A COMPARISON BETWEEN TWO APPROACHES USED FOR DETERMINISTIC MODELLING OF METABOLIC PROCESSES AND OF GENETIC REGULATORY CIRCUITS IN LIVING CELLS

Gheorghe MARIA*[1], Cristina MARIA[2], Carmen TOCIU[3]

*The paper is pointing-out, by referring simple examples reviewed by Maria [1-2], the conceptual differences between two modelling approaches used for developing deterministic dynamic models of metabolic biochemical processes in living cells. The reviewed examples concern the modelling framework of cell metabolic pathways by using continuous variable ordinary differential (**ODE**) dynamic models based on the process mechanism. Two approaches are discussed: I) the default Constant Volume Whole-Cell **(CVWC)** classical **ODE** models, that ignore the cell volume exponential increase during the cell growth, and ii) the holistic variable-volume whole-cell (**VVWC**) models which explicitly account for the cell-volume growth, with preserving the cell isotonicity. To support the superiority of the **VVWC** approach, the reader is referred to additional examples on the deterministic modelling of the gene expression regulatory modules (**GERM**), and of genetic regulatory circuits (**GRC**) in living cells given by [2], with using the same biochemical engineering principles, and rules of the nonlinear system control theory.*

Keywords: systems biology; bioinformatics; cell metabolism deterministic modelling; homeostatic regulation; gene expression regulatory modules; linking **GERM**s

Abbreviations

ATP	adenosin-triphosphate	**GMO**	Genetic modified organisms
ADP	adenosin-diphosphate	**GRC**	Genetic regulatory circuits
AMP	adenosin-monophosphate	**M**	mRNA
CCM	Central carbon metabolism	**MCA**	metabolic control analysis
CVWC	Constant Volume Whole-Cell modelling approach	**ODE**	Ordinary differential equation set
G	generic gene	**P**	Generic protein
GERM	Gene expression regulatory module	**P.I.**	Performance indices
QSS	Quasi steady-state	**VVWC**	Variable-volume whole-cell modelling

[1]. Professor, Faculty of Applied Chemistry and Materials Science, University POLITEHNICA of Bucharest, Romania; corresponding author, e-mail: gmaria99m@hotmail.com

[2] Senior Research Eng., National Inst. Research & Development in Environmental Protection (NIRDEP), Bucharest, Romania.

[3] Senior research PhD Eng., (NIRDEP), Bucharest, Romania.

1. Introduction

Living cells are organized, self-replicating, evolvable, self-adjustable, and responsive biological systems to environmental stimuli able to convert raw materials (substrates/nutrients) from the environment into additional copies of themselves.

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 [1-2]. Relationships between structure, function and regulation in complex cellular networks are better understood at a low (component) level rather than at the highest-level [5].

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 (Crampin and Schnell [9])**.** 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 (Maria [1])**.**

Cells have a very complex but hierarchic organization (structural, functional, and temporal, Fig. 1-left):

i) the *structural hierarchy* includes all cell components from simple molecules (nutrients, saccharides, fatty acids, aminoacids, simple metabolites), 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. Lodish [10]). 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, following a functional hierarchy.

ii) *functional hierarchy* is in accordance to the species structure; e.g. sources of energy (ATP, ADP, AMP), reaction intermediates, TF-s. Lodish [10] 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, of enzymes' activity, gene expression over a **GERM**, etc.). More complex functions, such as regulatory networks, synthesis networks, or metabolic cycles can be built-up using the *building blocks* rules of the *Synthetic Biology* **(**Heinemann and Panke [11]**)**. This is why, the modular **GRC** dynamic models, of an adequate mathematical representation, seem to be the most comprehensive mean for a rational design of **GRC** with desired behaviour (Sotiropoulos and Kaznessis [12]**)**. Such a building blocks cell structure is computationally very tractable when developing cell reduced dynamic models for various metabolic sub-processes, such as: regulatory functions of **GERM**s and of **GRC**s, enzymatic reaction kinetics, energy balance for ATP/ADP/AMP renewable systems, electron donor systems for NADH, NADPH, FADH, FADH2 renewable components, or functions related to the metabolism regulation (regulatory components / reactions of metabolic cycles, gene transcription and translation); genome replication, **GERM** (protein synthesis, storage of the genetic information, etc.), functions for cell cycle regulation (nucleotide replication and partitioning, cell division). When modelling **GRC**s, it is to consider the limited number of interacting **GERM**s, one gene interacting with no more than 23-25 [13].

iii) the wide-separation of time constants of the metabolic reactions in the cell systems is called *time hierarchy*. Reactions are separated in slow and fast according to their time constant. 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 used in building-up cell dynamic models in a way that intermediates are produced in a minimum quantity and consumed only by irreversible reactions. The stationary or dynamic perturbations are treated by maintaining the cell components homeostasis (steady-states), with minimizing the recovering or transition times after each perturbation [2].

A central part of such cell models concerns self-regulation of metabolic processes belonging to the central carbon metabolism (**CCM)**, via **GRC**-s. 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) [2,11]**.** Simulation of gene expression**,** and of **GRC** makes possible *in-silico* design of **GMO** that possess desired properties. By inserting **GRC**-s into organisms, one may create a large variety of mini-functions / tasks in response to external stimuli.

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	
Cell wall, Nucleic acids, Coenzymes		Metabolic cycles, pathways, Transcription, Translation	Metabolism regulation
Peptidoglycan, Membrane, Protein cplx., Nucleotides	Succession of events	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			
Raw materials (nutrients)	←Temporal	Source of energy and material	
←Structural Hierarchy→	Hierarchy→	←Functional Hierarchy→	

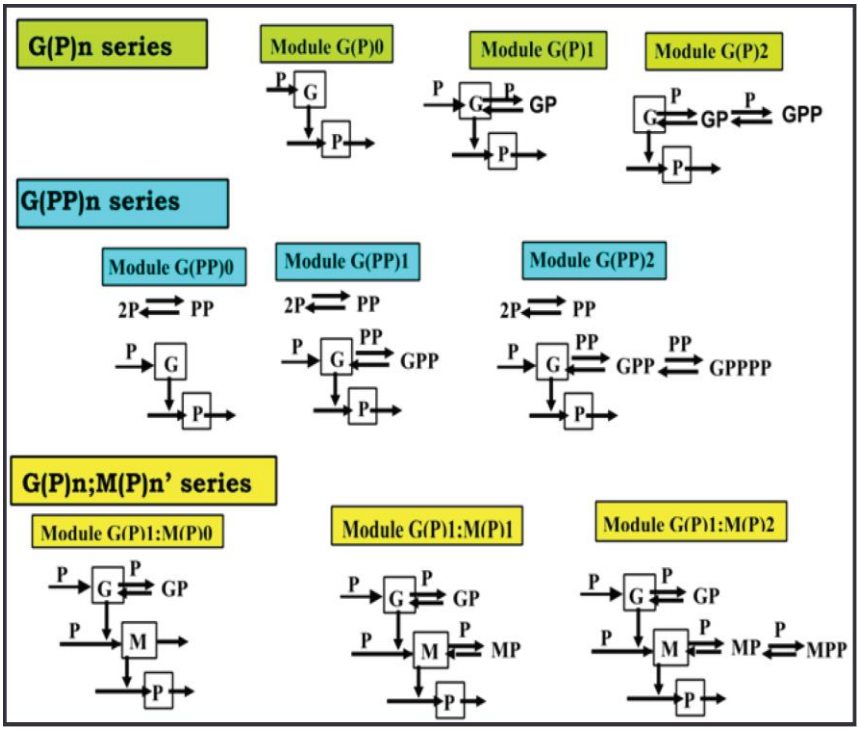

Fig. 1. (up) The hierarchical organization of living cells [1]. (down) The library of GERM for a generic protein P synthesis. G (DNA)= gene encoding P. Horizontal arrows indicate reactions; the vertical arrows indicate catalytic actions; M= mRNA; PP= effectors [1,3].

"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 microorganisms, such as the productivity, which is referred to as *Metabolic Engineering* **(**Bailey [14]**,** Nielsen [15]**,** Stephanopoulos [7]**).** 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" (Visser [17]**).** In this context, **GRC** dynamic models are powerful tools in developing re-design strategies of modifying genome and gene expression seeking for new properties of the mutant cells in response to external stimuli (Maria [2]**).** Examples of such **GRC** modulated functions include:

- 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 [22];
- 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 [32];
- **GRC** oscillator producing regular fluctuations in network elements and reporter proteins, and making the **GRC** to evolve among two or several **QSS** [18];
- specific treatment of external signals by controlled expression such as amplitude filters, noise filters or signal / stimuli amplifiers [21];
- **GRC** signalling circuits and cell-cell communicators, acting as 'programmable' memory units.

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 Styczynski and Stephanopoulos [6].

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**) (Visser [17], Styczynski and Stephanopoulos [6], Maria [18]), but also to the genetic regulatory system (Maria [2]) tightly controlling the metabolic processes reported significant progresses over the last decades in spite of the lack of structured experimental kinetic information. Being rather based on sparse information from various sources, and unconventional identification / lumping algorithms [2-3]**,** 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 (**GMO**), that is desired 'motifs' in response to the external stimuli [2]**.**

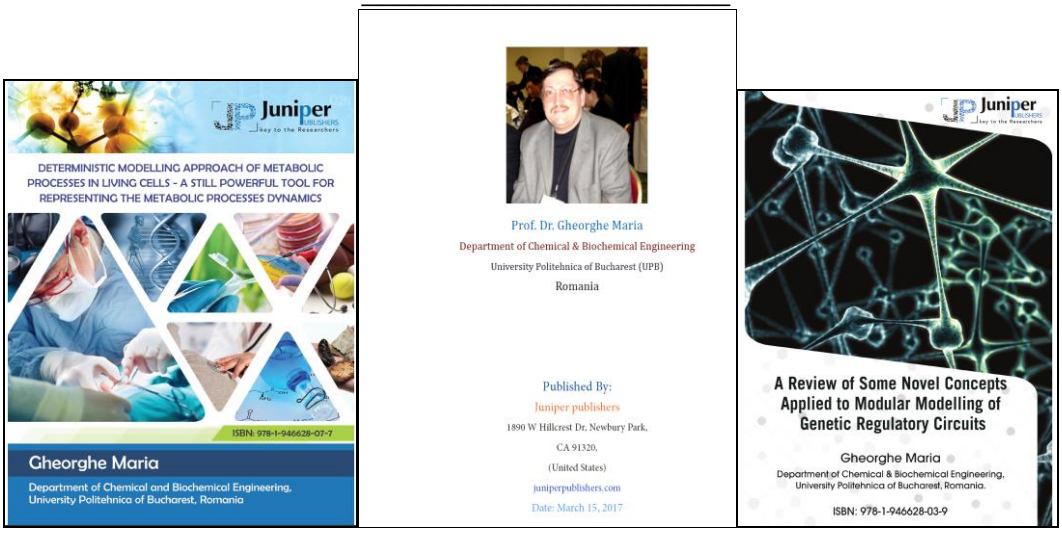

Fig. 2. Covers of the Maria e-book [1] (left), and of the e-book [2] (right) published with Juniper publ., Newbury Park, California 91320 (USA), 2017.

In fact, all the rules and algorithms used by the deterministic modelling of **CCM** and **GRC**, discussed in the works of Maria [1-3] 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) [4] is one of the modern tools which uses advanced mathematical **s**imulation models for *in-silico* design of micro-organisms 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 (Stelling [5]). A review of mathematical model types used to describe metabolic processes is presented by Maria [3], Styczynski and Stephanopoulos [6], and Stephanopoulos [7]. 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, and/or stochastic variables, in a modular construction, 'circuit-like' network, or compartmented simulation platforms (Crampin and Schnell [9], Maria [3], Bower and Bolouri [20]). 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.

1) Structure-oriented analyses or *topological models* ignore some mechanistic details and the process kinetics and use the only *network topology* to quantitatively characterize to what extent the metabolic reactions determine the fluxes and metabolic concentrations (Heinrich and Schuster [8]). The so-called 'metabolic control analysis' (MCA) is focus on using various types of sensitivity coefficients (the so-called 'response coefficients'), which are quantitative measures of how much a perturbation (an influential variable) affects the cell-system states [e.g. reaction rates, metabolic fluxes (stationary reaction rates), species concentrations] around the cell steady-state (**QSS**). The systemic response of fluxes or concentrations to perturbation parameters (i.e. the 'control coefficients'), or of reaction rates to perturbations (i.e. the 'elasticity coefficients') have to fulfil the 'summation theorems', which reflect the network structural properties, and the 'connectivity theorems' related to the properties of single enzymes vs. the system behaviour. Originally, MCA has been introduced to quantify the rate limitation in complex enzymatic systems. MCA have been followed by a large number of improvements, mainly dealing with the control analysis of the stationary states, by pointing-out the role of particular reactions and cell components in determining certain metabolic behaviour. MCA methods are able to efficiently characterize the metabolic network robustness and functionality, linked with the cell phenotype and gene regulation. MCA allows a rapid evaluation of the system response to perturbations (especially of the enzymatic activity), possibilities of control and self-regulation for the whole path or some subunits. Functional subunits are metabolic subsystems, called 'modules', such as amino acid or protein synthesis, protein degradation, mitochondria metabolic path, etc. (Kholodenko[19]). By ignoring the cell process dynamics and using only a linearized representation of the cell system, the MCA reported a limited utilisation for *in-silico* GMO design on a math model basis [3].

2) 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 (sub)systems (Maria [1-2]). Conventional dynamic models, based on ordinary differential (**ODE**) species mass balance, with a *mechanistic (deterministic) 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 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 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. Model reduction rules are presented by Maria [1-2, 28-29]). 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 paths (Maria [18]), and to reflect the species interconnectivity or perturbation effects on cell growth (Maria [2]). 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 (Bower and Bolouri [20]). Representation of metabolic process kinetics is made usually by using rate expressions of extended Michaelis-Menten or Hill type (Maria [1,18,22]). 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 (Maria [3], Hucka [23]). Oriented and unified programming languages have been developed (SBML, JWS, see Maria [1]) to include the bio-system organization and complexity in integrated platforms for cellular system simulation (E-Cell, V-Cell, M-Cell, A-Cell, see Maria [1,3]). Models from this category, among other advantages, can perfectly represent the cell response to continuous perturbations, and their structure and size can be easily adapted according to 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 (CRGM-database [24]; NIH-database [25], EcoCyc [26], KEGG [27]).

3) In the *Boolean approach*, variables can take only discrete values. Even if less realistic, such an approach is computationally tractable, involving networks of

genes that are either "on" or "off" (e.g. a gene is either fully expressed or not expressed at all) according to simple Boolean relationships, in a finite space. Such a coarse representation is used to obtain a first model for a complex biosystem including a large number of components, until more detailed data on process dynamics become available. 'Electronic circuits' structures (Maria [1]) have been extensively used to understand intermediate levels of regulation. Due to the very large number of states, and of TFs involved in the gene expression, the Boolean variables topological **GRC** models are organized in clusters, modules, disposed on multi-layers [1]. But, still they cannot reproduce in detail molecular interactions with slow and continuous responses to perturbations, eventually being abandoned.

4) *Stochastic models* replace the 'average' solution of continuous-variable **ODE** kinetics (e.g. species concentrations) by a detailed random-based simulator accounting for the exact number of molecules present in the system. Because the small number of molecules for a certain species is more sensitive to stochasticity of a metabolic process than the species present in larger amounts, simulation via continuous models sometimes can lack of enough accuracy for random process representation (as cell signalling, gene mutation, etc.). Monte Carlo simulators are used to predict individual species molecular interactions, while rate equations are replaced by individual reaction probabilities, and the model output is stochastic in nature. Even if the required computational effort is very high, such models are useful to simulate system dynamics when species spatial location is important [1].

By applying various modelling routes, successful structured models have been elaborated to simulate various regulatory mechanisms [2-3,30-34]. In fact, as mentioned by Crampin and Schnell [9], 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 even incomplete information, and cell global regulatory properties [9,28].

The scope of this paper is to pointing-out, by referring the simple examples reviewed by Maria [1-2], the main conceptual differences between two modelling approaches used when developing deterministic dynamic models of metabolic cell biochemical processes. Specifically, the modelling framework of cell metabolic pathways by using continuous variable **ODE** dynamic models based on the process mechanism, concerns two different conceptual approaches: I) the default **CVWC**, that ignores the cell volume exponential increase during the cell growth, and ii) the holistic **VVWC** models, which explicitly account for the cell-volume growth, with preserving the cell isotonicity. To support the superiority of the **VVWC** approach, the reader is referred to additional examples on the deterministic modelling of the gene expression regulatory modules

(**GERM**), and of genetic regulatory circuits (**GRC**) in living cells given by [2], with using biochemical engineering principles, and rules [37].

2. Deterministic modelling alternative

Even if complicated and, often over-parameterized, the continuous variable dynamic deterministic **ODE** models of **CCM** and **GRC**-s present a significant number of advantages, being able to reproduce in detail the molecular interactions, the cell slow or fast continuous response to exo/ando-geneous continuous perturbations. (Maria [3]**,** Styczynski and Stephanopoulos [6])**.** Besides, the use of **ODE** kinetic models presents the advantage of being computationally tractable, flexible, easily expandable, and suitable to be characterized using the tools of the nonlinear system theory (Banga [38]**,** Heinrich and Schuster [8]), accounting for the regulatory system properties, that is: dynamics, feedback / feedforward, and optimality. And, most important, such **ODE** kinetic modelling approach allows using the strong tools of the classical (bio)chemical engineering modelling, that is (Maria [1]):

i) molecular species conservation law (stoichiometry analysis; species **ODE** mass balance set);
ii) atomic species conservation law (atomic species mass balance);
iii) thermodynamic analysis of reactions (Haraldsdottir et al. [35]), including quantitative assignment of reaction directionality, set of equilibrium reactions, the Gibbs free energy balance analysis, set of cyclic reactions, identification of species at quasi-steady-state;
iv) improved evaluation of steady-state metabolic flux distributions (i.e. stationary reaction rates) that provide important information for metabolic engineering **(**Zhu et al. [36]**)**;
v) application of lumping rules to **ODE** models (species and/or reaction lumping [28-29])**.**

As classified by Maria [1-3], the **ODE** deterministic models have been developed in two alternatives below discussed: **CVWC** and **VVWC**.

2.1. Alternative (A)

The default Constant Volume Whole-Cell **(CVWC)** classical continuous variable **ODE** dynamic models do not explicitly consider the cell volume exponential, increase during the cell growth. When the continuous variable **CVWC** dynamic models are used to model the cell enzymatic processes, the default-modelling framework eq. (1) is that of a constant volume and, implicitly, of a constant osmotic pressure (π), eventually accounting 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. The **CVWC** formulation results from the species concentration definition of Cj = nj/V, leading to the default kinetic model:

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

where: Cj = (cell-)species j concentration; V = system (cell) volume; nj = species j number of moles; rj = j-th reaction rate; s(i,j) = stoichiometric coefficient of the species "j*"* (individual or lumped) in the reaction "i"; t = time; j= 1,…,ns = number of cell species (individual or lumped); k = rate constant vector; i = 1,…,nr = number of reactions.

The above formulation assumes a homogeneous volume with no inner gradients or species diffusion resistance. The used reaction rate expressions for the metabolic reactions are usually those of extended Michaelis-Menten or Hill type. Being very over-parameterized and strongly nonlinear, parameter estimation of such models in the presence of multiple constraints translates into a mixed integer nonlinear progeamming problem (MINLP) difficult to be solved because the searching domain is not convex [2-3].

Such a **CVWC** dynamic model might be satisfactory for modelling many cell subsystems, but not for an accurate modelling of cell **GRC** and holistic cell properties under perturbed conditions, or the division of cells, by distorting very much or even misrepresenting the prediction results, as exemplified by Maria [2].

2.2. Alternative (B)

As an alternative, Maria [2-3] promoted over the last 15 years the holistic "variable-volume whole-cell" (**VVWC**) modelling framework by explicitly including in the model constraint equations accounting for the cell-volume growth, and by keeping constant the cell-osmotic pressure (to not damage the cell membrane), while the continuous **ODE** model was re-written either in terms of species moles or of species concentrations, as following [2]:

$$\frac{dC_j}{dt}=\frac{1}{V}\frac{dn_j}{dt}-DC_j\,;\frac{1}{V}\frac{dn_j}{dt}=r_j\,;\text{(j=1,..no. of species)},\ D=\frac{d\left(ln(V)\right)}{dt}, \quad (2)$$

because:

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

where: *V* = cell volume (in fact cytosol volume); nj= species j number of moles; rj = j-th reaction rate; *D* = cell-content dilution rate, i.e. cell-volume logarithmic growing rate; species inside the cell are considered individually or lumped; t = time.

The (2-3) mass-balance formulation is that given by Aris [39] for the (bio)chemical reacting systems of variable-volume. In the **VVWC** formulation of

the cell dynamic cell model an additional constraint must be also considered to preserve the system isotonicity (constancy of the osmotic pressure π) under isothermal conditions. This constraint should be considered together with the **ODE** model (2-3), that is the Pfeiffers' law of diluted solutions [40] adopted and promoted by Maria [2-3]:

$$V(t)=\frac{RT}{\pi}\sum_{j=1}^{ns} n_j(t) \tag{4}$$

which, by derivation and division with V leads to [2]:

$$D=\frac{1}{V}\frac{dV}{dt}=\left(\frac{RT}{\pi}\right)\sum_{j}^{ns}\left(\frac{1}{V}\frac{dn_j}{dt}\right), \tag{5}$$

In the above relationships, T = absolute temperature, and R = universal gas constant, V = cell (cytosol) volume. As revealed by the Pfeffer's law eqn. **(4)** in diluted solutions [40], 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 all cell species (individual or lumped). The ($RT/\ \pi$) term can be easily deducted in an isotonic cell system, from the fulfilment of the following invariance relationship derived from **(4)**:

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

As another observation, from **(5)** it results that the cell dilution is a complex function ***D(C,k)*** being characteristic to each cell and its environmental conditions.

Relationships **(5-6)** are important constraints imposed to the **VVWC** cell model **(2-3),** eventually leading to different simulation results compared to the **CVWC** cell kinetic models that neglect the cell volume growth and isotonic effects (see an example given by Maria [2]).

On the contrary, application of the default classical **CVWC ODE** kinetic models of eqn. (1) type with neglecting the isotonicity constraints presents a large number of inconveniences, related to ignoring lots of cell properties (discussed in detail by Maria [2]), that is:

- the influence of the cell ballast in smoothing the homeostasis perturbations;
- the secondary perturbations transmitted via cell volume following a primary perturbation;
- more realistic evaluation of **GERM** regulatory performance indices (P.I.-s), allowing their optimisation following Fig. 4 objectives.

- the more realistic evaluation of the recovering/transient times after perturbations;
- loss of the intrinsic model stability;
- loss of the self-regulatory properties after a dynamic perturbation, etc.

The basic equations and hypotheses of a **VVWC** model are presented in Fig. 3. Even if all cell regulation mechanisms are not fully understood, metabolic regulation at a low-level is generally better clarified. By using (bio)chemical engineering rules and concepts, the developed conventional (*deterministic*) dynamic models, approached in this paper, based on **ODE** kinetics of continuous variables, and on a *mechanistic description* of cell reactions taking place among individual species [including proteins, mRNA, DNA, transcription factors **TF**-s, intermediates, etc.] have been proved to be a convenient route to analyse continuous metabolic cell processes and perturbations (see [1-2] for examples).

In the dynamic models, only essential reactions are retained, species and reactions often are being included as lumps, the model complexity depending on measurable variables and available information. Such reduced **VVWC** kinetic models can be useful to analyse the cell regulatory functions, the **CCM**, treatment of both stationary and dynamic perturbations, cell cycles, oscillatory metabolic paths [1-2], by analyzing the species interconnectivity or perturbation effects**.** Examples of structured deterministic **VVWC** cell models are discussed by Maria [1-2,16,18,21, 41], thus completing the reviewed considerations.

3. Modular modelling of GRC

One successful application of **VVWC** models with continuous variables is those of simulating the regulatory properties of individual **GERM**, and of **GRC** comprising several linked **GERM**-s (no more than 23-25 [13]).

A review of the systematic and comprehensive approaches in modelling the dynamics of **GERM**-s, and of **GRC-s** including chains of **GERM**s based on **VVWC** deterministic models and (bio)chemical engineering concepts and principles was presented by Maria [2](some simplified **GERM** representations are given in Fig. 1-right for a generic pair G/P, that is an encoding gene / and its expressed protein).

Maria [1,2] also exemplifies how such dynamic models of continuous variables, still remain powerful tools for representing lot of metabolic processes dynamics. Such an approach takes the advantage of using well-known mathematical tools and numerical calculus algorithms, as well as (bio)chemical engineering concepts and tools to characterize the kinetics of the cell metabolic processes. This involves application of the classical modelling techniques, algorithmic rules, and nonlinear system control theory and rules to characterize the self-regulation of cell metabolic processes.

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

Notations: 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; "s" index indicates the stationary state.

Fig.3. The variable cell-volume whole-cell (VVWC) dynamic modelling framework and its basic hypotheses [2-3].

Index	Goal	Objective Expression
stationary regulation	Min	R_{ss} = ([P]s - [P]ns)/ [P]ns ;
stationary regulation	Max	$A_{unsync} = k_{syn} \times k_{decline}$
stationary regulation	Min	$S^{i}_{NutPj} = \left[(\partial C_i / C_{is}) / (\partial C_{Nut_j} / C_{Nut_{js}}) \right]_S$
stationary regulation	Min	$S^{i}_{k_j} = \left[(\partial C_i / C_{is}) / (\partial k_j / k_j) \right]_S$
dynamic regulation	Min	$R_D = Max(\mathrm{Re}(\lambda_i))$; $\mathrm{Re}(\lambda_i) < 0$
dynamic regulation	Min	τ_j ; τ_P
regulatory robustness	Min	$(\partial R_D / \partial k)$
species interconnectivity	Min	$AVG(\tau_j) = average\ (\tau_j)$
species interconnectivity	Min	$STD(\tau_j) = st.dev.(\tau_j)$
QSS stability(note a)	Min	$\mathrm{Re}(\lambda_i) < 0$; for all i
QSS stability strength (note a)	Min	$Max(Re(\lambda_i))$
QSS stability strength(note b)	Min	$\lvert \lambda_{A_i} \rvert < 1$

Notations: "n"= nominal value; "s" = stationary value; A = monodromy matrix; τ_j= species "j" recovering time; Nut= nutrient; Re= real part; AVG= average; STD= standard deviation; C_j = species "j" concentration; R_D = dynamic regulatory (recovering) index; QSS = quasi-steady-state; P denotes the key-protein expressed in the analysed GERM.

Fig. 4. The regulatory efficiency performance indices P.I.-s proposed to evaluate the perturbation treatment efficiency by a GERM following the definitions of Maria [3]. Abbreviations: Min = to be minimized; Max = to be maximized. Note: k(syn) and k(decline) refers to the $\rightarrow$ P$\rightarrow$ overall reactions.

Examples of **GRC** models are also provided for the case of i) *in-silico* re-design of the *E. coli* cloned bacterium metabolism by using a **VVWC** structured dynamic model for simulating the mercury uptake efficiency controlled by the **GRC** responsible for the *mer*-operon expression, and ii) *in-silico* derivation of an adjustable structured **VVWC** model to characterize genetic switches with application in designing of a large number of genetically modified micro-organisms (**GMO**) with applications in medicine, such as therapy of diseases (gene therapy), new devices based on cell-cell communicators, biosensors, etc.

4. Conclusions

As a general conclusion, the (bio)chemical engineering principles and modelling rules are fully applicable to modelling cellular metabolic processes. This involves application of the classical modelling techniques (mass balance, thermodynamic principles), algorithmic rules, and nonlinear system control theory. 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 paper pointed-out how the novel **VVWC** deterministic modelling approach promoted by Maria [1-2] has been proved to be a superior alternative to get adequate numerical simulators of the cell metabolism to be used for *in-silico* design of **GRC** and **GMO** with desirable characteristics, with important applications in industry (production of vaccines, biosyntheses optimization), or in medicine (gene therapy).

R E F E R E N C E S

[1] *Maria, G.,* Deterministic modelling approach of metabolic processes in living cells - a still powerful tool for representing the metabolic process dynamics, Juniper publ., Newbury Park, California(USA), 2017, ISBN 978-1-946628-07-7(USA). https://juniperpublishers.com/ebook-info.php

[2] *Maria, G.,* A review of some novel concepts applied to modular modelling of genetic regulatory circuits, Juniper publ., Newbury Park, California(USA), 2017, ISBN (USA) 978-1-946628-03-9; https://juniperpublishers.com/ebook-info.php

[3] *Maria, G.,* Modular-Based Modelling of Protein Synthesis Regulation, Chemical and Biochemical Engineering Quarterly, 19, 2005, pp. 213-233.

[4] *UG,* Centre of Math. Applied to Life Sciences, Univ. Glasgow (UK), 2017. http://www.gla.ac.uk/research/az/cmals/research/systemsbiologystochasticmodelling/

[5] *Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D.,* Metabolic network structure determines key aspects of functionality and regulation, Nature 420, 2002, pp. 190.

[6] *Styczynski, M.P., Stephanopoulos, G.,* Overview of computational methods for the inference of gene regulatory networks, Chem. Eng., 29, 2005, pp. 519-534.

[7] *Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J.,* Metabolic Engineering. Principles and Methodologies, Academic Press, San Diego, CA, 1998.
[8] *Heinrich, R., Schuster, S.,* The regulation of cellular systems, Chapman & Hall, London, 1996.
[9] *Crampin, E.J., Schnell, S.,* New approaches to modelling and analysis of biochemical reactions, pathways and networks, Progress in Biophysics & Molecular Biology, 86, 2004, pp. 1-4.
[10] *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.
[11] *Heinemann, M., Panke, S.,* Synthetic Biology - putting engineering into biology Bioinformatics, 22, 2006, pp. 2790-2799.
[12] *Sotiropoulos, V., Kaznessis, Y.N.,* Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes, BMC Syst Biol., volume 1, art. no. 7, 2007.
[13] *Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R., Collins, J.J.,* Proc. Natl. Acad. Sci. USA., 101, 2004, pp. 8414–8419.
[14] *Bailey, J.E.,* Toward a science of metabolic engineering, Science, 252, 1991, pp. 1668–1675.
[15] *Nielsen, J.,* Metabolic engineering: techniques for analysis of targets for genetic manipulations, Biotechnol. Bioeng., 58, 1998, pp. 125–132.
[16] *Maria, G., Luta, I.,* Structured cell simulator coupled with a fluidized bed bioreactor model to predict the adaptive mercury uptake by *E. coli* cells, Computers & Chemical Engineering, 58, 2013, pp. 98-115.
[17] *Visser, D., Schmid, J.W., Mauch, K., Reuss, M., Heijnen, J.J.,* Optimal re-design of primary metabolism in Escherichia coli using linlog kinetics, Metabolic Engineering, 6, 2004, pp. 378–390.
[18] *G. Maria*, Insilico derivation of a reduced kinetic model for stationary or oscillating glycolysis in *Escherichia coli* bacterium, Chemical and Biochemical Engineering Quarterly, 28, 2014, pp. 509-529.
[19] *Kholodenko, B.N., Schuster, S., Garcia, J., Westerhoff, H.V., Cascante, M.,* Biochimica Biophysica Acta, 1379, 1998, pp. 337-352.
[20] *Bower, J.M., Bolouri, H.,* Computational modeling of genetic and biochemical networks, MIT Press, Cambridge, 2001.
[21] *Maria, G.,* A dynamic model to simulate the genetic regulatory circuit controlling the mercury ion uptake by E. coli cells, Revista de Chimie, 61(2), 2010, pp. 172-186.
[22] *Maria, G.,* Extended repression mechanisms in modelling bistable genetic switches of adjustable characteristics within a variable cell volume modelling framework, Chemical and Biochemical Engineering Quarterly, 28, 2014, pp. 83–99.
[23] *Hucka, M., et al.,* The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models, Bioinformatics, 19, 2003, 524-531.
[24] *CRGM-Database*, Centro di Ricerca in Genetica Molecolare, Istituto di Istologia e Embriologia Generale, University of Bologna, 2002, http://apollo11.isto.unibo.it/Cells.htm
[25] *NIH Database*, National Institute of Health database, 2004, http:// molbio.info.nih.gov/ molbio/ db.html.
[26] *EcoCyc*, Encyclopedia of Escherichia coli K-12 genes and metabolism, SRI Intl., The Institute for Genomic Research, Univ. of California at San Diego, 2005 http://ecocyc.org/

[27] *KEGG*, Kyoto encyclopedia of genes and genomes, 2011, http://www.genome.jp/kegg/pathway.html
[28] *Maria, G.*, A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems, Chemical and Biochemical Engineering Quarterly, 18, 2004, 195-222.
[29] *Maria, G.*, Relations between apparent and intrinsic kinetics of programmable drug release in human plasma, Chem. Eng. Sci. 60, 2005, pp. 1709-1723. doi:10.1016/j.ces.2004.11.009.
[30] *McAdams, H.H., Arkin, A.*, Stochastic mechanisms in gene expression Proc. Natl. Acad. Sci. USA, 94, 1997, pp. 814-819.
[31] *McAdams, H., Shapiro, L.*, Circuit simulation of genetic networks, Science 269, 1995, pp.650.
[32] *Tyson, J.J., Novak, B.*, Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions, J. theor. Biol., 210, 2001, pp. 249-263.
[33] *Heinrich, R., Rapoport, S.M., Rapoport, T.A.*, Metabolic regulation and mathematical models, Prog. Biophys. Mol. Biol., 32, 1977, pp. 1-82.
[34] *Somogyi, R., Sniegoski, C.A.*, Modeling the complexity of genetic network: understanding multigene and pleiotropic regulation, Complexity, 1, 1996, pp. 45-63.
[35] *Haraldsdottir, H.S., Thiele, I., Fleming, R.M.T.*, Quantitative Assignment of Reaction Directionality in a Multicompartmental Human Metabolic Reconstruction, Biophys J., 102, 2012, pp. 1703–1711.
[36] *Zhu, Y., Song, J., Xu, Z., Sun, J., Zhang, Y., Li, Y., Ma, Y.*, Development of thermodynamic optimum searching (TOS) to improve the prediction accuracy of flux balance analysisBiotechnology and Bioengineering, 110, 2013, pp. 914-923.
[37] *Maria, G.*, Application of (bio) chemical engineering principles and lumping analysis in modelling the living systems, Current Trends in Biomedical Engineering & Biosciences, 1*(4)*, 2017, CTBEB.MS.ID.555566(2017).
[38] *Banga, J.*, Optimization in computational systems biology, 6-th Simon Stevin lecture on optimization in engineering, Leuven-Heverlee, NL, 2008, Also in *BMC Systems Biology*, 2, 2008, pp. 47. doi:10.1186/1752-0509-2-47.
[39] *Aris, R.*, Elementary chemical reactor analysis, Prentice-Hall, New Jersey, 1969.
[40] *Wallwork, S.C., Grant, D.J.W.*, Physical Chemistry, Longman, London, UK, 1977.
[41] *Maria, G.*, In-silico design of Genetic Modified Micro-organisms (GMO) of industrial use, by using Systems Biology and (Bio)Chemical Engineering tools, Juniper publ. Inc., Simi Valley, California 93063, (USA) 2018, in-press, https://juniperpublishers.com/ebook-info.php