system flexibility (due to the reduced number of considered intermediates and species interactions); possibility to get multiple (rival) reduced models of proximate characteristics for the same cell system, difficult to be delimited; loss in the model prediction capabilities; lumped model parameters can lack of physical meaning; a loss / alteration of systemic / holistic properties (e.g. cell system stability, multiplicity, sensitivity, regulatory characteristics).

Starting from an available extended kinetic model, classical chemical engineering rules can be applied to reduce their structure [8] aiming at:

(i) Reducing the list of species, by eliminating unimportant components and/or lumping some species (by using various measures, e.g. small values for the product of target species of index i lifetime $LT_i = -1/J_{ii}$ and its production rate r_i, where the Jacobian elements are $J_{ik} = \partial f_i(C,k)/\partial C_k$, where f_i are the right-side functions of the ODE kinetic model (1);

(ii) Reducing the list of reactions, by eliminating unimportant side-reactions and/or assuming quasi-equilibrium for some reaction steps (or using sensitivity measures of rate constants, such as ridge selection, principal component analysis, time-scale separation, etc.);

(iii) Decomposing the kinetics into fast and slow 'parts' allowing a separate study and application of the quasi-steady-state-approximation (QSSA) to reduce its dimensionality [8].

When the ODE kinetic model is linear in parameters, then the reduction procedure of Maria [13] can be applied by preserving the system Jacobian invariants (eigenvalues, eigenvectors).

Due to the modular functional organization of the cell, 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 [8]. 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 to recreate 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. Such a procedure is very suitable for modelling genetic regulatory circuits (GRC) by linking gene expression regulatory modules (GERM-s) in such a way to maintain the cell homeostasis, that is to maintain relatively invariant species concentrations despite perturbations. [4,7-8,14-15].

A potential application of lumped modular GRC 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. By inserting new GRC-s into organisms, one may 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 considered as building blocks because they may be extracted, replicated, altered, and spliced into new biological organisms. Combination of induced motifs in modified cells one may create potent applications in industrial and medical fields, e.g. the production of biosensors used in medicine or environmental engineering applications. Design of modular GRC-s must account for some properties (see also below): 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 [8].

The emergent *Synthetic Biology* [4,9] 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. By assembling functional parts of an existing cell, such as promoters, ribosome binding sites, coding sequences and terminators, protein domains, or by designing new gene regulatory networks on a modular basis, it is possible to reconstitute an existing cell or to produce novel biological entities with new properties. One particular application of such dynamic models is the study of genetic regulatory circuits (GRC), in order to predict the way in which biological systems are self-regulated and respond to signals. The emergent field of such efforts is the so-called *gene circuit engineering* and a large number of examples [7] 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 [4]: toggle-switch, i.e. mutual repression control in two gene expression modules, and creation of decision making branch points between on/off states according to the presence of certain inducers; hysteretic GRC behaviour, that is a bio-device able to behave in a history-dependent fashion, in accordance to the presence of a certain inducer in the environment; GRC oscillator producing regular fluctuations in network elements and reporter proteins, and making the GRC to evolve among two or several quasi-steady-states; specific treatment of external signals by controlled expression such as amplitude filters, noise filters or signal / stimuli amplifiers; GRC signalling circuits and cell-cell communicators, acting as programmable memory units. [4].

One key element to such cell dynamic simulators is the adopted kinetic model of the GERM-s. from the large number of proposed mechanistic models from literature [4,7,9,14,16]. When constructing a GRC for a certain cell metabolic pathway, there are two problems which must be addressed properly: I) how to choose the suitable GERM structures of the GRC chain, by screening among alternatives [4,9,14,16,17-19] based on their regulatory properties (i.e. quantitative performance indices P.I.) matching with the experimental data; and ii) what rules to be applied to link such GERM-s to reproduce the cell system holistic behaviour.

The aim of this paper is to review and to extend the GRC modelling investigation of Maria [4,7-9,14,15] by pointing-out some milestones that should be considered when developing effective GRC in a WCVV modelling approach. While the first part of the study will present the general concepts related to WCVV modelling, the second part will

offer some examples by using two GERM-s of simple structure extracted from the *lac operon* of *E. coli* [8].

WCVV modelling framework

From the mathematical point of view, development of a (bio)chemical kinetic model of continuous variables for a (bio)chemical process implies writing a set of differential mass balances of the system state, i.e. biological/chemical species of concentrations C vector, that is by default:

$$\frac{1}{V(t)}\frac{dn_j}{dt} = \sum_{i=1}^{s} \upsilon_{ij} r_i(\boldsymbol{n}/V, \boldsymbol{k}, t) = f_j(\boldsymbol{C}, \boldsymbol{k}, t)\ ;\ ns = \text{size}(\boldsymbol{C});\ np = \text{size}(\boldsymbol{k}); \quad (1)$$

where:

C_j = cell-species j concentration;
V = cell volume;
n_j = species j number of moles;
r_j = j-*th* reaction rate;
n_s = number of species inside the cell (individual or lumped);
t = time;
π= osmotic pressure;
T = temperature;
R = universal gas constant.

When continuous variable dynamic models are used, the default modelling framework (1) is that of a constant volume/osmotic pressure system, accounting for the cell-growing rate as a *decay* rate of key-species (often lumped with the degrading rate) in a so-called *diluting* rate. Such a representation might be satisfactory for many applications, but not for accurate modelling of cell regulatory / metabolic processes under perturbed conditions, or for division of cells, distorting the prediction quality [4].

The living cell is a system of variable volume: from simple to double during the cell cycle. In a whole-cell modelling framework [6] of variable-volume WCVV [4,7-8], all cell species should be considered (individually or lumped) all species net reaction rates contributing to the cell volume increase (see the proof in the Appendix). As the cell volume is doubling during the cell cycle, this volume variability can not be neglected in the cell kinetic models describing various cell metabolic processes taking place in the cytosol. Thus, a metabolic kinetic model in a variable-volume approach should be written in a different form, defined by Aris [20] for chemical reacting systems, and Grainger and Gaffney [23] for biological systems, and promoted by Maria [4,7-9,14,15] for cell systems, that is:

$$\frac{dC_j}{dt} = \frac{1}{V}\frac{dn_j}{dt} - DC_j\ ;\ \frac{1}{V}\frac{dn_j}{dt} = r_j\ ,\ j = 1,\ldots,n_s\ ;\ D = \frac{d(ln(V))}{dt} \quad (2)$$

where:

C_j= cell-species j concentration;
V = cell volume;
n_j= species j number of moles;
r_j= j-th reaction rate;
D = cell-content dilution rate, i.e. cell-volume logarithmic growing rate;
n_s= number of species inside the cell (individual or lumped);
t = time.

The hypotheses of such a cell model are given in the Appendix.

The definition of D results from the way by which it was deducted for variable volume reacting systems [7, 20]:

$$\frac{dC_j}{dt} = \frac{d}{dt}\left(\frac{n_j}{V}\right) = \frac{1}{V}\frac{dn_j}{dt} - C_j\frac{d(ln(V))}{dt} = \frac{1}{V}\frac{dn_j}{dt} - DC_j = h_j(C,k,t)\ ;\ D = \frac{d(ln(V))}{dt} \quad (3)$$

The system Jacobian at steady-state (index s) is:

$$J_C = \left(\partial \boldsymbol{h}(\boldsymbol{C},\boldsymbol{k})/\partial \boldsymbol{C}\right)_s \quad (4)$$

Additionally, for an isotonic (of constant osmotic pressure π) and isothermal cell system a supplementary constraint is added, that is the Pfeiffers'law of diluted solutions [4,7,8,21-23]:

$$V(t) = \frac{RT}{\pi}\sum_{j=1}^{n_s} n_j(t) \quad (5)$$

which, by derivation and division with V is leading to [4,7,8]:

$$D = \frac{1}{V}\frac{dV}{dt} = \left(\frac{RT}{\pi}\right)\sum_{j}^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right) \quad (6)$$

As revealed by the Pfeffer's law eqn. (5) in diluted solutions [22], the volume dynamics D variable is linked to the molecular species dynamics under isotonic and isothermal conditions by means of relationship (6). Consequently, the dilution D results as a sum of reacting rates of all cell species (individual or lumped), The RT/π term can be easily deducted in an isotonic cell system, from the fulfilment of the following invariance relationship derived from the Pfeffer's law in diluted solutions:

$$V(t) = \frac{RT}{\pi}\sum_{j=1}^{n_s} n_j(t) \Rightarrow \frac{RT}{\pi} = \frac{V(t)}{\sum_{j=1}^{n_s} n_j(t)} = \frac{1}{\sum_{j=1}^{n_s} C_j} = \frac{1}{\sum_{j=1}^{n_s} C_{jo}} = constant \quad (7)$$

If this WCVV modelling concept is not applied, the default classical constant volume ODE kinetic modelling of type eqn. (1) has been applied, with a large number of inconveniences, related to ignoring lots of intra-cell effects, such as: the influence of the cell ballast in smoothing the homeostasis perturbations; the secondary perturbations transmitted via cell volume following a primary perturbation; the more realistic evaluation of GERM P.I.-s, and of the recovering/transient times after perturbations, etc. [4,7,8].

When elaborating a cell-regulation model, the cell-volume growing rate effect is essential for accounting several effects: continuous diluting of the cell content, cell-*resistance* to small perturbations in the internal/external components due to the inertial 'big volume' effect, indirect effect due to perturbations in concentrations on the cell-metabolism through the induced changes in the volume growing rate. The volume-growth diluting effect acts as a continuous stationary perturbation for the concentration-levels, and formally can be assimilated with a first order decay rate of all cellular species during a cell-cycle.

It follows that the cell-volume variation during a cell-growth cycle is an essential term to be accounted in every cell-modelling attempt to obtain more satisfactory predictions. Although most of reported models, both deterministic or stochastic, ignore or diminishes such effects and build-up 'constant' cell-volume equations written in terms of species concentrations, the new elaborated regulatory/whole cell-models over the last couple of years accounted the cell-volume growth in an explicit way [4,7-9,14,15,24-28].

For instance, Sewell et al. [17] included the volume-diluting effect only for the protein-concentrations through a formally first-order decay rate. Such an approach, even being satisfactory for rapid and simple predictive purposes, suffers of two major disadvantages: (i) the 'decay' rate is not considered for all the species in order to avoid high model complexity, and (ii) the 'decay' rates can report different rate constants for various species, when in reality the same diluting constant rate is reported for all the species; (iii) the 'inertial' cell-volume / large copynumber effect to smooth perturbations cannot be simply and naturally included in a constant-volume cell-model.

The same fictive decay-rate approach has been reported by Tomita et al. [5] in developing an E-cell continuous differential model with including a larger number of genes and proteins. As mentioned by Tomita [6], *the E-cell system also accepts user-defined reactions, making it capable of handling many other phenomena such as diffusion and variable cell volume.* Using the EcoCyc and KEGG databases, the authors simulate the dynamics of 127 genes/proteinic system for the *M. genitalium* cell. However, this model suffers from several drawbacks such as lack of autocatalysis effects during a cell-cycle, by considering any replication of the genome, and any cell-division process. Recently, the authors reported some improvements of the E-cell model with including the osmotic pressure balance and volume cell growth without specifying details [29].

Other models, such of Gibson and Bruck [30], avoid including the cell-volume increase effects when considering only first-order reaction terms into the model equations. However, the authors signalled that such an approximation *can create a large calculation error.*

In the present study a different way is approached when modelling a regulatory cell-module, by explicitly including separate equations for the cell-volume growth and cell-osmotic pressure, while the continuous ODE model was re-written in terms of species moles. The cross-autocatalytic effects are also included when separate protein and gene synthesis catalytic paths are considered. Moreover, external cell-factors are better accounted by separately considering the protein and gene *raw-materials* (see below formulations of GERM-s). It is also to observe that, from definition of D in eq. (3), it results the cell volume dynamics:

$$V(t) = V_o \exp(+D \,.\, t) \tag{8}$$

Cell volume doubles over the cell cycle period (t_c), with an average logarithmic growing rate of $D=\ln(2V_o/V_o)/t_c=(2)/t_c$. For stationary balanced growing conditions, species synthesis rates are equal to first-order dilution rates (DC_j), leading to time-invariant species concentrations (i.e. homeostatic conditions, $(dC_j/dt)_s = 0$), that is;

$$\left(\frac{dC_j}{dt}\right)_s = \left(\frac{1}{V}\frac{dn_j}{dt}\right)_s - D_s C_{js} = h_{js}(\boldsymbol{C_s}, \boldsymbol{k}, t) = 0\;;\; j = 1,\ldots,n_s\;;\; D_s = \left(\frac{RT}{\pi}\right)\sum_j^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right)_s \tag{9}$$

The hypotheses of such a WCVV kinetic model are given in the Appendix.

Rate constant estimation.

In the WCVV differential model, the large number of rate constants are estimated by using several methods. If the stationary cell species concentration vector C_s is known (for the individual or lumped components considered in the kinetic model), the rate constant vector k results by solving the nonlinear algebraic set (9), for every cell subsystem (e.g. a GERM), by using an effective procedure [31, 42]. As the (RT/p) term is known from the initial condition, and the number of model parameters is usually higher than the number of observed cell species, supplementary optimization rules can be applied to determine some rate constants, by imposing optimum regulatory criteria for GERM-s, such as minimum recovering time of the stationary concentrations (homeostasis) after a dynamic ('impulse'-like) perturbation in a key-species [7], and using effective solvers [31]:

$$\left[\hat{\boldsymbol{k}}, \hat{\boldsymbol{C}}_s\right] = \arg Min(\tau_P)\text{, subjected to fulfilment of eq. (9), } \left[\hat{\boldsymbol{k}}, \hat{\boldsymbol{C}}_s\right] > 0\text{, and:}$$

$$\sum_{i=0}^{n}\left[G(P_i)\right] = constant\;;\; \sum_{i=0}^{n}\left[G(PP_i)\right] = constant\;;\; \sum_{i=0}^{n}\left[L(O_i)\right] = constant\text{, etc.}$$

$$[L]_{active}/[L_{total}] = 1/2\text{, etc.;}\; \left(\sum_j^{all} C_j\right)_{cell} = \left(\sum_j^{all} C_j\right)_{env} = constant \tag{10}$$

where τ_p has been evaluated by applying a 10%$[P]_S$ impulse perturbation and by determining the recovering time with a tolerance of 1%$[P]_S$; L_i (e.g. enzymes or even genes G, P, M, etc.) is a GERM component at which regulatory element O, TF, R acts (fig. 1). To estimate $\left[\hat{k}, \hat{\hat{C}}_s\right]$, other regulatory global properties can also be used together with the constraints eqns. (7,9,10) [7,32]. The reverse reaction rate constants in the rapid buffer reactions of GERM-s, of type G+P ⇔ GP, are adopted at values five to seven orders of magnitude higher than *D* (see the proof of Maria [7]). That is because fast buffering reactions are close to equilibrium and have little effect on metabolic control coefficients. As a consequence, rate constants of such rapid reactions are much higher than those of the core synthesis and dilution rates.

A WCVV model formulation presents an important number of advantages [4,7,8,14]:

- the estimated rate constants are more realistic comparatively with those derived from constant-volume model formulations (usually used in modelling cell biochemical systems) due to the considered cell regulatory properties;

- some simplifications, such as dilution terms defined for only key species are removed, and all species are treated on the same basis;

- species inter-connectivity (i.e. the degree to which a perturbation in one component influences others) is better characterized by including direct interrelations (via common reactions and intermediates) but also indirect relationships via the common cell-volume to which all species contribute (eq. 6-7);

- possible perturbations in the volume size and osmotic pressure can be also considered;

- perturbations applied to components of large concentrations lead to an important perturbation of the cell volume, which in turn lead to large perturbations of other cell component stationary concentrations (i.e. the so-called *secondary* or *indirect* perturbations); vice-versa, perturbations in species of low levels will have a low effect on the cell volume, and then a small secondary effect on other components, because:

$$V_{perturb} / V_o = \left(\sum n_j\right)_{perturb} \Big/ \left(\sum n_j\right)_o ;$$

- cells of large content (large *ballast*) diminishes the indirect perturbations (the so-called 'inertial' effect, or perturbation smoothing); the ballast effect shows how all cell components are interconnected via volume changes;

- the derived performance indices P.I.-s of GERM-s under WCVV conditions [7,8,14] present more realistic estimates comparatively with those derived from the classical constant-volume kinetic models which tend to overestimate the P.I.-s [16].

The only disadvantages of a WCVV kinetic model result from:

- a larger computation effort to identify the model parameters from the stationary species concentrations, and solving the nonlinear set eq. (9) (which sometimes can present multiple solutions, difficult to be discriminated);

- in a WCVV kinetic model all species (individual or lumped) have to be included in the model, because all contributes to the volume dynamics. In such a manner, the number of rate constants increases leading to a corresponding increase in the identification effort.

Such WCVV kinetic models with continuous variables, usually characterize metabolic processes during the cell balanced growing phase (ca. 80% of cell cycle). When cell reaches a critical size and a certain level of the surface-area-to-volume ratio, the division phase begins, lasting the last 20% of cell cycle. Over this phase (not analysed here), specialized proteins constrict the cell about its equator, thus leading to cell division. The duplicated content is thus partitioned, more or less evenly, between daughter cells. To model such a phase, supplementary terms must be added to explicitly account for the cell membrane dynamics [26].

Modelling a gene expression regulatory circuit (GERM)

Protein synthesis by gene expression is a highly regulated process to ensure a balanced and flexible cell growth under indefinitely variate environmental conditions. How this very complex process occurs is partially understood, but a multi-cascade control with negative feedback loops seems to be the key element. Enzymes catalyzing the synthesis are allosterically regulated by means of positive or negative effector molecules (transcriptional factors TF), while cooperative binding and structured cascade regulation (of the gene transcription and translation) amplify the effect of a change in an exo/endo-geneous inducer. Gene expression is also highly regulated to flexibly respond to the environmental stress The metabolic regulator features are determined by its ability to efficiently vary species flows and concentrations under changing environmental conditions so that a stationary state of the key metabolite concentrations can be maintained inside the cell. [7]

To model such a complex metabolic regulatory mechanisms at a molecular level with ODE kinetic models, Sewell et al [17], Savageau [18], Hlavacek and Savageau [19]; Maria [7,16] proposed simple mechanistic structures by using a modular approach, useful in simulating the hierarchical organization of cell regulatory networks.

Concerning the protein synthesis, this process is presumably regulated by a complex homeostatic mechanism that controls the expression of the encoding genes. On the other hand, cells contain a large number of proteins of well-defined functions, but strongly interrelated to ensure an efficient metabolism and cell growth under certain environmental conditions. Proteins interact during the synthesis and, as a consequence, the homeostatic systems perturb and are perturbed by each other. To understand and simulate such a complex regulatory process, the modular approach is preferred, being based on coupled semi-autonomous regulatory groups (of reactions and species), linked to efficiently cope with cell perturbations, to ensure system homeostasis, and an equilibrated cell growth. Various types of kinetic modules can be analyzed individually as mechanism, reaction pathway, regulatory characteristics, and effectiveness. As a limited number of regulatory module types govern the protein synthesis, it is computationally convenient to step-by-step build-up the modular regulatory network (GRC) by applying certain principles and rules to be further discussed, and then adjusting the network global properties. Accordingly, it is desirable to focus the metabolic regulation and control analysis on the regulatory/control features of functional GERM subunits than to limit the analysis to only kinetic properties of individual enzymes acting over the synthesis pathway.

The modular approach to analyse the gene expression assumes that the reaction mechanism and stoichiometry of various types of kinetic modules are known, while the involved species are completely observable and measurable. Such a hypothesis is rarely fulfilled due to the inherent difficulties in generating reliable experimental (kinetic) data for each individual metabolic subunit.

However, incomplete kinetic information can be incorporated by performing a suitable model lumping [31], or by exploiting the cell and module global optimal properties during identification steps. The regulatory modules can be constructed relatively independent to each other, but the linking procedure has to consider common input/output components, common linking reactions, or even common species. Rate constants can be identified separately for each module, and then extrapolated when simulating the whole regulatory network, by assuming that linking reactions are relatively slow comparatively with the individual module core reactions. In such a manner, linked modules are able to respond to changes in common environment and components such that each module remains fully regulated.

When elaborating a protein synthesis regulatory module (i.e. a gene expression regulatory module, GERM), different degrees of simplification of the process complexity can be followed. A GERM is a semi-autonomous regulatory group of reactions and species, linked to efficiently cope with cell perturbations, to ensure system homeostasis, and an efficient gene expression.

For instance, the gene expression (see schema of fig. 1) can be translated into a modular structure of reactions, more or less extended, accounting for individual or lumped species. At a generic level, in the simplest representation (fig. 1, up), the protein (P) synthesis rate can be adjusted by the 'catalytic' action of the encoding gene (G). The catalyst activity is in turn allosterically regulated by means of 'effector' molecules (O, P, or R) reversibly binding the catalyst G via fast and reversible reactions (the so-called 'buffering' reactions). These simple regulation scheme can be further detailed in order to better reproduce the experimental data, with the expense of a supplementary effort to identify the module kinetic parameters. For instance, a two-step cascade control of P-synthesis model also includes the M = mRNA transcript encoding P (fig. 1 center). The effector (O), of which synthesis is controlled by the target protein P (fig. 1 center), can allosterically adjust the activity of G and M, i.e. the catalysts for the transcription and translation steps of the gene expression. In such a cascade schema, the rate of the ultimate reaction is amplified, depending on the number of cascade levels and catalysis rates (fig. 1 down). More complex regulatory modules have been elaborated [4,15], and used in developing genetic regulatory circuits GRC) following a similar route to *translate* from the *language* of molecular biology to that of mechanistic chemistry, by preserving the structural hierarchy and component functions. Once elaborated, such a modular structure can be modelled by using a continuous variable ODE kinetic model under a WCVV framework, and then analysed as functional efficiency by means of some quantitative performance indices below described.

The correctly WCVV kinetic modelling of GERM-s by reproducing the linked GERM regulatory efficiency in a GRC is an essential step in describing the cell metabolism regulation via the hierarchically organized GRC-s (where key-proteins play the role of regulatory nodes). Also, it is to be underlined that the gene expression is a highly self- / cross- regulated and mutually catalyzed process by means of the produced enzymes / effectors.

As the cell regulatory systems are module-based organized, complex feed-back and feed-forward loops are employed for self- or cross-activation / repression of interconnected GERM-s, leading to different interaction alternatives (directly/inversely, perfect/incomplete, coupled/uncoupled connections) of a gene with up to 23–25 other genes. (Maria, 2014), to ensure the key-species homeostasis, holistic and local regulatory properties of the enzymatic reactions. While Maria [4,7,9,14,16], Sewell et al [17], Savageau [18], Hlavacek and Savageau [19] used reduced GERM structures of 10-14 reactions, that ensures a satisfactory tradeoff between model simplicity and its predictive quality [8], more sophisticated constructions are proposed in the literature [9].

As an example, Kaznessis [33, 34] designed a bistable switch genetic circuit, by using two gene modules extracted from the *lac operon* of *E. coli*. The transcriptional

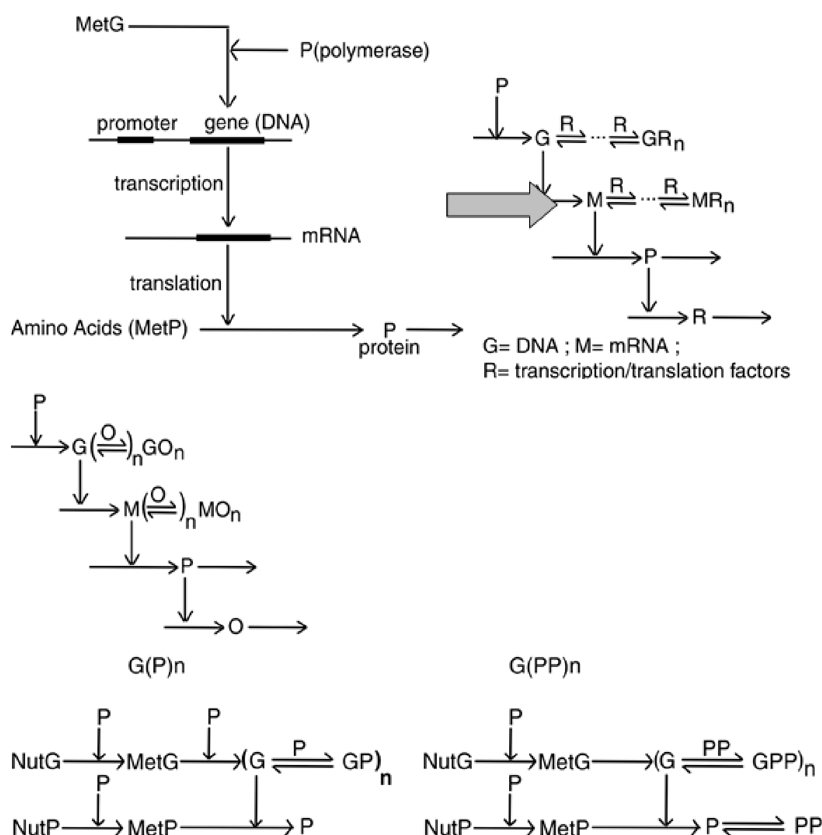

Fig.1. Protein P synthesis - simplified representations of a generic gene expression regulatory module (GERM). The horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicate an assumed concentration invariance of these species; G = gene encoding P; M = mRNA). The right structure corresponds to a [G(R)n; M(R)n] module type. *Up-row*: the simplified representation of the gene expression two-steps: transcription and translation (left), and (right) a simplified reaction schema of the gene expression [7,16]. The right model corresponds to a [G(R)n; M(R)n] GERM type. *Center* (adapted from Maria, 2005): Protein P synthesis - simplified representation of type [G(O)n; M(O)n] of the self-regulated expression module (horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicate an assumed concentration invariance of these species; G = DNA gene encoding P; M = mRNA; O = allosteric effectors). *Down-row*: two types of GERM simplified representations for protein synthesis: [G(P)n] (left) and [G(PP)n] (right)(Maria, 2005); horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicate an assumed concentration invariance of these species)

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 schema (fig. 1), including dimeric self-repressors (PP, or OO; [8]) and mutual repression following the presence in excess of one of the activating inducers, can also be illustrated by means of simpler representations of Yang et al. [35], and Maria [7,16]. The advantage of such a modular approach is the possibility to adapt the model size according to the available information, or to use the same GERM structure to model several gene expressions. Modular approach can also be useful in simulating the hierarchical organization of the cell regulatory networks.

The modular approach assumes that the reaction mechanism and stoichiometry of the kinetic module are known, while the involved species are completely observable and measurable. Such a hypothesis is rarely fulfilled due to the inherent difficulties in generating reliable experimental (kinetic) data for each individual metabolic subunit. However, incomplete kinetic information can be incorporated by performing a suitable model lumping, or by exploiting the cell and module global properties during identification steps. The regulatory modules can be constructed relatively independent to each other, but the linking procedure has to consider common input/output components, common linking reactions, or even common species. Rate constants can be identified separately for each module, and then extrapolated when simulating the whole regulatory network, by assuming that linking reactions are relatively slow comparatively with the individual module core reactions. In such a manner, linked modules are able to respond to changes in common environment and components such that each module remains fully regulated [7]. The advantage of such a modular approach is the possibility to reduce the system model complexity and the size of the identification problem, by understanding, for instance, the gene expression response to a perturbation as the response of a few genetic regulatory loops instead of the response of thousands of genetic circuits in the metabolic pathway [14].

To easily study and compare GERM regulatory efficiency, Sewell et al.[17], Yang et al.[35], and Maria [4,7-9,14-16] proposed various types of hypothetical GERM simplified reaction pathway designed to ensure homeostatic regulation of a generic protein-gene (P/G) pair synthesis, with a large number of exemplifications from *E. coli* (some of them are displayed in fig. 1, 2).

These simplified representations include the essential nutrient lumps (NutP, NutG), metabolites (MetP, MetG), and intermediates involved in the reactions controlling the transcriptional and translation steps of the G expression and P synthesis. The module nomenclature, proposed by Yang et al. [35] of type $[L_1(O_1)n_1;...;L_i(O_i)n_i]$ includes the assembled regulatory units $L_i(O_i)n_i$. One unit *i* is formed by the component L_i [i.e. enzymes/TF/P or even gene G, or M (mARN), etc.] at which regulatory element acts, and $n_1 = 0,1,2,\ldots$ number of *effector*/TF species O_i (i.e. P, PP, PPPP, O, OO, OOOO, R, RR, RRRR, etc.) binding the *catalyst* L. For instance, a *G(P)5* unit includes five successive binding steps of G with the product P, that ia G + P ⇔ GP + P ⇔ GPP + P⇔GPPP + P ⇔ GPPPP + P ⇔ GPPPPPP, all intermediate species GP, GPP, GPP, GPPP, GPPPP, GPPPPP being inactive catalytically, while the mass conservation law is all time fulfilled, i.e. $\sum_{i=0}^{5}[G(P_i)]$=constant. Such a representation accounts for the protein concentration diminishment due to the cell-growth dilution effect, but could also include protein degradation by proteolysis.

The G(P)n type of units, even less realistic (fig. 1), represent the simplest GERM used as control mechanism against which all others are compared. In a G(P)0 module (fig. 3), there are only two main synthesis chains. P is a permease that catalyses the import of NutG and NutP from the environment, and a metabolase that converts them into cellular metabolites MetG and MetP. P is also a polymerase that catalyses the synthesis of G from MetG. Gene G, symbolizing the genome of the cell, functions as catalyst for the synthesis of P from MetP. The result is that G and P syntheses are mutually autocatalytic. In *G(P)0* there are no regulatory elements (no buffering reactions

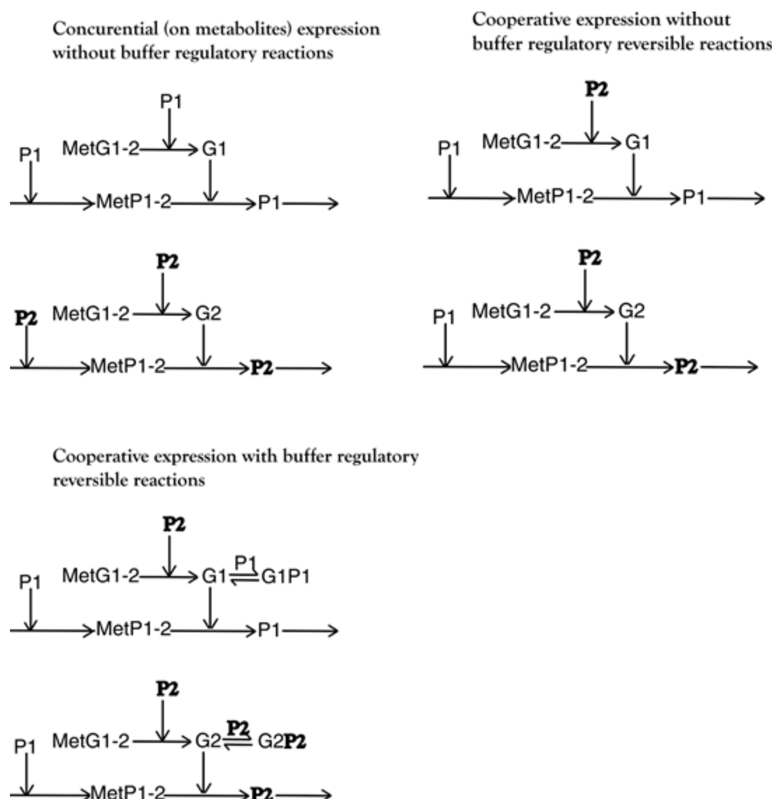

Figure 2. Investigate alternative kinking of two GERM-s: The horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicate an assumed concentration invariance of these species). There is no direct connectivity between the two GERM-s. *Up-left*: Competitive (on common metabolites) linking of [G1(P1)0]+[G2(P2)0]; *Up-right:* 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. *Down:* cooperative linking with buffer reversible regulatory reactions to modulate the G1, G2 catalytic activity in the modules [G1(P1)1]+[G2(P2)1]

to control the G activity). In *G(P)1*, the negative feedback control of transcription is realized by P itself (as effector), via a rapid buffering reaction, ,G+P⇔GP leading to the catalytically inactive GP. As proved [7,16], the maximum regulatory efficiency at steady-state (index 's') corresponds to $[G]_S/[G]_{total}$ =1/2, when the maximum regulation sensitivity vs. perturbations in $[P]_S$ is reached [17]. Further allosteric control of G activity, leading to inactive species $[GP_n]$, amplifies the regulatory efficiency of the module. As an example, prokaryotes commonly bind multiple copies of transcription factors as a means of promoting cooperative effects and thus improving regulatory effectiveness [35]. For instance, *dnaA* is an auto-regulated protein and at least five copies can bind to *dnaA* gene in *E. coli.* [35, 36].

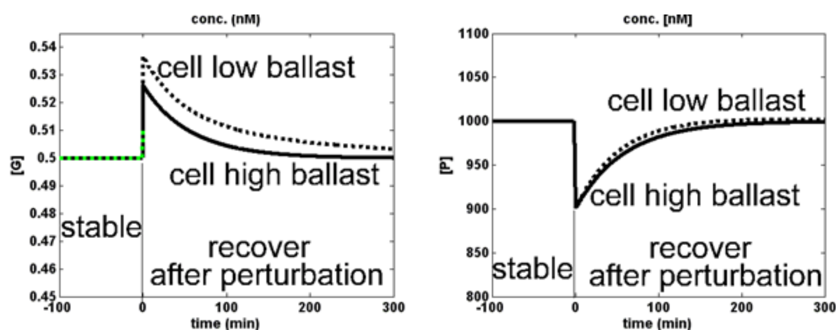

Fig. 3. Exemplification of the cell content ballast effect on the species recovering times to homeostasis, in the case of a [G(P)1] gene expression module. Gene G (left) and its encoding protein P(right) recovery trajectories after a -10% impulse perturbation in the [P]s = 1000 nM at t=0. Solid line trajectories correspond to a high ballast cell, while the dash line trajectories to a low ballast cell. The species concentrations in nM are given in the table 2 of [44]

The $G(PP)_m$ units reflect better the regulatory loops in which multiple copies of effectors (proteins and transcription factors TF) bind to promoter sites on the DNA that control expression of gene G encoding P (see exemplifications from *E. coli* by Yang et al. [35]. The control is better realized by including a supplementary P dimerization step before the buffering reactions. This explains why most of transcription factors bind as oligomers (typically dimers or tetramers) and why they typically bind in multiple copies [7, 35, 37].

Module [G(P)n; M(P)n'] (fig. 1 up-right, and center) tries to reproduce more accurately the transcription / translation cascade of reactions during the gene expression, by including an allosteric control at two levels of catalysis: on G (i.e. DNA) and on M (i.e. mRNA). M is synthesized from nucleotides under G catalysis, and then, P is synthesized in a reaction catalyzed by M (translation). Such a supplementary control of mRNA activity is proved to be a more effective means of regulating protein synthesis [7,16,35,38].

It is also to mention the way by which the rate constants in the rapid buffering reactions are estimated, that is for the *effector* reaction type [7]:

$$L+O \underset{k_{diss}}{\overset{k_{bind}}{\rightleftarrows}} LO\,;\; K_{LO}=\frac{k_{bind}}{k_{diss}}=\frac{C_{LO,s}\left(1+\dfrac{D_s}{k_{diss}}\right)}{C_{L,s}C_{O,s}}\,;\text{ (WCVV model).} \quad (11)$$

As discussed by Kholodenko et al. [39], fast buffering reactions are close to equilibrium and have little effect on metabolic control coefficients. As a consequence, rate constants of such rapid reactions are much higher than those of the core synthesis and dilution rates. To reduce the size of the unknown vector in (9), large values of can be postulated (5 to 7 orders of magnitude higher) [7]

Appendix. WCVV modelling hypotheses and effect of a perturbation on the cell-volume and species concentrations.

In a reacting system of constant volume, the classical (bio)chemical formulation of species mass balance lead to the process kinetic model of the form of eqn. (1). By contrast, in a variable-volume WCVV, all cell species should be considered (individually or lumped) all contributing to the cell volume increase (see the proof below). As the cell volume is doubling during the cell cycle, this volume variability can not be neglected in the cell kinetic models describing various cell metabolic processes taking place in the cytosol. Thus, a metabolic kinetic model in a WCVV approach should be written in the form underlined by Aris [20], and promoted by Maria [4,7-10, 13-16,24,25,43], that is eq. (2). The definition of *D* results from the way by which eqn. (2) was deducted [7,20], see eqn. (3).The system Jacobian is $J_C=(\partial h(C,k)/\partial C)_s$. The rate constants k can be estimated from the known stationary concentration vector C_s, by solving the nonlinear algebraic set (9), for every GERM type.

So, the so-called *cell-content dilution rate D* term is in fact cell-volume logarithmic growing rate. To better understand the importance of involving D in the cell kinetic model, and its significance and link to the reactions taking place into the cell, including the protein synthesis regulation, one considers a continuous-variable ordinary differential equations (ODE) model (2) of a standard formulation. Additionally, if the constant osmotic pressure is imposed to the cell system (isotonic system) to maintain the cell-membrane integrity, but neglecting the inner-cell gradients, one considers an additional constraint (7) to the cell kinetic model introduced by Grainger et al. [23], and Pfeiffer et al. [21], which, by derivation and division with V is leading to definition of Grainger et al. [23] eqn. (6). Consequently, the dilution *D* results as a sum of reacting rates of all cell species (individual or lumped), while the RT/π term can be easily deducted in a isotonic cell system, from the fulfilment of the following invariance relationship (7) derived from the Pfeffer's law in diluted solutions. Relationships (6-7) are important constraints of the cell model (3), eventually leading to different simulation results compared to cell ODE kinetic models that neglect the cell volume growth and isotonic effects [4,7,16].

If one assumes a constant *D* over the cell cycle, then, from eqn. (6) it results an exponential increase of the cell volume during its balanced growth: $V(t)=V_o\exp(+D\,.\,t)$.

Such a variable-volume formulation (3) is suitable to accurately model the cell growth and its division [26]. Such a model formulation allows studying the various GERM regulatory efficiency and the response of coupled GERM-s to continuous perturbations in the environment, and also the 'inertial' effect of the cell-'ballast' vs. continuous changes in cell and environment [7,8]. As ca. 80% of the cycle period is the growing phase [26], and assuming a quasi-constant osmotic pressure and a constant volume

growing logarithmic rate, the generic cell model (3) can be considered satisfactory to study the regulatory network effectiveness. The model was proved to be also very effective to study the response to continuous perturbations in the environment of various genetic regulatory circuits including genetic switches [9,14,43], or expression of a certain operon like those for mercury uptake in gram-negative bacteria [15,24-25].

By adopting a continuous-variable model of (3) type to construct a hypothetical Whole Mechanical (deterministic) Cell (in the sense of Tomita et al.[5,6]) that maintain intracellular homeostasis while growing auto-catalytically using environmental nutrients present in variable amounts, the following hypotheses should be adopted:

(i) The cell system consists in a sum of hierarchically organized components, e.g. *metabolites, genes DNA, proteins, RNA, intermediates, etc.* (interrelated through transcription, translation and DNA replication); the cell is separated from the *environment (*with *nutrients)* by a *membrane.*

(ii) The cell is an isotherm open system with an uniform content (perfectly-mixed case); species behave ideally, and present uniform concentrations within cell. The cell system is not only homogeneous but also isothermal, and isotonic (constant osmotic pressure), with no inner gradients or species diffusion resistance

(iii) The cell is an open system interacting with the environment interacts through a semi-permeable membrane.

(iv) The membrane, of negligible volume, presents a negligible resistance to nutrient diffusion; the membrane dynamics is neglected in the cell model, being assumed to follow the cell growing dynamics.

(v) When studying an individual P-synthesis regulatory module, the other cell species are lumped together in the so-called *cell ballast* [4,7,9].

The inner osmotic pressure is constant, and all time equal with the environmental pressure, thus ensuring the membrane integrity ($\pi_{cyt} = \pi_{env} =$ constant). As a consequence, the isotonic osmolarity under isothermal conditions leads to the equality $RT/\pi_{cyt} = RT/\pi_{env}$, which, using eqn. (7) indicates that the sum of cell species concentrations must equal those of the environment, i.e. $\left(\sum_j^{all} C_j\right)_{cyt} = \left(\sum_j^{all} C_j\right)_{env}$. Even if in a real cell such equality is approximately fulfilled due to perturbations and transport gradients, and in spite of migrating nutrients from environment into the cell, the overall environment concentration is considered to remain unchanged. On the other hand, species inside the cell transform the nutrients into metabolites and react to make more cell components. In turn, increased amounts of polymerases are then used to import increasing amounts of nutrients. The net result is an exponential increase of cellular components in time, which translates, through isotonic osmolarity assumption, into an exponential increase in volume with time. The overall concentration of cellular components is time-invariant (homeostasis), because the rate at which cell-volume increases equals that at which overall number of moles increases, leading to a constant $\left(\sum_{j=1}^{n_s} N_j\right)/V$ ratio.

The species concentrations are usually expressed in nano-moles, being computed with the relationship [7]:

$$Concentration = \frac{no.\ of\ copies\ /\ cell}{N_A \times V_{cyt,o}} \qquad [A1]$$

where N_A is the Avogadro number. For instance, for an *E. coli* cell, with an approximate volume $V_{cyt,o} = 1.66 \times 10^{-15} L$ [35], it results for C_{Gs} a value of $1/\left(6.022 \times 10^{23}\right)\left(1.66 \times 10^{-15}\right) = 1\ nM\ \left(i.e.\ 10^{-9}\ mol/L\right)$.

(vi) Cell volume doubles over the cell cycle period (t_c), with an average logarithmic growing rate of Dln(2) / t_c. Under stationary growing conditions, it results an exponential volume growth given by: $V(t)=V_o \exp(+D \cdot t)$.

(vii) For stationary growing conditions, species synthesis rates are equal to first-order dilution rates (DC_j), leading to time-invariant species concentrations, i.e. homeostatic conditions of $(dC_j / dt)_s = 0$ Such a nonlinear algebraic set (9) is used to estimate the rate constants ***k*** from the known stationary concentration vector .

Importance of the WCVV approach in computing the secondary perturbations. When one applies an impulse perturbation to one of the cell-species (e.g. a protein denominated by P), that implies removal (by excretion), or addition (by import) of a certain copynumbers of P from (to) the cell, by applying (7) and keeping constant the cell osmotic pressure p and temperature, as an effect of perturbation according to (7) the cell-volume will immediately contract (or dilate) from V to V^*. Consequently, the species concentrations will vary from $C_j = n_j / V$ to $C_j^* = n_j / V^*$ irrespectively that the component suffered any copynumbers' variation. This effect due to the cell volume variation is called secondary perturbations or indirect perturbations [7]. The difference in the P copynumbers before and after perturbation, i.e. $(n_{P,o} - n_P^*)$ (the * denotes the perturbed state), can be easily calculated for a certain imposed final concentration for P, i.e. $C_P^* = n_P^* / V^*$. The final copynumbers of P, n*p, for an imposed C*p results by re-evaluating the species concentrations:

$$n_P^* = V_o C_P^* \frac{RT}{\pi} \left(\sum_{j=1}^{m} C_{jo} - C_{P,o} \right) / \left(1 - C_P^* \frac{RT}{\pi} \right) \qquad [A2]$$

To prove the relationships [A2], one considers that, before and after applying the P perturbation, the sum of copynumbers of all other species remains invariant, that is:

$$\sum_{j=1}^{m} n_{jo} - n_{P,o} = \sum_{j=1}^{m} n_j^* - n_P^* \qquad [A3]$$

On the other hand, volume relationship (7), written for initial and final states and for the same pressure π, is $V_o = \frac{RT}{\pi} \sum_{all} n_{jo}$, and $V^* = \frac{RT}{\pi} \sum_{all} n_j^*$. By multiplying [A3] with RT/ π constant value, then changing terms from left to right, and including the volume formula, one obtains the net volume variation:

$$\left(V^* - V_o\right) = \frac{RT}{\pi}\left(n_P^* - n_{P,o}\right) \qquad [A4]$$

By dividing [A4] with the product (V^*V), and then multiplying with n_p^*, one obtains:

$$\frac{n_P^*}{V_o} - \frac{n_P^*}{V^*} = \frac{n_P^*}{V^*} \frac{RT}{\pi} \left(\frac{n_P^*}{V_o} - \frac{n_{P,o}}{V_o} \right) \qquad [A5]$$

or by changing terms from left to right, leads to:

$$\frac{n_P^*}{V_o} - \frac{n_P^*}{V^*}\frac{RT}{\pi}\frac{n_P^*}{V_o} = \frac{n_P^*}{V^*} - \frac{n_P^*}{V^*}\frac{RT}{\pi}\frac{n_{P,o}}{V_o} \quad [A6]$$

or by re-arranging, one results:

$$\frac{n_P^*}{V_o}\left(1 - \frac{n_P^*}{V^*}\frac{RT}{\pi}\right) = \frac{n_P^*}{V^*}\frac{RT}{\pi}\left(1/\frac{RT}{\pi} - \frac{n_{P,o}}{V_o}\right) \quad [A7]$$

By introducing the invariant $\frac{RT}{\pi} = \frac{1}{\sum_{j=1}^{m} C_{jo}}$, eqn. (7) in the right side, and substituting with, $C_P^* = n_P^* / V^*$ one obtains:

$$\frac{n_P^*}{V_o}\left(1 - C_P^*\frac{RT}{\pi}\right) = C_P^*\frac{RT}{\pi}\left(\sum_{j=1}^{m} C_{jo} - C_{P,o}\right) \quad [A8]$$

Relationship [A8] is identical with [A2]. On the other hand, the volume relationship (7), written for initial and final states and for the same pressure π, is $V_o = \frac{RT}{\pi}\sum_{all} n_{jo}$, and $V^* = \frac{RT}{\pi}\sum_{all} n_j^*$, leading to: $V^*/V_0 = \sum_{i=1}^{m} n_j^* \Big/ \sum_{j=1}^{m} n_{j,o}$. So, the sum of concentrations into the cell is a conservative term, as proved by:

$$\sum_{i=1}^{m} C_j^* = \frac{\sum_{i=1}^{m} n_j^*}{V^*} = \frac{\sum_{i=1}^{m} n_j^*}{V_o\left(\sum_{i=1}^{m} n_j^* / \sum_{i=1}^{m} n_{j,o}\right)} = \frac{\sum_{i=1}^{m} n_{j,o}}{V_o} = \sum_{i=1}^{m} C_{j,o} = constant \quad [A9]$$

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. LODISH, H., BERK, A., MATSUDAIRA, P., KAISER, C.A., KRIEGER, M., SCOTT, M.P., ZIPURSKY, L., DARNELL, J., Molecular cell biology, Freeman & Co., New York, 2000.
2.SAURO, H.M., KHOLODENKO, B.N., Prog. Biophys. Mol. Biol., 86, 2004, p. 5-43.
3.MARIA, G., DAN, A., STEFAN, D.N., Chemical & Biochemical Engineering Quarterly, 24, no. 3, 2010, p.265-281.
4. MARIA, G., Chemical and Biochemical Engineering Quarterly, 21, 2007, p. 417-434.
5.TOMITA, M., HASHIMOTO, K., TAKAHASHI, K., SHIMIZU, T., MATSUZAKI, Y., MIYOSHI, F., SAITO, K., TANIDA, S., YUGI, K., VENTER, J.C., Bioinformatics, 15, 1999, p. 72-84.
6. TOMITA, M., Trends in Biotechnology, 19, 2001, p.205-210.
7. MARIA, G., Chemical and Biochemical Engineering Quarterly, 19, 2005, p. 213-233.
8. MARIA, G., Chemical and Biochemical Engineering Quarterly, 20, 2006, p. 353-373.
9. MARIA, G., Chemical & Biochemical Engineering Quarterly, 28, 2014, p. 83–99.
10.MARIA, G., Chemical & Biochemical Engineering Quarterly, 28, 2014, p. 509-529.
11. VISSER, D., SCHMID, J.W., MAUCH, K., REUSS, M., HEIJNEN, J.J., Metabolic Engineering, 6, 2004, p. 378–390.
12. ALON, U., An introduction to system biology. Design principles of biological circuits, Boca Raton, Chapman & Hall / CRC, 2007.
13. MARIA, G., Chemical Engineering Science, 60, 2005, p. 1709-1723.
14. MARIA, G., Asia-Pacific Journal of Chemical Engineering, 4, 2009, p. 916-928.
15. MARIA, G., LUTA, I., Computers & Chemical Engineering, 58, 2013, p. 98-115.
16.MARIA, G., Chemical and Biochemical Engineering Quarterly 17, 2003, p.99-117.
17. SEWELL, C., MORGAN, J., LINDAHL, P., J. theor. Biol., 215, 2002, p. 151-167.
18.SAVAGEAU, M.A., Math. Biosciences, 180, 2002, p. 237-253.
19.HLAVACEK, W.S., SAVAGEAU, M.A., J. Mol. Biol. 266, 1997, p. 538-558.
20.ARIS, R., Elementary chemical reactor analysis, New Jersey, Prentice-Hall, 1969.
21.PFEIFFER, T., SANCHEZ-VALDENBRO, I., NUNO, J.C., MONTERO, F., SCHUSTER, S., Bioinfomatics, 15, 1999, p. 251-257.
22. WALLWORK, S.C., GRANT, D.J.W., Physical Chemistry, Longman, London, 1977.
23. GRAINGER, J.N.R., GAFFNEY, P.E., WEST, T.T., J. theor. Biol. 21, 1968, p. 123-130.
24. MARIA. G., Chemical and Biochemical Engineering Quarterly, 23, 2009, p. 323-341.
25. MARIA, G., Rev. Chim. (Bucharest), **61**, no. 2, 2010, p.172
26. MORGAN, J.J., SUROVTSEV, I.V., LINDAHL, P.A., Jl. theor. Biology, 231, 2004, p. 581-596.
27. SUROVSTEV, I.V., MORGAN, J.J., LINDAHL, P.A., Journal of Theoretical Biology, 244, 2007, p. 154–166.
28. VILELA, M., MORGAN, J.J., LINDAHL, P.A., PLoS Computational Biology, 6, 2010, e1001036, p. 1-11.
29. KINOSHITA, A., NAKAYAMA, Y., TOMITA, M., ICSB-2001, 2-nd Int. Conf. on Systems Biology, California Institute of Technology, Pasadena (Ca.), Nov. 4-7, 2001.
30. GIBSON, M.A., BRUCK, J. A probabilistic model of a prokaryotic gene and its regulation, In: Computational modeling of genetic and biochemical networks (Bower, J.M. & Bolouri, H., Eds.), Cambridge (Mass.), MIT Press, 2001.
31. MARIA, G., Chemical and Biochemical Engineering Quarterly 18, 2004, p.195-222.
32.VAN SOMEREN, E. P., WESSELS, L. F. A., BACKER, E., REINDERS, M. J. T., Signal Processing, 83, 2003, p. 763-775.
33. KAZNESSIS, Y. N., Chemical Engineering Science 61, 2006, p. 940-953.
34. SALIS, H., KAZNESSIS, Y., Computers & Chemical Engineering, 29, 2005, p. 577-588.
35. YANG, Q., LINDAHL, P., MORGAN, J., J. theor. Biol., 222, 2003, p. 407-423.
36. SPECK, C., WEIGEL, C., MESSER, W., EMBO J., 18, 1999, p. 6169-6176.
37. PTASHNE, M., Cell Press & Blackwell Scientific Publs., Cambridge (Mass.), 1992, pp. 82.
38. HARGROVE, J. L., SCHMIDT F. H., FASEB Jl. 3, no. 12, 1989, p. 2360-2370.
39. KHOLODENKO, B. N., SCHUSTER, S., GARCIA, J., WESTERHOFF, H. V., CASCANTE, M., Biochimica Biophysica Acta 1379, no. 3, 1998, p. 337-352.
40. HEINRICH, R., SCHUSTER, S., The regulation of cellular systems, New York, Chapman & Hall, 1996.
41. VANCE, W., ARKIN, A., ROSS, J., Proceedings of the National Academy of Science USA, 99, 2002, p. 5816-5821.
42. SMIGELSCHI, O., MARIA, G., Hungarian Journal of Industrial Chemistry, 14, no. 4, 1986, p.453-462.
43. MARIA, G., Modelling bistable genetic regulatory circuits under variable volume framework, Chemical and Biochemical Engineering Quarterly, 21, 2007, p. 417-434.
44. MARIA, G., SCOBAN, A.G., Rev. Chim. (Bucharest), **67**, no. 1, 2018, (part 2 of this paper), in-press

Manuscript received: 15.02.2017

REVISTA DE CHIMIE
SI
REVISTA MATERIALE PLASTICE
POT FI ACCESATE GRATUIT
LA URMATOARELE ADRESE:

http://www.revistadechimie.ro/
http://www.revmaterialeplastice.ro/

FREE WEB ACCESS FOR
REVISTA DE CHIMIE
AND
MATERIALE PLASTICE

http://www.revistadechimie.ro/
http://www.revmaterialeplastice.ro/