

Comments on several review eBooks promoting a novel kinetic modelling framework of metabolic processes and of genetic regulatory circuits in living cells



Gheorghe Maria^{1,2*}

¹Dept. of Chemical and Biochemical Engineering, Politehnica University of Bucharest, Romania

²Romanian Academy, Chemical Sciences section, Romania

Submission: December 25, 2022; Published: January 12, 2023

*Corresponding author: Gheorghe Maria, Dept. of Chemical and Biochemical Engineering, Politehnica University of Bucharest, Romania, Romanian Academy, Chemical Sciences section, Romania; Email; gmaria99m@hotmail.com; web-page: https://en.wikipedia.org/wiki/Gheorghe_Maria

Abstract

The paper is pointing-out the main aspects reviewed by several eBooks denoted by (EB1-EB4), namely [1-4] respectively. These booklets in the Bioinformatics area are focus on the progresses made in the area of deterministic modelling of metabolic cell biochemical processes related to the central carbon metabolism (CCM) by using chemical and biochemical engineering (CBE) principles and rules [EB1-EB3], and [5]. The works EB1-EB2 reviewed a novel math (kinetic) modelling framework of metabolic pathways and of genetic regulatory circuits (GRC) with continuous variable ordinary differential mass balance sets (ODE) based on the process mechanism. Two approaches are discussed: I) the classic (default) "Whole-Cell Constant Volume" (WCCV) ODE models, that ignore the cell volume exponential increase during the cell growth, and ii) the holistic "whole-cell variable-volume" (WCVV) dynamic ODE models by explicitly accounting for the cell-volume growth, while preserving the cell isotonicity (and cell membrane integrity). While EB1-EB2 are focused on detailing the deterministic modelling of the individual gene expression regulatory modules (GERM), and of GRC-s in living cells by using the CBE principles, by employing a WCVV approach and rules of the nonlinear system control theory (NSCT), the EB3-EB4 are dealing with reviewing the ways to derive deterministic modular structured cell kinetic models (MSDKM) (with continuous variables, and based on cellular metabolic reaction mechanisms). These extended MSDKM are further linked to those of the bioreactor dynamic models (including macro-scale state variables), thus resulting hybrid structured modular dynamic (kinetic) models (HSMDM) proved to successfully solve more accurately difficult bioengineering problems.

The EB3-EB4 prove, by means of several demonstrative relevant examples the superiority of using MSDKM and HSMDM dynamic math models (within the WCVV modelling framework) when solving various bioengineering problems, dealing with: (a) *in-silico* (math model based) off-line optimization of the operating policy of various bioreactors types (e.g. the fed-batch bioreactors), and (b) *in-silico* design/screening/check GMO-s (genetically modified microorganisms) alternatives to be used for improving performances of various industrial bioprocess/bioreactors.

Abbreviations: ATP: Adenosin-Triphosphate; ADP: Adenosin-Diphosphate; AMP: adenosin-monophosphate; CBE: Chemical and biochemical engineering concepts/rules; CBE: Chemical And Biochemical Engineering Concepts/Rules; CCM: Central Carbon Metabolism Of a Cell; G: generic gene (DNA); EB1,EB2,EB3,EB4: the eBooks no. 1-4, corresponding to [1-4] respectively ; GERM: individual gene expression regulatory module; GMO: Genetically Modified Microorganisms; GRC: Genetic Regulatory Circuits; GRN: Genetic Regulatory Networks; HSMDM: Hybrid Structured Modular Dynamic (Kinetic) Models; M: mRNA; MCA: metabolic control analysis; MSDKM: Deterministic Modular Structured Cell Kinetic Model; NSCT: Rules Of The Control Theory Of Nonlinear Systems; ODE: Ordinary Differential Equation Set; P: Generic protein; PI: Regulatory performance indices of GERM-S; QSS: Quasi steady- state; TF: transcription factors; WCVV: Whole-cell variable volume modeling approach; WCCV: Whole cel constant volume kinetic modelling approach

Introduction

Biologically catalyzed reactions (with enzymes, or living cell cultures) can successfully replace complex chemical syntheses, being more selective, by using milder reaction conditions, and generating less waste. As proved by the recent literature, and the

eBooks EB1-EB3 (Figure 1) of Maria [1-3] respectively, and by the under way EB4 of Maria [4], the developed *in-silico* (math-model-based) numerical analysis of such biochemical/biological systems turned out to be a beneficial tool (i) to *in-silico* off-line determine optimal operating policies of complex multi-enzymatic

or biological reactors with a higher precision and predictability, or (ii) to *in-silico* design GMO-s of desired characteristics for various uses. This work presents a holistic 'closed loop' approach that facilitate the control of the *in vitro* through the *in silico* development of dynamic models for living cell systems, by deriving *deterministic modular structured* cell kinetic models (MSDKM) with continuous variables, and based on cellular metabolic reaction mechanisms.

The ever-increasing availability of experimental (qualitative and quantitative) information about the tremendous complexity of cell metabolic processes, stored in large bio-omics databanks (including genomic, proteomic, metabolomic, fluxomic cell data for various microorganisms, see the review of EB3), but also about the bioreactors' operation [3-4, 6] necessitates the advancement of a systematic methodology to organise and utilise these data.



Figure 1: cover of the e Books EB1 -EB3, that is [1-3] respectively, published with [juniper publisher, Irvine, CA(USA)].

This work is aiming to briefly review the concepts and rules developed in (EB1-EB4), by proving the feasibility and advantages of using the classic but also the novel principles and numerical tools of the CBE / NSCT to develop extended cell-scale MSDKM-s to adequately simulate the dynamics of reaction pathway modules which belong to CCM, and of GRCs/GRNs. These extended MSDKM will be further linked to those of the bioreactor dynamic models, thus resulting HSMDM structured models proved to successfully solve more accurately difficult bioengineering problems (EB3-EB4).

In such HSMDM, the cell-scale model part (including nano-level state variables) is linked to the biological reactor macro-scale state variables for improving the both model prediction quality and its validity range. By contrast, the current (classical/default) approach in biochemical engineering and bioengineering practice for solving design, optimization and control problems based on the math models of industrial biological reactors is to use *unstructured* Monod (for cell culture reactor) or Michaelis-Menten type (if only enzymatic reactions are retained) global kinetic models by ignoring a detailed representation of the metabolic cellular processes.

EB1-EB2 booklets presents some general concepts of CBE, of NSCT, and of Bioinformatics used to derive MSDKM and HSMDM

models, with continuous variables and based on cellular metabolic reaction mechanisms. Such extended structured cell math (kinetic) models consider, with a degree of detail suitable to the each approached case study, the cellular key-metabolic reactions, being able to adequately reproduce the cell key-species dynamics. These structured MSDKM models can satisfactorily represent the key steps of the CCM at a cell scale, by also including GRC-s responsible for the CCM syntheses regulation, besides reaction modules responsible for the synthesis of cellular metabolites of interest for the industrial biosynthesis. Special attention is paid to the conceptual and numerical rules used to construct various individual GERM-s kinetic models, but also various GRC-s (e.g. toggle-switch, amplitude filters, operons expression, etc.) [7-10] modular kinetic models from linking individual GERM-s [11,12].

At the same time, EB1-EB2 briefly described a novel WCVV kinetic modelling framework for metabolic GRC and CCM processes, introduced and promoted by the author in previous works [13-17]. These works point-out the features of the deterministic WCVV models, and their advantages when simulating GERM-s, and GRC-s dynamics in living cells, by contrast to the classical (default) WCCV modelling framework. Comparison is made by using the regulatory performance indices (P.I.-s) of GERM-s, inspired from the NSCT. The EB1-EB2, EB4 also present rules to link GERM-s when modelling GRC-s.

Self-replicating apparatus	Time scale separation (slow / fast manifolds)	Self-replication	Regul. net
Replisome, Partitioning apparatus, Z-ring		Nucleoid replication & partitioning, cell division	Cell cycle regulation
Nucleoid		Supercoil and organize genome	Gene expression regulation
Ribosomes, Genome, Energy harnessing apparatus	Intermediate characteristic time	Protein synthesis, Store genetic info, Harness energy	Metabolism regulation
Cell wall, Nucleic acids, Coenzymes	Succession of events	Metabolic cycles, pathways, Transcription, Translation	
Peptidoglycan, Membrane, Protein eplx., Nucleotides		Catalysis, Energy currency	Regulation of enzyme activity
Lipids, Proteins, Nucleosides	Catalysis, Hydrophobic effects		
Saccharides, Fatty acids, Aminoacids	Transient recovering time	Intermediates and building blocks for cell structures and functions	
Simple metabolites	←Temporal Hierarchy→	Source of energy and material	
Raw materials (nutrients)		←Functional Hierarchy→	

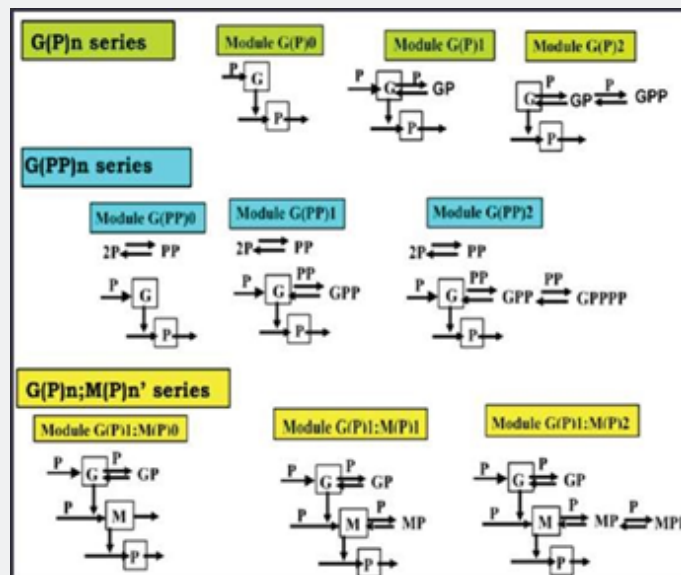


Figure 2: (top) The hierarchical organization of living cells [EB2]. (middle) The generic protein P synthesis – simplified representations of a generic GERM regulatory module (horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; G = gene encoding P; M = mRNA; R = repressor; In = inducer; Met = metabolites). See reviews of [EB1-EB4], and [5,9,14,33,49]. (down) The library of individual GERM lumped models used to simulate the cross-catalysed synthesis of a generic pair of protein P and its encoding gene G (DNA). Horizontal arrows indicate reactions; the vertical arrows indicate catalytic actions; M= mRNA; PP= effectors [EB1-EB2], [14].

EB3-EB4 booklets complete the theoretical aspects developed in EB1-EB2 by including related rules necessary to construct MSDKM and HSMDM models. Also, these booklets prove, by means of 5-6 demonstrative relevant examples, the superiority of using MSDKM and HSMDM dynamic math models vs. classical (default) unstructured/global dynamic models when solving difficult bioengineering or bioinformatics problems. As demonstrated, the *in-silico* numerical analysis of biochemical or biological processes by using MSDKM or HSMDM models are proved to be not only an essential but also an extremely beneficial tool for engineering evaluations aiming (i) to determine with a higher accuracy the optimal operating policies of complex multi-enzymatic reactors [6,18-22], or of bioreactors including the biomass adaptation to the variable bioreactor environment over hundreds of cell cycles [12,23-26], or even (ii) to easier and quickly simulate and analyse the performances/ characteristics of various *in-silico* design GMO-s alternatives, by using the “metabolic flux analysis” (MFA) [26-30], together with the gene-knock-out technique) [EB3-EB4],

and [12,30-32].

By contrast, by considering only the macroscopic key-variables of the process (biomass, substrate, and product concentrations), the *unstructured* (apparent, global) math (kinetic) models do not adequately reflect the metabolic changes of the bioreactor biomass, being inadequate to accurately predict the cellular response to the medium disturbances through the self-regulated cellular metabolism. Such classical global/unstructured dynamic models may be satisfactory for an approximate modeling of the biological process, but not for modeling of cellular metabolic processes, and they can not make any correlation between the bioreactor operation and the continuous adaptation of the biomass metabolism to the variable conditions of the bioreactor. Even worst, as proved by the author in EB1-EB4 and their previous papers [16,17,33,34], such global models may lead to biased and distorted conclusions about the GERM's performances, thus making difficult the modular constructions of GRC-s by linking individual GERM-s (Figure 4).

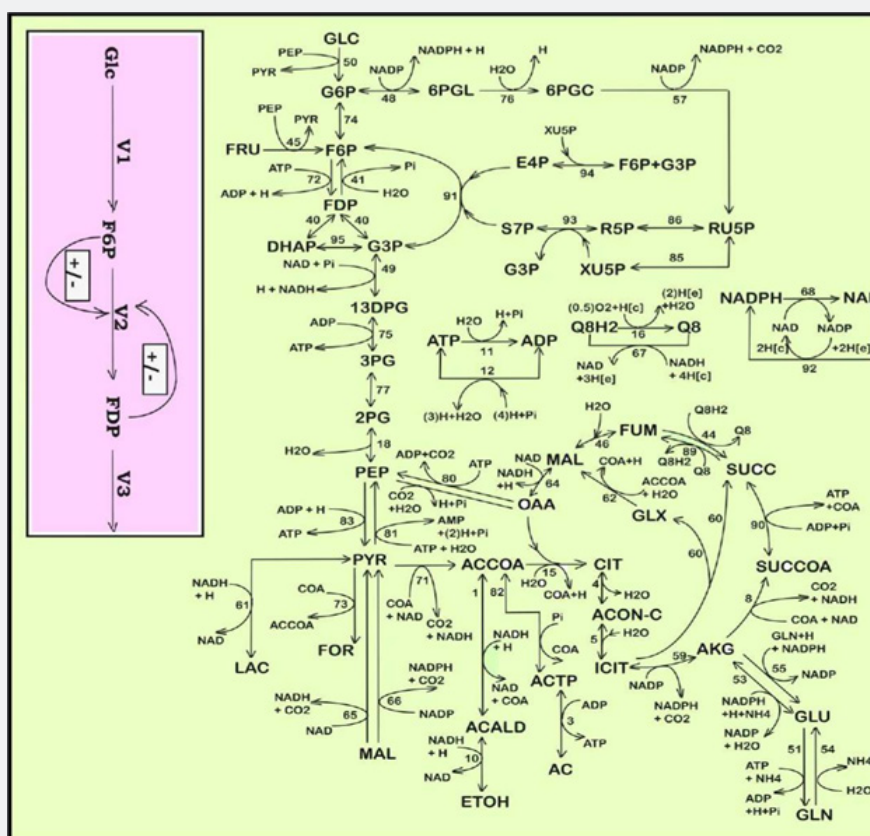
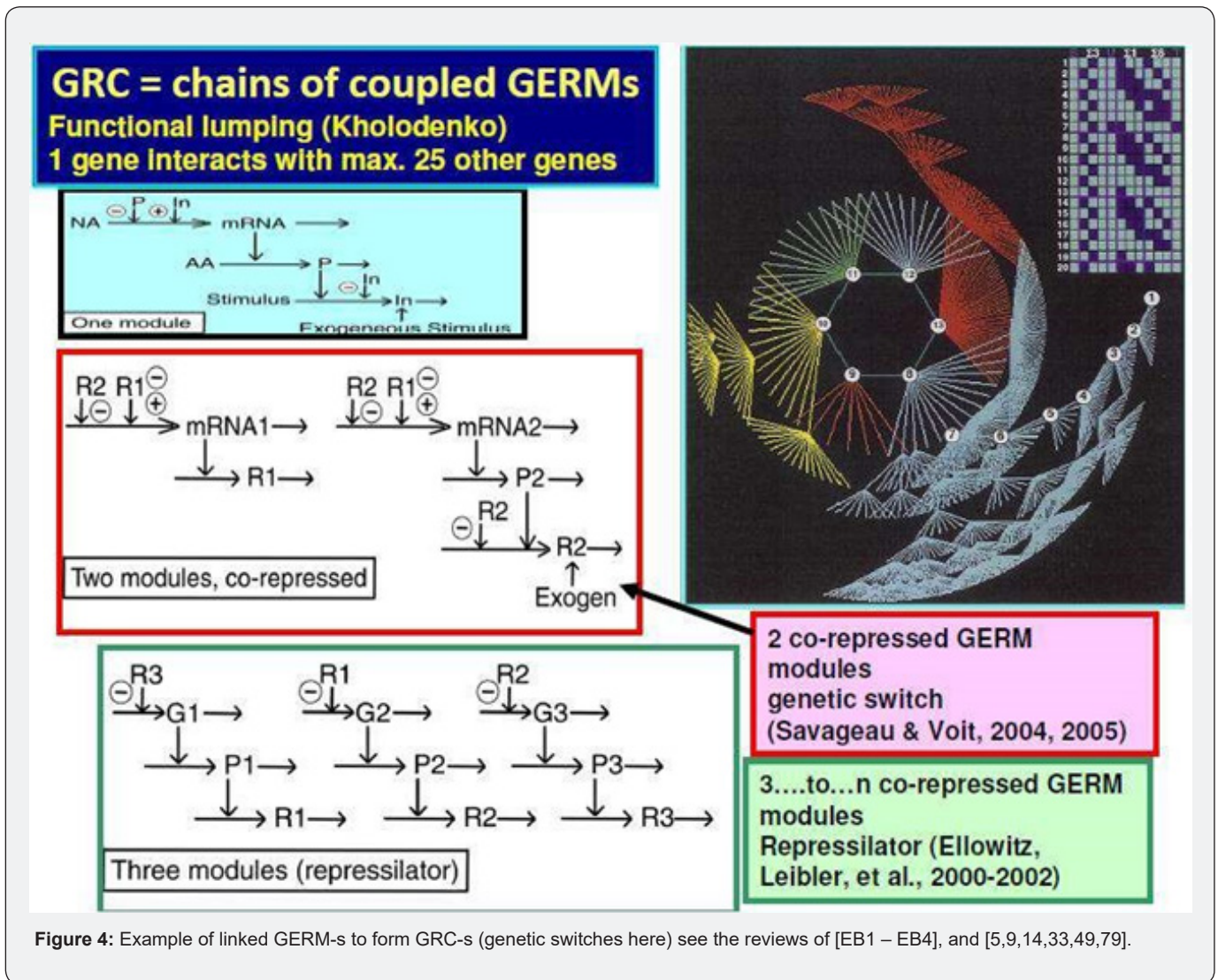


Figure 3: Simplified representation of the CCM pathway in *E. coli* [31,78] (the “wild” cell including the PTS-system). Fluxes characterizing the membranar transport [*Metabolite*(e) \leftrightarrow *Metabolite*(c)] and the exchange with environment have been omitted from the plot. See [31] for details and explanations regarding the numbered reactions. Notations: [e]= environment; [c]=cytosol. Adapted from [31] with the courtesy of CABEQ JI. The considered 72 metabolites, the stoichiometry of the 95 numbered reactions, and the net fluxes for specified conditions are given by [31]. The pink rectangle indicates the chemical node inducing glycolytic oscillations [7,44]. Notations, + and - denotes the feedback positive or negative regulatory loops respectively. GLC = glucose; F6P= fructose-6-phosphate; FDP = fructose-1,6-biphosphate; see the abbreviation list for species names; V1-V6 = lumped reaction rates indicated by [32]. Species notations are explained in the abbreviation list of [32].



In general, living cells are organized, self-replicating, evolvable, selfadjustable, and responsive biological systems to environmental stimuli able to convert raw materials (substrates/nutrients) from the environment into additional copies of themselves. (Figure 5-down/right, and in figure 2-top)

The structural and functional cell organization, including components and reactions, is extremely complex, comprising involving $O(10^{3-4})$ components, $O(10^{3-4})$ transcription factors (TF-s), activators, 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 [EB1-EB2] (Figure 2-top). Relationships between structure, function and regulation in complex cellular networks are better understood at a low (component) level rather than at the highest-level [35].

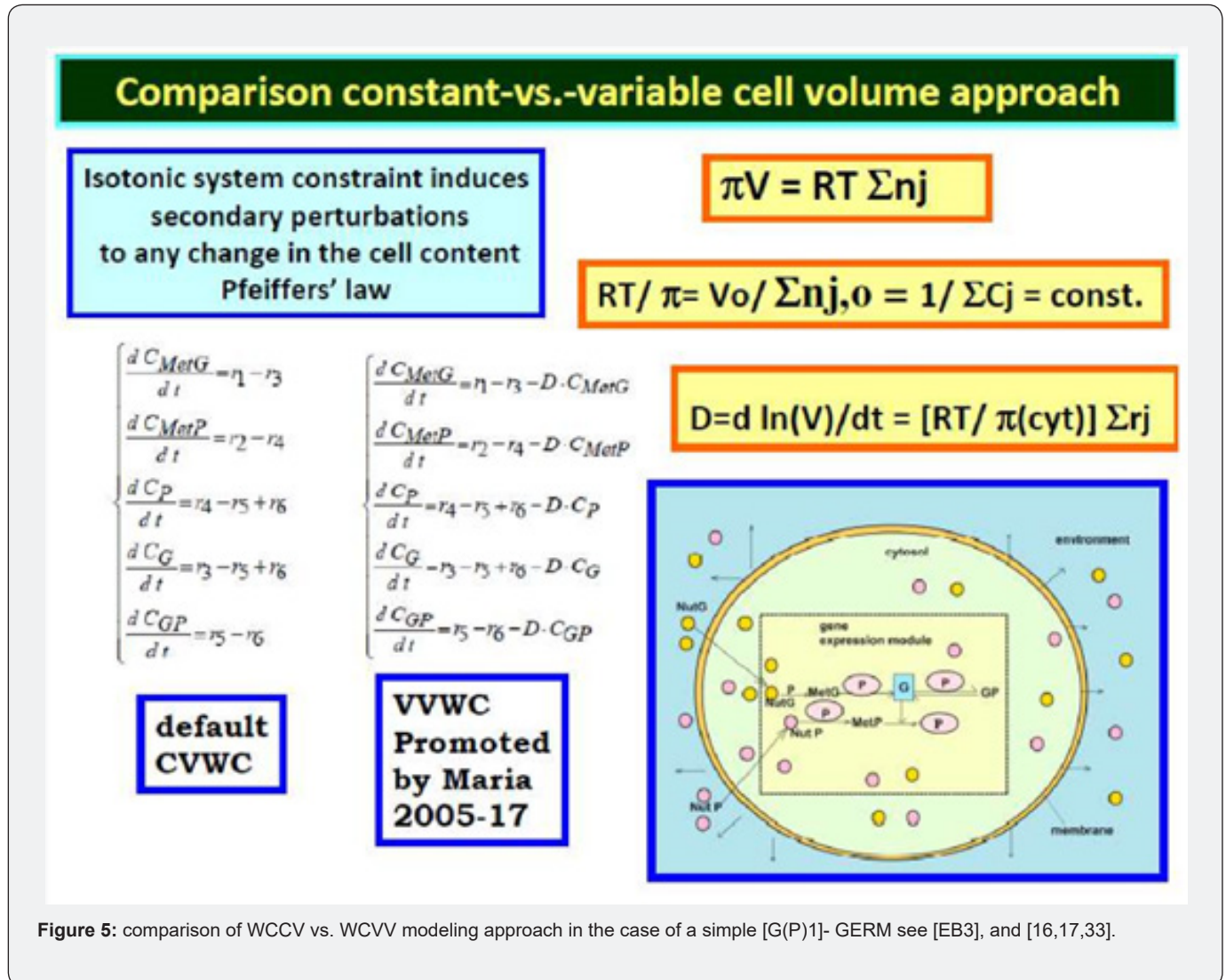
Cell regulatory and adaptive properties are based on *homeostatic* mechanisms, which maintain quasi-constant (QSS) the key-species concentrations and metabolites' output levels, by

adjusting the synthesis rates, by switching between alternative substrates, or development pathways. Cell regulatory mechanisms include allosteric enzymatic interactions and feedback in gene transcription networks, metabolic pathways, signal transduction and other species interactions [36]. In particular, protein synthesis homeostatic regulation includes a multi-cascade control of the gene expression with negative feedback loops and allosteric adjustment of the enzymatic activity (Figure 2middle/down) [EB1-EB4].

Cells have a very complex but hierarchic organization (structural, functional, and temporal, Figure 2-top), well described by [EB1-EB2]. A central part of such cell models concerns self-regulation of metabolic processes belonging to the CCM (Figure 3), via GRC-s (Figure 4). So, one application of such dynamic deterministic cell models is the study of GRC-s, for predicting ways by which biological systems respond to signals, or environmental perturbations. The emergent field of such efforts is the so-called "gene circuit engineering" and, a large number of examples have

been reported with *in-silico* re-creation of GRC-s conferring new properties to the mutant cells (i.e. desired 'motifs' in response to external stimuli) [EB1, EB3, EB4], and [37]. Simulation of gene expression (GERM-s), and of GRC makes possible *in-silico* design

of GMO-s that possess desired properties. By "inserting" the design GRCs into host organisms, one may create a large variety of GMO-s with minifunctions / tasks in response to external stimuli [EB3, EB4].



"With the aid of recombinant DNA technology, it has become possible to introduce specific changes in the cellular genome. This enables the directed improvement of certain properties of micro-organisms, such as the productivity, which is referred to as Metabolic Engineering [27,38,39]. This is potentially a great improvement compared to earlier random mutagenesis techniques, but requires that the targets for modification are known. The complexity of pathway interaction and allosteric regulation limits the success of intuition-based approaches, which often only take an isolated part of the complete system into account. Mathematical models are required to evaluate the effects of changed enzyme levels or properties on the system as a whole, using metabolic control analysis or a dynamic sensitivity

analysis" [40]. In this context, GRC dynamic models together with the metabolic flux balance analysis (FBA) based on MSDKM models are powerful tools for *in-silico* developing re-design strategies of modifying genome and gene expression seeking for new properties of the mutant cells in response to external stimuli (n.b. a metabolic flux is the enzymatic reaction under QSS /homeostatic cell conditions; EB1, [27]). Examples of such GRC modulated functions include [EB3, EB4]:

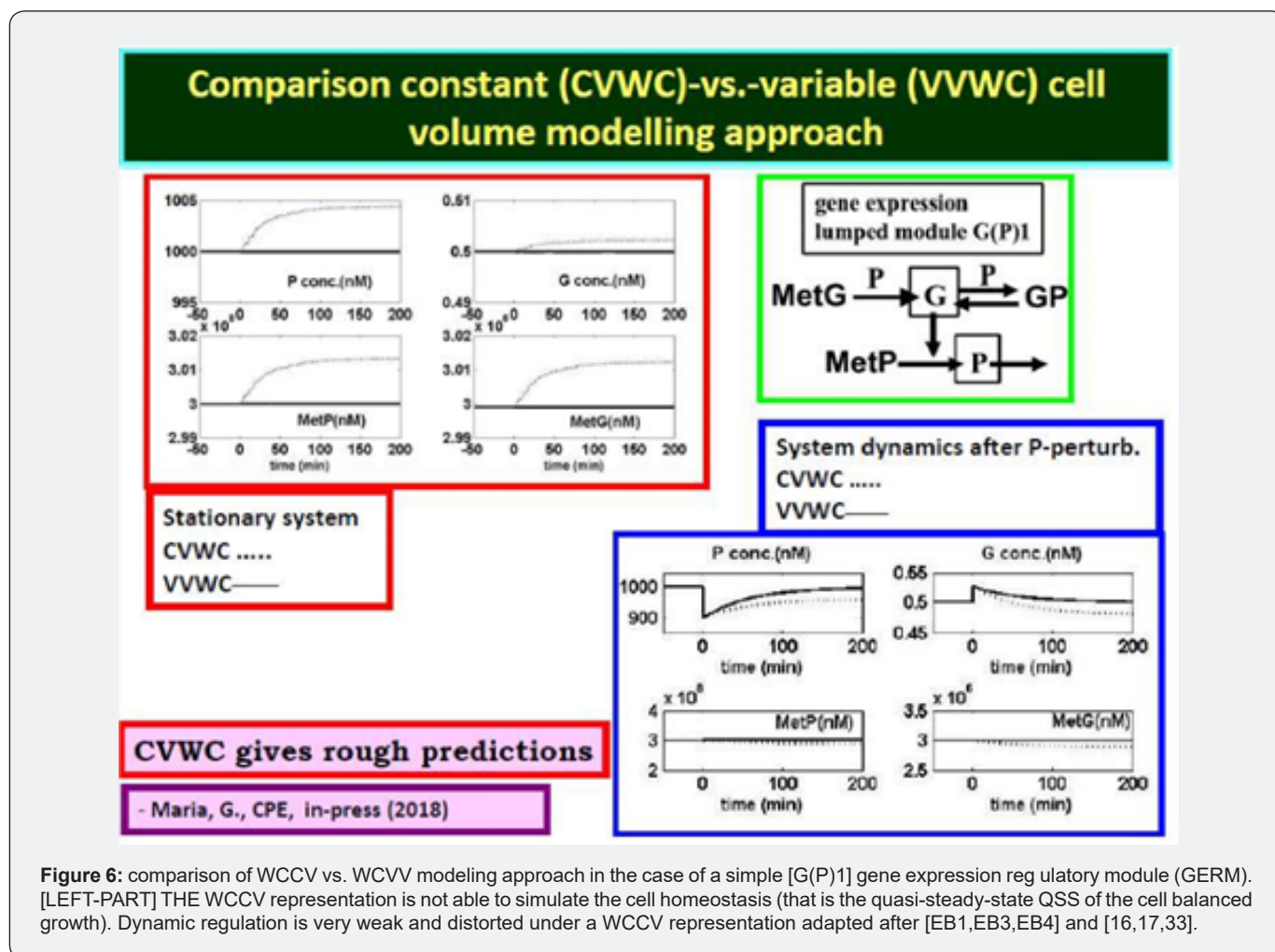
- a) **toggle-switch**, i.e. mutual repression control in two gene expression modules, and creation of decision-making branch points between on/off states according to the presence of certain inducers [EB3], and [9,10,41]

b) **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 [42]

c) **GRC oscillator** producing regular fluctuations in the CCM network elements, reporter proteins, and cell GRN nodes, and making the GRC to evolve among two or several QSS [7,8,43,44,45]

d) specific treatment of external signals by controlled expression such as **amplitude filters**, **noise filters** or signal / stimuli amplifiers [8,11,12]

e) **GRC signalling circuits** and cell-cell communicators, acting as 'programmable' memory units [EB3-EB4].



The development of dynamic models on a deterministic basis to adequately simulate *in detail* the cell metabolism self-regulation, cell growth, and replication for such an astronomical cell metabolism complexity is practical impossible due to lack of structured information and computational limitations. A review of some trials is presented by [8,12,32,46], and [EB1-EB4].

In spite of such tremendous modelling difficulties, development of *reduced* dynamic models to adequately reproduce such complex synthesis related to the central carbon metabolism (CCM) [8,32,40,46,47], but also to the genetic regulatory systems [EB1], and [11,12,24] tightly controlling the metabolic processes reported significant progresses over the last decades in spite of the lack of structured experimental kinetic information. Even if

being rather based on sparse information from various sources and unconventional identification / lumping algorithms [48-50], such structured deterministic kinetic models have been proved to be extremely useful for *in-silico* design of novel GRC-s conferring new properties/functions to the mutant cells, that is desired 'motifs' in response to the external stimuli [EB1-EB4].

In fact, all the rules and algorithms used by the deterministic modelling of CCM and GRCs, discussed in the works of Maria [EB1-EB4], and [11,12,14] belong to the new emergent field of *Systems Biology*. Systems Biology defined as "the science of discovering, modelling, understanding and ultimately engineering at the molecular level the dynamic relationships between the biological molecules that define living organisms" (Leroy Hood, Inst.

Systems Biology, Seattle) [51] is one of the modern tools which uses advanced mathematical simulation models for *in-silico* design of GMO-s that possess specific and desired functions and characteristics.

To model such a complex metabolic regulatory mechanism at a molecular level, two main approaches have been developed over decades: structure - oriented analysis, and dynamic (kinetic) models [35]. A review of mathematical model types used to describe metabolic processes is presented by [EB1-EB3], and [14,27,46]. From the mathematical point of view, various structured (mechanism-based) dynamic models have been proposed to simulate the metabolic processes and their regulation, accounting for continuous, discrete (Boolean), and/or stochastic variables, in a modular construction, 'circuit-like' network, or compartmented simulation platforms [EB1-EB2], and [14,36,52]. Each model type presents advantages but also limitations. Each theory presenting strengths and shortcomings in providing an integrated predictive description of the cellular regulatory network.

Among all these math modelling alternatives [EB1-EB3], the classical approach to develop *deterministic dynamic models* is based on a hypothetical reaction mechanism, kinetic equations, and known stoichiometry. This route meets difficulties when the analysis is expanded to large-scale metabolic networks, because the necessary mechanistic details and standard kinetic data to derive the rate constants are difficult to be obtained. However, advances in genomics, transcriptomics, proteomics, and metabolomics, lead to a continuous expansion of bioinformatic databases, while advanced numerical techniques, non-conventional estimation procedures, and massive software platforms reported progresses in formulating such reliable cell models. Valuable *structured dynamic models*, based on cell biochemical mechanisms, have been developed for simulating various CCM-(sub) systems [EB1-EB4]. Conventional dynamic models, based on ordinary differential sets (ODE) of species mass balance, with a *mechanistic (deterministic) description* of reactions taking place among individual species (proteins, mRNA, intermediates, etc.) have been proved to be a convenient route to analyse continuous metabolic / regulatory processes and how the cell cope with the continuous perturbations from environment.

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 deterministic models, usually only *essential* reactions and components are retained, the model complexity depending on the measurable variables and available information. To reduce the structure of such a model, 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. Kinetic model reduction/lumping rules are

presented by Maria [EB1-EB2], and [48,50,53]. Such kinetic models can be useful to analyse the regulatory cell-functions, both for stationary and dynamic perturbations, to model cell cycles and oscillatory metabolic pathways [7,8,43,44,45], and to reflect the species interconnectivity or perturbation effects on cell growth [EB1, EB2]. Mixtures of ODE kinetic models with discrete states (i.e. 'continuous logical' models), and of continuous ODE kinetics with stochastic terms can lead to promising mixed models able to simulate both deterministic and non-deterministic cell processes [52]. Representation of metabolic process kinetics is made usually by using rate expressions of extended Michaelis-Menten or Hill type [EB1-EB2], and [8,9]. To model in detail the cell process complexity with deterministic ODE models is a challenging and difficult task. The large number of inner cell species, complex regulatory chains, cell signalling, motility, organelle transport, gene transcription, morphogenesis and cellular differentiation cannot easily be accommodated into existing computer frameworks. Inherently, any model represents a simplification of the real phenomenon, while relevant model parameters are estimated based on the how close the model behaviour is to the real cell behaviour. A large number of software packages have been elaborated allowing the kinetic performance of enzyme pathways to be represented and evaluated quantitatively [14,54]. Oriented and unified programming languages have been developed (SBML, JWS, see [EB1-EB2]) to include the bio-system organization and complexity in integrated platforms for cellular system simulation (E-Cell, V-Cell, M-Cell, A-Cell, see [EB2], and [14]). Continuous variable deterministic models, among other advantages, models can perfectly represent the cell response to continuous perturbations, and their structure and size can be easily adapted based on the available -omics information. Such integrated simulation platforms tend to use a large variety of biological databanks including enzymes, proteins and genes characteristics together with metabolic reactions are reviewed by [EB1-EB2]. Among them it is to mention: CRGM-database [55]; NIH-database, [56]; EcoCyc [57]; KEGG, [58], etc.

By applying various modelling routes, successful structured models have been elaborated to simulate various regulatory mechanisms [EB1- EB4], and [14,42,59-62]. In fact, as mentioned by [36], a precondition for a reliable modelling is the correct identification of both topological and kinetic properties. As few (kinetic) data are present in a standard form, non-conventional estimation methods have been developed, by accounting for various types of information (even incomplete) and global cell (regulatory) properties [36, 48].

The scope of this paper is to point-out some of the main aspects reviewed by the works of Maria [EB1-EB3] (see their covers in Figure 1) concerning some novel concepts and (bio) chemical engineering rules applied to kinetic modelling of metabolic processes and, especially the modular modelling of gene expression regulatory modules (GERM), GRC-s, and other metabolic processes on a deterministic basis by using continuous variable dynamic models under the novel WCVV modelling

framework promoted by Maria [EB1-EB4], and [14]. All the [EB1-EB4] works of Maria [1-4] are pleading in the favour of developing mathematical models on a mechanistic (deterministic) basis to characterize the cell metabolic processes, with focussing on modelling the dynamics and properties of CCM, GERM, and of GRC under the novel WCVV modelling framework.

GRC comparative modelling using WCVV vs. WCCV

This section is aiming to exemplify, in a simple and meaningful way, the importance of using a WCVV modelling framework compared to the classical (default) WCCV models when simulating the main regulatory properties of GERM-s or GRC-s, by explicitly accounting for the cell-volume growth, and system thermodynamic isotonicity (constant osmotic pressure). Exemplification is made for the case of the simplest generic GERM model, of [G(P)1] type (see GERM nomenclature in the below sections and the subsequent references), with characteristics taken from *E. coli* cells [14,16,17,41,57,63,64], by mimicking the cell homeostasis and its response to dynamic perturbations. The paper subject importance is very high, as long as a large number of cell simulators are developed and used for practical applications in the biosynthesis industry, and in medicine. The isotonicity constraint is proving to be a natural way to preserve the homeostatic properties of the cell system [EB1-EB4], and [14,15,33], instead of imposing others constraints, such as “the total enzyme activity” and “total enzyme concentration” constraints [65,66].

A comparison of model prediction quality in the case of a GERM of [G(P)1] type (Figure 2, middle/down) modelled under WCCV or WCVV, clearly indicate that WCCV can lead to biased and distorted conclusions on GERM regulatory performances (that is its response to both stationary or dynamic perturbations), thus making difficult the modular construction of GRC-s by linking individual GERM-s (Figure 5& 6) [16,17,33].

WCCV modelling framework

For a system of chemical or biochemical reactions conducted in a cellular defined volume 'V' (assumed an open system of uniform content), the classical (default) formulation of the corresponding (bio)chemical kinetic models based on continuous variables (concentration vector 'C', or number of moles vector 'n') implies writing a set of ordinary differential equations (ODE) representing the mass balance of the considered system states (biological/chemical species index 'j', taken individually or lumped), in the following WCCV (whole-cell constant-volume) modelling formulation (with referring to the whole system volume)[67]:

$$\frac{1}{V(t)} \frac{dn_j}{dt} = \sum_{i=1}^{nr} v_{ij} r_i(n/V, k, t) = h_j(c, k, t) \quad 1(A)$$

$$\frac{d(n_j/V)}{dt} = \frac{dC_j}{dt} = \sum_{i=1}^{nr} v_{ij} r_i(n/V, k, t) = h_j(C, k, t) \quad 1(B)$$

Where C_j = species “j” concentration; n_j = species “j” number

of moles; v_{ij} = the stoichiometric coefficient of the species “j” in the reaction “i”; r_i = reaction “i” rate; nr = number of reactions in the analyzed system. The above formulation assumes a homogeneous constant volume with no inner gradients or species diffusion resistance into the cell. When continuous variable ODE dynamic models are used to model cell enzymatic/metabolic processes, the default-modelling framework Eq. (1A-B) is that of a constant volume and, implicitly, of a constant osmotic pressure (π) in isothermal systems (to ensure the cell membrane integrity), according to the assumed fulfilled Pfeffer’s law in diluted solutions (i.e. the cytosol system) [13,14,68]

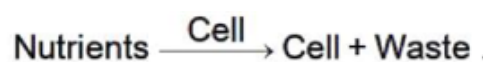
$$\pi V(t) = RT \sum_{j=1}^{n_s} n_j(t) \quad (2)$$

Where: T = absolute temperature, R = universal gas constant, V = cell (cytosol) volume; π = osmotic pressure; t = time; n_j = species “j” number of mole n_s = number of species in the analyzed system. To overcome this drawback, some WCCV models accounts for the cell-growing rate as a pseudo- ‘decay’ rate of key-species (often lumped with the degrading rate) in a so-called ‘diluting’ rate denoted here by an average D_s [see below eq.(3B) and eq.(4A) for its significance]. In fact, by ignoring the direct influence of the cell volume increase, the WCCV dynamic model cannot ensure the system isotonicity constraint fulfilment because the sum of species number of moles doubles over the cell cycle. Such a WCCV dynamic model might be satisfactory for modelling many cell sub-systems, but not for an accurate modelling of cell regulatory / metabolic processes under perturbed conditions, or for division of cells [68], distorting the prediction quality, as reviewed by [EB1-EB4], and [14,16,17,33]. Other researchers [66] tried to preserve the homeostatic properties of the cell system, not by imposing the isotonic constraint Eq. (2), but by means of “artificial” cell constraints, such as “the total enzyme activity” and “the total enzyme concentration” [65].

WCVV formulation

The WCVV (“whole cell variable cell volume”) modelling formulation is based by a couple of *hypotheses*, presented in this section (Figure 7). Life at its simplest level involves two major divisions of interacting molecular species called the *cell* and the *environment*. The environment consists of molecules dissolved in water and largely separated from the cell. In their simplest form, cells consist of hydrophilic molecules in aqueous volumes (*cytosol*), encapsulated by semi-permeable hydrophobic *membranes* composed of phospholipids and proteins [EB1-EB4], and [14,16,17,33].

Cellular components interact to catalyze the synthesis of more cells from environmental components called *nutrients*. Imported into the cell and transformed in *metabolites*. This auto-catalytic process is specified by the following overall auto-catalytic global lumped reaction:

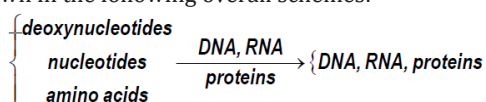


Mass Balance and State Equations	Remarks
$\frac{dC_j}{dt} = \frac{1}{V} \frac{dn_j}{dt} - D C_j = g_j(\mathbf{C}, \mathbf{k});$	continuous variable dynamic model representing the cell growing phase (ca. 80% of the cell cycle)
$\frac{1}{V} \frac{dn_j}{dt} = r_j(\mathbf{C}, \mathbf{k}); j = 1, \dots, n_s;$ $V(t) = \frac{RT}{\pi} \sum_{j=1}^{n_s} n_j(t);$	Pfeffer's law in diluted solutions
$D = \frac{1}{V} \frac{dV}{dt} = \left(\frac{RT}{\pi} \right) \sum_j \left(\frac{1}{V} \frac{dn_j}{dt} \right)$ $\frac{RT}{\pi} = \frac{V}{\sum_{j=1}^{n_s} n_j} = \frac{1}{\sum_{j=1}^{n_s} C_j} = \frac{1}{\sum_{j=1}^{n_s} C_{j0}} = \text{constant.}$	D = cell content dilution rate = cell volume logarithmic growing rate constant osmotic pressure (π) constraint
$\left(\begin{matrix} \text{all} \\ \sum_j C_j \end{matrix} \right)_{\text{cyt}} = \left(\begin{matrix} \text{all} \\ \sum_j C_j \end{matrix} \right)_{\text{env}}$	Derived from the isotonic osmolarity constraint
Hypotheses:	
<p>a. Negligible inner-cell gradients.</p> <p>b. Open cell system of uniform content.</p> <p>c. Semi-permeable membrane, of negligible volume and resistance to nutrient diffusion, following the cell growing dynamics.</p> <p>d. Constant osmotic pressure (the same in cytosol "cyt" and environment "env"), ensuring the membrane integrity ($\pi_{\text{cyt}} = \pi_{\text{env}} = \text{constant}$).</p> <p>e. Nutrient and overall environment species concentration remain unchanged over a cell cycle t_c.</p> <p>f. Logarithmic growing rate of average $D_s = \ln(2)/t_c$; volume growth of $V = V_0 e^{D_s t}$; t_c = duration of the cell cycle.</p> <p>g. Homeostatic stationary growth of $\left(dC_j / dt \right)_s = g_j(\mathbf{C}_s, \mathbf{k}) = 0$.</p> <p>h. Perturbations in cell volume are induced by variations in species copy numbers under the isotonic osmolarity constraint: $V_{\text{perturb}} / V = \left(\sum n_j \right)_{\text{perturb}} / \left(\sum n_j \right)$.</p>	

Notations: T = absolute temperature; R = universal gas constant; V= cell (cytosol) volume; π = osmotic pressure; c_j = cell species j concentration; n_j = species j number of moles; r_j = j-th reaction rate; t = time; k =rate constant vector.

Figure 7: The WCVV dynamic modelling framework and it's basic hypotheses [EB1, EB], [14].

As long as excess nutrients are available, this auto-catalysis causes cell populations to increase exponentially. The volume of a newborn cell doubles during its cell cycle. Cells contain nucleic acids (DNA, RNA, or both) and proteins, interrelated through the processes of transcription, translation and DNA replication. Taken together, these *metabolic processes* are mutually autocatalytic, as shown in the following overall schemes:



DNA and protein are co-catalysts for RNA synthesis from ribo-nucleotides. In turn, RNA and proteins (enzymes) are co-catalysts for the synthesis of proteins from amino acids, of DNA (from the monomeric units called deoxyribonucleotides),

RNA, and other proteins. The substrates for these processes (deoxyribonucleotides, nucleotides, amino acids etc.) are *metabolites*, synthesized from imported environmental *nutrients* through complex metabolic pathways [69].

“At this point, it is to strongly emphasize that living cells are systems of variable volume. They double their volume during the cell cycle. For chemical or biochemical systems of variable-volume, another formulation is more appropriate, being given by Aris [70] for chemical reacting systems. Such CBE modelling tools were translated and promoted in developing structured math models of cell processes (that is CCM and GRC-s) by Maria [EB1-EB4], and [5,13,14,15,16,17,33,41]. Such a novel kinetic modelling framework of cell systems by also including the cell isotonicity, and the variable cell-volume constraints in the so-called WCVV

modelling of metabolic dynamic processes (for details, see the above references). In mathematical terms, the species mass balance Eq.(1) should be re-written in the following form:

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

$$C(t=0) = C_s$$

Where: C = cell species concentration vector; t = time; k = math model constants; 's' index = at steady-state; h = cell kinetic model functions. The variable "D" is the logarithmic growing rate of the cell volume, also known as the "cell dilution rate". It is defined by the following relationship:

$$D = \frac{d(\ln(V))}{dt} \quad 3(B)$$

There are two possibilities to calculate the cell dilution 'D' necessary for solving the model Eq. (3A). The simplest, but not the accurate one, is to use a value averaged over the whole cell cycle, that is:

$$\frac{1}{V} \frac{dV}{dt} = D_s, \text{ leading to } V(t) = V_0 \exp(D_s t) \quad (4A)$$

By accounting the cell double volume at the end of the cell-cycle, then the constant D_s can be a-priori evaluated by using the following relationship (for cells of known cell cycle):

$$\ln\left(\frac{2V_0}{V_0}\right) = D_s t_c, \text{ and } D_s = \ln(2) / t_c \quad (4B)$$

The second alternative, and the more rigorous way to evaluate the cell dilution 'D' is those to impose a constraint accounting for the cell-volume growth while preserving a constant osmotic pressure and membrane integrity. Thus, by derivation of the Pfeffer's law Eq. (2) in respect to V, and by division to V, one obtains the 'isotonic' dilution rate D_i [EB1], and [14,16,17,33]:

$$D_i = \frac{1}{V} \frac{dV}{dt} = \left(\frac{RT}{\pi} \right) \sum_j \left(\frac{1}{V} \frac{dn_j}{dt} \right) \quad (5)$$

It is to observe in Eq. (5) that the cell content dilution rate D_i is linked to the all species (taken individually or lumped) reaction rates via the isotonicity constraint. As species reaction rates varies during the cell cycle, it clearly results that formulation Eq. (5) offers a more accurate estimation of the (variable) cell dilution at any time. Such a system isotonicity constraint is more 'natural' and eventually includes "the total enzyme activity" and "the total enzyme concentration" constraints suggested by Komasilovs et al. [66].

In the above relationships eqns.(2, 5), the following notations have been used: T = absolute temperature, R = universal gas constant, V = cell (cytosol) volume. As revealed by the Pfeffer's law eqn.(2) in diluted solutions [71], and by the eq.(5), the volume dynamics is directly linked to the molecular species dynamics under isotonic and isothermal conditions. Consequently, the cell dilution 'D' results as a sum of reacting rates of the 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 eqn.(2):

$$V(t) = \frac{RT}{\pi} \sum_{j=1}^{ns} n_j(t) \Rightarrow \frac{RT}{\pi} = \frac{V(t)}{\sum_{j=1}^{ns} n_j(t)} = \frac{1}{\sum_{j=1}^{ns} C_j} = \frac{1}{\sum_{j=1}^{ns} C_{jo}} = \text{constant} \quad (6)$$

The basic hypotheses of the WCVV dynamic models of type Eqs. (3-6) are briefly presented in the below paragraphs. These formulations are valid over ca. 80% of the cell cycle representing the balanced cell growth before its division [68].

The whole chemical/biochemical cell processes are called 'cell metabolism', defined as: Metabolism is the set of life-sustaining chemical transformations within the cells of living organisms. The three main purposes of metabolism are the conversion of food/fuel to energy to run cellular processes, the conversion of food/fuel to building blocks for proteins, lipids, nucleic acids, and some carbohydrates, and the elimination of nitrogenous wastes. These enzyme-catalyzed reactions allow organisms to grow and reproduce, maintain their structures, and respond to their environments [69].

The basic equations and *hypotheses* of a deterministic WCVV simplified cell model (with continuous variables) presented in this work, also called a "mechanical cell", are discussed by Maria [EB1-EB4], [14,33], and summarized in (Figure 7). To better underline the WCVV models hypotheses, a couple of issues should be explained, as followings [EB1-EB4]:

a) Genes (generically denoted by G), and the encoded proteins (generically denoted by P) are in a mutually auto-catalytic relationship: the synthesis of P is catalyzed by G, and vice-versa (directly or indirectly), in the so-called GERM-s (see the GERM library of Figure 2-down).

b) During its cell cycle, the cell volume grows continuously, but preserving a constant osmotic pressure.

c) The regulatory mechanisms to achieve the gene expression modelled by lumped GERM-s, and the internal homeostasis are explained in detail by [EB1-EB4], and [14,16,17,33].

d) The cell WCVV model assumes an ideal system, that is: isothermal and 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 isotonic (constant osmotic pressure), with no inner gradients or species diffusion resistance.

e) The cell is an open system interacting with the environment through a semipermeable membrane. To better reproduce the GERM properties interconnected with the rest of the cell, the other cell species are lumped together in the so-called "cell ballast" [EB1-EB4], and [14,16,17,33]. The cell ballast has an important influence on the GERM performance indices (see below) through the common cell volume to which all species contributes.

f) The inner osmotic pressure (π cyt) is constant, and

all time equal with the environmental pressure, thus ensuring the membrane integrity ($\pi_{\text{cyt}} = \pi_{\text{env}} = \text{constant}$ [EB1, EB3]). 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 permeases 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 [$V = V_0 \exp(+D_s \cdot t)$] [see eqn.(3B,4A,4B,5)].

g) Due to the “D” term in eq.(3A, 3B), the cell content reports a continuous dilution, that is a species concentration decline due to the continuous increase of the denominator of the expression

h) $C_j = n_j(t)/V(t)$. In spite of that, concentrations of key species remain constant because the numerator (copynumbers) increases at the same rate with the denominator. So, the overall concentration of cellular components is time-invariant at the cell homeostasis (i.e. quasi-steady-state, or balanced growth).

i) Species concentrations at the cell level are usually expressed in nanomoles, being computed with the relationship of Maria [14]

$$\text{Concentration} = \frac{\text{no. of copies/cell}}{N_A \times V_{\text{Cyt},0}} \quad (7)$$

j) where N_A is the Avogadro number. For instance, for an *E. coli* cell with an approximate volume of $V_{\text{Cyt},0} = 1.66 \cdot 10^{-15}$ L [64], concentration of one generic gene G copynumber is: $[G]_s = (1/(6.022 \times 10^{23})(1.66 \times 10^{-15})) = 1 \text{ nM}$ (that is 10^{-9} mol/L).

k) Under quasi-stationary growing conditions (QSS), from eq.(1A, 3A) it results that species “j” synthesis rates (r_j) must equal to first-order dilution rates ($D_s C_{j,s}$), leading to the time-invariant (index ‘s’) species concentrations $C_{j,s}$, i.e. the homeostatic conditions (corresponding to a balanced steadystate growth). Under QSS cell growing conditions, the ODE model mass balance eq. (3A) is leading to the following nonlinear algebraic mass balance set:

$$\left(\frac{dC_j}{dt}\right)_s = \left(\frac{1}{V} \frac{dn_j}{dt}\right)_s - D_s C_{j,s} = h_{j,s}(C_s, k, t) = 0; \left(\frac{1}{V} \frac{dn_j}{dt}\right)_s = r_{j,s} \quad (8)$$

$j = 1, \dots, n_s$ (no. of species); where $D_s = \left(\frac{RT}{\pi}\right) \sum_j \left(\frac{1}{V} \frac{dn_j}{dt}\right)_s$

This QSS mass balance eq.(8) is used to estimate the rate constants ‘k’ by using the known experimental stationary species concentration vector C_s , with also imposing some constraints to ensure the optimal properties of the cell system. Some examples are given by [EB1-EB4], and [9,10,11,12,14,15,16,17,33,41,72,73].

$$dC/dt = h(c, k); C(0) = C_s; dA/dt = J_c A; A(0) = I; J_c = (\partial h(c, k) / \partial C)_s \quad (11)$$

j) It is to observe that, in a continuous variable cell kinetic model, species concentrations can present fractional values. When treated deterministically, fractional copy numbers must be loosely interpreted either as time-invariant average in a population of cells, or as a time-dependent average of single cells. For other types of cell kinetic models (with stochastic, or Boolean variables, topological, etc.) see the review of [EB1-EB3], and [14].

k) A metabolic kinetic model in a WCVV approach should be written in the

form eq. (3-6). In such a formulation, all cell species should be considered (individually or lumped), because all species net reaction rates contribute to the cell volume increase [see eq.(6)]. As the cell volume is doubling during the cell cycle, this continuous volume variation can not be neglected.

The simplest representation of the core of such a ‘mechanical cell’ is shown in (Figure 5-down-right). It exists in an environment consisting of two nutrients NutG and NutP. The cell contains one gene (lumped genome), and protein (lumped proteome) [11,12] in a mutually autocatalytic relationship, two lumped metabolites (MetG and MetP) used in the synthesis of the G/P pair, and various regulatory elements promoting internal homeostasis. A membrane is presumed to demarcate the cell from its environment, but is not an explicit component of the system.

Advantages of using the WCVV kinetic modelling framework in living cells:

As an important observation, “from eqn. (5) it results that the cell dilution is a complex function $D(C, k)$ being characteristic to each cell and its environmental conditions. Relationships eq.(5-6) are important constraints imposed to the WCVV cell model eq.(3A-B), eventually leading to different simulation results compared to WCCV models that neglect the cell volume growth and isotonic effects (see some examples given by [EB1-EB4], and [16,17,33]).

On the contrary, application of the default classical WCCV-ODE kinetic models of eqns. (1A-B) type with neglecting the isotonicity constraints presents a large number of inconveniences, related to ignoring lots of cell properties (discussed in detail by [EB1-EB4], and [16,17,33]), that is:

- 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-s regulatory performance indices (P.I.s, see [EB1-EB4], and [16,17,33])
- the more realistic evaluation of the recovering/transient times after perturbations (see Figure 5, and Figure 6)[33]

- e) loss of the intrinsic model stability;
- f) loss of the self-regulatory properties after a dynamic perturbation, etc.”

The multiple advantages of the WCVV modelling framework are discussed, and exemplified by Maria [EB1-EB4], and [13-17,33,34,49]. In short, the novel proposed modelling concept/framework WCVV, widely promoted in a large number of applications by Maria (see the above references) allows to derive cell kinetic models, in a holistic approach, well reproducing the cell processes homeostasis, and the individual/holistic GRC regulatory properties, by including in a natural way constraints related to the cell system isotonicity, and the variable-volume in relationship to the species reaction rates, and the lumped proteome/genome replication [EB1-EB4], and [11,12,14,16,17,33]. Such an isotonicity constraint is required to ensure the cell membrane integrity, but also to preserve the homeostatic properties of the cell system, not by imposing ‘the total enzyme activity’, or the ‘total enzyme concentration’ constraints used by the classical (default) constant-volume cell modelling approach (WCCV). As proved by Maria et al. [33], compared to the classical WCCV models, the WCVV novel modelling framework is leading to a more accurate simulation of several cell metabolic effects, such as: relationships between the external conditions, species net synthesis reactions, osmotic pressure, cell content (ballast) influence on smoothing

the continuous perturbations in external nutrient concentrations, the more realistic representation of GERM regulatory modules, etc.

This shortly presented WCVV holistic modelling framework proposed, extrapolated, and widely promoted in a large number of applications in bioengineering and bioinformatic by Maria (see the above references) was proved to be more accurate and present a large number of advantages briefly reviewed in the (Table 1).

Other aspects related to the extended MSDKM model under WCVV approach, such as a) numerical rules to reduce its structure; b) estimation of the kinetic model rate constants from the QSS mass balance eq.(8) with known (from experiments) stationary species concentration vector C_s , with also imposing some optimal regulatory properties to the math-modelled cell system, and c) some examples are given by [EB1-EB4]

GERM-s regulatory performance indices (P.I.):

”As proved in previous works [EB1-EB3], the performances indices (P.I.-s) of GERM-s of [G(P)n] type in (Figure 2-down), are as better as the number ‘n’ of buffer reactions increases (Figure 14).

Also, Maria [EB1-EB3] proved that when P is acting as a TF, its efficiency is better if it is present in a dimeric form (PP), in GERM-s of [G(PP)n] type in (Figure 2-down, and Figure 8).

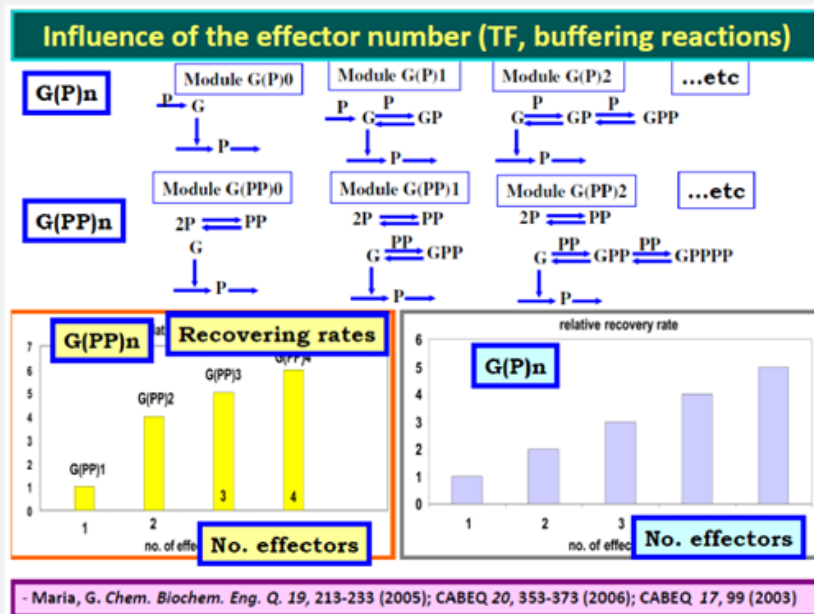


Figure 8: Influence of the number of effectors in a GERM on their regulatory performance indices P.I.-s [EB2-EB3].

Maria [EB1-EB3] also proved that the GERM regulatory efficiency is better if TF is a dimmer PP acting at both G and M levels of the expression (middle and down-rows of Figure 2-down), thus

developing a cascade control scheme of the expression where transcription and translation regulatory steps are separately considered, that is GERM-s of [G(PP)n;M(PP)n'] type.

Following an extended study [EB1-EB3], it was concluded that, as the number of effectors increases in GERM-s as their P.I.-s are better, the right GERM choice being case dependent (see [11,12] as an example). Simulation with a large number of GERM types lead to important conclusions. Some of them are presented in this section.

Perturbations of the species steady-state (homeostatic) concentrations are caused by environmental processes. In a GERM case, these processes tend to increase or decrease the key-protein stationary level [P]s. These processes occur in addition to those of the “core” system (G/P pair replication over the cell cycle).

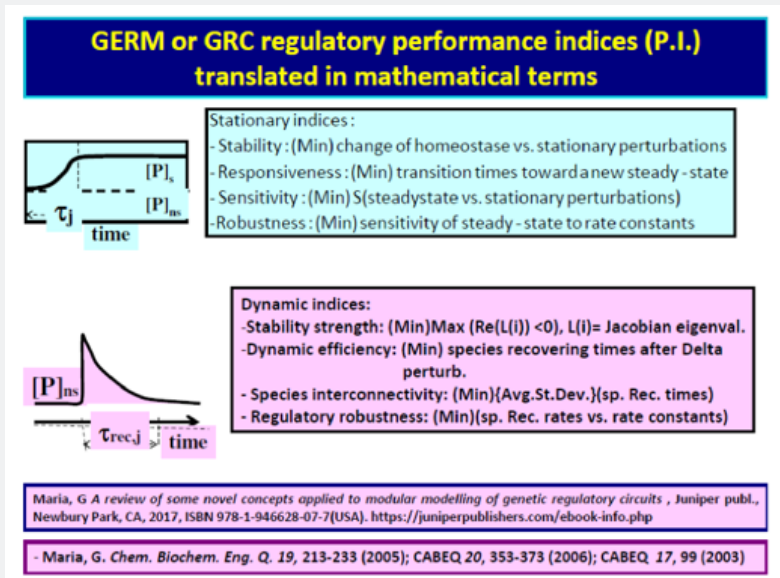


Figure 9: Some regulatory performance indices P.I.-s of GERM, following a stationary (up-blue- rows), or a dynamic (down-pink- rows) [EB1-EB4].

Index	Goal	Objective Expression
stationary regulation	Min	$R_{ss} = ([P]_s - [P]_{ns}) / [P]_{ns}$
stationary regulation	Max	$A_{sync} = k_{syn} \times k_{decline}$
stationary regulation	Min	$S_{NIP_j}^i = [(\partial C_i / C_{is}) / (\partial C_{NIP_j} / C_{NIP_{jz}})]_S$
stationary regulation	Min	$S_{kj}^i = [(\partial C_i / C_{is}) / (\partial k_j / k_j)]_S$
dynamic regulation	Min	$R_{ij} = Max(Re(\lambda_i)), Re(\lambda_i) < 0$
dynamic regulation	Min	$\tau_j; \tau_P$
regulatory robustness	Min	$(\partial R_D / \partial k)$
species interconnectivity	Min	$AVG(\tau_j) = average(\tau_j)$
species interconnectivity	Min	$STD(\tau_j) = st.dev.(\tau_j)$
QSS stability(note a)	Min	$Re(\lambda_i) < 0, \text{ for all } i$
QSS stability strength (note a)	Min	$Max(Re(\lambda_i))$
QSS stability strength(note b)	Min	$ \lambda_{A_j} < 1$

Figure 10: The main regulatory efficiency performance indices (P.I.-s) proposed by Maria [EB1-EB4], [14] to evaluate the treatment efficiency of dynamic/stationary (Figure 9) perturbations inside cell by a GERM following the definitions of [14]. Abbreviations: Min = to be minimized; Max = to be maximized. Note: k(syn) and k(decline) refers to the $\rightarrow P \rightarrow$ overall reaction. See the notations and details of [EB1- EB3].

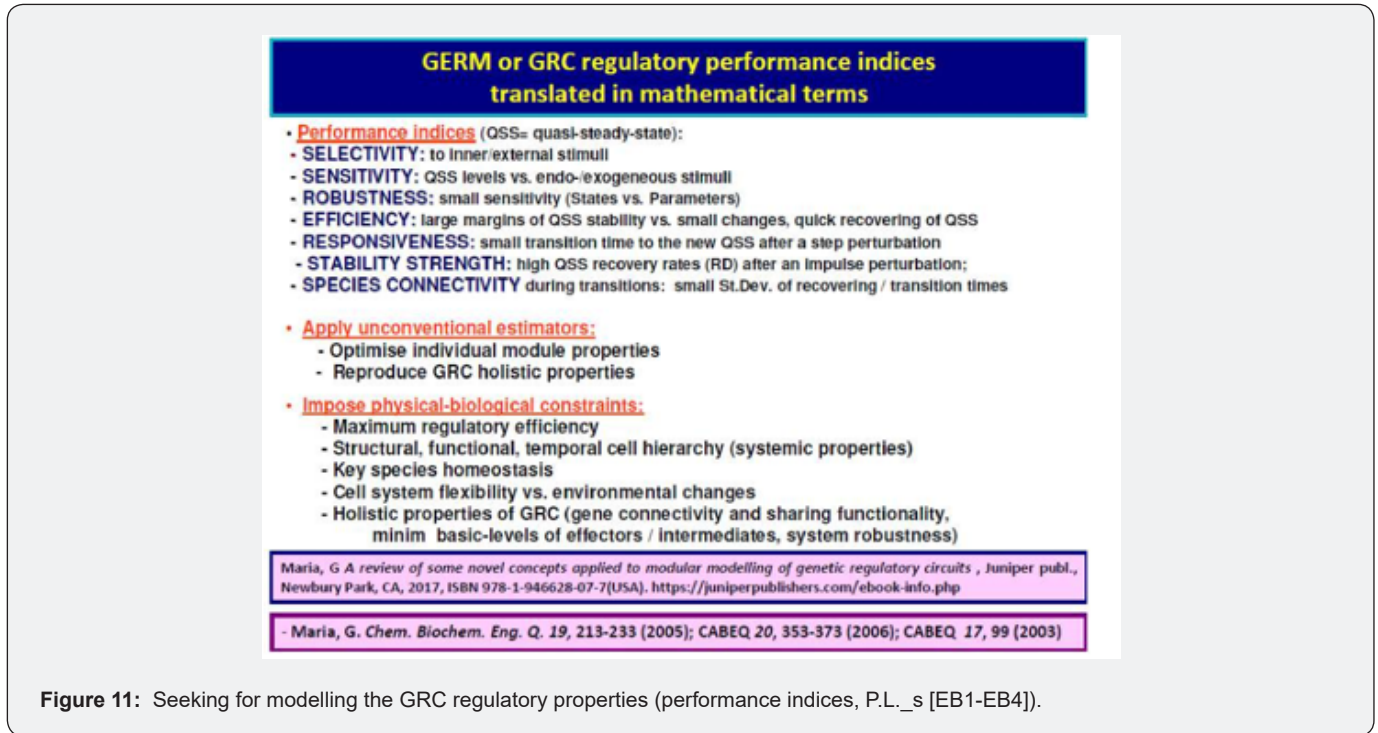


Figure 11: Seeking for modelling the GRC regulatory properties (performance indices, P.L._s [EB1-EB4]).

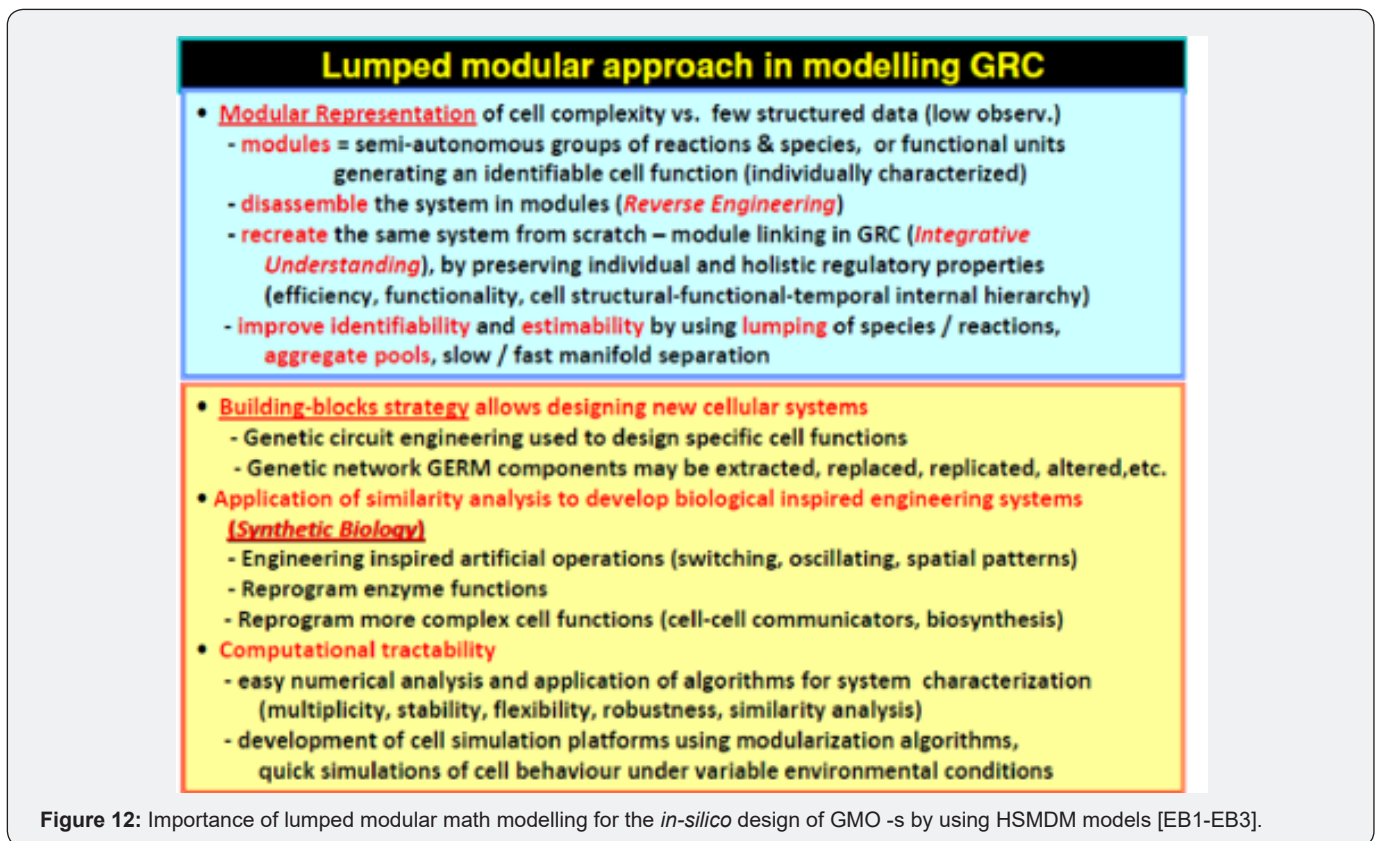


Figure 12: Importance of lumped modular math modelling for the *in-silico* design of GMO -s by using HSMDM models [EB1-EB3].

GERM or GRC regulatory performance indices P.I.-s are of two types [EB1-EB3]: *stationary* and *dynamic*, according to the perturbation type (Figure 9), and Figure 13). Briefly the P.I.-s are presented in the (Figure 10), together with the associated

optimization objective (goal), for a general nonlinear dynamic cell model described by eq. (3A-B). See also an intuitive display in (Figure 9, and Figure 11) Detailed information is given by Maria [EB1-EB3], and [14]. The monodromy matrix A, necessary

to express the species QSS-level stability 'strength' is evaluated together with the cell process ODE model eqn.(3A-B), by using the differential relationship eq (11).

Modelling individual GERM-s under WCVV formulation:

In order not to overly complicate the HSMDM models that also include GRC-s, it is necessary to have a "library" of kinetic

models to represent individual GERM-s, to be used for build-up GRC-s of desirable properties (e.g. genetic switches, operon expression, etc.). See some case studies of [EB1-EB4]. Obviously, the selection of the most suitable GERM to be included in the GRC chain depends on its regulation performances (that is, the so-called selfregulation performance indices (P.I.-s), related to the GERM type.

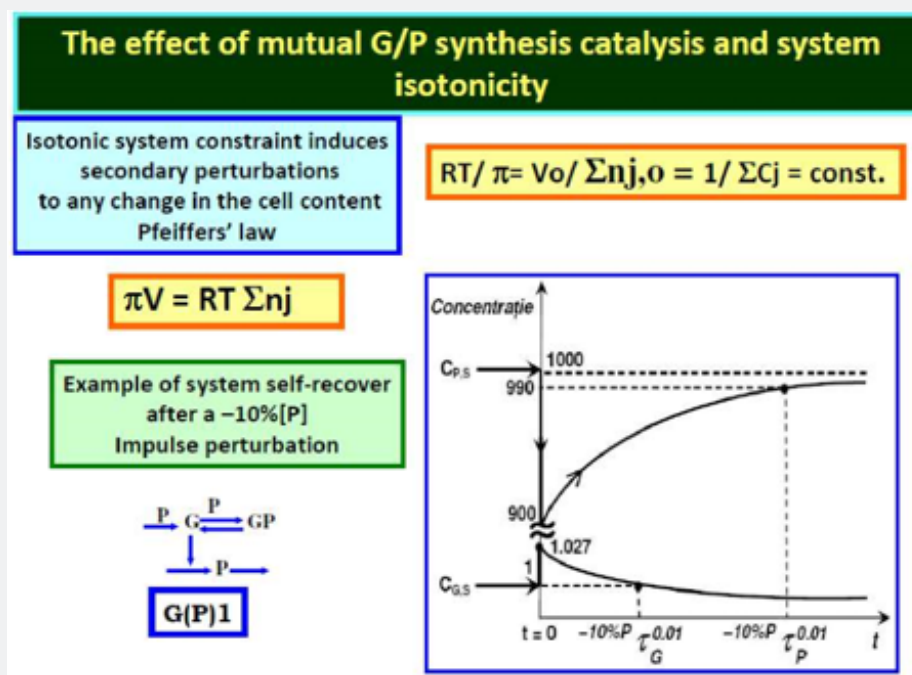


Figure 13: The effect of the G/P mutual self-catalysis, and of the isotonicity in the case of a simple [G(P)1] gene expression regulatory module GERM [EB1-EB3], [16,17]. Species QSS recovery after a 10% dynamic (impulse-like) perturbation in [P]s. Notations: π = cell osmotic pressure; V = cell volume; T= temperature; G = generic gene; P = the protein encoded by G; C(j) = cell species concentrations; „s” index = at QSS; „o” index = initial; $\tau_P^{0.01}$ = species P recovering time of its QSS, [P]s., with a 1% precision (i.e. 0.01).

This chapter briefly presents the main GERM-s models proposed by Maria [EB1-EB3], and [74] used in the construction of HSMDM-s, in terms of the reduced reaction scheme, kinetic model, and their associated P.I.-s.

As experimentally proved in the literature [EB1-EB4], the GRC-s (or GRN-s) „that control the synthesis of all proteins (enzymes) in the cell, present a modular construction, every operon (a cluster of genes under the control of a single promoter) including a variable number of interacting GERM-s. However, it is well-known that one GERM interacts with no more than other 23-25 GERM-s [75], while most of GERM structures are repeatable. Consequently, when developing the GRC *in-silico* analysis and reaction schemes / kinetic models, the modular approach is preferred due to several advantages: (i) A separate analysis of the constitutive GERM-s in conditions that mimic the stationary or perturbed cell growth; (ii) The GERM modules are then *in-silico* linked to construct the target GRC of an optimized regulatory

efficiency that ensures key-species homeostasis and cell network holistic properties (Figure 12, and Figure 4). (iii) *In-silico* investigations of GERM-s and GRC-s characteristics focus on the tight control of gene expression, the quick dynamic response, the high sensitivity to specific inducers, and the GRC robustness (i.e. a low sensitivity vs. undesired inducers). Such advanced regulatory structures must ensure the homeostasis (quasistationarity) of the regulated key-species, and quick recovery (with a trajectory of minimum amplitude) after a dynamic (impulse-like) or stationary (step-like) perturbation of one of the involved metabolites or nutrients [EB1-EB3], and [74] (Figure 13, and Figure 14).

However, to not complicate the resulted simulation model when coupling GERM chains in complex GRC-s, simple GERM dynamic models have been proposed and investigated by various researchers [EB1-EB3], and

[9,27,39,40,63,76], with Hill-type, or pseudo-Hill-type activation [EB1-EB3].

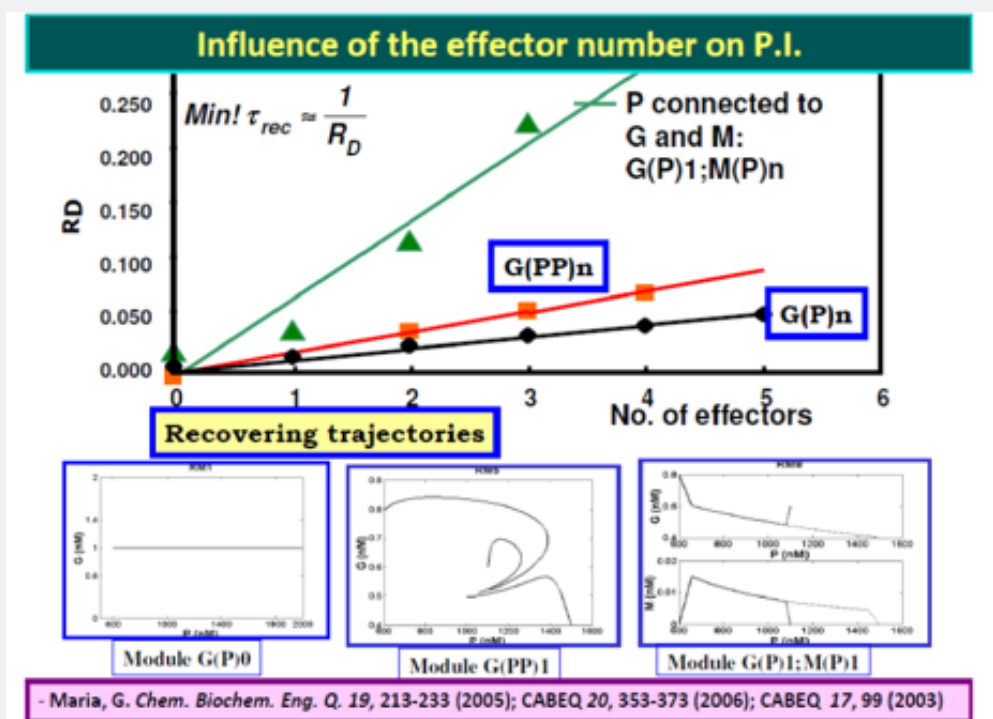


Figure 14: Influence of the GERM's number of effectors on some of their properties, i.e. the QSS recovering rate (top), and the amplitude / species recovering trajectory after a 10% dynamic perturbation in the [P]s [EB1-EB3], [74].

To make this rule easier, Maria [EB1-EB3] elaborated a library with reduced representations of GERM-s (Figure 2-down) to be used for every particular case. Of course, these individual GERM modules differ by the regulatory performance indices (P.I.) to be further defined, in response to stationary or dynamic perturbations into the cell, or in the environment.

These simplified deterministic models of lumped GERM and structures have been proved to adequately represent complex GRC-s (Figure 4). The simplest GERM structure with one regulatory element is those denoted by [G(P)1], one of the better regulatory efficiency [G(PP)1] (Figure 2-middle).

The generic [G(P)1] regulatory module (schematically represented in (Figure 2-down, the row-up) refers to the synthesis of a generic protein P and the simultaneous replication of its encoding gene G. The lumped [G(P)1] model (Figure 5) includes only one regulatory element (a so-called "effector", that is a fast 'buffer' reversible reaction $G + P \rightleftharpoons GP(\text{inactive})$) (Figure 2-down), aiming at controlling the P synthesis rate and its homeostatic QSS level. The following notations have been used: G = active part of the gene encoding protein P; GP = inactive part of the gene encoding protein P; MetG, MetP = lumped DNA and protein precursor metabolites, respectively.

In such a generic lumped construction, the protein P and its encoding gene G mutually catalyses the synthesis of each other. The protein P is the 'control node' playing multiple roles in

such a simplified lumped representation. Thus, P is a permease leading to the import of nutrients NutG, NutP in the cell, but also a metabolase converting the nutrients into precursors MetG and MetP of the G and P respectively. Protein P is also a polymerase catalysing the gene replication. And, finally, the protein P is also a transcriptional factor (TF) by dynamically adjusting the catalytic activity of the G by means of a very rapid 'buffer' regulatory reaction $G + P \rightleftharpoons GP(\text{inactive})$. When P is produced in excess, it reversibly inactivates more amount of G, which in turn, will slow-down the P synthesis. When P is produced in too low amounts, the regulatory process goes backwards.

The module nomenclature used in (Figure 2-down) for such GERM models, proposed by [14,77] is those of $[L_i(O_i)n_i; \dots; L_i(O_i)n_i]$. It includes the assembled regulatory units $L_i(O_i)n_i$. One unit 'i' is formed by the component $L(i)$ (e.g. enzymes or even genes G, P, M, etc.) at which regulatory element acts, and $n(i) = 0, 1, 2, \dots$ number of 'effectors'/TF, generically denoted by species $O(i)$ (that is 'effectors' like P, PP, PPPP, R, RR, RRRR, etc.) binding the 'catalyst' L. For instance, a [G(P)2] unit of (Figure 2-down) includes two successive binding steps of G with the product P, that is $G + P \rightleftharpoons GP + P \rightleftharpoons GPP$, all intermediate species GP, GPP, being inactive catalytically, while the mass conservation law is all time fulfilled [1-3], i.e.

$$\sum_{i=0}^2 [G(P)] = \text{constant}$$

Such a representation accounts for the protein concentration diminishment due to the cell-growth dilution effect, but could also include protein degradation by proteolysis. It is also to observe that such GERM lumped models try to account essential properties of the gene expression, that is a highly self- / cross- regulated and mutually catalyzed process by means of the produced enzymes / effectors. As depicted in (Figure 2-down) for the [G(P)1] module case, the protein P synthesis is formally catalysed by its encoding gene G. In turn, P protein formally catalyse the G synthesis, but also modulate the G catalyst activity (via the fast buffering reaction $G + P \rightleftharpoons GP$).

Even if such a generic [G(P)1] regulatory module is more complex, by including a large number of reactions involved in the regulation of the gene expression (schematically presented by [EB1-EB3]), it was proved (see the case study of [EB4]) to satisfactory reproduce the dynamics of complex GRC-s in HSMDM models.

As proved by Maria [EB1-EB4], [15] this simplified formulation of GERMs in (Figure 2-down) implicitly ensures the *homeostatic regulation* of the gene expression and G/P mutual auto-catalysis of their synthesis.

An extended discussion on the GERM types and their ranked/ordered regulatory efficiency is given by [EB1-EB3].

Conclusion

As a general conclusion, the CBE and NSCT principles and modelling rules are fully applicable to math modelling cellular metabolic processes, by developing extended and valuable of using MSDKM and HSMDM dynamic math models (within the

WCVV modelling framework). This involves application of the classical modelling techniques (mass balance, thermodynamic principles), algorithmic rules, and modern CBE concepts [80]. The metabolic pathway representation with continuous and/or stochastic variables remains the most adequate and preferred representation of cell processes, the adaptable-size and structure (reaction, species) of the lumped model depending on available information and model utilisation scope.

The developed MSDKM and HSMDM have been proved to be very useful when solving various bioengineering problems (a) to *in-silico* (math model based) off-line optimize the operating policy of various types of bioreactors, and (b) to *in-silico* design/screening/check some GMO-s of industrial use able to improve the performances of several bioprocess/bioreactors.

GRC representations combining *Reverse Engineering* and *Integrative Understanding* [EB1-EB4] allows *in-silico* design of GRC-s inducing specific cell motifs of GMO-s. Examples given by [EB1-EB4] includes: Genetic switches of adjustable certainty, sensitivity to exo-/endogeneous stimuli, responsivity, regulatory efficiency; Metabolism behaviour of wild or cloned cells with plasmids. Even if the cell simulators still present lot of drawbacks and limited adequacy, they become more and more valuable tools in designing GMO-s with desirable characteristics, or for obtaining micro-organisms cloned with desirable plasmids with important applications in industry (production of vaccines), or in medicine (gene therapy). As mentioned by the late Dr. G.E.P. Box (Professor of statistics at the Univ. of Wisconsin, and a pioneer in the areas of quality control, time series analysis, design of experiments, and Bayesian inference): "All models are wrong, but some are useful."

Table 1: Some advantages when using the holistic WCVV framework when modelling GRC-s [EB1-EB4], and [16,17,33].

a.	the role of the high cell-ballast in "smoothing" the perturbations of the cell homeostasis;
b.	the secondary perturbations transmitted via the cell volume;
c.	the system isotonicity constraint reveals that every inner primary perturbation in a key-species level (following a perturbation from the environment) is followed by a secondary one transmitted to the whole cell via cell volume;
d.	allows comparing the regulatory efficiency of various types of GERM-s;
e.	allows a more realistic evaluation of GERM performance indices [33]
f.	allows studying the recovering/transient intervals between steady-states (homeostasis) after stationary perturbations;
g.	allows studying conditions when the system homeostasis intrinsic stability is lost
h.	allows studying the self-regulatory properties after a dynamic/stationary perturbation, etc.
i.	allows simulate with a higher accuracy the plasmid-level effects in cloned cells [11,12].

References

1. Maria G (2017) A review of some novel concepts applied to modular modelling of genetic regulatory circuits, Juniper publ Irvine, CA, USA.
2. Maria G (2017) Deterministic modelling approach of metabolic processes in living cells - a still powerful tool for representing the metabolic process dynamics, Juniper publ., Irvine, CA, USA.
3. Maria G (2018) In-silico design of Genetic Modified Micro-organisms (GMO) of industrial use, by using Systems Biology and (Bio)Chemical Engineering tools, Juniper publ Irvine, CA USA,
4. Maria G (2023) Hybrid modular kinetic models linking cell-scale structured CCM reaction pathways to bioreactor macro-scale state variables. Applications for solving bioengineering problems, Juniper publ. Irvine, CA, USA in-press.

5. Maria G (2017) Application of (bio) chemical engineering principles and lumping analysis in modelling the living systems, *Current Trends in Biomedical Engineering & Biosciences* 1(4): 555-566.
6. Maria G (2012) Enzymatic reactor selection and derivation of the optimal operation policy, by using a model-based modular simulation platform. *Comput Chem Eng* 36(10): 325-341.
7. Maria G (2020) In-silico determination of some conditions leading to glycolytic oscillations and their interference with some other processes in *E. coli* cells. *Frontiers in Chemistry* 8: 526679-526693.
8. Maria G (2014) Insilico derivation of a reduced kinetic model for stationary or oscillating glycolysis in *Escherichia coli* bacterium. *Chemical & Biochemical Engineering Quarterly* 28(4): 509-529.
9. Maria G (2014) Extended repression mechanisms in modelling bistable genetic switches of adjustable characteristics within a variable cell volume modelling framework. *Chemical & Biochemical Engineering Quarterly* 28(1): 35-51.
10. Maria G (2009) Building-up lumped models for a bistable genetic regulatory circuit under whole-cell modelling framework. *Asia-Pacific Journal of Chemical Engineering* 4: 916-928.
11. Maria G (2010) A dynamic model to simulate the genetic regulatory circuit controlling the mercury ion uptake by *E. coli* cells, *Revista de Chimie (Bucharest)* 61(2): 172-186
12. Maria G ,Luta I (2013) Structured cell simulator coupled with a fluidized bed bioreactor model to predict the adaptive mercury uptake by *E. coli* cells. *Computers & Chemical Engineering* 58: 98-115.
13. Maria G, Morgan J, Lindahl P (2002) Kinetic simulation of hypothetical "Mechanical Cells" that maintain intracellular homeostasis while growing autocatalytically on environmental nutrients present in variable amounts. *Jl Theor Biology*.
14. Maria G (2005) Modular-Based Modelling of Protein Synthesis Regulation. *Chemical and Biochemical Engineering Quarterly* 19: 213-233
15. Maria G (2006) Application of lumping analysis in modelling the living systems -A trade-off between simplicity and model quality, *Chemical and Biochemical Engineering Quarterly* 20: 353-373.
16. Maria G , Scoban AG (2017) Setting some milestones when modelling gene expression regulatory circuits under variable-volume whole-cell modelling framework. 1 Generalities *Revista de Chimie (Bucharest)* 68(12): 3027-3037.
17. Maria G, Scoban AG (2018) Setting some milestones when modelling gene expression regulatory circuits under variable-volume whole-cell modelling framework - 2 Case studies. *Revista de Chimie(Bucharest)* 69(1): 259-266.
18. Maria G, Crisan M (2017) Operation of a mechanically agitated semi-continuous multi-enzymatic reactor by using the Pareto-optimal multiple front method. *J Process Control* 53: 95-105.
19. Maria G (2020) Model-based optimization of a batch reactor with a coupled bienzymatic process for mannitol production. *Computers & Chemical Engineering* 133: 106628-106635.
20. Maria G, Peptanaru IM (2021) Model-based optimization of mannitol production by using a sequence of batch reactors for a coupled bi-enzymatic process - A dynamic approach. *Dynamics-Basel* 1: 134-154.
21. Maria G, Renea L, Maria C (2022) Multi-objective optimization of the fed-batch bi-enzymatic reactor for mannitol production, *Dynamics-MDPI* 2: 270-294.
22. Straathof AJJ, Adlercreutz P (2005) *Applied biocatalysis*. Harwood Academic Publ. Amsterdam.
23. Maria G (2020) Model-based optimization of a fed-batch bioreactor for mAb production using a hybridoma cell culture. *Molecules-Basel - Organic Chemistry* 25(23): 5648-5674.
24. Maria G, Renea L (2021) Tryptophan production maximization in a fed-batch bioreactor with modified *E. coli* cells, by optimizing its operating policy based on an extended structured cell kinetic model. *MDPI Bioengineering-Basel* 8(12): 210-247.
25. DiBiasio D (1989) Introduction to the control of biological reactors. In: Shuler ML. (ed.), *Chemical engineering problems in biotechnology*. American Institute of Chemical Engineers New York USA pp. 351-391.
26. Miskovic L, Tokic M, Fengos G, Hatzimanikatis V (2015) Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes. *Current Opinion in Biotechnology* 36: 146-153.
27. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) *Metabolic Engineering. Principles and Methodologies*. Academic Press San Diego CA.
28. Maranas C, Hatzimanikatis V (2005) Preface - Cell modelling. *Comp Chem Eng* 29: 405-406.
29. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* 3 Article number 121: 1-18.
30. Hatzimanikatis V, Floudas CA, Bailey JE (1996) Analysis and design of metabolic reaction networks via Mixed-Integer Linear Optimization. *AIChE Journal* 42(5): 1277-1292.
31. Maria G, Xu Z, Sun J (2011) Multi-objective MINLP optimization used to identify theoretical gene knockout strategies for *E. coli* cell. *Chemical & Biochemical Engineering Quarterly* 25: 403-424.
32. Maria G (2021) A CCM-based modular and hybrid kinetic model to simulate the tryptophan synthesis in a fed-batch bioreactor using modified *E. coli* cells. *Computers & Chemical Engineering* 153: 107450-107466.
33. Maria G, Gijiu CL, Maria C, Tociu C, Mihalachi M, et al. (2018) Importance of considering the isotonic system hypothesis when modelling the selfcontrol of gene expression regulatory modules in living cells. *Current Trends in Biomedical Engineering & Biosciences* 12(2).
34. Maria G, Maria C, Tociu C (2018) A comparison between two approaches used for deterministic modelling of metabolic processes and of genetic regulatory circuits in living cells. *UPB Sci Bull Series B Chemie* 80(1): 127-144,
35. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2022) Metabolic network structure determines key aspects of functionality and regulation *Nature* 420: 190.
36. Crampin EJ, Schnell S (2004) New approaches to modelling and analysis of biochemical reactions, pathways and networks, *Progress in Biophysics & Molecular Biology* 86:1-4.
37. Heinemann M, Panke S (2006) *Synthetic Biology - putting engineering into biology* *Bioinformatics* 22: 2790-2799.
38. Bailey JE (1991) Toward a science of metabolic engineering *Science* 252: 1668-1675.
39. Nielsen J (1998) Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol Bioeng* 58: 125-132.
40. Visser D, Schmid JW, Mauch K, Reuss M, Heijnen JJ (2004) Optimal redesign of primary metabolism in *Escherichia coli* using linlog kinetics. *Metabolic Engineering* 6: 378-390.

41. Maria G (2007) Modelling bistable genetic regulatory circuits under variable volume framework. *Chemical and Biochemical Engineering Quarterly* 21: 417-434.
42. Tyson JJ, Novak B (2001) Regulation of the eukaryotic cell cycle: molecular antagonism hysteresis and irreversible transitions. *J theor Biol* 210: 249-263.
43. Maria G, Gijiu CL, Maria C, Tociu C (2018) Interference of the oscillating glycolysis with the oscillating tryptophan synthesis in the *E. coli* cells. *Computers & Chemical Engineering* 108: 395-407.
44. Maria G, Mihalachi M, Gijiu CL (2018) Model-based identification of some conditions leading to glycolytic oscillations in *E. coli* cells. *Chemical and Biochemical Engineering Quarterly* 32: 523-533.
45. Maria G, Mihalachi M, Gijiu CL (2018) In silico optimization of a bioreactor with an *E. coli* culture for tryptophan production by using a structured model coupling the oscillating glycolysis and tryptophan synthesis. *Chemical Eng Res and Design* 135: 207-221.
46. Styczynski MP, Stephanopoulos G (2005) Overview of computational methods for the inference of gene regulatory networks. *Comp Chem Eng* 29: 519-534.
47. Maria G (2021) A CCM-based modular and hybrid kinetic model to simulate the tryptophan synthesis in a fed-batch bioreactor using modified *E. coli* cells. *Computers & Chemical Engineering* 153: 107450-107466.
48. Maria G (2004) A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chemical and Biochemical Engineering Quarterly* 18: 195-222.
49. Maria G, Gijiu CL, Crişan M, Maria C, Tociu C (2019) Model-based re-design of some genetic regulatory circuits to get Genetic Modified Micro-organisms (GMO) by using engineering computational tools (a mini-review). *Current Trends in Biomedical Engineering & Biosciences* 18(3): 555988.
50. Maria G (2019) Numerical methods to reduce the kinetic models of (bio)chemical processes. *Printech Publ. Bucharest* p. 815.
51. (2017) UG, Centre of Math. Applied to Life Sciences. Univ Glasgow (UK).
52. Bower JM, Bolouri H (2001) *Computational modeling of genetic and biochemical networks*. MIT Press, Cambridge.
53. Maria G (2005) Relations between apparent and intrinsic kinetics of programmable drug release in human plasma. *Chem Eng Sci* 60: 17091723.
54. Hucka M, et al. (2003) The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524-531.
55. (2002) CRGM-Database, Centro di Ricerca in Genetica Molecolare, Istituto di Istologia e Embriologia Generale. University of Bologna.
56. (2004) NIH Database, National Institute of Health database.
57. EcoCyc (2005) *Encyclopedia of Escherichia coli K-12 genes and metabolism*, SRI Intl., The Institute for Genomic Research. Univ of California at San Diego.
58. KEGG (2011) Kyoto encyclopedia of genes and genomes.
59. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA* 94: 814-819.
60. McAdams H, Shapiro L (1995) Circuit simulation of genetic networks. *Science* 269: 650.
61. Heinrich R, Rapoport SM, Rapoport TA (1977) Metabolic regulation and mathematical models. *Prog Biophys Mol Biol* 32 : 1-82.
62. Somogyi R, Sniegoski CA (1996) Modeling the complexity of genetic network: understanding multigene and pleiotropic regulation. *Complexity* 1: 45-63.
63. Allen GC, Kornberg A (1991) The priB gene encoding the primosomal replication n-protein of *E. coli*. *JBiological Chemistry* 266: 11610-11613.
64. Kubitschek HE(1990) Cell volume increase in *Escherichia coli* after shifts to richer media. *Journal of Bacteriology* 172: 94-101.
65. Heinrich R, Schuster S (1996) *The regulation of cellular systems* New York. Chapman & Hall.
66. Komasilovs V, Pentjuss A, Elsts A, Stalidzans E (2017) Total enzyme activity constraint and homeostatic constraint impact on the optimization potential of a kinetic model. *Biosystems* 162: 128-134.
67. Levenspiel O (1999) *Chemical Reaction Engineering*. 3-rd Ed Wiley, New York.
68. Morgan JJ, Surovtsev IV, Lindahl PA (2004) A framework for whole-cell mathematical modelling. *Jl theor Biology* 231: 581-596.
69. Kimball JW (2022) Cell metabolism, course at Tufts and Harvard Univ. May 15.
70. Aris R (1969) *Elementary chemical reactor analysis*. Prentice-Hall New Jersey.
71. Wallwork SC, Grant DJW(1977) *Physical Chemistry* Longman, London, UK.
72. Maria G (2009) A whole-cell model to simulate the mercuric ion reduction by *E. coli* under stationary and perturbed conditions. *Chemical and Biochemical Engineering Quarterly* 23(3): 323-341.
73. Hudder B, Yang Q, Bölfing B, Maria G, Morgan JJ (2002) Computational Modeling of Iron Metabolism in Mitochondria, NIST Conference on 'Systems Biology Approaches to Health Care: Mitochondrial Proteomics. Sept 17-18 Gaithersburg (USA) p.1.
74. Maria G (2003) Evaluation of protein regulatory kinetics schemes in perturbed cell growth environments by using sensitivity methods, *Chemical and Biochemical Engineering Quarterly* 17(2): 99-117.
75. Kobayashi H, Kaern M, Araki M, Chung, K., Gardner TS, et.al (2004) Programmable cells: Interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA* 101: 8414-8419.
76. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, et.al (2002) Untangling the wires A strategy to trace functional interactions in signalling and gene networks. *Proc Natl Acad Sci USA* 99: 12841-12846.
77. Yang Q, Lindahl P, Morgan J (2003) Dynamic responses of protein homeostatic regulatory mechanisms to perturbations from steady state *J theor Biol* 222L: 407-423.
78. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci* 97: 5528-5533.
79. Maria G, Maria C, Renea L (2022) Application of (bio-)chemical engineering concepts and rules in Bioinformatics. Review of a CCM-based modular and hybrid kinetic model used to simulate and optimize a bioreactor with genetically modified cells, *Current Trends in Biomedical Engineering & Biosciences* 20(3): 556039.
80. Woinaroschy A (2016) A paradigm-based evolution of chemical engineering. *Chinese Journal of Chemical Engineering* 24: 553-557.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/CTBEB.2023.21.556057](https://doi.org/10.19080/CTBEB.2023.21.556057)

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>