A Review of Some Novel Concepts Applied to Modular Modelling of Genetic Regulatory Circuits

Gheorghe Maria

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

A Review of Some Novel Concepts Applied to Modular Modelling of Genetic Regulatory Circuits

Prof. Dr. Gheorghe Maria

Department of Chemical & Biochemical Engineering

University Politehnica of Bucharest (UPB)

Romania

Dr. Gheorghe Maria is currently professor in Chemical & Biochemical Reaction Engineering with the UPB - University Politehnica of Bucharest (Romania). He received the PhD in 1987 in chemical engineering at UPB (supervisor Prof. Dr. ing. Raul Mihail). On 1982 he joined ICECHIM (catalysis, and chemical & biochemical energetics Institute, as a senior researcher), and on 1990 he joined UPB as a lecturer. On 1992 he come in Switzerland for working as Assistant Professor with ETH Zürich (with the late Prof. David W.T. Rippin group of Process System Engineering). On 1997 he return to Romania becoming Associate Professor and then full Professor (1999) with UPB. His research interests concern the fields of (bio)chemical reactor and kinetic modelling, biochemical engineering and bioinformatics, risk analysis of chemical plants, modelling of cell metabolic processes, gene expression and regulatory circuits, and drug release kinetics. Over the past 25 years he participated to various national or intl. Projects (more than 15), making short research stages/visitingship at ETH Zürich (3-months on 1997, SNSF fellow on environmental risk analysis), Univ. des Saarlandes (3-months on 1999, DAAD fellow on modelling complex enzymatic kinetics), TU Erlangen (3-months on 2000, catalytic membrane reactors), Texas A&M University (2002-2003, res. scientist on modelling gene expression and gene regulatory networks, synthetic biology), TU Braunschweig (2006) and TU Hamburg (3-months on 2009, DFG and DAAD fellows on simulating bacteria resistance to environmental pollutants), Tianjin Inst. Ind. Biotechnology China (2-months on 2010, *in-silico* searching for gene knockout strategies for *E. coli* cell, synthetic biology). He presented more than 31 invited Lectures at various Universities in EU, CAN, USA, China (among them: Princeton Univ. 1994, Texas A&M University 2002, EPFL Lausanne 1997, Queen's Univ. Kingston 1994 Canada, BASF Ludwigshafen 1996). He authored 6 books, 5 teaching books, 6 book chapters, 130 papers in peer reviewed ISI international journals and univ. journals, and 77 in intl. conference proceedings, with more than 1000 citations (h-index 17 and i10 index 35; ResearchGate score = 34.74). Among them, 29 papers have been published in Chem. Eng. journals with the highest IF>3. According to the Romanian ranking system, he reported high scores signing NP=97 ISI papers as principal author, with a cummultative impact factor of FIC= 135.04, with more than 1000 Citations (Googleschoolar). He is a reviewer for many (bio)chemical engineering journals (25). He co-chaired or was member of the organizing Committees of 16 international conferences (among them: 5th Int. Conference on Computational Bioengineering (ICCB-5), 11-13 September, 2013, Leuven, Belgium; ROMPHYSCHEM-15, 15-th International Conference of Physical Chemistry, 11-13 September, 2013, Bucharest; 13th Edition of Academic Days of Timisoara, June 13-14, 2013, Timişoara (Romania); ELSEDIMA 10th and 11th International Conference (Environmental Legislation, Safety Engineering and Disaster Management), 18-19 September 2014 and 26-28 May 2016, Cluj-Napoca (RO).

• On 1985 he award the Prize of the Romanian Academy for kinetic studies and scalep / plant design for the methanol-to-gasoline process at Brazi petrochemical works (Ploiesti, Romania). He is member in the scientific/editorial board of Chem. & Biochemical Eng. Q. (Zagreb), Revista de Chimie (Bucharest), The Scientific Bulletin of University POLITEHNICA of Bucharest, Bulletin of Romanian Chemical Engineering Society, ECOTERRA Journal of Environmental Research and Protection (edited by the Romanian Society of Environmental Sciences and Engineering, Cluj-Napoca Romania). He was also an expert for various EU and national research programs, being also a PhD supervisor at UPB (from 2008, in Chem. & Biochem. Engineering topics), with 5 PhDs finalized, and 4 PhDs in progress. He participated as researcher or director to a large number of national and international grants. Among them are to be mentioned:

• SNSF Swiss-Romanian Project Project 7IP 050 113 / 1997-1998 on: 'Ecological Design and Operation of Chemical Processes', at ETH Zürich;

• NATO Grant no. 974850-99/1999-2001 on 'Identification, Optimal Monitoring and Risk Limits for Wastewater Biological Treatment Plants', at Universidade da Porto, Portugal, 1999-2001;

• NIH Project 2002, Department of Chemistry and Biochemistry, Texas A&M University, College Station, USA, on the theme: 'Methodology to construct and simulate molecular-level mechanisms by which living systems grow and divide';

• NIH Project 2002, Texas A&M University, College Station, USA, on 'Kinetics of Programmable Drug Release in Human Plasma';

• DFG Grant SFB-578/2006, Development of biotechnological Processes by Integrating Genetic and Engineering Methods, at TU Braunschweig, Germany;

• National CNCSIS Project nr. 1543/2008-2011 (IDEI) on: 'A nonlinear approach to conceptual design and safe operation of chemical processes' ("O abordare neliniara a problemelor de proiectare conceptuala si de operare in conditii de siguranta a proceselor chimice");

• European Commission Project through European Regional Development Fund and of the Romanian state budget, project POSCCE-O2.1.2-2009-2, ID 691 / 2010-2013, "New mesoporous alumnosilicate matterials for controlled release of biological active substances" (Noi materiale din clasa aluminosilicatilor mezoporosi pentru eliberare controlata de substante biologic active).

Complete list of publications on his Web-page: https://sites.google.com/site/gheorghemariasite/ ORCID ID= J-4840-2012

Supplementary information also on: http://sicr.ro/wp-content/uploads/2017/01/BRChES_2_2016.pdf

Contents

1. Abstract

The relatively novel concept of "whole-cell" simulation of cell metabolic processes has been reviewed to prove its advantages when building-up dynamic models of modular structures that can reproduce complex metabolic syntheses inside living cells. The more realistic "whole-cell-variable-volume" (VVWC) approach is exemplified when developing kinetic representations of the gene expression regulatory modules (GERM) that control the protein synthesis and homeostasis of metabolic processes. In the first part of the paper, the general concepts and particularities of the VVWC modelling are presented. In the second part, both past and current experience with constructing effective GERM models is reviewed, together with some rules used when linking GERM-s to build-up models for optimized globally efficient genetic regulatory circuits (GRC).

By using quantified regulatory indices evaluated vs. simulated dynamic and stationary environmental perturbations, the study exemplifies with using typical GERM-s from *E. coli*, how this VVWC modelling methodology can be extended: i) to characterize a GERM module regulatory efficiency; ii) to build-up modular GRC models of various complexity; iii) to prove feasibility of the cooperative vs. concurrent GRC math-constructions that ensures an efficient gene expression, cell system homeostasis, protein functions, and a balanced cell growth during the cell cycle; iv) to prove the effect of the whole-cell content ballast in smoothing the effect of internal/external perturbations on the system homeostasis. Exemplifications of such modular GRC models are presented for the case of i) *in-silico* re-design of the *E. coli* cloned bacterium metabolism by using a structured dynamic model for simulating the mercury uptake efficiency controlled by the GRC responsible for the *mer*-operon expression, and ii) *in-silico* deriving an adjustable structured model to characterize genetic switches with application in designing of a large number of genetically modified micro-organisms (GMO) with potential applications in medicine, such as therapy of diseases (gene therapy), new devices based on cell-cell communicators, biosensors, production of vaccines, etc.

2. Keywords: Systems biology; Bioinformatics; Kinetic modelling; Cell protein synthesis; Homeostatic regulation; Gene expression regulatory modules (GERM); Linking GERM-s; Genetic regulatory circuits (GRC)

3. Abbreviations: WC: Whole-Cell; GMO: Genetically Modified Micro-Organisms; TF: Transcription Factors; VVWC: Variable-Volume Whole-Cell; ODE: Ordinary Differential Equation Set; P.I: Performance Indices; QSS: Quasi-Steady-State; CVWC: Constant-Volume Whole-Cell; Nut: Nutrients; Met: Metabolites; GCE: Gene Circuit Engineering; GS: Genetic Switches

4. The VVWC Modelling Framework

4.1. Introduction

Living cells are organized, self-replicating, evolvable, and responsive to environment biological systems able to convert environmental nutrients in duplicates of the cell content, in order to replicate the cell content in exactly one cell cycle. The structural and functional cell organization, including components and reactions, is very complex. Relationships between structure, function and regulation in complex cellular networks are better understood at a low (component) level rather than at the highest-level Stelling, et al. [1,2]. Cell regulatory and adaptive properties are based on *homeostatic* mechanisms, which maintain quasi-constant key-species concentrations and 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 & Schnell [3]. 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 Kholodenko[4]; Sewell, et al. [5].

Their structure is highly sophisticated, 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. The cell is highly responsive to the environmental stimuli and highly evolvable by self-changing its genome/proteome and metabolism (that is the stoichiometry and the rates / fluxes of the enzymatic reactions) to get an optimized and balanced growth by using minimum resources (nutrients/substrates).

Because the GRC-s are responsible for the control of the cell metabolism, the adequate kinetic modelling of the constitutive GERM-s, but also the adequate representation of the linked GERM regulatory efficiency in a GRC is an essential step in describing the cell metabolism regulation via the hierarchically organized GRC-s (where key-proteins play the role of regulatory nodes). Eventually, such models allow simulating the metabolism of modified cells.

The development of dynamic models to adequately reproduce such complex synthesis related to the central carbon metabolism [Maria 2014, [93] but also to the genetic regulatory system tightly controlling such 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 [6,87]. However, such structured models are 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 external stimuli Heinemann & Panke[7]; Salis & Kaznessis[8]; Kaznessis[9]; Atkinson[10]; Klipp et al. [11]; Chen & Weiss[12]; Tian & Burrage [13]; Sotiropoulos & Kaznessis [14]; Tomshine & Kaznessis [15]; Zhu [16]; Maria et al. [95].

Cells have a hierarchic organization (structural, functional, and temporal):

A. The *structural hierarchy* includes all cell components from simple molecules (nutrients, saccharides, fatty acids, aminoacids, simple metabolites), then macromolecules or complex molecules (lipids, proteins, nucleotides, peptidoglycans, coenzymes, fragments of proteins, nucleosides, nucleic acids, intermediates), and continuing with well-organized nano-structures (membranes, ribosome, genome, operons, energy harnessing apparatus, replisome, partitioning apparatus, Z-ring, etc. [Lodish, et al. [17]). To ensure self-replication of such a complex structure through enzymatic metabolic reactions using nutrients (Nut), metabolites (Met), and substrates (glucose/fructose, N-source, dissolved oxygen, and micro-elements), all the cell components should be associated with specific functions into the cell, following an functional hierarchy.

B. The *functional hierarchy* depends on the species structure; e.g. sources of energy (ATP, ADP, AMP), reaction intermediates, TF-s. Sauro & Kholodenko [18] provided examples of biological systems that have evolved in a *modular* fashion and, in different contexts, perform the same basic functions. Each module, grouping several cell components and reactions, generates an identifiable function (e.g. regulation of a certain reaction, gene expression, etc.). More complex functions, such as regulatory networks, synthesis networks, or metabolic cycles can be built-up using the *building blocks* rules of the *Synthetic Biology* (Heinemann & Panke [7]). 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 the regulatory GRC with desired behaviour (Sotiropoulos & Kaznessis [14]). By chance, such a building blocks cell structure is computationally very tractable when developing cell reduced dynamic models, by defining and characterizing various metabolic sub-processes, such as: regulatory functions of the gene expression and genetic regulatory circuits (GRC), enzymatic reaction kinetics, energy balance functions for ATP/ADP/AMP renewable system, electron donor systems of the NADH, NADPH, FADH, FADH2 renewable components, hydrophobic effects; or functions related to the metabolism regulation (regulatory components / reactions of the metabolic cycles, gene transcription and translation); genome replication / gene expression regulation (protein synthesis, storage of the genetic information, etc.), functions for cell cycle regulation (nucleotide replication and partitioning, cell division). In the case of modelling GRC-s, by chance the number of interacting GERM-s is limited, one gene interacting with no more than 23-25 Kobayashi, et al. [19] .

C. The wide-*separation of time constants* of the metabolic reactions in the cell systems is called *time hierarchy*. Thus, the reactions are separated in slow and fast according to their time constant; in fact, only fast and slow reactions are of interest, while the very slow processes are neglected or treated as parameters (such as the external nutrient or metabolite evolution). Aggregate pools (combining fast reactions) are usually used in building-up cell dynamic models in a way that intermediates are produced in a

minimum quantity and consumed only by irreversible reactions. All cell processes obey a certain succession of events, while stationary or dynamic perturbations are treated by maintaining the cell components homeostasis, by minimizing the recovering or transition times after perturbations.

A central part of such cell models concerns self-regulation of the metabolic processes via GRC-s. Consequently, one particular application of such dynamic cell models is the study of GRC-s, in order to predict ways in which biological systems respond to signals, or environmental perturbations. The emergent field of such efforts is the so-called '*gene circuit engineering*' (GCE [97]) and a large number of examples have been reported with *in-silico* re-creation of GRC-s conferring new properties/functions to the mutant cells (i.e. desired 'motifs' in response to external stimuli) Heinemann & Panke [7]; Salis & Kaznessis [8]; Kaznessis [9]; Atkinson [10]; Klipp, et al. [11]; Chen & Weiss [12]; Tian & Burrage[13]; Sotiropoulos & Kaznessis[14]; Tomshine & Kaznessis [15]; Zhu [16]. By using simulation of gene expression, the GCE *in-silico* design organisms that possess specific and desired functions. By inserting new GRC-s into organisms, one may create a large variety of mini-functions / tasks (or desired 'motifs') in response to external stimuli.

"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 [20]; Nielsen [21]; Stephanopoulos, et al. [22]). 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, et al. [23]. 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 [89,90,93]; Heinemann & Panke [7]; Salis & Kaznessis [8]; Kaznessis [9]; Atkinson [10]; Klipp, et al. [11]; Chen & Weiss [12]; Tian & Burrage [13]; Sotiropoulos & Kaznessis [14]; Tomshine & Kaznessis [15]; Zhu, et al. [16]). Examples of such GRC modulated functions include:

i. 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.

ii. 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.

iii. GRC oscillator producing regular fluctuations in network elements and reporter proteins, and making the GRC to evolve among two or several quasi-steady-states.

iv. Specific treatment of external signals by controlled expression such as amplitude filters, noise filters or signal / stimuli amplifiers.

v. GRC signalling circuits and cell-cell communicators, acting as 'programmable' memory units.

The difficult task to design complex biological circuits with a *building blocks strategy* can be accomplished by properly defining the basic cell components, functions, structural organisation. Because many cell regulatory systems are organized as "modules" Kholodenko et al. [24], it is natural to model GRC-s using a *modular* approach (Maria [6,85,87,89,90,93]). Further analyses including engineered GRC-s can lead to predict / design desirable cell characteristics, that is Kaznessis[9]: a tight control of gene expression, i.e. low-expression in the absence of inducers and accelerated expression in the presence of specific external signals; a quick dynamic response and high sensitivity to specific inducers; GRC robustness, i.e. a low sensitivity vs. undesired inducers (external noise). Through the combination of induced motifs in modified cells one may create potent applications in industrial, environmental, and medical fields (e.g. biosensors, gene therapy). Valuable implementation tools of the design GRC in real cells have been reported over the last years Heinemann & Panke [7].

The *modular* GRC dynamic models, of an adequate mathematical representation, seem to be the most comprehensive mean for a rational design of the regulatory GRC with desired behaviour (Sotiropoulos & Kaznessis [14]; Maria [6,87]). However, the lack of detailed information on reactions, rates and intermediates make the extensive representation of the large-scale GRC difficult for both deterministic and stochastic approach (Maria [87]; Tian & Burrage [13]). When continuous variable dynamic models are used, the default framework is that of a constant volume / osmotic pressure system, accounting for the cell-growing rate as a 'decay' rate of key-species (often lumped with the degrading rate) in a so-called 'diluting' rate. Such a representation might be satisfactory for many applications, but not for accurate modelling of cell regulatory / metabolic processes under perturbed conditions, or for division of cells, distorting the prediction quality.

The variable-volume modelling framework VVWC detailed in this paper, with explicitly linking the volume growth, external conditions, osmotic pressure, cell content ballast and net reaction rates for all cell-components, is proved as being more promising in predicting local and holistic properties of the metabolic network (Maria [6,87]; Morgan,et al. [25]) while the classical one tends to over-estimate some of the regulatory dynamic properties (Maria [6,85,87,88]; Yang, et al.[26]).

To model such an astronomically complex cell system with a detailed kinetic model is practically impossible even if expandable bio-molecular data are continuously added in –omics databanks. However, as underlined by Tomita, et al. [27,28], "whole-cell (WC) simulation of metabolic processes with mechanistic kinetic models of continuous variables, represents the grand challenge of the 21st century". Such a huge effort is justified by the very large number of immediate applications: design genetically modified micro-organisms (GMO) with desirable characteristics to be used in industry (new biotechnological processes, production of vaccines).

A large number of GMO potential applications are in medicine, such as therapy of diseases (gene therapy), new devices based on cell-cell communicators, biosensors, etc. The so-called "*silicon cell*" generic concept includes all dynamic modelling efforts to adequately represent the "cell hierarchical control and regulation analysis" (not covered by the flux balance analysis, or metabolic engineering) leading to derive "new tools in bioprocess engineering, i.e. the *silicon cell* (cf. www.siliconcell. net), that is a collection of computer replica of processes in living organisms, which should be linked up to produce models of larger networks. They can be also used to play with and engineer biological processes on line, *in silico*." (Bruggeman & Westerhoff [29]; Westerhoff [30]; Myers [97]).

As underlined by Tomita [28] "Computer models and *in silico* experiments are necessary to understand and predict phenotypes of the cell, especially when they are polygenic phenotypes. After all, most biological and pathological phenomena in which the pharmaceutical industry has a great interest, such as cancer and allergy, are polygenic." Various mathematical models of different types have been developed over decades. For instance, for modelling GRC-s, the most common are the topological (not-dynamic), but also deterministic / mechanistic dynamic models with discrete, continuous, or stochastic variables ([Styczynski & Stephanopoulos [31]; Maria [87]), each one presenting advantages and disadvantages. The structure-oriented analyses ignore some mechanistic details and the process kinetics, and use the only network topology, the so-called 'Metabolic Control Analysis' (MCA) being focused on using various types of sensitivity coefficients (the so-called 'response coefficients'), which are quantitative measures of how much environmental perturbations (influential variable x_j) affects the cell-system states y_j (e.g. r = reaction rates, J = fluxes, C = concentrations) in a vicinity of the steady-state (QSS, of index 's'), i.e. [$S(y_i; x_j) = (\eth y_i / \eth y_{is}) / (\eth x_j / \eth x_{js})$]. The systemic response of fluxes (i.e. stationary metabolic reaction rates), or of concentrations to perturbation parameters (i.e. the 'control coefficients'), or of reaction rates to perturbations (i.e. the 'elasticity coefficients') have to fulfil the 'summation theorems', which reflect the network structural properties, and the 'connectivity theorems' related to the properties of single enzymes vs. the system behaviour [review of Shifton [32]; Heinrich & Schuster [33].

The emergent *Synthetic Biology* (Benner & Sismour [34]) "interpreted as the engineering-driven building of increasingly complex biological entities" (Heinemann & Panke [7]), aims at applying engineering principles of systems design to biology with the idea to produce predictable and robust systems with novel functions in a broad area of applications (Voit [35]; Heinemann & Panke [7]) such as therapy of diseases (gene therapy), design of new biotechnological processes, new devices based on cell-cell communicators, biosensors, etc. By assembling functional parts of an existing cell, such as promoters, ribosome binding sites, coding sequences and terminators, protein domains, or by designing new GRC-s on a *modular* basis, it is possible to reconstitute an existing cell or to produce novel biological entities with new properties.

Encouraging results have been reported for the design of artificial gene networks (Sprinzak & Elowitz [36]) for reprogramming signalling pathways (Dueber, et al. [37]) for refactoring of small genomes (Chan, et al. [38]) or for re-design of metabolic fluxes with using switching genes (Lebar, et al. [39]; Selvarasu, et al. [40]). By assembling functional parts of an existing cell, such as promoters, ribosome binding sites, coding sequences and terminators, protein domains, or by designing new gene regulatory networks on a modular basis, it is possible to reconstitute an existing cell (the so-called "integrative understanding") or to produce novel biological entities with new properties (Heinemann & Panke, [7]).

The genetic components may be considered as "building blocks" because they may be extracted, replicated, altered, and spliced into the new biological organisms. The *Synthetic Biology* is in direct connection with the *Systems Biology* focus on the cell organization, the former being one of the main tools in the *in-silico* design of GMO-s. In such topics, the metabolism characterization by means of lumped but adequate cell models plays a central role, as underlined by the following definition "*Synthetic Biology* is 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, president of the Institute for Systems Biology, Seattle, USA, cited by Banga [41].

Applications of GERM chain dynamic simulators in *Synthetic Biology* field are immediate, as long as GRC-s controlling the cell metabolism allow *in-silico* re-programming the cell metabolism by means of modified GRC properties leading to GMO of desirable characteristics. Among essential GRC structures used in this respect are to be mentioned the genetic-switches (decision-making branch points between on/off states according to the presence of inducers), oscillators (cell systems evolving among two or several quasi-steady-states), signal/external stimuli amplifiers, amplitude filters, signal transduction circuits (specific treatment of external signals by controlled gene expression), etc. Modular construction of GRC-s must account for some individual (local) but also for holistic properties of the cell considered by the whole-cell modelling approach (see below chapter), such as: a tight control of gene expression (i.e. low-expression in the absence of inducers and accelerated expression in the presence of specific external signals); a quick dynamic response and high sensitivity to specific inducers (Maria [6]).

Even if all cell regulation mechanisms are not fully understood, metabolic regulation at a low-level is generally better clarified. Based on that, conventional dynamic models (based on Ordinary Differential Equation ODE kinetics) using *continuous variables*, also approached in this paper, with a mechanistic description of reactions taking place among individual species [including proteins, mRNA, DNA, transcription factors TF-s, intermediates, etc.] 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 models, only essential reactions are retained, species and reactions often were being included as lumps, the model complexity depending on measurable variables and available

information. An important problem to be considered in such a lumped modelling approach is the distinction between the qualitative and quantitative process knowledge, stability and instability of involved species, the dominant fast and slow modes of process dynamics, reaction time constants, macroscopic and microscopic observable elements of the state vector. Such reduced kinetic models can be useful to analyse the regulatory cell-functions, and the treatment of both stationary and dynamic perturbations, the cell cycles, oscillatory metabolic paths, and lot of cell biosyntheses related to the central carbon metabolism (Maria [87,93,94]), by reflecting the species interconnectivity or perturbation effects on cell growth (Maria, [88]).

To reduce the modelling effort, and due to the lack of structured kinetic data, reduced cell kinetic models are used according to the available –omics/experimental data and utilization scope. These models of modular construction, including lumped species and/or reactions (Maria [87]) present the advantage of being easily extensible, the rate constants being estimated from stationary data (cell culture at homeostasis, under a balanced growth) and by imposing optimal properties to the GERM or GRC (see below chapters). It is to observe that, "traditionally, kinetic metabolic models are based on mechanistic rate equations, which are derived from in-vitro experiments. However, due to large differences between in-vivo and in-vitro conditions, it is unlikely that the in-vitro obtained parameters are valid in-vivo" [23].

"Thus, the kinetic parameters must be adjusted, using data on in-vivo metabolite levels and fluxes obtained in dynamic experiments. Due to the complexity of mechanistic rate equations, which often contain a considerable amount of parameters, this requires a large experimental and mathematical effort." (Visser et al. [23]). Such an approach is computational tractable, a large number of algorithms from chemical engineering, and non-linear system control theory being available.

Application of systematic math-lumping rules to metabolic processes must account for physical significance of lumps, species interactions, and must preserve the *systemic/ holistic* properties of the metabolic pathway. The only separation of components and reactions based on the time-constant scale (as in the modal analysis of the *Jacobian* of the ODE model) has been proved to be insufficient (Maria [86,53]). The work with reduced kinetic models of cell syntheses and GRC-s, even if computationally very convenient, presents some inherent disadvantages, that is: multiple reduced model structures might exist difficult to be discriminated; a loss of information is reported on certain species, on some reaction steps, and a loss in system flexibility (given by the no. of intermediates and species interactions); a loss in the model prediction capabilities; a lack of physical meaning of some model parameters / constants thus limiting its robustness and portability; alteration of some cell / GRC holistic properties (stability, multiplicity, sensitivity).

Even if complicated and, often overparameterized, the continuous variable dynamic ODE models of 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.

(Styczynski & Stephanopoulos [31]; Maria [87]). 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 (Heinrich & Schuster [33]; Banga [41]), accounting for the regulatory system properties that are: 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:

(i) Molecular species conservation law (stoichiometry analysis; species differential mass balance set).

(ii) Atomic species conservation law (atomic species mass balance).

(iii) Thermodynamic analysis of reactions (quantitative assignment of reaction directionality), Haraldsdottir, et al. [42]; set equilibrium reactions; Gibbs free energy balance analysisset cyclic reactions; find species at quasi-steady-state; improved evaluation of steady-state flux distributions that provide important information for metabolic engineering (Zhu, et al. [43]).

(iv) Allows application of ODE model species and/or reaction lumping rules (Maria [53,86]).This paper aims to review some own results on modular modelling of GRC-s by using continuous variable dynamic models under the novel VVWC modelling framework.

The study is focused on an extended analysis of gene expression regulatory modules (GERM), which are the constitutive units of GRC-s. The study will point-out some milestones that should be considered when developing effective GRC-s in the promoted VVWC modelling approach. The general concepts and particularities related to VVWC modelling are exemplified by using generic GERM-s of simple structure. In the first part of the study, general VVWC modelling concepts are presented. In the second part, particularities and applications of such modular GRC models are presented by using simple GERM-s for the case of

a) Re-design the cell metabolism in *E. coli* using a structured model of the GRC involving the *mer-operon* expression for mercury uptake.

b) Derive an adjustable structured model to characterize genetic switches of adjustable performances.

4.2. Variable-volume whole-cell (VVWC) modelling framework

For a system of chemical or biochemical reactions conducted in a (cell) closed volume V, the classical formulation of the corresponding (bio) chemical kinetic model based on continuous variables (concentration vector C, or number of moles vector n) implies writing an ODE mass balance model including the considered system states (biological/chemical species), in the following Constant Volume Whole-Cell (CVWC) default formulation:

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

Where: C_j = (cell-) species j concentration; V = system (cell)

volume; n_j = species j number of moles; r_j = j-th reaction rate; υ_{ij} = stoichiometric coefficient of the species "j" (individual or lumped) in the reaction "i"; t = time np = number of model parameters (k =rate constant vector); nr= number of reactions. The above formulation assumes a homogeneous volume with no inner gradients or species diffusion resistance.

When continuous variable dynamic models are used to model cell enzymatic processes, the default-modelling framework eq. (1) is that of a constant volume and, implicitly of a constant osmotic pressure, 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 $C_j = n_j / V$, leading to the kinetic model:

$$\frac{1}{V(t)}\frac{dn_j}{dt} = \sum_{i=1}^{ns} \upsilon_{ij} r_i(n/V,\mathbf{k},t) = h_j(\mathbf{C},\mathbf{k},t) \; ; \; \frac{d(n_j/V)}{dt} = \frac{dC_j}{dt} = \sum_{i=1}^{ns} \upsilon_{ij} r_i(n/V,\mathbf{k},t) = h_j(\mathbf{C},\mathbf{k},t) \; ...(2)$$

Such a CVWC dynamic model might be satisfactory for modelling many cell subsystems, but not for an accurate modelling of cell regulatory / metabolic processes under perturbed conditions, or for division of cells, distorting the prediction quality, as exemplified by Maria [6,84,85,87,89,90,93]; Morgan et al. [25], and by the below example of chapter 4.2.2.

However, deterministic modelling approach, with using continuous variables are one of the most used models in modelling dynamic cell processes (review of Maria [87]). That is because the stochastic models are much more difficult to be solved, being also computationally very intensive. On the other hand, systems solved deterministically can be viewed loosely (and with caution) as reflecting the average of an ensemble of cells (Axe & Bailey [44]).

The mass-balance formulation in the chemical reacting systems of variable-volume, is that given by Aris [45]:

$$\frac{dC_j}{dt} = \frac{1}{V}\frac{dn_j}{dt} - DC_j \; ; \; \frac{1}{V}\frac{dn_j}{dt} = r_j \; ; \; j = 1,\ldots,n_s \text{ (no. of species), where:}$$

$$D = \frac{d(ln(V))}{dt} \; , \; \ldots\ldots(3)$$

Because: $\frac{dC_j}{dt} = \frac{d}{dt}\left(\frac{n_j}{V}\right) = \frac{1}{V}\frac{dn_j}{dt} - C_j\frac{d(ln(V))}{dt} = \frac{1}{V}\frac{dn_j}{dt} - DC_j = h_j(C,k,t)$ $\ldots\ldots\ldots(4)$

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

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 biological / biochemical systems the variable-volume formulation (3-4) of Aris [45] was re-written and adopted by Grainger, et al. [46], and later promoted in modelling various cell subsystems by also including isotonicity constraints in the so-called VVWC modelling framework by Maria [87,93].

To derive the variable-volume VVWC modelling framework, Maria [87,93] adapted the "whole-cell" concept of Tomita [28], and adopted additional hypotheses and constraints of Table 1. Thus, in a VVWC model of (3-4) type, a hypothetical "Whole Mechanical (deterministic) Cell" has to be defined in the sense of Tomita, et al. [27,28], and of Maria [87,93] to ensure a balanced auto-catalytic growth of the cell by maintaining intracellular homeostasis, by using environmental nutrients present in

variable amounts. Basically, the following hypotheses should be adopted in a VVWC modelling framework (Table 1):

Table 1: Variable cell-volume whole-cell VVWC dynamic modelling framework and its basic hypotheses [adapted from Maria [87] by the courtesy of CABEQ Jl.].

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)
$\frac{1}{V}\frac{dn_j}{dt} = r_j(\mathbf{C},\mathbf{k}) \; ; \; j = 1,\ldots,n_s;$	
$V(t) = \frac{RT}{\pi}\sum_{j=1}^{n_s} n_j(t)$	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
$\frac{RT}{\pi} = \frac{V}{\sum\limits_{j=1}^{n_S} n_j} = \frac{1}{\sum\limits_{j=1}^{n_S} C_j} = \frac{1}{\sum\limits_{j=1}^{n_S} C_{jo}}$ constant.	constant osmotic pressure (π) constraint
$\left(\sum\limits_j^{all} C_j\right)_{cyt} = \left(\sum\limits_j^{all} C_j\right)_{env}$	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 $\left(dC_j/dt\right)_s = g_j(\mathbf{C}_s,\mathbf{k}) = 0$.	
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.

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

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

III. The cell is an isothermal system with a 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.

IV. The cell is an open system interacting with the environment through a semi-permeable membrane.

V. 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" (Maria [87,89,93]).

The inner osmotic pressure (π_{cyt}) is constant, and all time equal with the environmental pressure, thus ensuring the membrane integrity ($\pi_{cyt} = \pi_{env}$ = constant). As a consequence, the isotonic osmolarity under isothermal conditions leads to the equality RT / π_{cyt} = RT / π_{env} which, following the below eqn. (11), indicates that the sum of cell species concentrations must equal those of the environment, i.e. $(\sum_j^{all} C_j)_{cyt} = (\sum_j^{all} C_j)_{env}$. Otherwise, the osmosis will eventually lead to an equal osmotic pressure $\pi_{cyt} = \pi_{en}$. Even if, in a real cell, such equality is approximately fulfilled due to perturbations and transport gradients, and in spite of migrating nutrients from environment into the cell, the overall environment concentration is considered to remain unchanged. On the other hand, species inside the cell transform the nutrients into metabolites and react to make more cell components. In turn, increased amounts of polymerases are then used to import increasing amounts of nutrients. The net result is an exponential increase of cellular components in time, which translates, through isotonic osmolarity assumption, into an exponential increase in volume with time (below observation vi). The overall concentration of cellular components is time-invariant (homeostasis), because the rate at which cell-volume increases equals that at which overall number of moles increases, leading to a constant $(\sum_{j=1}^{ns} n_j)/V$ ratio. The species concentrations at the cell level are usually expressed in nano-moles, being computed with the relationship Concentration = no.of copies per cell/$N_{A/}V_{cyt,o}$ [eqn. 5] (Maria [87]):

where N_A is the Avogadro number. For instance, for an *E. coli* cell, with an approximate volume $v_{cyt,0}$ = 1.66*10^{-15}L (Kubitschek [47]), it results a value of:

C_{Gs} = 1/(6.022*10^{23})(1.66*10^{-15}) = 1 nM(i.e.10^{-9} mol/L).

vi) Cell volume doubles over the cell cycle period (t_c), with an average logarithmic growing rate of D= ln(2)/t_c [resulted from integrating the defined D = d(ln(V))/dt in eq. (3)]. Under stationary growing conditions, that is a constant D over the cell cycle, integration of this relationship indicates an exponential increase of the cell volume, that is V(t) = V_0exp(+D·t)

vii) Under stationary growing conditions, species synthesis rates (r_j) must equal to first-order dilution rates D_sC_{js}, leading to time-invariant (index "s") species concentrations C_{js} , i.e. homeostatic conditions of $(dC_j / dt)_s$ = 0. Under such a balanced steady-state growth, the below nonlinear algebraic mass balance set [see also eq. (10)]:

$(dc_j/dt)_s$ =(1/V dn_j / dt)$_s$ -D_sC_{js} = $h_{js}(C_s,k,t)$= 0; j = 1,....,n_s(no. of species),

D_s = (RT/π) $\sum_j^n$s (1 / v dn_j/dt)$_s$ (6)

Is used to estimate the rate constants k from the known stationary concentration vector C_s, with also

imposing some optimal properties of the cell system (Maria [6,84,87,89,90,91,92,93]) (see below chapter 4.3).

viii) It is to observe that, in a continuous variable metabolic 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, see the review of Maria [87].

Thus, a metabolic kinetic model in a VVWC approach should be written in the form (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 (eq. 9-10). As the cell volume is doubling during the cell cycle, this continuous volume variation cannot be neglected.

However, to not complicate too much the VVWC dynamic model, usually a reduction in the number of cell species and reactions by common lumping rules is usually performed (Maria [6,53,86]). Such a model reduction strategy of the metabolic kinetic models present a series of disadvantages, such as: a loss in model adequacy, and in the simulated system flexibility (due to the reduced number of considered intermediates and species interactions); an increased possibility to get multiple (rival) reduced models of proximate characteristics for the same cell system, difficult to be delimited; a loss in the model prediction capabilities; lumped model parameters can lack physical meaning; a loss / alteration of systemic / holistic properties (e.g. cell system stability, multiplicity, sensitivity, regulatory characteristics). Here can be mentioned only a few of the classical chemical engineering rules used for reducing an extended kinetic model (Maria [6,53,86]):

(i)   Reduce the list of reactions, by eliminating unimportant side-reactions and/or assuming quasi-equilibrium for some reaction steps; use sensitivity measures of rate constants to detect the redundant part of the model (e.g. ridge selection, principal component analysis, time-scale separation, etc.);

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

(iii) Decompose the kinetics into fast and slow 'parts' allowing application of the quasi-steady-state-approximation (QSSA) to reduce its dimensionality (Maria [53,87]).

(iv)   When the ODE kinetic model is linear in parameters, then the reduction procedure of Maria [86] can be successfully applied by preserving the system Jacobian invariants (eigenvalues, eigenvectors).

Due to the modular functional organization of the cell, a worthy route to develop reduced models is to base the analysis on the concepts of 'reverse engineering' and '*integrative understanding*' of the cell system (review of Maria [6]). Such a rule allows disassembling the whole system in parts (modules) and then, by performing tests and applying suitable numerical procedures, to define rules that allow recreating the whole

and its characteristics by reproducing the real system. Such an approach, combined with derivation of lumped modules, allows reducing the model complexity by relating the cell response to certain perturbations to the response of few inner regulatory loops instead of the response of thousands of gene expression and metabolic circuits. Such a procedure is very suitable for modelling GRC-s by linking GERM models in such a way to maintain the cell homeostasis that is to maintain relatively invariant species concentrations despite perturbations (Maria [6,84,87,89-93]).

The so-called "cell-content dilution rate" D term is in fact the cell-volume logarithmic growing rate, its definition resulting from the way by which eqn. (3-4) were deducted for a variable-volume reacting system by Aris [45]; Maria [87]:

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

$$D = \frac{d(ln(V))}{dt}$$; Where $\frac{1}{V}\frac{dn_j}{dt} = r_j(C,k), j=1,...,n_s$ (no. of cell species) (7)

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

$$J_C = \left(\partial h(C,k)/\partial C\right)_s \quad ....(8)$$

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 (1-7), that is the Pfeiffers'law of diluted solutions (see Wallwork & Grant [48]) adopted and promoted by Maria [6,84, 87,89-93], that is:

$$V(t) = \frac{RT}{\pi}\sum_{j=1}^{n_s} n_j(t) \quad ....(9)$$

Which, by derivation and division with V leads to (Maria [6,87]):

$$D = \frac{1}{V}\frac{dV}{dt} = \left(\frac{RT}{\pi}\right)\sum_{j}^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right), \quad ....(10)$$

In the above relationships, T = absolute temperature and R = universal gas constant, V= cell (cytosol) volume. As revealed by the Pfeffer's law eqn. (9) in diluted solutions (Wallwork & Grant [48]), and by the eq. (10), 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 term RT/π can be easily deducted in an isotonic cell system, from the fulfilment of the following invariance relationship derived from (9):

$$V(t) = \frac{RT}{\pi}\sum_{j=1}^{n_s} n_j(t) \Rightarrow \frac{RT}{\pi} = \frac{V(t)}{\sum\limits_{j=1}^{n_s} n_j(t)} = \frac{1}{\sum\limits_{j=1}^{n_s} C_j} = \frac{1}{\sum\limits_{j=1}^{n_s} C_{jo}} = constant \quad ....(11)$$

As another observation, from (10) it results that the cell dilution is a complex function $D(C,k)$ being characteristic to each cell and its environmental conditions.

Relationships (10-11) are important constraints imposed to the VVWC cell model (3), eventually leading to different simulation results compared to the CVWC cell kinetic models that neglect the cell volume growth and isotonic effects (see the below example 4.2.2.).

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 (below discussed): the influence of the cell ballast in smoothing the homeostasis perturbations; the secondary perturbations transmitted via cell volume following a primary perturbation; the more realistic evaluation of GERM P.I.-s, and of the recovering/transient times after perturbations; loss of the intrinsic model stability; loss of the self-regulatory properties after a dynamic perturbation, etc.

The VVWC formulation (3-11) was proved to be also suitable to accurately model the cell growth and its division (Morgan, et al. [25]). Such a model formulation allows studying various regulatory properties of GERM-s, and the response of coupled GERM-s to dynamic / stationary continuous perturbations in the environment, and also the 'inertial' effect of the cell-'ballast' vs. continuous changes in cell and environment (Maria [6,84, 87,89-93]) . As ca. 80% of the cycle period is the growing phase and, assuming a quasi-constant osmotic pressure and a constant volume growing logarithmic rate, the generic cell model (3-4) can be considered satisfactory to study the GRC effectiveness. The model was proved to be also very effective to study the response to continuous perturbations in the environment of various GRC-s including genetic switches (Maria [89,90,93]), or expression of certain operons like those for mercury uptake in gram-negative bacteria (Maria [84,91,92]).

4.2.1. The importance of the VVWC approach in computing the "secondary perturbations".

When elaborating the kinetic model for whatever GERM, the cell-volume growing rate is an essential issue to account for, due to several reasons:

a) The continuous dilution of the cell content (that is concentration decline due to the continuous increase of the denominator of $C_j = n_j(t)/V(t)$; in spite of that, concentrations of key species remain constant because the numerator (copynumbers) increases at the same rate with the denominator.

b) The increased cell homeostatic (steady-state) concentrations 'resistance' vs. small perturbations in the level of some internal or external components due to the inertial 'big volume' or 'cell content 'ballast' effect (to be further discussed).

c) Indirect effect of perturbations in concentrations on the cell-metabolism transmitted via induced changes in the volume growing rate (10).

d) The volume-growth diluting effect acts as a continuous stationary perturbation of the concentrations, and can formally be assimilated with a first order decay rate in (7) of all cellular species during the cell-cycle.

For instance, when applied an impulse perturbation to one of the cell-species (e.g. a generic protein denominated by P), this implies removal by excretion, or the import of certain copynumbers of P from (to) the cell. In turn, such a variation

in the P-copynumbers involves, according to (11), by keeping constant the cell osmotic pressure π and the temperature, that the cell-volume will immediately contract (or dilate) from V to V*. Consequently, all species concentrations will vary from $C_j = n_j / V$ to $C_j^* = n_j / V^*$ irrespectively if they exhibit any change in copynumbers. This effect due to the cell volume variation is called "secondary perturbations" or "indirect perturbations". The difference in the P copynumbers before and after perturbation, i.e. $(n_{P,o} - n_P^*)$ (the * denotes the perturbed state), can be easily calculated for a certain imposed final concentration for P. The final copynumbers of P, n_P^*, for an imposed $C_P^* = n_P^* / V^*$ results by re-evaluating the species concentrations:

$$n_P^* = V_o C_P^* \frac{RT}{\pi} (\sum_{j=1}^{m} C_{jo} - C_{P,o}) / (1 - C_P^* \frac{RT}{\pi}) \dots\dots\dots(12)$$

To prove the relationships (12), one considers that, before and after applying the perturbation in P of $(C_{P,o} - C_P^*)$, the sum of copynumbers of all other species remains invariant, that is:

$$\sum_{j=1}^{m} n_{jo} - n_{P,o} = \sum_{j=1}^{m} n_j^* - n_P^* \quad \dots\dots\dots(13)$$

On the other hand, the volume given by eqn. (11), written for initial and final states and for the same pressure π, becomes $V_o = \frac{RT}{\pi} \sum_{all} n_{jo}$, and $V^* = \frac{RT}{\pi} \sum_{all} n_j^*$, respectively. By multiplying (13) with the constant RT/π, then changing terms from left to right, and including the volume formula, one obtains the net volume variation:

$$\left(V^* - V_o\right) = \frac{RT}{\pi}\left(n_P^* - n_{P,o}\right) \quad \dots\dots\dots (14)$$

By dividing (14) with the product (V*V), and then multiplying with n_P^*, one obtains:

$$\frac{n_P^*}{V_o} - \frac{n_P^*}{V^*} = \frac{n_P^*}{V^*} \frac{RT}{\pi}\left(\frac{n_P^*}{V_o} - \frac{n_{P,o}}{V_o}\right) \quad \dots\dots\dots(15)$$

Now, by changing terms from left to right, one obtains:

$$\frac{n_P^*}{V_o} - \frac{n_P^*}{V^*} \frac{RT}{\pi} \frac{n_P^*}{V_o} = \frac{n_P^*}{V^*} - \frac{n_P^*}{V^*} \frac{RT}{\pi} \frac{n_{P,o}}{V_o} \quad \dots\dots\dots(16)$$

By re-arranging, one obtains the results:

$$\frac{n_P^*}{V_o}\left(1 - \frac{n_P^*}{V^*} \frac{RT}{\pi}\right) = \frac{n_P^*}{V^*} \frac{RT}{\pi}\left(1 / \frac{RT}{\pi} - \frac{n_{P,o}}{V_o}\right) \quad \dots\dots\dots(17)$$

By introducing the invariant $\frac{RT}{\pi} = \frac{1}{\sum_{j=1}^{m} C_{jo}}$ from eqn. (11) in the right side, and substituting with $C_P^* = n_P^* / V^*$, one obtains:

$$\frac{n_P^*}{V_o}\left(1 - C_P^* \frac{RT}{\pi}\right) = C_P^* \frac{RT}{\pi}\left(\sum_{j=1}^{m} C_{jo} - C_{P,o}\right) \quad \dots\dots\dots(18)$$

Relationship (18) is identical to (12), q.e.d. On the other hand, the volume relationship (11), written for initial and final states and for the same pressure π, is $V_o = \frac{RT}{\pi} \sum_{all} n_{jo}$, and $V^* = \frac{RT}{\pi} \sum_{all} n_j^*$, is leading to: $V^* / V_0 = \sum_{i=1}^{m} n_j^* / \sum_{j=1}^{m} n_{j,o}$. So, the sum of concentrations into the cell is a conservative term, as imposed by the isotonicity constraint and proved by the below relationship:

$$\sum_{i=1}^{m} C_j^* = \frac{\sum_{i=1}^{m} n_j^*}{V^*} = \frac{\sum_{i=1}^{m} n_j^*}{V_o\left(\sum_{i=1}^{m} n_j^* / \sum_{i=1}^{m} n_{j,o}\right)} = \frac{\sum_{i=1}^{m} n_{j,o}}{V_o} = \sum_{i=1}^{m} C_{j,o} = constant \quad \dots\dots\dots(19)$$

It follows that the cell-volume variation during the cell-growth cycle is an essential term to be considered in any cell process model to obtain more realistic predictions of the process dynamics. While most of reported models, both deterministic or stochastic (see Maria [6,87,89]) ignore such effects and build-up CVWC models written in terms of species concentrations, the new elaborated VVWC models over the last couple of years accounted for the cell-volume growth in an explicit way but also linked to the system osmotic pressure (see Maria [6,84,87,89-93]) become a promising alternative, as underlined by the below discussed advantages.

For instance, Sewell, et al. [5] included the volume-diluting effect only for the protein-concentrations through a formally first-order decay rate. Such an approach, even being satisfactory for rapid and simple predictive purposes, suffers of two major disadvantages:

i. The 'decay' rate is not considered for all the species in order to avoid high model complexity.

ii. The 'decay' rates can report different rate constants for various species, when in reality the same diluting constant rate is reported for all the species.

iii. The 'inertial' cell-volume / large copynumber effect to smooth perturbations cannot be simply and naturally included in such a CVWC model.

The same fictive decay-rate approach has been reported by Tomita, et al. [27] in developing an 'E-cell' continuous differential model with including a larger number of genes and proteins. As mentioned by Tomita [28], "the E-cell system also accepts user-defined reactions, making it capable of handling many other phenomena such as diffusion and variable cell volume". By using the EcoCyc [49] and KEGG [50] databases, the authors simulate the dynamics of 127 genes/proteins system for the *M. genitalium* cell. However, this model suffers from several drawbacks such as lack of autocatalysis effects during a cell-cycle, by considering any replication of the genome, and any cell-division process. Recently, the authors reported some improvements of the E-cell model with including the osmotic pressure balance and volume cell growth without specifying details (Kinoshita, et al. [51]).

Other models, such as Gibson & Bruck [52] avoid including the cell-volume increase effects when considering only first-order reaction terms in the cell model. However, the authors signalled that such an approximation can "create a large calculation error".

In the present study the VVWC models used for modelling GRC-s by linking GERM chains, explicitly include constraint equations accounting for the cell-volume growth and by preserving the same cell-osmotic pressure, while the continuous-variable ODE model was re-written either in terms of species moles or of species concentrations (3-11). The cross-autocatalytic effects can also be included when protein and gene synthesis catalytic paths are considered. Moreover, external cell-factors are better accounted by separately considering the protein and gene "metabolite raw-materials" (see below formulations of GERM-s). It is also to observe that, from the definition of D in eq. (3) it results, for a stationary growth of constant Ds, the cell volume dynamics:

$$V(t) = V_O \exp(+D_s \cdot t) \quad \dots\dots\dots(20)$$

Cell volume doubles over the cell cycle period (t$_c$), with an

average logarithmic growing rate of $D_s = \ln(2V_0/V_0)/t_c = \ln(2)/t_c$. For stationary balanced growing conditions, the species synthesis rates r_{js} must equal to first-order dilution rates ($D_s C_{js}$), leading to time-invariant species concentrations (characterizing the homeostatic conditions, $((dC_j/dt)_s = 0$), that is:

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

j= 1,..., n_s (no. of species); where $D_s = \left(\frac{RT}{\pi}\right)\sum_j^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right)_s$ ....(21)

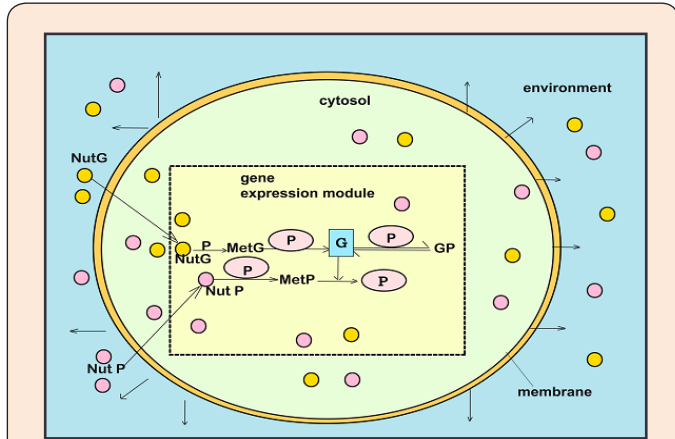## 4.2.2. A simple example to illustrate superiority of VVWC vs. CVWC models.

Figure 1

The reaction scheme of a generic gene G expression. The regulatory module of G(P)1 type was used to exemplify the synthesis of a generic P protein in the *E. coli* cell by Maria [87]. Figure adapted from Maria [87] by courtesy of CABEQ Jl. To improve the system homeostasis stability, that is quasi-invariance of key species concentrations (enzymes, proteins, metabolites), despite of perturbations in nutrients Nut*, and metabolites Met*, or of internal cell changes, a very rapid buffering reaction G + P <===> GP(inactive) has been added. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; G = gene encoding protein P; MetG, MetP = lumped DNA and protein precursor metabolites respectively.

In order to simply illustrate the discrepancy in predictive capabilities between VVWC and CVWC kinetic models, and how deceptive can the predictions of such CVWC kinetic models be, one considers the regulatory module of type G(P)1 of a generic gene G expression illustrated in a simplified form in (Figure 1) and (Figure 6) The reaction scheme and reaction rate expressions for the G(P)1 gene expression module are those given by the below model (20). The kinetic model formulated in CVWC terms (using the notation n_j = [nano-moles] of species j) is the following:

$$\begin{cases} NutG + P \xrightarrow{k_1} MetG + P, & r_1 = k_1 C_{NutG} C_P \\ NutP + P \xrightarrow{k_2} MetP + P, & r_2 = k_2 C_{NutP} C_P \\ MetG + P \xrightarrow{k_3} G + P, & r_3 = k_3 C_{MetG} C_P \\ MetP + G \xrightarrow{k_4} P + G, & r_4 = k_4 C_{MetP} C_G \\ P + G \xrightarrow{k_5} GP, & r_5 = k_5 C_P C_G \\ GP \xrightarrow{k_6} G + P, & r_6 = k_6 C_{GP} \end{cases} \quad \begin{cases} \frac{1}{V}\frac{dn_{MetG}}{dt} = r_1 - r_3 \\ \frac{1}{V}\frac{dn_{MetP}}{dt} = r_2 - r_4 \\ \frac{1}{V}\frac{dn_P}{dt} = r_4 - r_5 + r_6 \\ \frac{1}{V}\frac{dn_G}{dt} = r_3 - r_5 + r_6 \\ \frac{1}{V}\frac{dn_{GP}}{dt} = r_5 - r_6 \end{cases} \quad \begin{cases} \frac{dC_{MetG}}{dt} = r_1 - r_3 \\ \frac{dC_{MetP}}{dt} = r_2 - r_4 \\ \frac{dC_P}{dt} = r_4 - r_5 + r_6 \\ \frac{dC_G}{dt} = r_3 - r_5 + r_6 \\ \frac{dC_{GP}}{dt} = r_5 - r_6 \end{cases}$$

....(4.2.2.a)

For comparison, the kinetic model of this system written in the VVWC formulation is the following:

$$\begin{cases} \frac{dC_{MetG}}{dt} = r_1 - r_3 - D \cdot C_{MetG} \\ \frac{dC_{MetP}}{dt} = r_2 - r_4 - D \cdot C_{MetP} \\ \frac{dC_P}{dt} = r_4 - r_5 + r_6 - D \cdot C_P \\ \frac{dC_G}{dt} = r_3 - r_5 + r_6 - D \cdot C_G \\ \frac{dC_{GP}}{dt} = r_5 - r_6 - D \cdot C_{GP} \end{cases} \quad \text{........................(4.2.2.b)}$$

Where the cell-content dilution rate D is estimated by means of the relationship:

$D = \frac{RT}{\pi}\sum_{j=1}^{n_s}\frac{1}{V}\frac{dn_j}{dt}$, which, in the present case, translates in:

$$D = \frac{RT}{\pi}(r_1 + r_2 - r_5 + r_6) \quad \text{.....................(22)}$$

Notations: C_j = cell-species j concentration [nM]; V = cell volume [L]; n_j = amount of species j [n-moles]; r_j= j-th reaction rate [n-moles/L/min]; D = cell-content dilution rate [1/min]; π = osmotic pressure [atm]; T = temperature [K]; R = universal gas constant [L.atm/n-moles/K]; n_s = number of species inside the cell; t = time [min].

The rate constants of (4.2.2.b) have been evaluated by solving the following nonlinear algebraic set obtained from the cell steady-state condition at the homeostasis, that is:

$$\begin{cases} 0 = r_{1s} - r_{3s} - D_s \cdot C_{MetGs} \\ 0 = r_{2s} - r_{4s} - D_s \cdot C_{MetPs} \\ 0 = r_{4s} - r_{5s} + r_{6s} - D_s \cdot C_{Ps} \\ 0 = r_{3s} - r_{5s} + r_{6s} - D_s \cdot C_{Gs} \\ 0 = r_{5s} - r_{6s} - D_s \cdot C_{GPs} \end{cases}$$

with: $D_s = \frac{RT}{\pi}(r_{1s} + r_{2s} - r_{5s} + r_{6s})$ ....(23)

Table 2: *E. coli* cell homeostatic (stationary) characteristics used in the G(P)1 modelling example of chapter 4.2.2. Adopted values from Maria [87,89] by the courtesy of CABEQ Jl.].

Species	Homeostatic level [nM]
Lump $C_{NutG,s}$	$3*10^6$ (adopted)
Lump $C_{NutP,s}$	$3*10^8$ (adopted)
Lump $\sum_j C_{MetG_j,s}$ (note a)	approx. 10^6
Lump $\sum_j C_{MetP_j,s}$	$3*10^8$ (adopted)
$C_{P,S}$	1000 (adopted)
$C_{G,,S} = C_{G,P,S} = 0.5$	(adopted; see note b)
t_c	Cell life cycle of 100 min
$D_s = \ln 2 / t_c$.	Cell-volume logarithmic growing rate (average) [1/min]
V_o, Cell (cytosol) initial volume	1.660434503 10^{-15} (L)

Footnotes:

(a) Evaluated from the isotonicity constraint $\sum_{all\ j}^{cell} C_j = \sum_{all\ j}^{env} C_j = C_{NutG} + C_{NutP} = \sum_j C_{MetG_j} + \sum_j C_{MetP_j,s}$

(b) Adopted to ensure the maximum responsiveness of the GERM steady-state, [see discussion of Maria [85,87].

Table 3: Estimated rate constants for the model G(P)1 in the VVWC formulation for the example of chapter 4.2.2. The units of the 1-st order reactions are min^{-1}, while those of the 2-nd order are nM^{-1}min^{-1}. Species stationary concentrations are those of Table 2. V denotes the cell (cytosol) volume.

Reaction	Rate expression	Estimated Rate Constant
$NutG + P \rightarrow MetG + P$	$r_1 = k_1(n_{NutG}/V_{env})(n_P/V)$	6.93909711*10^{-4}
$NutP + P \rightarrow MetP + P$	$r_2 = k_2(n_{NutP}/V_{env})(n_P/V)$	7.63259052*10^{-7}
$MetG + P \rightarrow P + G$	$r_3 = k_3(n_{MetG}/V)(n_P/V)$	6.93147180*10^{-10}
$MetP + G \rightarrow G + P$	$r_4 = k_4(n_{MetP}/V)(n_G/V)$	1.38629436*10^{-3}
$P + G \rightarrow GP$	$r_5 = k_5(n_P/V)(n_G/V)$	1.000000069*10^{+2}
$GP \rightarrow P + G$ (note a)	$r_6 = k_6(n_{GP}/V)$	1*10^{+5} (note a)

Footnote:
(a) Adopted value much larger than , see the discussion of Maria [87]

("s" index refers to the steady-state). To exemplify the prediction discrepancies between the CVWC and VVWC formulations for this case study, one considers as a numerical example an arbitrary G(P)1 system from the *E. coli* cell under balanced growth with the initial homeostatic concentrations given in (Table 2). The obtained rate constants are given in (Table 3). The used solver was the routine "solve" of Maple8^{TM}- package. By simulating one cell cycle by using these estimated parameters and the VVWC model (4.2.2b) and (22) one obtains the species copynumber dynamics into the cell plotted in (Figure 2).
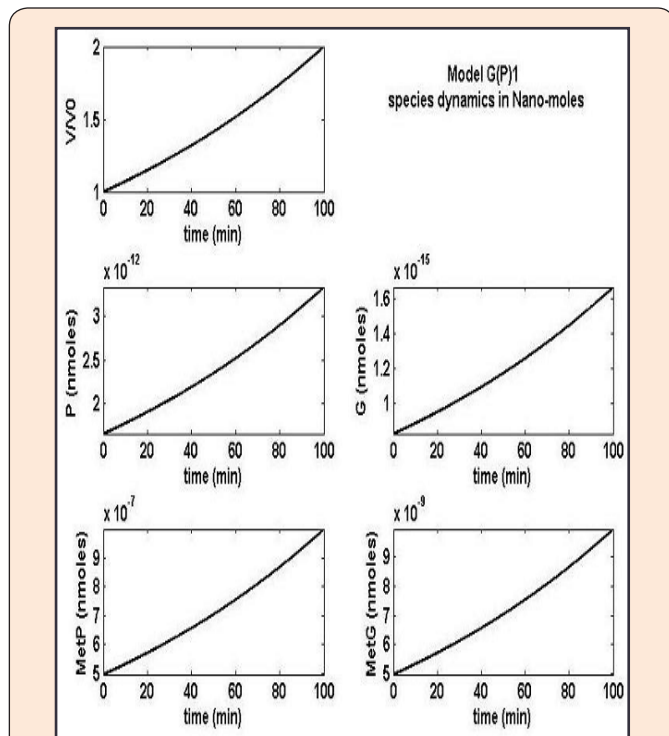
Dynamics of species copy numbers n$_i$ during the cell cycle predicted by the VVWC model for the example of chapter 4.2.2 concerning the G/P pair replication using a regulatory module of G(P)1 type.
The *E. coli* cell species homeostatic concentrations are those of Table 2.

Dynamics of species concentrations (in nM) during the cell cycle predicted by the constant volume CVWC model (-------- , with no dilution D term in the model), compared to those predicted by the variable volume VVWC model (_____ , with dilution D term included in the model). Example used in chapter 4.2.2 concerning the G/P pair replication using a regulatory module of G(P)1 type. The *E. coli* cell species homeostatic concentrations are those of Table 2.
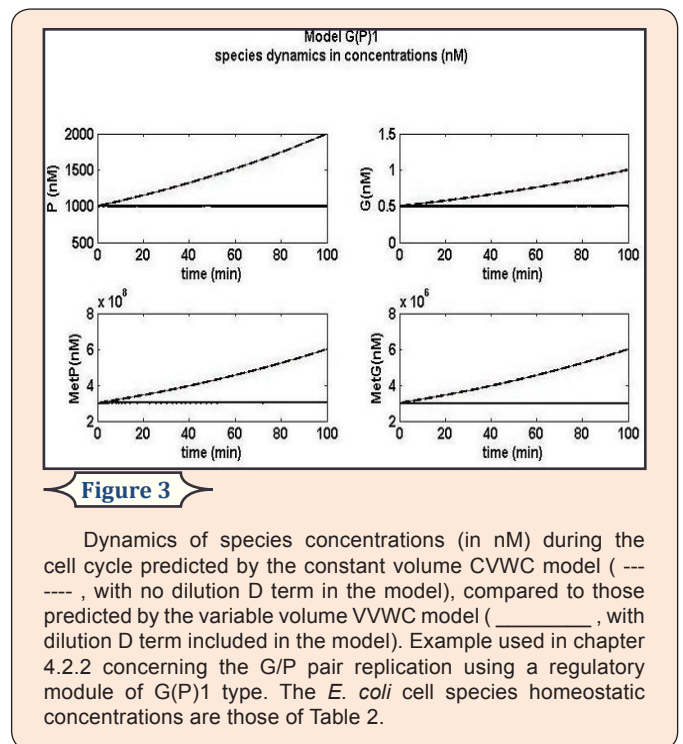
It is to observe that, while the cell volume doubles, the species copynumbers double as well. If the species dynamics is plotted in terms of concentrations (referred to the cell cytosol volume), the predicted trajectories given in (Figure 3) by the CVWC, and VVWC are very different. The VVWC model correctly reproduces the system homeostasis, that is the species quasi-constant concentrations because both nominator and denominator of the fraction $C_j(t) = n_j(t)/V(t)$ are doubling at the same rate. By contrast, the CVWC model predictions are wrong, the predicted species concentrations having the same shape and relative growth as with those of the copynumbers trajectories.

Due to such a realistic representation of a GERM, the VVWC model better reflects the GERM regulatory properties after a dynamic (impulse-like) or a stationary (step-like) internal or external perturbation in one of the module-species (see below discussion over the next paragraphs 5.2.1-5.2.10).

4.3. Modelling a gene expression regulatory module (GERM) in a VVWC modelling framework

One of the very promising applications of VVWC dynamic models with continuous variables is the study of genetic regulatory circuits (GRC-s), in order to predict the way by which biological systems are self-regulated and respond to signals (Maria [6,84,87,89-93]). The emergent field of such efforts is the so-called '*gene circuit engineering*' and a large number of examples (review of Maria [87]) have been reported with *in-silico* creation of novel GRC-s conferring new properties/ functions to the mutant cells (i.e. desired 'motifs' in response to external stimuli), such as: 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 (Maria [89,90,93]); hysteretic GRC behaviour, that is a bio-device able to behave in a history-dependent fashion, in accordance to the presence of a certain inducer in the environment; GRC oscillator producing regular fluctuations in network elements and reporter proteins (Maria [94]), and making the GRC to evolve among two or several quasi-steady-states; specific treatment of external signals by controlled expression such as amplitude filters (Maria [84,91,92]), noise filters or signal / stimuli amplifiers; GRC signalling circuits and cell-cell communicators, acting as 'programmable' memory units. The subject is of tremendous importance as long as GRC-s are the essential parts used to re-design the cell metabolism. Here it is worth to mention the modular construction of GRC-s. Sauro & Kholodenko [18]; Kholodenko, et al. [24] provided examples of biological systems that have evolved in a modular fashion and, in different contexts, perform the same basic functions. In fact, a GRC presents a modular construction including chains of GERM-s. Each GERM groups the cell components and reactions that are linked to generate an identifiable function (e.g. regulation of a certain reaction, gene expression, etc.).

More complex functions, such as regulatory networks, synthesis networks, or metabolic cycles can be built-up using the *building blocks* rules Heinemann & Panke [7]. The modular organization of cell regulatory systems is computational very tractable. Moreover, it is well-known that one gene expression interacts with no more than other 23-25 GERM-s (Kobayashi et al. [19]), while most of GERM structures are repeatable.

Consequently, in developing the GRC analysis, the modular approach is preferred due to various advantages: a separate analysis of the constitutive GERM-s in conditions that mimic the stationary and /or perturbed cell growth; investigation of module links used to construct the whole GRC of an optimized regulatory efficiency that ensures key-species homeostasis and network holistic properties; investigation of GRC characteristics such as 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 (quasi-stationarity) 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 metabolite or nutrient.

The key element to such cell GRC dynamic simulators is the adopted kinetic model of the GERM-s from the large number of mechanistic models proposed in literature (Maria [85,87,89,90,93]; Savageau[54]; Hlavacek & Savageau [55]; Wall, et al. [56]; Salvador & Savageau [57]; Atkinson, et al. [10]; Sewell, et al. [5]). When constructing a GRC for a certain cell metabolic pathway, there are two problems, which must be addressed properly:

A. How to choose the suitable GERM structures of the GRC chain, by screening among existing alternatives, by selecting from simple reduced structures of Maria [87,89,90,93]; Sewell, et al. [5]; Savageau [54]; Hlavacek & Savageau [55], and using reduced GERM structures with no more than 10-14 reactions, thus ensuring a satisfactory trade-off between model simplicity and its predictive quality Maria [6]. Even if more sophisticated constructions are proposed in the literature Maria [93], such a GERM selection must be based on their regulatory properties (i.e. quantitative performance indices P.I. below defined in the chapter5.1.) matching with the experimental data; and

B. What rules to be applied to link such GERM-s to reproduce the cell system holistic behaviour.

In constructing a complex protein (enzyme) synthesis regulatory network there are some important issues to also be considered:

I. There are three modelling and control levels of the regulatory circuit: a) a single gene expression module GERM (local, or individual level); b) the GRC including several GERM-s having in the regulation nodes proteic complexes resulted from proteins (enzymes) interactions that promote a catalytically efficient sequence of reactions over the so-called 'channelling intermediate metabolites' Sorribas, & Savageau [58]; Gabaldon & Huynen, [59], and c) the whole-cell replication regulation (to be included in the whole-cell modelling concept, VVWC).

II. The following concepts derived from the experimental evidence should be included in the VVWC models of GERM-s and GRC-s:

(i) All regulatory performance indices P.I.-s should be optimal at both GERM and GRC levels. That is because all cell metabolic processes occur with optimal performances to ensure the cell replication over an exact cell cycle. So, the metabolic reactions must occur with maximum reaction rates, with using minimum of resources (substrates, energy), and producing minimum amount of reaction intermediates; reactions and key-species homeostasis should be less influenced by the environmental perturbation even if involving simple GRC-s with a preferable cascade control of the gene expression that minimize the transition or recovering times of a quasi-steady-state QSS (see the proof of Maria [6] when computing the species sensitivities vs. environmental nutrient levels $S(C_j; Nut_i) = \partial \ln C_j / \partial \ln Nut_i$ for various GRC structures; see also chapter 5.2.6];

(ii) In all VVWC cell models (Maria [6,84,87,89-93]) it should be also considered the genome and proteome replication (cell "ballast") big influence on the cell metabolism, on the volume increase, and on the treatment of perturbations, in a simple way, by means of a lumped GERM. Consideration of the lumped proteome and genome replication in all VVWC cell models is mandatory in order to fulfil the isotonicity constraint. In such a VVWC formulation, all cell species should be considered (individually or lumped), because all species net reaction

rates contribute to the cell volume increase [eq. (9-10)]. Some examples are presented in chapters 6-7 with modelling certain GRC-s by also considering the genome and proteome replication (Maria [84,87,89-93]).

C. The modular GRC dynamic models, of an adequate mathematical representation, seem to be the most comprehensive mean for a rational design of the regulatory GRC with desired behaviour (Sotiropoulos & Kaznessis [14]). However, the lack of detailed information on reactions, rates and intermediates make the extensive representation of the large-scale GRC difficult for both deterministic and stochastic approach (Tian & Burrage[13]).

In spite of the gene expression process complexity, to be easy to use, most of the GERM kinetic models proposed in the literature present a simplified structure, by including only lumped species and reactions, but conferring optimised regulatory P.I.-s to the GERM regulatory efficiency (see below chapter 5.1.). Besides, such GERM models must fulfil some constraints, that is: the metabolic reaction stoichiometry; reaction rates must be maximal, but with rate constants limited by the diffusional processes and in agreement with the thermodynamic equilibrium steps; the total enzyme (proteine) content of the cell is limited by the isotonicity condition; also the total cell energy (ATP) and reducing agent (NADH) resources are limited; the reaction intermediate level must be minimum; the cell model at homeostasis must be stable, reaching again the steady-state after termination of a perturbation (see Heinrich & Schuster [33] for details). Most of the mentioned aspects will be below discussed and exemplified.
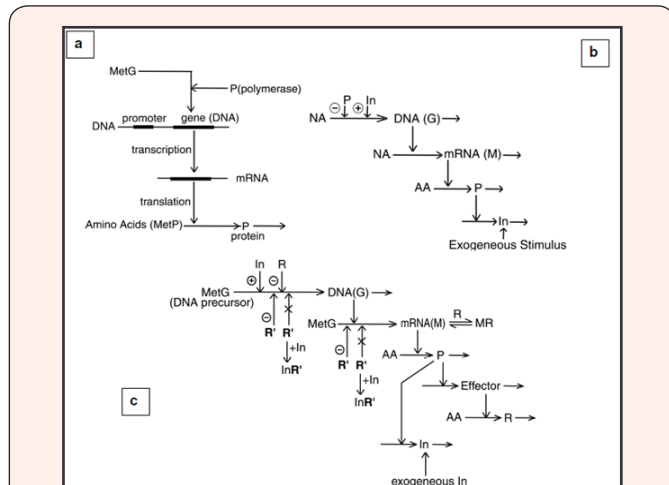

Figure 4

(b and c) Simplified representations of the regulatory module (GERM) for a generic gene G expression (a) with perfectly coupled enzyme/regulator P expression. Such GERM models are further used to construct various GRC models. Notations: In= inducer; AA= aminoacids; horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; G = gene encoding protein P; M = mRNA; R, R' = transcriptional factors (repressors); In = inducer; MetG = DNA precursor metabolites. The enzyme (protein P) interacts with the inducer In for controlling the transcription rate by means of feedback positive ± or negative regulatory loops. Figures adapted from the rough descriptions of [35,60,54, 56], and from Maria [87,93] by courtesy of CABEQ Jl.

Protein synthesis by gene expression is a highly regulated process to ensure a balanced and flexible cell growth under

indefinitely variate environmental conditions. How this very complex process occurs is partially understood, but a multi-cascade control of the transcription and translation steps (see some simplified GERM representations in (Figures 4-6), with negative feedback loops seems to be the key element. Enzymes catalyzing the synthesis are allosterically regulated by means of positive or negative effector molecules (transcriptional factors TF), while cooperative binding and structured regulation amplify the effect of a change in an exo/endo-geneous inducer. Gene expression is also highly regulated to flexibly respond to the environmental stress. The metabolic regulator features are determined by its ability to efficiently vary species flows and concentrations under changing environmental conditions so that a stationary state of the key metabolite concentrations can be maintained inside the cell (Maria [87]).
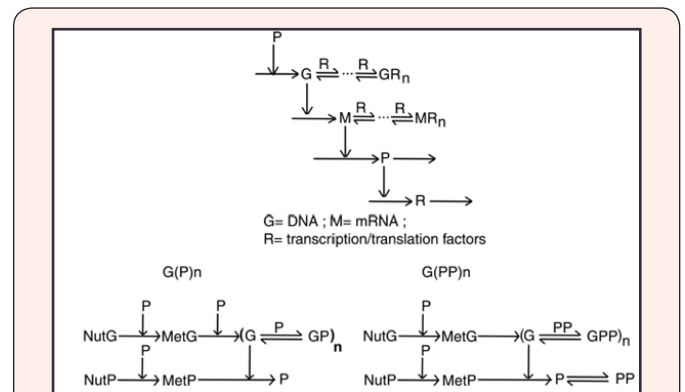

Figure 5

Protein P synthesis. Simplified representations of a generic gene expression G/P regulatory module (GERM). The horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicates an assumed concentration invariance of these species; G= gene encoding P; M= mRNA. Up-row: simplified representation of the gene expression self-regulation over the transcription and translation steps. The model corresponds to a [G(R)n; M(R)n] regulatory module type of Figure 6. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicates an assumed concentration invariance of these species; G= DNA gene encoding P; M= mRNA; R= allosteric effectors of the transcription / translation. Down-row: two types of GERM simplified representations of the self-regulated protein P synthesis: [G(P)n] (left) and [G(PP)n] (right). Adapted from Maria [6,87] by courtesy of CABEQ Jl.

The protein synthesis is regulated by a complex homeostatic mechanism that controls the expression of the encoding genes. On the other hand, cells contain a large number of proteins of well-defined functions, but strongly interrelated to ensure an efficient metabolism and cell growth under certain environmental conditions. Proteins interact during the synthesis and, as a consequence, the homeostatic systems perturb and are perturbed by each other. To understand and simulate such a complex regulatory process, the modular approach is preferred, being based on coupled semi-autonomous regulatory groups (of reactions and species), linked to efficiently cope with cell perturbations, to ensure system homeostasis, and an equilibrated cell growth.

The modular approach to analyse the gene expression assumes that the reaction mechanism and stoichiometry of various types of kinetic modules are known, while the involved

species are completely observable and measurable. Such a hypothesis is rarely fulfilled due to the inherent difficulties in generating reliable experimental (kinetic) data for each individual metabolic subunit. However, incomplete kinetic information can be incorporated by performing a suitable model lumping (Maria [53,86]), or by exploiting the cell and module global optimal properties during identification steps (Lodish, et al. [17]; Maria [89]). The regulatory modules can be constructed relatively independent to each other, but the linking procedure has to consider common input/output components, common linking reactions, or even common species (Maria [6,87]). Rate constants can be identified separately for each module, and then extrapolated when simulating the whole regulatory network, by assuming that linking reactions are relatively slow comparatively with the individual module core reactions (Maria [93]). In such a manner, linked modules are able to respond to changes in common environment and components such that each module remains fully regulated. "The advantage of such a modular approach is the possibility to reduce the system model complexity and the size of the identification problem, by understanding, for instance, the gene expression response to a perturbation as the response of a few genetic regulatory loops instead of the response of thousands of genetic circuits in the metabolic pathway." (Maria [90]).

Various types of kinetic modules can be analyzed individually as mechanism, reaction pathway, regulatory characteristics, and effectiveness. As a limited number of regulatory module types govern the protein synthesis, it is computationally convenient to step-by-step build-up the modular regulatory network (GRC) by applying certain principles and rules to be further discussed, and then adjusting the network global properties. Accordingly, it is desirable to focus the metabolic regulation and control analysis on the regulatory/control features of functional GERM subunits than to limit the analysis to only kinetic properties of individual enzymes acting over the synthesis pathways.

The difficulty to precise the very large number of parameters in complex GRC-s leads to account for lumped representations, such as gene clustering and path structure reduction based on various system properties (stability, sensitivity, multiplicity). Consequently, to model such individual GERM-s and some simple GRC-s, at a molecular level, Sewell, et al. [5]; Savageau [54]; Hlavacek & Savageau [55]; Maria [85,87] proposed simple mechanistic structures by using a modular approach, useful in simulating the hierarchical organization of cell regulatory networks. Some simple GERM models are given in (Figures 4-6).

While Sewell, et al. [5], and Maria [6,84,85,87,89-93] used ODE kinetic models with continuous variables based on classical rate expressions (of Michaelis-Menten, or Hill type), Savageau [54], Hlavacek & Savageau [55] and Savageau &Voit [60] ; Voit [35] used simple rate expressions of power-law type (the so-called S-systems) obtained by recasting the elementary steps and intermediates in a lumped representation including apparent rate constants and reaction orders. Such apparent power-law models with fractional orders of reactions can generate a biased representation of the real process and suffer of a series of inconveniences related to distorted GERM properties (multi-stationarity, sub-optimal P.I.-s). The mechanistic based GERM
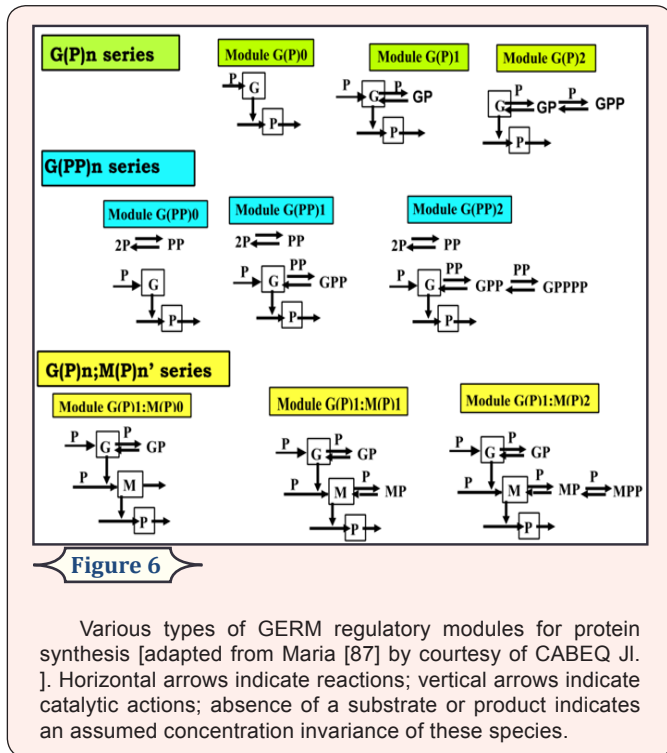
models seem to be more robust (Maria [87,89,93]), flexible and easily adaptable (chapter 6).

On the other hand, a too advanced lumping can lead to diminish some network properties (local stability strength, efficient responsiveness, flexibility; see below P.I. measures). One disadvantage of using continuous variable formulations is the possibility of translating fractional concentrations to fraction of copynumbers. For instance, for a born *E. coli* cell volume of , V[cyt,o] = $1.66 \cdot 10^{-15}$ L, one gene G copy number translates to 1 nmol L^{-1} concentration, while a concentration of [G] = 0.5 nmol L–1 must be interpreted either as an average of time-invariant in a population of cells (e.g. half of all cells containing 1 copy number of G), or as a time-dependent average for a single cell (e.g. that cell contains 1 copy number of G half of the time).

When elaborating a protein synthesis GERM model, different degrees of simplification of the process complexity can be followed. For instance, the gene expression (Figure 4) can be translated into a modular structure of reactions, more or less extended, accounting for individual or lumped species. At a generic level, in the simplest representation (Figure 5), the protein (P) synthesis rate can be adjusted by the 'catalytic' action of the encoding gene (G). The catalyst activity is in turn allosterically regulated by means of 'effector' molecules (P, or R) reversibly binding the catalyst via fast and reversible reactions (the so-called 'buffering' reactions). This simplest regulation scheme can be further detailed in order to better reproduce the real process, with the expense of a supplementary effort to identify the module kinetic parameters. For instance, a two-step cascade control of P-synthesis includes the M = mRNA transcript encoding P (Figure 5). The effector (R), of which synthesis is controlled by the target protein P, can allosterically adjust the activity of G and M, i.e. the catalysts for the transcription and translation steps of the gene expression. In such a cascade schema, the rate of the ultimate reaction is amplified, depending on the number of cascade levels and catalysis rates. More complex regulatory modules can be elaborated (Figure 4) following a similar route to 'translate' from the 'language' of molecular biology to that of mechanistic chemistry, by preserving the structural hierarchy and component functions. Once elaborated, such a modular structure can be modelled by using one of the previously described alternatives, and then analysed as functional efficiency by means of some defined performance indices.

When elaborating a protein synthesis regulatory module (i.e. a gene expression regulatory module, GERM), different degrees of simplification of the process complexity can be followed. A GERM is a semi-autonomous regulatory group of reactions and species, linked to efficiently cope with cell perturbations, to ensure system homeostasis, and an efficient gene expression. For instance, to easily study and compare GERM regulatory efficiency, Sewell, et al. [5]; and Yang, et al. [26] proposed various types of hypothetical GERM simplified reaction pathway designed to ensure homeostatic regulation of a generic protein-gene (P/G) pair synthesis (Figures 4-6). These structures have been studied by using the classical CVWC approach with exemplifications from *E. coli* cells. On the contrary, Maria [6,84,85,87,89,90,91,92,93] studied similar structures properties but using the novel VVWC approach.

Figure 6

Various types of GERM regulatory modules for protein synthesis [adapted from Maria [87] by courtesy of CABEQ Jl.]. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicates an assumed concentration invariance of these species.

4.3.1. Some simple GERM models

Simplified representations of (Figures 4,5) include the essential nutrient lumps (NutP, NutG) used to obtain the protein and DNA precursor metabolites (MetP, MetG) respectively, and intermediates involved in the reactions controlling the transcriptional and translation steps of the P synthesis. The module nomenclature of such GERM models, proposed by Yang, et al. [26]; Maria, [87] is those of $[L_1(O_1)n_1 ;...; L_i(O_i)n_i]$ that includes the assembled regulatory units $L_i(O_i)n_i$. One unit i is formed by the component Li (e.g. enzymes or even genes G, P, M, etc.) at which regulatory element acts, and n_i =0,1,2,... number of 'effector'/TF species O_i (i.e. 'effectors' P, PP, PPPP, R, RR, RRRR, etc) binding the 'catalyst' L. For instance, a G(P)2 unit of (Figure 6) includes two successive binding steps of G with the product P, that is G + P <==> GP + P <==> GPP, all intermediate species GP, GPP, being catalytically inactive, while the mass conservation law is all time fulfilled, i.e. $\sum_{i=0}^{2} [G(P_i)]$ = constant. Such a representation accounts for the protein concentration diminishment due to the cell-growth dilution effect, but could also include protein degradation by proteolysis. It is also to observe that such GERM models try to account essential properties of the gene expression, which is a highly self- / cross- regulated and mutually catalyzed process by means of the produced enzymes / effectors. As depicted in (Figure 6) 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 <===> GP).

The GERM model structure can be extended according to experimental information and accounting for individual or lumped species. For instance, at a generic level, in the simplest representation (Figure 4, up), the protein (P) synthesis rate can be adjusted by the 'catalytic' action of the encoding gene (G) (Figure 5, down). The catalyst activity is in turn allosterically

regulated by means of 'effector' molecules (P, or R; Figure 5) reversibly binding the catalyst G via fast and reversible reactions (the so-called 'buffering' reactions). These simple regulation schemes can be further detailed in order to better reproduce the experimental data, with the expense of a supplementary effort to identify the module kinetic parameters.

For instance, a two-step cascade control of P-synthesis model also includes the M = mRNA transcript encoding P (Figure 5, up). The effector (R), of which synthesis is controlled by the target protein P (Figure 5), can allosterically adjust the activity of G and M, i.e. the catalysts for the transcription and translation steps of the gene expression. In such a cascade scheme, the rate of the ultimate reaction is amplified, depending on the number of cascade levels and catalysis rates.

More complex regulatory modules have been elaborated (Maria [84,89,92,93]) and used in developing genetic regulatory circuits (GRC) following a similar route to 'translate' from the 'language' of molecular biology to that of mechanistic chemistry, by preserving the structural hierarchy and component functions. Once elaborated, such a modular structure can be modelled by using a continuous variable ODE kinetic model under a VVWC framework, and then analysed as functional efficiency by means of some quantitative performance indices (P.I.-s) below described in chapter 5.1.

Similar GERM structures have been used by Wall, et al. [56] and Elowitz & Leibler [61] (Figure 7), and Maria [84,89,90,91,92,93] to model some GRC-s such as genetic-switches (GS) or genetic stimuli amplifiers. Concerning the GERM model size, it depends on its use and properties. As the cell regulatory systems are module-based organized, complex feed-back and feed-forward loops are employed for self- or cross-activation / repression of interconnected GERM-s, leading to different interaction alternatives (directly/inversely, perfect/ incomplete, coupled/uncoupled connections) of a gene with up to 23–25 other genes." Kobayashi et al. [19], to ensure the key-species homeostasis, holistic and local regulatory properties of the enzymatic reactions. While Maria [84,87,89,90,93]; Sewell, et al [5], Savageau [54], Hlavacek & Savageau [55] used reduced GERM structures of 10-14 reactions, that ensures a satisfactory tradeoff between model simplicity and its predictive quality more sophisticated constructions are proposed in the literature (Maria 93]).

As an example, Salis &Kaznessis [8], and Kaznessis [9] designed a bistable genetic circuit, by using two GERM-s extracted from the lac operon of E. coli. The transcriptional regulation is modelled by using a stochastic approach accounting for 40 reactions and 27 species (reduced model) or 70 reactions and 50 species (extended model). Such a regulatory schema (Figures 4-7), including dimeric self-repressors (PP, or RR,) and mutual repression following the presence in excess of one of the activating inducers, can also be illustrated by means of simple representations of Yang, et al. [26], and Maria [6,87,89,90,93]. The advantage of such a modular approach is the possibility to adapt the model size according to the available information, or to use the same GERM structure to model several gene expressions. Modular approach can also be useful in simulating the hierarchical organization of the cell regulatory networks.

The G(P)n type of units (Figure 6), even less realistic, represent the simplest GERM used as control mechanism against which all others are compared. In a G(P)0 module (Figure 6), there are only two main synthesis chains. P is a permease that catalyses the import of NutG and NutP from the environment, and a metabolase that converts them into cellular metabolites MetG and MetP. P is also a polymerase that catalyses the synthesis of G from MetG. Gene G, symbolizing the genome of the cell, functions as catalyst for the synthesis of P from MetP. The result is that G and P syntheses are mutually autocatalytic.

In G(P)0 there are no regulatory elements. In G(P)1, the negative feedback control of transcription is realised by P itself (as effector), via a rapid buffering reaction, G + P <===> GP, leading to the catalytically inactive GP. As proved (Maria [85,87]), the maximum regulatory efficiency at steady-state (index 's') corresponds to [G]s/[G]total =1/2, when the maximum regulation sensitivity vs. perturbations in $[P]_s$ is reached Sewell, et al. [5]. Further allosteric control of G activity, leading to inactive species [GPn], amplifies the regulatory efficiency of the module. As an example, prokaryotes commonly bind multiple copies of transcription factors as a mean of promoting cooperative effects and thus improving regulatory effectiveness (Yang, et al. [26]). The control is better realised by including a supplementary P dimerization step before the buffering reactions that is GERM of type G(PP)n. This explains why most of transcription factors bind as oligomers (typically dimers or tetramers) and why they typically bind in multiple copies (Yang, et al. [26]).

For instance, *dnaA* is an auto-regulated protein and at least five copies can bind to *dnaA* gene in *E. coli.* (Yang, et al.[26]; Speck, et al. [62]). Also, in *E. coli* the monomeric $\pi^{35.0}$ protein of plasmid R6K forms dimers that bind to the operator of the *pir* gene that encodes the protein, and represses its own synthesis (Yang, et al. [26]; Chen, et al. [63]). The λ repressor transcription factor is a dimer, three copies of which bind the operator region of the gene that it regulates (Yang, et al.[26]; Ptashne [96]).

The G(PP)n types of GERM models (Figure 6) better reflect the regulatory loops in which multiple copies of effectors (proteins and transcription factors TF) bind to promoter sites on the DNA that control expression of gene G encoding P (see exemplifications from E. coli by Yang, et al. [26]. The control is better realized if a supplementary P dimerization step is included before the buffering reactions. This explains why most of transcription factors bind as oligomers (typically dimers or tetramers) and why they typically bind in multiple copies (Maria [87]; Yang, et al. [26]; Ptashne [64]).

Module [G(P)n; M(P)n'] (Figures 4-6) tries to reproduce more accurately the transcription / translation cascade of reactions during the gene expression, by including an allosteric control at two levels of catalysis: on G (i.e. DNA) and on M (i.e. mRNA). M is synthesized from nucleotides under G catalysis, and then, P is synthesized in a reaction catalyzed by M (translation). Such a supplementary control of mRNA activity is proved to be a more effective mean of regulating protein synthesis. (Maria [85,87]; Yang, et al. [26]; Hargrove & Schmidt [65]).

It is also to mention the way by which the rate constants in the rapid buffering reactions are estimated, that is for the "effector" reaction type (Maria [87]):

$$L + O \; \underset{\xleftarrow{k_{diss}}}{\xrightarrow{k_{bind}}} \; LO \; ; \; K_{LO} = \frac{k_{bind}}{k_{diss}} = \frac{C_{LO,s}(1+\frac{D_s}{k_{diss}})}{C_{L,s}C_{O,s}} \; ; \text{ (VVWC model)}.$$
....(24)

As discussed by Kholodenko, et al. [66], fast buffering reactions are close to equilibrium and have little effect on metabolic control coefficients. As a consequence, rate constants of such rapid reactions are much higher than those of the core synthesis and dilution rates. To reduce the size of the unknown vector during rate constant estimation of the GERM model, large values of k_{diss} >> D_s can be postulated (5 to 7 orders of magnitude higher) (Maria [87]).
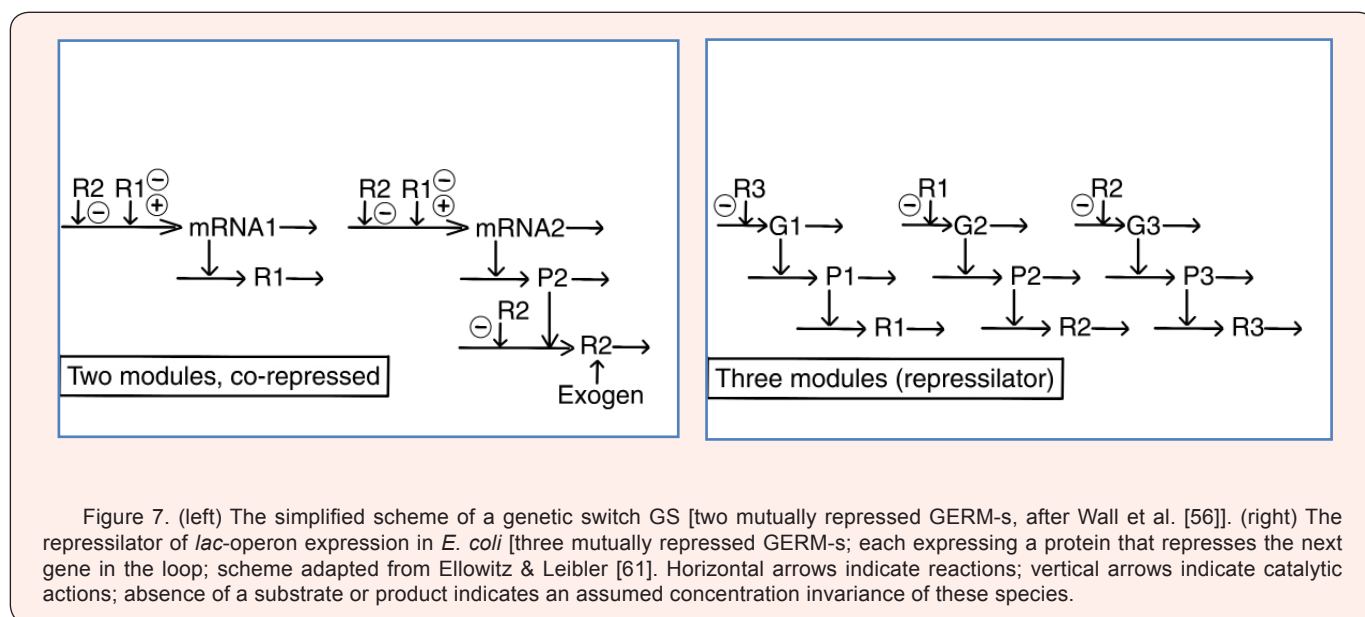
Figure 7. (left) The simplified scheme of a genetic switch GS [two mutually repressed GERM-s, after Wall et al. [56]]. (right) The repressilator of *lac*-operon expression in *E. coli* [three mutually repressed GERM-s; each expressing a protein that represses the next gene in the loop; scheme adapted from Ellowitz & Leibler [61]. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicates an assumed concentration invariance of these species.

Savageau &Voit [60] ; Savageau [54]; Hlavacek & Savageau [55], Voit [35]; Atkinson, et al. [10]; Wall, et al. [56] used similar GERM structures to construct and study properties of genetic switches (Figure 7), but modelled using ODE kinetic sets including rate expressions of power-law type (the so-called S-systems) obtained by recasting the elementary steps and

intermediates in a lumped representation including apparent rate constants and reaction orders. Such apparent power-law models with fractional orders of reactions produced a biased representation of the real process and suffered of a series of inconveniences related to:

I. distorted GERM properties (multi-stationarity, sub-optimal P.I.-s);

II. a very limited predictive capacity and flexibility due to the empirical / apparent rate expressions, and the inherent reduced number of considered intermediates and species interactions;

III. lack of physical meaning of model parameters making the model impossible to be generalized;

IV. a loss / alteration of systemic / holistic properties of GRC, (related to cell system stability, multiplicity, sensitivity, regulatory characteristics).

By contrast, the mechanistic based GERM models in a VVWC framework (Maria [87,89,90,93]) seem to be more robust, flexible and easily adaptable to different case studies (examples in chapters 6-7).

4.3.2. Rate constant estimation in GERM-s and GRC-s

Under stationary (homeostatic) growing conditions of the cell, species synthesis rates (r_j) must equal to first-order dilution rates $D_s C_{js}$ in (3-4), leading to time-invariant (index "s") species concentrations C_{js}, i.e. homeostatic conditions of $\left(dC_j / dt\right)_s = 0$. Under such a balanced growth, the resulted nonlinear algebraic mass balance set is the following [see also eq. (10)]:

$$\left(\frac{dC_j}{dt}\right)_s = \left(\frac{1}{V}\frac{dn_j}{dt}\right)_s - D_s C_{js} = h_{js}(C_s, k, t) = 0 \quad ; \quad j = 1,\ldots,n_s(\text{no. of species}),$$

$$D_s = \left(\frac{RT}{\pi}\right)\sum_j^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right)_s \qquad \ldots\ldots\ldots\ldots\ldots\ldots(25)$$

This nonlinear set is used to estimate the rate constants k (and even some unobservable $C_{j,s}$) for every cell subsystem (a GERM in the present case), from the known stationary concentration vector C_s (individual or lumped components considered in the kinetic model), by using an effective procedure (e.g. routine "solve" of Maple8™- package). As the (RT/π) term is known from the initial condition, and the number of model parameters is usually higher than the number of observed cell species, supplementary optimization rules must be applied to determine some rate constants, by imposing optimum regulatory criteria for GERM-s, such as minimum recovering time of the stationary concentrations (homeostasis) after a dynamic ('impulse'-like) perturbation in a key-species (Maria [87]), by using effective solvers (Maria [53]), or a quick action of the buffering reactions to get the fastest recovering time after an external / internal perturbation; smallest sensitivity of the key-species homeostatic levels vs. external perturbations in the nutrient levels; stability highest homeostasis strength, etc. (see below chapters 5.1-5.2.10) (Maria [6,84,87,89-93]). Estimation rule is based on the fulfilment of the stationary condition (25), system invariants (mass balance equations), and on imposing optimum regulatory criteria, formulated as following (Maria [53,86,87]):

$$\left[\hat{k},\hat{C}_S\right] = \arg Min\left(\tau_P\right), \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26a)$$

subjected to:

$$\left(\frac{dC_j}{dt}\right)_s = \left(\frac{1}{V}\frac{dn_j}{dt}\right)_s - D_s C_{js} = h_{js}(C_s, k, t) = 0 \;;\text{j} = 1,\ldots,\text{ns(no.of species)},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26b)$$

$$D_s = \left(\frac{RT}{\pi}\right)\sum_j^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right)_s \quad \text{(VVWC cell model at steady-state)}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26c)$$

$$\left[\hat{k},\hat{C}_s\right] > 0 \text{ , (physical significance requirement)}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26d)$$

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

$[L]_{active} / [L_{total}] = 1/2$, (maximum dynamic efficiency, chapter 5.1; (Maria [85]; Yang, et al. [26]; Sewell, et al.[5]);

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26e)$$

$$\left(\sum_j^{all}C_j\right)_{cell} = \left(\sum_j^{all}C_j\right)_{env} = constant \quad \text{(isotonic system fulfilling the}$$
hypothesis $\pi_{cyt} = \pi_{env}$=constant of the VVWC model, chapter4.2).

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26f)$$

In the above formulation, the τ_p is the recovering time of the stationary key-P-protein concentration [P]s expressed in that GERM. The τ_p has been evaluated by applying a ±10%[P]s impulse perturbation and by determining by simulation with the VVWC GERM model the recovering time with a tolerance of 1%[P]s (Maria [6,87,89,90,93]). The estimation problem has a significant degree of freedom because the nonlinear set (26b) is under-determined. Species Li [e.g. enzymes P, or even genes G (DNA), M (mRNA), (Figures 4-6)] denotes a GERM component at which regulatory element / transcriptional factor TF (P, R) acts. To estimate $\left[\hat{k},\hat{C}_s\right]$, other regulatory global properties can also be used together with the constraints of (26) (Maria [87]; Van Someren et al. [67]. The reverse reaction rate constants in the rapid buffer reactions of GERM-s, of type G+P < === > GP, are adopted at values five to seven orders of magnitude higher than the $D_S = \ln 2 / t_c$ (see the proof of Maria [87]). That is because fast buffering reactions are close to equilibrium and have little effect on metabolic control coefficients. As a consequence, rate constants of such rapid reactions are much higher than those of the core synthesis in the GERM and than that of the dilution rate.

VVWC model formulation presents important advantages vs. classical CVWC models such as (Maria [6,87,89,90,93]):

A. the estimated rate constants are more realistic comparatively with those derived from constant-volume model formulations, due to the considered cell regulatory properties.

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

C. species inter-connectivity (i.e. the degree to which a perturbation in one component influences other GERM model components) is better characterized by including direct

interrelations (via common reactions and intermediates) but also indirect relationships via the common cell-volume to which all species contribute [see eq. (6-7,9-11)].

D. Possible perturbations in the volume size and osmotic pressure are also considered.

E. perturbations applied to components of large concentrations lead to an important perturbation of the cell volume, which in turn lead to large perturbations of other cell component stationary concentrations (i.e. the so-called 'secondary' or 'indirect' perturbations); vice-versa, perturbations in species of low levels will have a little effect on the cell volume, and then a small secondary effect on the other components, because: $V_{perturb}/V_o = \left(\sum n_j\right)_{perturb} \Big/ \left(\sum n_j\right)_o$ (see chapter 4.2.1 and 5.2.5).

F. cells of large content (large 'ballast') tend to diminish the effect of environmental perturbations (the so-called 'inertial' effect, or perturbation smoothing); the ballast effect is an expression of how all cell components are interconnected via cell volume changes (see chapter 5.2.5).

G. the derived performance indices P.I.-s of GERM-s under a VVWC formulation present more realistic estimates comparatively with those derived from the classical CVWC kinetic models which tend to overestimate these P.I.-s (Maria [6,85,87,89,90]).

The only disadvantages of a VVWC kinetic models result from:

a. a larger computation effort to identify the model parameters from the stationary species concentrations, and for solving the nonlinear set eq. (25) (which sometimes can present multiple solutions, difficult to be discriminated);

b. In a VVWC kinetic model of a GERM, or GRC, all species (individual or lumped) have to be considered (including the lumped genome, proteome, and metabolome), because all these lumps contribute to the volume and dynamics via the isotonicity constraint. In such a manner, the number of rate constants increases [due to the lumped GERM accounting for the genome, and proteome replication introduced by Maria [84,89,90,91,92,93]; see chapters 6-7] leading to a corresponding increase in the identification effort.

Such VVWC kinetic models with continuous variables can fairly reproduce metabolic processes occurring during the cell balanced growing phase (ca. 80% of the cell cycle). When the cell reaches a critical size and a certain level of the surface-area-to-volume ratio, the division phase begins, lasting the last 20% of the cell cycle. Over this phase (not analysed here), specialized proteins constrict the cell about its equator, thus leading to cell division. The duplicated content is thus partitioned, more or less evenly, between daughter cells. To model such a phase, supplementary terms must be added to explicitly account for the cell membrane dynamics (Morgan, et al. [25]).

5. Some Rules Used For Modelling Genetic Regulatory Circuits (GRC)

Due to the huge complexity of cell processes, including the tight regulation of the metabolic reactions through the enzymes' level expressed during the cell cycle, early attempts to model such complex GRC-s (by using the *in-silico* tools of the emergent systems biology) have been reported after 2000's when a considerable increase in the omics data (EcoCyc [49] and KEGG [50] databases) and advances in high-throughput experiments have been reported. But really fast advances in systems biology and bioinformatics have been reported when there have been constructed more and more adequate cell models by using valuable tool from <u>border areas</u>, like nonlinear system control theory; chemical and biochemical engineering modelling theory, that is:

a. Molecular species conservation law (stoichiometry analysis; species differential mass balance set).

b. Atomic species conservation law (atomic species mass balance).

c. Thermodynamic analysis of reactions (quantitative assignment of reaction directionality),Haraldsdottir, et al. [42]; set equilibrium reactions; Gibbs free energy balance analysisset cyclic reactions; find species at quasi-steady-state; improved evaluation of steady-state flux distributions that provide important information for metabolic engineering Zhu, et al. [43].

d. Application of ODE model species and/or reaction lumping rules (Maria [53,86]).

These tools have made possible the development of some milestone works, such as: Heinrich & Schuster [33]; Torres & Voit [69]; Brazhnik, et al. [70]; Cornish-Bowden [71] ; Stelling [1]; Stephanopoulos et al. [22]; etc.], to mention only a few of them.
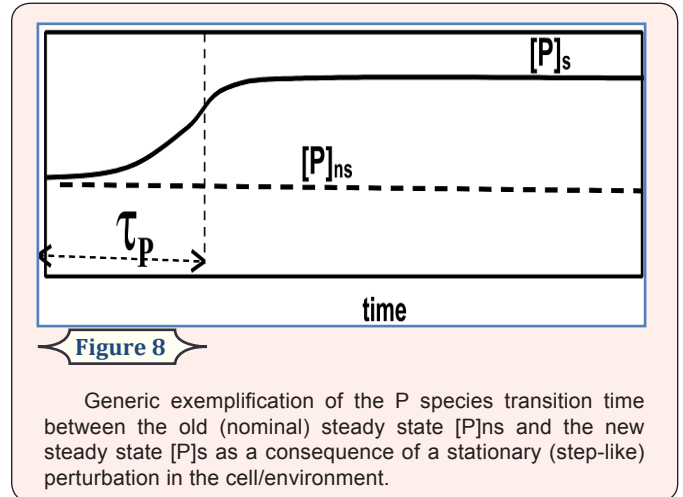
However, to properly analyse, and compare the GERM regulatory performances to select the best structure for a certain case study, it is absolutely necessary to define quantitative performance indices (P.I.) of the GERM. At this point, the nonlinear system control theory offers an invaluable aid. In this chapter, the main P.I.-s are defined based on the previous studies of Maria [6,85,87,89,90,91,93].

These P.I.-s roughly fall in two categories of indices, defined for stationary ('step' like) or dynamic ('impulse' like) continuous perturbations of key-species stationary concentrations. When analysing a GERM regulating the synthesis of a certain P-protein, it is necessary to also consider the random perturbations that appeared due to the interactions of the P-protein synthesis with other metabolic processes, or due to environmental changes in the nutrient levels. All these determine a GERM response that tends to maintain the key-component functions and species homeostasis. The module regulatory efficiency depends on the GERM regulatory structure, species inter-connectivity, quasi-steady-state (QSS) characteristics, cell size, and perturbation magnitude to be below discussed and exemplified. The characteristics and the P.I.-s of various GERM structures are essential when constructing GRC-s made up of chains of GERM -s.

5.1. Define performance indices (P.I.) of a GERM to homeostatically regulate a gene expression under a deterministic VVWC modelling approach

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 [P]s. These processes occur in addition to those of the "core''system (G/P replication over the cell cycle)

5.1.1. Define stationary perturbations and stationary efficiency

Figure 8

Generic exemplification of the P species transition time between the old (nominal) steady state [P]ns and the new steady state [P]s as a consequence of a stationary (step-like) perturbation in the cell/environment.

Stationary perturbations refer to permanent modifications in the levels of the external nutrients or of the internal metabolites, leading to new stationary component concentrations inside the cell. Referring to a target protein P in a GERM, the regulatory module tends to diminish the deviation [P]s - [P]ns between the 'nominal' QSS (unperturbed set-point, of index 'ns') and the new QSS reached after perturbation (the new setpoint [P]s see Figure 8). Equivalently, the P-synthesis regulatory module will tend to maintain [P]ns within certain limits, [P]min ≤ [P]ns ≤ [P]max. A relative R_{ss}=±10% maximum deviation has been proposed by Sewell, et al. [5] and Yang, et al. [26] for an effective GERM. A measure of the species "i" steady-state concentration $(C_{i,s})$ 'resistance' to various stationary perturbations (in rate constants, k_j, or in nutrient concentrations, $C_{Nut,j}$) is given by the magnitude of relative sensitivity coefficients at QSS, i.e. $S(C_i; k_j)$ and $S(c_i; Nut_j)$ respectively, where $S(State; Perturbation) = \partial(State)/\partial(Perturbation)$ are the state sensitivities vs. perturbations Varma, et al. [72].

5.1.1.1. Responsiveness

Responsiveness to exo/endogeneous signaling species of the analysed GERM or GRC can be represented by the small transient times τ_j necessary for a species j QSS-level to reach a new QSS (with a certain tolerance) after applying a stationary external stimulus (Maria [90]). Consequently, the P.I. measure of the GERM efficiency to move fast to a new QSS is given by the duration of the transition time τ_p (given for a generic species P in Figure 8) necessary of this species old QSS level to reach the new steady-state concentration. Another regulatory P.I., that is $A_{unsync} = k_{syn}*k_{decline}$, has been introduced to illustrate the maximum levels of (unsynchronized) stationary perturbations in synthesis or

consumption rates of a key-species 'tolerated' by the cell within defined limits Sauro & Kholodenko [18]. For instance, in the case of the P-species these rate constants belong to the synthesis and degradation lumped reactions $\xrightarrow{k_{syn}} P \xrightarrow{k_{\deg}}$.

5.1.1.2. Stationary efficiency

Another stationary P.I. is related to the small sensitivities $S(C_i; Nut_j)$ of the key-species levels C_i vs. changes in the external nutrient levels Nut_j. These sensitivities are computed from solving a nonlinear algebraic set (26b-c) obtained by assuming QSS conditions of the ODE kinetic model, and known nominal species stationary concentrations C_s. Then, differentiation of the steady-state conditions eqn. (26b) leads to the evaluation of the state sensitivity vs. nutrient levels, i.e. $S(C_i; Nut_j) = \left(\partial C_i / \partial C_{Nut_j} \right)_s$ by solving the following linear set ("s" index denotes stationary condition) :

$$\left[\frac{\partial h_i(\boldsymbol{C}, C_{Nut}, \boldsymbol{k})}{\partial C_i} \right]_s \left[\frac{\partial C_i}{\partial \boldsymbol{C}_{Nut_j}} \right]_s + \left[\frac{\partial h_i(\boldsymbol{C}, C_{Nut}, \boldsymbol{k})}{\partial \boldsymbol{C}_{Nut_j}} \right]_s = \boldsymbol{0} \quad \text{.......................(27)}$$

In the previous relationship, the ODE model Jacobian $\boldsymbol{J_C} = \left[\partial h_i / \partial C_j \right]_s$ was numerically evaluated for the cell-system stationary-state (26b), by using the Matlab-package symbolic and numerical calculus facilities.

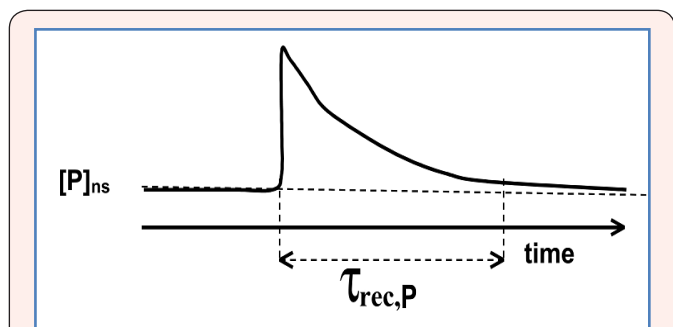### 5.1.2 Define dynamic perturbations and dynamic efficiency

Generic exemplification of the species P recovering time of the steady state [P]ns after a dynamic (impulse-like) perturbation in the cell/environment.

Dynamic perturbations refer to instantaneous changes in the concentration of one or more cell components that arise from a process lasting an infinitesimal time (impulse-like perturbation). After perturbation, the system recovers and returns to its stable nominal state QSS (see Figure 9 for a generic P-protein case). The computed recovering times τ_{rec}, necessary to cell components to return to their stationary concentrations (with a tolerance of 1-5% proposed by Maria [87]) may differ from one species to another depending on how effective are their corresponding regulatory circuits.

Recovery rates are properties of all interactions within the system rather than of the individual elements thereof Heinrich & Schuster [33]. In terms of the evolution and stability of component QSS concentrations included in a dynamic cell system expressed by an ODE model (2) or (6), these properties can be evaluated from the analysis of the eigenvalues λ_i (i = no. of species) of the linearized model Jacobian matrix $J_C = \left(\partial \boldsymbol{h}(\boldsymbol{C}, \boldsymbol{k}) / \partial \boldsymbol{C} \right)_s$ of elements $J_{ik} = \partial h_i(\boldsymbol{C}, \boldsymbol{k}) / \partial C_k$. If small perturbations of a steady state C_s are considered then, this steady state is asymptotically stable if the

real parts of the Jacobian eigenvalues are all negative $Re(\lambda_i) < 0$ (Liao & Lightfoot [73]; Heinrich & Schuster [33]; Hofmeyr [74]). If the system is stable then, it reaches the same QSS after cessation of a dynamic impulse-like perturbation, or it reaches another QSS after cessation of a stationary step-like perturbation.

Here are to be mentioned the works of Maria [85,87] and Sewell, et al. [5] that proved that the optimum concentrations in the "buffering" reactions of GERM-s involving the active and inactive forms of the "catalyst" ensuring the maximum regulation *dynamic efficiency* vs. perturbations (see below) are those of , , or [G] = [GP], [Gi] = [G$_i$ P$_j$ P$_j$], or [M] = [MP], [M$_i$] = [M$_i$ P$_j$ P$_j$] etc. (Figure 6).

5.1.2.1 Recovering time (dynamic efficiency)

A P.I. measure of the GERM efficiency to fast recover the key-species stationary concentrations is given by the time τ_j necessary to the species "j" to recover its steady-state concentration with an assumed tolerance (of 1% as proposed by Maria [87]). Formally, the recovering trajectory and rate (denoted by R_D) can be approximated from the solution of the linearized model of the cell system Heinrich & Schuster [33]:

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

$$\text{..(28)}$$

where: C = concentration vector; λ_i = eigenvalues of the system Jacobian matrix at QSS, $\boldsymbol{J_C} = \left(\partial \boldsymbol{h}(\boldsymbol{C}, \boldsymbol{k}) / \partial \boldsymbol{C} \right)_S$; b_i, d_i = constants depending on the system characteristics at stationary conditions; t = time. The recovering rate R_D reflects the recovering properties of the regulated key-P synthesis by the GERM system. In a simple way, the species "j" recovering times $\tau_j \sim 1/R_D$ and trajectories $C_j(t)$ can be obtained by simulating the GERM system dynamics (using the GERM model) after applying a small impulse perturbation of the species steady-state of ±10% $C_{j,s}$ and determining the recovering time until the steady-state $C_{j,s}$ is reached with a 1% tolerance (Maria [85]). Species recovering trajectory and amplitude are both very important. As proved by Maria [85], the GERM-s display very different recovering trajectories and amplitudes. The most effective are the GERM-s ensuring the smallest amplitude of the recovering pathway, thus not disturbing the other cell metabolic processes. As underlined by Maria [85], the recovering trajectories in the G/P phase plane is more linear for the efficient GERM-s, presenting a lower amplitude, thus not disturbing other cell processes.

5.1.2.2. Regulatory robustness

The regulatory robustness of a GERM model is defined by Maria [87] as being the property to realize (Min) $(\partial R_D / \partial \boldsymbol{k})$, where R_D denotes the key-species recovering rates, while k = rate constant vector (depending on the micro-organism). This P.I. can be considered a systemic regulatory property, as long as GERM species levels are able to modify the apparent reaction rates. In fact, the cell metabolic network robustness and functionality are linked to the cell phenotype and gene regulation scheme.

5.1.2.3. The effect of the no. of regulatory effectors (n)

By definition, the GERM models include an adjustable number of "regulatory effectors", that is: n in the *G(P)n* or *G(PP) n* series; and n and n' in the *[G(P)n; M(P)n']* series (Figure 6). As

proved by Maria [6,85,87], and by Yang, et al. [26], a quasi-linear relationship of P.I. function of no. of regulatory effectors (n) can be derived for every GERM type, of the form:

$$\text{P.I.} = a_0 + \Sigma_i a_i n_i , \qquad ....(29)$$

Where P.I.= regulatory performance index, such as R_D, AVG(τ_j), STD(τ_j), stability strength, etc. (see below; AVG= average; STD= standard deviation); n_i = number of effectors (P, PP, R, etc.) acting in the i-th allosteric regulatory unit $L_i(R_i)n_i$; a_o, a_i = constants related to the P.I. and module type. Such a dependence can also be observed in the (Figure 10) Roughly, Maria [6,87,89,93] proved that:
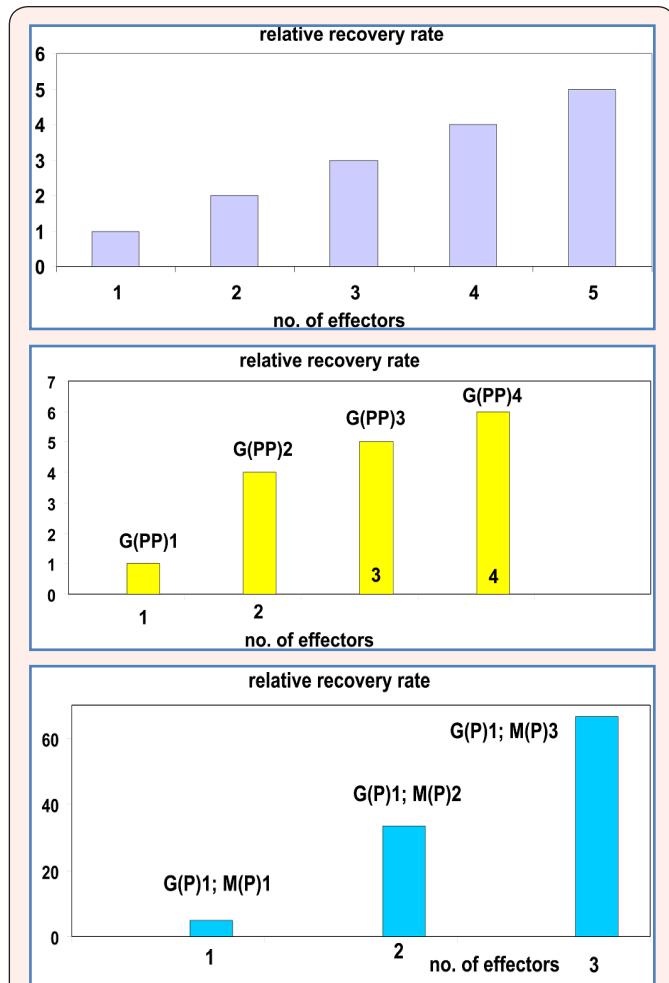

Figure 10

Dependence of the GERM dynamic regulatory efficiency (expressed by the P key species relative recovering rate RD after a ±10% impulse perturbation of its stationary level [P]s) for various GERM types, as function of the number n of buffering reactions (i.e. no. of 'effectors'). Adapted from the results of Maria [85,87]; Yang et al. [26]; [58]; Gabaldon & Huynen [59]; Ferrell [76]. Up-plots refer to the G(P)n type of gene expression regulatory modules. Middle-plots refer to the G(PP)n type of gene expression regulatory modules; Down-plots refer to the [G(P)1; M(P)n] type of gene expression regulatory modules.

a. P.I. improves ca. 1.3-2 times (or even more) for every added regulatory unit to the module. Multiple regulatory units lead to an average recovering time AVG (τ_j) of all GERM species much lower than the cell cycle period t_c, under a constant logarithmic volume growing rate, D = ln(2)/t_c.

b. Combinations of regulatory schemes and units (with different effectors) can improve the regulatory P.I.s.

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

5.1.2.4. Species interconnectivity

Species interconnectivity in a modular regulatory schema of reactions can be viewed as a degree to which they 'assist' each other and 'cooperate' during the GERM system recovering. Cell species connections appear due to common reactions, or common intermediates participating to chain reactions, or from the common cell volume to which all cell species contribute (under constant osmotic pressure, eq. 6,9,10, and VWC model hypotheses in chapter 4.2. and Table 1. Vance, et al. [75] reviewed and proposed several quick experimental - computational rules to check a reaction schema via species inter-connectivities.

By inducing experimental perturbations to a (bio)chemical system, by means of tracers, or by fluctuating the inputs of the system, one can measure the perturbation propagation through the consecutive / parallel reaction path. Then, various techniques can determine the "distance" among observed species, and rules to include this information in elaborating a reaction schema. In the present study, one proposes a similar approximate measure of species interconnectivity related to the species recovering-times after a dynamic perturbation, that is: AVG(τ_j) and STD(τ_j), i.e. the average and the standard deviation of the species individual recovering times τ_j. As AVG and STD are larger, as the cell dynamic regulatory effectiveness is lower, species less interconnected, and components recover more disparately (scattered). The higher the number of effectors and buffering reactions, the better these dynamic regulatory indices of the GERM are (Maria [85,87,89,90,93]).

5.1.2.5. GERM system stability

The QSS stability of a GERM, that is the system's capacity to recover the QSS after cessation of a dynamic perturbation, can be predicted by analysing the QSS of the dynamic GERM ODE model (2) or (6). This property can be evaluated from the analysis of the eigenvalues λ_i (i = no. of species) of the linearized model Jacobian matrix of elements $J_{ik} = \partial h_i(\boldsymbol{C}, \boldsymbol{k})/\partial C_k$. The QSS is asymptotically stable if the real parts of the Jacobian $J_C = (\partial \boldsymbol{h}(\boldsymbol{C}, \boldsymbol{k})/\partial \boldsymbol{C})_s$ eigenvalues λ_i are all negative, that is Re(λ_i) < 0 0 for all i, (Liao & Lightfoot [73]; Heinrich & Schuster [33]; Hofmeyr [74]; Varma, et al. [72]).

If the system is stable then, it reaches the same QSS after cessation of a dynamic impulse-like perturbation, or it reaches another QSS after cessation of a stationary step-like perturbation. Here it is to mention two important observations:

i. A characteristic of the VVWC models including the Pfeiffers'constraint (9-11) is that they are always stable (intrinsic stability), because, as proved by Morgan et al. [25], always Max(Re(λ_i)) = -D .

ii.   By contrast, one fundamental deficiency of the classical CVWC model formulations is the lack of the intrinsic stability of the cell system model, because these models do not include the Pfeiffers'constraint (9-11). Consequently, the GERM regulatory mechanism for recovering the homeostasis (illustrated in (Figure 3,11,12,13) and chapter 5.2.4) is no longer working and the CVWC model becomes invalid.

5.1.2.6. The steady-state C_s stability strength

This GERM property is related to the strong capacity of the regulatory system to 'resist' to large external / internal perturbations, thus maintaining the system steady-state C_s and determining very quick recovering paths. As with all other P.I.-s, this GERM property is related to the GERM system characteristics. Basically, as $Max(Re(\lambda_i))<0$ is smaller as this QSS is more stable. Here, the eigenvalues λ_i of the Jacobian $J_C=(\partial h(C,k)/\partial C)_s$ are evaluated at a checked QSS C_s.

In a more systematic approach, the steady-state C_s stability strength can also be associated to an index against periodic oscillations of key-species synthesis. This index can be evaluated from the linearized form of the system model, by calculating the monodromy matrix A(T) after a checked period T of time, that is Maria [85]:

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

(30)

For a stable C_s, i.e. $\left|\lambda_{A_i}\right|<1$, as $\left|\lambda_{A_i}\right|$ are smaller, as the stability of the C_s state is stronger and that QSS recovers faster after a small dynamic perturbation. Here, λ_{A_i} denotes the eigenvalues of the A(T) matrix, while I = identity matrix. In other words:

QSS stability strength involves: (Min) $Max(Re(\lambda_i))$, $Re(\lambda_i)$ < 0; (Min) $\left|\lambda_{A_i}\right|<1$.

(31)

To summarize, the regulatory efficiency P.I.-s proposed to evaluate the perturbation treatment by a GERM or by a chain of GERM-s, are given in (Table 4)

5.2 Efficiently linking GERM-s in a VVWC modelling framework

When linking GERM-s to construct a GRC reproducing a certain function of the cell, there are two contrary trends: on one hand is using simple GERM structures to reduce the model identification computational effort; on the other hand, it is important to use simple, but effective and flexible GERM-s able to reproduce individual enzyme-synthesis, but also holistic properties of the GRC (complex examples are provided in chapters 6-7).

To exemplify in this chapter the GERM linking analysis in a simple way, one considers an *E. coli* cell, in a balanced growth under isothermal and isotonic conditions, with a cell cycle period of t_c = 100 min, and a quasi-constant logarithmic growing rate of $D_s = \ln 2/t_c$. The nominal concentrations of the individual and lumped cell species correspond to a cell of "high ballast', and are given in (Table 5), being similar to those of Maria [87], i.e. $C_{NutG,s}=3\times10^6\,nM$, $C_{NutP,s}=3\times10^8\,nM$ (nM= nano-molar, i.e. 10^{-9} mol/L

concentration). As only a few numbers of individual species are accounted in the model, the cell "ballast" is mimicked by adopting high levels for metabolite concentrations according to Ecocyc [49] information, i.e.:

$$\sum_j C_{MetPj,s} =3\times10^8\,nM \;\; ; \sum_j C_{MetGj,s} =C_{NutG,s}+C_{NutP,s} - \sum_{j\neq MetGj}^{cell} C_{js} \approx 3\times10^6\,nM \,.$$

....(32)

Nutrient concentrations in the environment are assumed to be constant during the cell cycle.

5.2.1. Ranging the number of transcription factors TF and buffering reactions

For selecting the suitable GERM structure that fits the available data, the first problem to be solved is related to the number of buffering reactions of type G + P <===> GP or M + P <===> MP (Figure 1,6) necessary to be included in the model. Evaluation of P.I.-s for a large number of GERM structures (Maria [6,85,87,89,90,93]) indicates that the dynamic regulatory efficiency of $[G(P)n]$ modules is nearly linearly increasing with the number (n) of buffering reactions (eq. 29, and Figures 6, 10). Moreover, the plots of (Figure 10) reveal that this increase is more pronounced in the case of $[G(PP)n]$ model structures using dimeric TF (that is PP instead of simple P), and also for $[\,G(P)n\,;\,M(P)n'\,]$ modules that use a control scheme in cascade of the gene expression.

Such a module efficiency ranking concerns not only the dynamic efficiency, but also other P.I.-s, discussed in the previous paragraphs, such as the stationary regulatory effectiveness; low sensitivity vs. stationary perturbations; stability strength of the homeostatic QSS, and species recovering trajectories more linear in the G/P phase plane and of a lower amplitude.

To summarize, when selecting a suitable GERM to be included in a GRC the following issues are to be considered (Maria [6,85,87,89,90,93]):

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

b.   The catalyst activity control at a single enzyme level (i.e. G(P)0, G(PP)0, [G(P)n;M(P)0] structures, that is lacking of buffering reactions able to modulate the gene G and M catalytic activity) appears to be of lowest regulatory efficiency.

c.   Multiple copies of effector molecules (i.e. R, P in Figures 5,6), which reversibly and sequentially (allosterically) bind the catalyst (G, M) in negative feedbacks, improve the regulation effectiveness.

d.   A structured cascade control of the "catalyst" activity, with negative feedback loops at each level as in the $[\,G(P)n\,;\,M(P)n'\,]$ model series, improves regulation and amplifies the effect of a change in a stimuli (inducer). The rate of the ultimate reaction is amplified, depending on the number of cascade levels and catalysis rates. As an example in (Figures 5,6) by placing regulatory elements (P, R) at the level of mRNA (i.e. species M), and at the level of DNA (i.e. species G) in the $[\,G(P)n\,;\,M(P)n'\,]$ model is highly effective.

e.   The nearly linear increase of GERM P.I.-s with the number n_i of effectors (P, PP, R) acting in the i-th allosteric unit $L_i(R_i)n_i$

of buffering reactions applied at various level of control of the gene expression, is valid for both dynamic and stationary P.I.-s of the (Table 4)

Table 4: The regulatory efficiency performance indices P.I.-s proposed to evaluate the perturbation treatment efficiency by a GERM following the definitions of Maria [87]. Abbreviations: Min = to be minimized; Max = to be maximized. Note: k_{syn} and $k_{decline}$ refers to the --> P --> overall reaction.

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	$S^i_{NutP_j} = \left[(\partial C_i / C_{is})/(\partial C_{Nut_j} / C_{Nut_{js}}) \right]_S$
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	$(\partial R_D / \partial \boldsymbol{k})$
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	$Max(\mathrm{Re}(\lambda_i))$
QSS stability strength(note b)	Min	$\left| \lambda_{A_i} \right| < 1$

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.

Footnotes:

(a) λ_i = i-th eigenvalues of the model Jacobian matrix $J_C = (\partial \boldsymbol{h}(\boldsymbol{C},\boldsymbol{k})/\partial \boldsymbol{C})_s$ defined by eq. (30).

(b) see eq. (30) and Maria [85] for the monodromy matrix A calculation;

λ_{Ai} = i-th eigenvalue of the monodromy matrix A.

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

g. Combinations of regulatory schemes and units (with different effectors) might improve the regulatory P.I.-s [chapter. 5.2.3, 5.2.6, 5.2.7].

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

5.2.2 The effect of the mutual G/P synthesis catalysis

One essential aspect of the $[G(P)n]$, $[G(PP)n]$, and $[G(P)n\ ; M(P)n']$ kinetic models of GERM is the mutual catalysis of G and its encoding protein P synthesis. If one adds the VVWC modelling constraints eqn. (6,10) and the requirement of getting a maximum dynamic responsiveness and efficiency by keeping $[G]_S = [G(P)]_S = [G(PP)]_S = ...= [G(P)_n]_S$ discussed in chapter 5.1.2. This direct and indirect link of G and P syntheses ensures a quick recovery of both stationary [G]s and [P]s after any small perturbation. To prove this in a simple way, one considers a synthesis of the G/P pair in a GERM of [G(P)1] or a [G(P)0] type (Figures 1,6). After estimating the rate constants from solving the stationary model equations by using the homeostatic concentrations of (Table 5) (high ballast cell case), one determines each GERM dynamic efficiency by applying a –10% impulse perturbation in the [P]s = 1000 nM at an arbitrary t=0. The obtained recovering trajectories of P and G obtained by model simulations are plotted in (Figure 11) for the [G(P)1]. The plots reveal a very good regulatory efficiency of the [G(P)1], both G, and P species presenting relatively short recovering rates, and negligible for the other species (Table 5). These plots reveal in a simple way the self-regulation of the G/P pair synthesis: after the impulse perturbation leading to the decline of [P]s from 1000 nM to 900 nM, the very fast buffering reaction G + P <===> GP leads to restore the active G, whose concentration quickly increases to [G]= 1.027 nM; as a consequence, the synthesis rate of P increases leading to a fast P recovering rate which, in turn, contributes to the recovering of G-lump steady-state. For comparison, as revealed by the results displayed in (Table 6) the dynamic efficiency of the module [G(P)0] is much lower, species recovering their QSS over longer transient times. Also, the species connectivity is better in the [G(P)1] compared to [G(P)0], being reported smaller STD (τ_j). Consequently, removal of the buffering reaction that automatically adjusts the "catalytic activity" of G, will: decrease the species inter-connectivity (increasing the standard deviation of the recovering times); will increase the species recovering times; will increase the sensitivities of the species steady-state vs. external nutrients (Table 6).
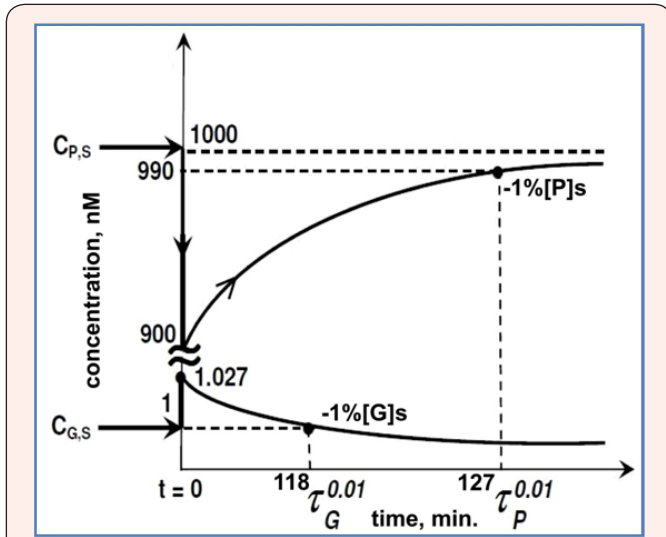
Figure 11

Exemplification of the self- and mutual G/P pair catalysis and regulation after a –10% impulse perturbation in the [P]s = 1000 nM at an srbitrary time t=0 for a generic GERM of [G(PP)1] type (example discussed in the chapter 4.2.2). Simulated results have been generated by using the cell nominal stationary conditions of Table 5 (for a high ballast cell, with adopting [G]s = 1 nM).

As expected, the P.I.-s of the GERM depend not only on I) the no. of effectors (buffering reactions), but also on ii) the TF type (P, or PP), and even more iii) on the used control scheme (i.e. simple or in cascade).

Table 5: The nominal stationary homeostatic (QSS) species concentrations of the analysed *E. coli* cell, and species recovering times in a GERM of [G(P)1] type after a –10% impulse perturbation in the key-protein stationary [P]s produced at an arbitrary moment t=0. The cell initial volume is of Vcyt,0 = 1.66*10⁻¹⁵ L. Values are adapted from the example of Maria [87]. (Applications are commented in the chapter 5.2). nM = nano-molar.

Species	Low Ballast hypothetical Mutant cell (nM)		High Ballast Cell (nM) (*)	
	QSS conc. $C_{j,s}$ (nM)	Recovery Time (min.)	QSS conc. $C_{j,s}$ (nM)	Recovery Time (min.)
Lump $C_{NutP,s}$	3000	NG	$3*10^8$	NG
Lump $C_{NutG,s}$	3000	NG	$3*10^6$	NG
Lump $\sum_j C_{MetG_{j,s}}$	~2000	NG	~ $3*10^6$	NG
Lump $\sum_j C_{MetP_{j,s}}$	3000	NG	$3*10^8$	NG
$C_{P,s} = [P]_s$	1000	103	1000	133
$C_{G,s} = [G]_s$	0.5	223	0.5	93
$C_{GP,s}$	0.5	246	0.5	93
$\sum_j C_j$	12001		~6.06*10⁸	

Footnotes: The lump $\sum_j C_{MetG_{j,s}}$ results from the isotonicity constraint

$$\overset{cell}{\underset{all\ j}{\sum}} C_{j,s} = \overset{env}{\underset{all\ j}{\sum}} C_{j,s} = C_{NutG,s} + C_{NutP,s} = \sum_j C_{MetG_{j,s}} + \sum_j C_{MetP_{j,s}}$$

The considered cell cycle is of t_c=100 min.; the cell-volume logarithmic growing rate is Ds=ln(2)/tc. The Max(Re(λ_j)) < 0 indicates a stable QSS homeostasis of the cell, where λ_j are the Jacobian eigenvalues of the VVWC kinetic model of the GERM. The model rate constants have been estimated from solving the stationary form of the model (that is for $dC_j / dt = 0$) with known stationary species concentrations $C_{j,s}$ displayed in this table. The G + P <===> GP buffer reaction was considered with the reverse reaction rate constant of 10^5 1/min Maria [6]. Lumps NutP and NutG denote substrates used in the synthesis of metabolites MetP and MetG respectively, further used for the P and G synthesis. Notations: G= a generic gene (DNA) from *E. coli* cell; P= the protein encoding G; M= mRNA; GP= the inactive complex of G with the TF; C_j = species "j" concentration; cyt= cytoplasma; "o"= initial; 's' index refers to the QSS; NG = neligible. (*) Species concentrations of the high ballast cell correspond to the *E. coli* cell K-12 strain EcoCyc [49]; Allen & Kornberg [83]; details of Maria [89].
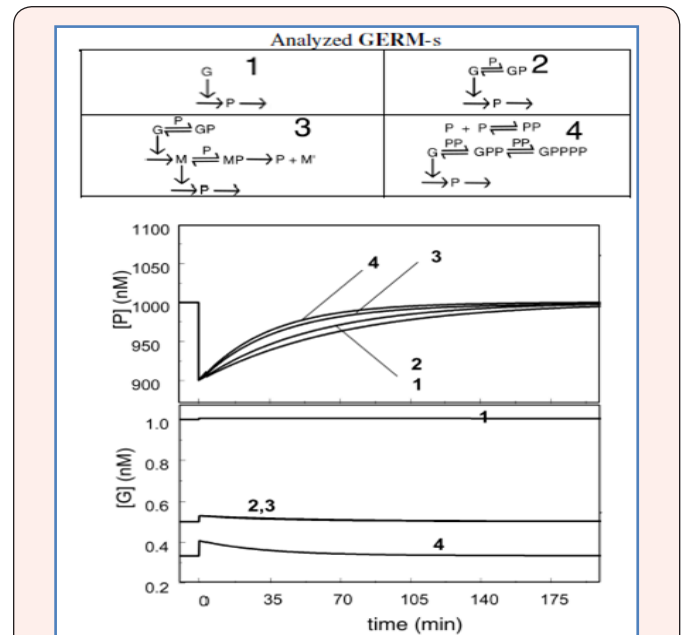

Figure 12

Exemplification of the self- mutual G/P catalysis and regulation after a –10% impulse perturbation in the [P]s = 1000 nM at t= 0 for various types of generic GERM-s. Key-species dynamics is generated by simulations by using the following GERM models (see schemes from the top): 1= [G(PP)0], 2 = [G(P)1], 3 = [G(P)1; M(P)1], 4= [G(PP)2]. Plots adapted by using the examples of Maria [85,87]. The cell nominal (stationary) condition is those of Table 5 (for a high ballast cell, with [G]s = 1 nM).

To exemplify these issues, one considers the same G/P gene expression example with the species homeostatic stationary concentrations given in (Table 5) (high ballast cell case). For comparison, one considers the gene encoding gene G expression by means of GERM-s of various structures given in (Figure 12), that is *[G(P)0]* without mutual catalysis, *[G(P)1]* with mutual catalysis and one buffering reaction, or *[G(PP)2]* with dimeric *TF=PP*, or *[G(P)1 ; M(P)1]]* with mutual catalysis and a cascade control via buffering reactions at the level of G and M. The rate constants have been estimated by solving the stationary form of the GERM model with the stationary concentrations given in

(Table 5) (high ballast cell case). Additionally, the requirement of getting a maximum dynamic responsiveness and efficiency discussed in chapter 5.1.2, leads to adopt *[M]s = [M(P)]s* = 0.5 nM and *[G]s = [G(PP)]s = [G(PPPP)]s* = 1/3 nM. The resulted recovering trajectories of the G and P species after a –10% impulse perturbation in the [P]s = 1000 nM at an arbitrary t=0, are comparatively presented in (Figure 12).

It is to remark that the incomplete *[G(P)0]*, reports the worst dynamic efficiency, with very slow recovering tendencies in (Figure 12). Better performances are reported by *[G(P)1]*. Even better regulatory efficiency is reported by the cascade control of separately considered transcription and translation in the *[G(P)1 ; M(P)1]* module. The best QSS recovering efficiency is reported by the [G(PP)2] module that uses two buffering reactions and a dimeric PP as TF, quickly synthesized in a small amount (of optimal level [PP]s =0.01 nM determined together with the model rate constants to ensure an optimal P.I.).

All the above analyzed GERM-s has been modelled in a VVWC framework (eq. 6-10). For such VVWC kinetic models it is to remark the way by which the variable cell-volume plays an important role to species inter-connectivity (direct or indirect via the cell volume) in the same GERM regulatory module or among linked modules. Even if species connectivity can be expressed in several ways (Vance, et al. [75]; Maria [87]), it is directly dependent on the manner by which species in a GERM or in a GRC recover more or less independently after a perturbation.

Table 6: Comparison of the species recovering rates for two GERM-s used for the G/P pair replication. One GERM is of [G(P)1] type; the other GERM is of [G(P)0] type [the cell nominal conditions are those of the high ballast cell given in the [Table 5], but with [G]s = 1 nM]. Example adapted from Maria [87,90]. The results are commented in the chapter 4.2.2.

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

Notations G= a gene (DNA); P= a protein; M= mRNA; GP= the inactive complex of G with P; NutP and NutG are substrates used in the synthesis of metabolites MetP and MetG used for P and G synthesis; C_j= species "j" concentration; AVG = average of the species recovering times; STD = standard deviation of the species recovering times; NG= negligible.

When the species connectivity increases, they recover with a more comparable rate (or equivalently, over the same time), by 'assisting' each other to cope with a perturbation (see the comparison of species recovering times in Table 6). By contrary, when the species are more disconnected, they recover in a more disparate way, and the GERM presents weaker P.I.-s. (including not only larger species recovery times $(\tau_{rec})_j$, but also larger state sensitivities to external nutrients).

Thus, the mutual autocatalysis appears to *interconnect* the GERM key-components such that they are regulated more as a unit than would otherwise be the case. Interconnectivities (the degree to which a perturbation in one component influences others) may arise from a direct connection between components (e.g. when they are involved in the same chain of reactions), or from an indirect connection (via cell volume changes for an isotonic system).

Our analysis indicates that mutual auto-catalysis is a particularly strong type of interaction that unifies the regulatory response, and they serve to "smooth" the effects of perturbations. It also suggests a way to quantitatively evaluate interconnectivities between all cellular components: each component could be perturbed one at a time, and recovery rates or some other measure of regulatory effectiveness could be evaluated for all components. The resulting relationships thus reflecting the holistic properties of the GRC-s.

5.2.3 The effect of cell system isotonicity

The effect of the isotonicity constraint eqn. (6-11) of a VVWC cell model can be easily proved in the same way as done in the previous chapter 5.2.2. By simulating a GERM with a *[G(P)1]* model type, the effect of applying a –10% impulse perturbation in the key-protein homeostatic level [P]s = 1000 nM at an arbitrary time t=0, the effect on the key-species (G, P) can be observed in the (Figure 11) while species recovering times are given in the (Table 5).

By contrast, in a CVWC cell model formulation, when the isotonicity constraint is missing from the model, the key-species do not recover. By contrast, as revealed by the simulations with the *[G(P)1]*, the system isotonicity imposes relatively short recovering rates for the key-species, and negligible for the other GERM species present in a large amount (lumped nutrients and metabolites). As proved by the examples of chapter 5.2.2., 5.2.4., 5.2.5., the VVWC models, including the "cell ballast" effect, and the G/P mutual autocatalysis, are more flexible and adaptable to environment constructions, being able to better represent the influence of the environmental changes on the cell homeostasis.

5.2.4 The importance of the adjustable regulatory TF-s in a GERM

As proved by the example given in chapter 5.2.2 and by Maria [85,87], dimeric TF-s, such as PP in *[G(PP)n]*, instead of simple P in *[G(P)n]* leads to several conclusions:

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

The dynamic regulatory efficiency (defined in Table 4) decreases in the order:

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

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

The stationary regulatory efficiency decreases in the order:

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

....(33)

5.2.5 The effect of the cell ballast on the GERM efficiency

When constructing more or less simplified VVWC cell models, it is important to know what is the minimum level of simplification to not essentially affect the holistic properties of the cell. This paragraph proves why it is essential to include in a VVWC model the so-called "*cell ballast*", that is the sum of concentrations of all species, which are not accounted in the ODE mass balance of the GRC model. Basically the isotonic constraint imposes all species (individual or lumped) to be accounted in the cell model, because species concentrations and rates are linked through the common cell volume. As proved by the example of chapter 5.2.2., in such VVWC cell model constructions, the recovery rates are properties of all interactions within the system rather than of the individual elements thereof Heinrich & Schuster [33].

However, another important question derived from the isotonicity constraint refers to the degree of importance of the cell content (ballast) for the cell reactions and resistance to perturbations. In other words, the P.I.-s of a GERM are the same in a "rich" cell of high cell content (ballast), compared to those from a "poor" cell of low cell content (ballast)? The answer is no. To simply prove that, one considers a GERM of G(P)1 type placed in an *E. coli* cell with two different nominal conditions given in (Table 5): a high-ballast cell, and a low-ballast cell.

To not complicate these models, the lumped gene and protein metabolites $\sum_j C_{MetG_{j,s}}$, $\sum_j C_{MetP_{j,s}}$ respectively have been considered. Being present in a large amount, these components also play the role of cell ballast, their concentrations being set to values much larger than those of the other cell species (Table 5). Simulations allowed obtaining the species trajectories after a –10% impulse perturbation in the key-protein [P]s of 1000 nM applied at an arbitrary time t=0. These recovering trajectories are presented in (Figure 13).

The selection of appropriate lumped $\sum_j C_{MetP_{j,s}}$ and $\sum_j C_{MetP_{j,s}}$ lumped   will lead to understanding their effect on the cell self-

regulatory properties. Low concentrations relative to the total number of other molecules in the cell afforded shorter $(\tau_{rec})_P$ for the key-protein. For instance, in [G(P)1], with $\Sigma_j C_{MetP_{j,s}} = 3000\,nM$; $\Sigma_j C_{MetG_{j,s}} = 2000\,nM$, and $\Sigma_j C_j = 12001\,nM$, the resulted $(\tau_{rec})_P = 103$ min, and $(\tau_{rec})_G = 223$ min after a –10% impulse perturbation in the [P]s of 1000 nM at an arbitrary t = 0 (Table 5). Whereas for $\Sigma_j C_{MetP_{j,s}} = 3 \times 10^8\,nM$; $\Sigma_j C_{MetG_{j,s}} = 3 \times 10^6\,nM$, and $\Sigma_j C_j = 6.06 \times 10^8\,nM$, the resulted $(\tau_{rec})_P = 133$ min, and $(\tau_{rec})_G = 93$ min after a –10% impulse perturbation in the [P]s of 1000 nM at an arbitrary t=0 (Table5.).

We refer to this as the *Inertial Effect*. It arises because the invariance relationships described above require that larger rate constants for P and G synthesis be used to counterbalance lower [MetP] and [MetG], and these constants are determinants for key-species recovering rates $(\tau_{rec})_j$ after a perturbation. On the other hand, when metabolite concentrations were low, perturbation of cell volume was greater than when they were high (volume increase plots not presented here). The attenuation of perturbation-induced volume changes by large metabolite concentrations is called the *Ballast Effect*. Ballast diminishes the indirect perturbations otherwise seen in concentrations of all cellular components. Thus, [G] was perturbed far less, as a result of an impulse perturbation in [P], for the cell containing higher metabolite concentrations than for that containing lower metabolite concentrations (Figure 13). Thus, increasing metabolite concentrations attenuates the impact of perturbations on all cellular components but negatively influences recovery times.
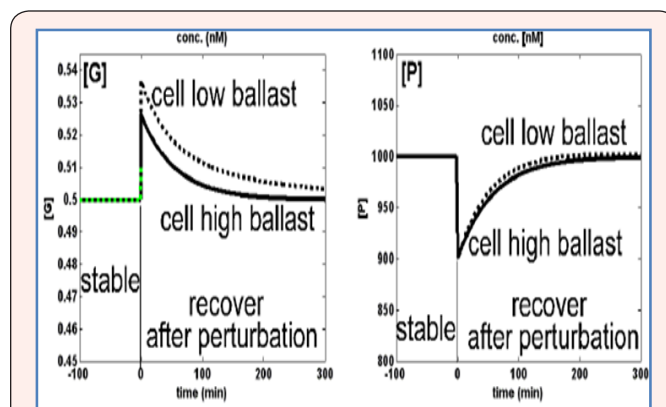
Exemplification of the cell content ballast effect on the species recovering times to homeostasis, in the case of a [G(P)1] gene expression regulatory module from two different *E. coli* cell mutants: one with a high ballast (content), and another with a low ballast. Examples adapted from Maria [89]. Recovery trajectories of the gene G (left) and of its encoding protein P (right) after a –10% impulse perturbation in the [P]s = 1000 nM at the arbitrary time t=0. Solid line trajectories (_____) correspond to a high ballast cell, while the dotted line (•••••) trajectories to a low ballast cell. Simulations are based on the *E. coli* cell species concentrations given in the Table 5.

In fact, the so-called 'ballast effect' shows how all components of the cell are interconnected via volume changes. It represents another holistic property of cells, and it is only evident with only variable-volume VVWC modelling framework. Its importance is related to the magnitude of perturbations and the total number

of species in a cell. For a single perturbation in real cells, the "*Ballast effect*" will be insignificant due to the large number of total intracellular species. However, the sum of all perturbations experienced during a cell cycle might be significant.

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

One important issue to be solved when linking GERM-s to construct a GRC is the degree of detail of the adopted GERM-s to accurately reproduce the GRC regulatory properties. The examples discussed below and by Maria [6] revealed that more important than the number of considered species in the regulatory loops is the selected GERM regulatory scheme, able to render the GRC holistic synchronized response to environmental perturbations. Consequently, when developing a suitable VVWC kinetic model of a GRC, it is important to adopt a suitable reduced model structure by means of an acceptable trade-off model simplification-vs.-model quality (adequacy).

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

To exemplify how a suitable trade-off between GRC model simplicity and its capabilities can be obtained, one considers the problem of adequate and efficient linking of two GERM-s (related to the expression of G1/P1 and G2/P2 pairs) such that the resulted GRC to present optimal P.I.-s. To solve this problem, Maria [6] compared two linking alternatives:

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

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

The GERM linking in $[G1(P1)1] + [G2(P2)1]$ is as following: the expressed P1 in $[G1(P1)1]$ is the metabolase that converts NutG2 in MetG2 and NutP in MetP2 in the $[G2(P2)1]$. In turn, the expressed P2 in $[G2(P2)1]$ is the polymerase that converts MetG1 in G1 in the module $[G1(P1)1] + [G2(P2)1]$.

The GERM linking in $[G1(P1)1 ; M1(P1)1] + [G2(P2)1 ; M2(P2)1]$ is as following: the expressed P1 in $[G1(P1)1 ; M1(P1)1]$ is the metabolase that converts NutG in MetG2 and NutP in MetP2 in the $[G2(P2)1 ; M2(P2)1]$. In turn, the expressed P2 in $[G2(P2)1 ; M2(P2)1]$ is the polymerase that converts MetG1 in G1 in the module $[G1(P1)1 ; M1(P1)1]$.

Tests have been made by using the nominal conditions of (Table 5), the high ballast cell case, with [P1] s = 1000 nM, [P2]s = 100 nM, [G1]s = [G2]s = 0.5 nM, $\sum_j C_{MetP_{j,s}} = 3 \times 10^8 \ nM$; $\sum_j C_{MetG_{j,s}} = 3 \cdot 10^6$ nM. By evaluating various P.I.-s of the GRC, including the two linked GERM-s, the following conclusions are

derived: in spite of a slightly more complex structure (14 vs. 10 individual and lumped components, and two more buffering reactions), the GRC variant B presents much better P.I.-s, that is (values not presented here):

I)   key-species shorter recovering times after an impulse perturbation;

II)   lower AVG and STD species connectivity indices;

III)   species QSS concentrations lower sensitivity vs. environmental perturbations.

Thus, the right choice of the GERM structures in a GRC is an essential modelling step. This example proves how, with the expense of a little increase in the model complexity (4 additional species and 2 buffering reactions), the cascade control of the gene expression in modules of $[G(P)n ; M(P)n']$ type (Figure 6) presents superior regulatory properties suitable for designing robust GRC-s, with easily adjustable properties via model parameters, including a better species synchronization when coping with perturbations (i.e. low AVG, STD indices).

5.2.7 Cooperative vs. concurrent linking of GERM-s in GRC and species interconnectivity

When coupling two or more GERM modules into the same cell, such as the nutrients, and metabolites in the G/P syntheses are roughly the same (Figures 1, 5down). The modelling problem is what alternative should be chosen? A competitive scheme (due to the common substrate, i.e. Met $[G(P)0]$ and Met $[G(P)1]$), or a cooperative scheme, the two GERM-s assisting each other? For exemplification, one considers the problem of adequate and efficient linking of two GERM-s, related to the expression of G1/P1 and G2/P2 pairs. By using simple [G(P)0] or [G(P)1] modules, there are tested three alternatives of module coupling illustrated in (Figure 14), that is:

Variant A: *Competitive* expression (competition on using the common metabolites) of the type $[G1(P1)0] + [G2(P2)0]$;

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

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

The tests were performed by using the VVWC modelling framework, and the nominal high-ballast cell condition of (Table 5). Simulations lead to very interesting conclusions Maria [87]:

A. In Variant A, one links two modules $[G1(P1)0] + [G2(P2)0]$, both ensuring regulation of the two proteins (P1, P2) synthesis, in an *concurential disconnected* way (Figure 14). For this hypothetic system, synthesis of P1/G1 and P2/G2 from metabolites is realized with any interference between modules (C_{P1s} = 1000nM, C_{P2s} = 100nM, C_{G1s} = 1nM, C_{G2s} = 1nM). The only connection is due to the common cell volume to which both protein syntheses contribute. If one checks the system stability, by applying a ±10% impulse perturbation in C_{P1s}, it results an *unstable system*, evolving toward the decline and *disappearance of one of the proteins* (i.e. those presenting the lowest synthesis

rate). Consequently, the homeostasis condition is not fulfilled, the cell functions cannot be maintained, and the disconnected protein synthesis results as an *unfeasible* and *less plausible* GERM linking alternative. "

B. In Variant B, the simple *cooperative* linking of $[G1(P1)0]$ + $[G2(P2)0]$ modules in (Figure 14) ensures specific individual functions of each protein, i.e. P1 lumps both the permeases and metabolases functions, while P2 is a polymerase.
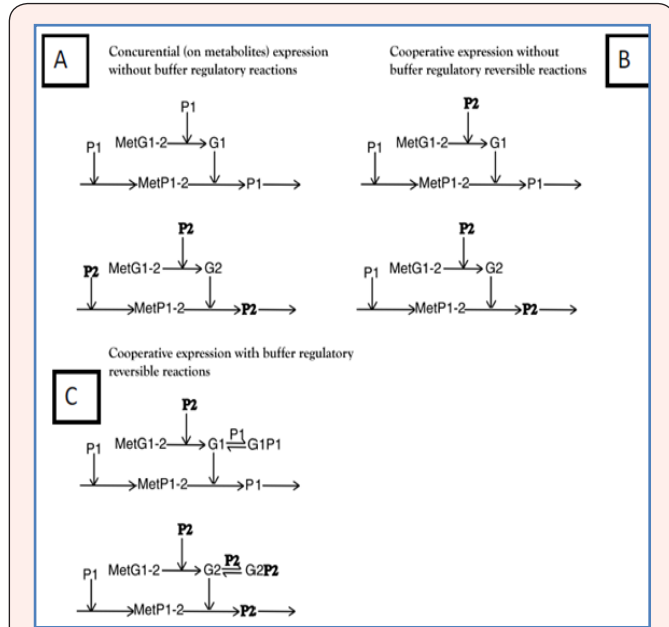

Figure 14

Investigate alternative linking of two GERM-s. The horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicates an assumed concentration invariance of these species. (Alternative A) There is no direct connectivity between the two GERM-s. Competitive linking of [G1(P1)0]+[G2(P2)0] is due to the common metabolites (substrate); (Alternative B) Simple cooperative linking of [G1(P1)0]+[G2(P2)0] modules. P1 is permease and metaboase for both GERM-s; P2 is polymerase for replication of both G1 and G2 genes. (Alternative C) is a cooperative linking with supplementary buffer reversible regulatory reactions to modulate the G1, and G2 catalytic activity in each of the modules [G1(P1)1]+[G2(P2)1]. (Example commented in chapter 5.2.7.). Plots adapted from Maria [87] by courtesy of CABEQ Jl.

C.   In Variant C, the simple *cooperative* linking of $[G1(P1)0]$ + $[G2(P2)0]$ system of Variant B has been *improved* by adding simple effectors for gene activity control. In the cooperatively linked system , thus resulting the system $[G1(P1)1]$ + $[G2(P2)1]$, of (Figure 14), where the effectors P1 and P2 act in two buffering reactions, G1+P1 < ===> G1P1, and G2+P2 < ===> G2P2, respectively, with the stationary states $C_{G1,s}$ = $C_{G1P1,s}$ =1/2 nM, and $C_{G2,s}$ = $C_{G2P2,s}$ =1/2 nM ensuring maximum dynamic P.I.-s.

The same rule of linking GERM-s can continue in the same way, for instance, also involving $[G(PP)n]$ modules, where the effectors being the dimers PP, acting in *n* buffering reactions of the type, G+PP < ===> GPP < ===> ...<===> GPn, with the stationary states $C_{G,s} = C_{GP_1,s} = C_{GP_2,s} = ... = C_{GP_n,s} = 1/(n+1)$ nM. The model rate constants should be estimated from the species stationary concentration vector , and by imposing regulatory optimal characteristics discussed in chapter 4.3.2. Besides, stationary

levels of active and inactive forms of catalyst should be adopted, $C_{L,s} = C_{TF_1,s} = C_{TF_2,s} = ... = C_{TF_n,s} = 1/(n+1)$, as discussed in chapter 4.3.1. Besides, the dissociation constant of the L(TF)n complex in the buffering reactions $k_{diss} \gg D_s$ has been adopted, e.g. $10^5 - 10^7 \cdot D_s$, being much higher than other rate constants of the GERM (Maria 87]). In subsequent works, Maria [6,87,89,90,93] also proved that optimization of the GERM P.I.-s with the criterion (26a-f) leads to small values for $C_{PP,s}$ (i.e. the active parts of dimeric TF-s).

The stability and the dynamic regulatory characteristics of all three GRC systems have been determined by studying the QSS-recover after a ±10% $C_{P1,s}$ impulse perturbation. The results, presented by Maria [87], reveal the following aspects concerning the systems A, B, C:

a. All three systems are stable $\left[\max\left(\text{Re}\left(\lambda_j\right)\right) = -D < 0\right]$ (where λ_j are the eigenvalues of the ODE model Jacobian. Systems B and C recover faster after a dynamic perturbation in C_{P1s} . It results that the *cooperative* module linking is superior to the competitive alternative A, being only one *viable* solution that ensures the system homeostasis. The B and C alternatives are superior because they preserve specific functions of each protein inside the cell. C alternative presents the best P.I.-s from all three checked alternatives due to the additional regulatory effector of type G+P < ===>GP.

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

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

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

In the same way, the regulatory network GRC design procedure can be continued, by accounting for new proteins (and their corresponding GERM-s). For instance, in the simplified representation of (Figure 14), a 3-rd GERM for P3 synthesis can be added to Variant C, by allocating specific functions to the P1, P2, P3 proteins, as follows: P1 and P3 lumps permease and metabolase enzymes, which ensure nutrient import inside the cell, and their transformation in gene-metabolites (MetG1-MetG3) and protein-metabolites (MetP1-MetP3) respectively. Protein P2 lumps polymerases able to catalyze the genes G1, G2, G3 production. If one considers the simplest effector case, the resulted GRC includes three modules $[G1(P1)1]+[G2(P2)1]+[G3(P3)1]$, which regulate the synthesis of P1, P2 and P3, in a *cooperative interconnected* way with preserving the protein individual functions.

5.2.8 The optimal value of TF

It is self-understood that, in a realistic VVWC model, the holistic properties of the cell and of the analysed GRC should be preserved, and modulated via model structure and parameters. One of the cell modelling principles postulates that the concentration of intermediates used in the GRC-s should be maintained at a minimum level to not exhaust the cell resources, but at the same time at an optimal value to maximize the GRC

P.I.-s. Such optimal [TF]s are obtained by optimization eqn. (26). An example was provided by Maria [90] in the case of a genetic switch (GS) in *E. coli* cell, modelled under the VVWC approach. The two considered self- and cross-repressing gene expression modules are of type $[G2(P2P2)I(P3P3)I]+[G3(P3P3)IG3(P2P2)I]$. These well chosen GERM-s allowed adjusting the regulatory properties of the VVWC (i.e. switch certainty, good responsivity to inducers, good dynamic and stationary efficiency). Besides, based on adequate math models, Maria [90] proved that there exists an optimal level of the TF-s [P2P2]s, or [P3P3]s that are associated to the optimal holistic regulatory properties of the GRC (low sensitivity vs. external nutrients, but high vs. inducers), and that these TF-s are rather dimmers than monomeric molecules. These *in-silico* obtained results have been confirmed by the literature data Ferrell [76].

5.2.9 Some rules to be followed when linking GERM-s

Cell GRC-s and, in particular, those involved in some protein synthesis regulation, are poorly understood. The modular approach of studying the regulation path, accounting for its structural and functional organization, seems to be a promising route to be followed. Because a limited number of GERM types exist, individual GERM-s can be separately analysed, as above checked for efficiency in conditions that mimic the stationary and perturbed cell growing conditions. Efficient GERM of regulatory indices of (Table 4) are then linked accordingly to certain rules to mimic the real metabolic process, by ensuring the overall GRC efficiency, system homeostasis, and protein individual functions. Module linking rules are not fully established, but some principles discussed in chapter 5.2 should be respected. The hierarchically organised network includes a large number of compounds with strong interactions inside a module and weaker interactions among modules, so that the whole cell system efficiency can be adjusted. By testing several ways to link GERM-s, Maria [87] advanced some rules:

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

b)   When linking GERM-s, the main questions arise on the connectivity mechanism and on the cooperative vs. uncooperative way in which proteins interact over the parallel/consecutive metabolic path Atkinson, et al. [10]; Wall, et al. [56]; Hlavacek & Savageau [55]; Maria [85,87]. In spite of an apparent 'competition' for nutrient consumption, protein synthesis is a closely cooperative process, due to the specific role and function of each protein inside the cell (see chapter 5.2.7). In a *cooperative linking*, common species (or reactions) are used for a cross-control (or cross-catalysis) of the synthesis reactions. Thus, the system stability is strengthened, while species inter-connectivity is increased leading to a better treatment of perturbations.

c)   Protein interactions are very complex, being part of the cell metabolism and distributed over regulatory network nodes. There are many nodes with few connections among proteins and

a small, but still significant, number of nodes with many proteic interactions. These highly connected nodes tend to be essential to an organism and to evolve relatively slowly. At a higher level, protein interactions can be organized in 'functional modules', which reflect sets of highly interconnected proteins ensuring certain cell functions.

Specific proteins are involved in nutrient permeation (permeases), in metabolite synthesis (metabolases), or in gene production (polymerases). In general, experimental techniques can point-out molecular functions of a large number of proteins, and can identify functional partners over the metabolic pathways. Moreover, protein associations can ensure supplementary cell functions. For instance, enzyme associations (like dimeric or tetrameric TF-s) lead to the well-known 'metabolic channelling' (or tunnelling) process, that ensures an efficient intermediate transfer and metabolite consecutive transformation without any release into the cell bulk phase Maria [87].

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

e)   A natural strategy for building complex and realistic cell models is to analyse independent functional modules or groups of closely interacting cellular components, and then link them. The VVWC approach may facilitate this strategy. Each module could be modelled as a separate "entity" growing at the actual rate of the target cell. The volume of the newborn cell and the environment characteristics could match those of the target. To allow this, and to reproduce the cell ballast effect, lumped molecular species could be defined into each cell where a GERM is tested, in amounts equal to those of the target cell minus those due to the components of the module. Thus, each tested cell carrying a certain defined GERM-s would grow at the same observed rate. As a result, linking GERM-s would be a seamless process requiring only that the ballast level to be kept at its experimental level.

f)   The VVWC modelling approach demonstrates that each and every component of a cell affects, and is affected by, all other cellular components. Indirect interconnectivities arise because all components in a cell contribute to cell volume, and cell volume influences component concentrations. Thus, perturbations in one component reverberate throughout the cell. The importance of these indirect relationships will vary with the diversity and complexity of cellular components. Increasing numbers add ballast to the cell, minimizing these indirect relationships, while increasing diversity allows individual metabolites to be present at lower concentrations, improving dynamic responses of GERM-s and GRC to perturbations. Another issue, thus far unexamined, is how specific types of interconnectivities affect the regulatory behaviour of cells. This could be probed using experimental methods developed by Vance, et al. [75] to deduce connectivities in biochemical pathways from the effects of impulse perturbations.

g) When modelling complex operon structures, simple GERM structures should be adopted to not complicate too much the VVWC model. The default GERM is the $[G1(P1)1]$. But, according to the experimental data and interactions among genes and proteins, more complicated GERM constructions can be elaborated, as those described in chapter 7 given by Maria [84,89,93].

5.2.10 The effect of cascade control on the GERM efficiency

Among GERM-s reviewed and tested under VVWC, the most significant are the [G(P)n] of effectiveness nearly linearly increasing with the number (n) of buffering reactions (chapter 5.2.1). Due to their simple structure such GERM-s are most suitable to construct complex GRC-s (chapters 6-7). On the next place, the [G(PP)n] are also favorites by presenting a more pronounced regulatory efficiency due to the used dimeric TF-s (that is PP). The most effective are the GERM-s with a cascade control of the expression, by means of buffering reactions applied at both gene G, and mRNA (M) catalyst level, that is [G(P)n ; M(P)n'] (Figures 1,4-6). Maria [6,85,87] *in-silico* proved the superiority of the $[G(P)n; M(P)n']$ gene expression structures. The conclusions are the following:

(i) the very rapid buffering reactions, such as

G + P <===> GP+ P <===> GPP ......<===> GPn, or
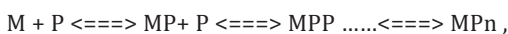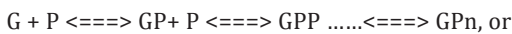
M + P <===> MP+ P <===> MPP ......<===> MPn ,

have been proved to be very effective regulatory elements, by quickly adjusting the active/inactive

G/GP/GPP/GPn or M/MP/MPP/MPn ratios thus efficiently coping with the perturbations (chapters 5.2.1-5.2.2, 5.2.4, 5.2.6, 5.2.8)

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

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

[G(P)n;M(P)m] (n+m regulatory elements), etc.

Roughly, the obtained improvement of the P.I. per regulatory element is of ca. 1.3 (under VVWC modelling; see chapter 5.1.2.3), while the same improvement is of 2.5 under CVWC modelling framework (Maria [6,85,87,89]). It clearly appears that the VVWC modelling framework is more realistic, the default CVWC approach tending to over-estimate the P.I.-s of GERM-s.

5.2.11 Use structured cell models to optimize some cell performances by identifying theoretical gene knockout strategies

Beside simulation of cell GRC-s, another objective of the structured cell simulators is to identify genome modifications leading to the improvement of some optimization objectives.For instance, Maria [95] used a reduced *E. coli* cell kinetic model of 95 reactions and 72 metabolites to *in-silico* determine what genes should be removed (the so-called 'gene knockout' procedure) to realize maximization of both biomass and succinate production. The used multi-objective Pareto-front procedure was applied by also accounting of the stoichiometric constraints. Problem solutions indicated concomitant removal of 2-4 genes (*in-silico* indicated by the used procedure). Due to the very high complexity of the problem (its solution requiring dozens of hours of computing time), and the lack of information, optimization is usually made using a simplified CVWC model.

6. Example of Modeling Bistable Genetic switches of Adjustable Characteristics under a Deterministic VVWC Modeling Approach and Using Multiple Repression Steps

As detailed in the present study, one essential application of the modular GRC-s VVWC models is the simulation of various genetic circuits. Among them, the genetic switches (GS) are particularly attractive because such a toggle-switch realizes the mutual repression control in two gene expression modules that create decision-making branch points between on/off states according to the presence of certain external inducers. In fact, GS re-direct the cell metabolism to better adapt to environmental changes Maria [89,90,93]; Savageau [54,60]; Hlavacek & Savageau [55]; Wall, et al. [56]; Atkinson, et al. [10] presented various principles to construct GS models. However, in contrast to the large number of CVWC models of GS from literature (review of Maria [89,90,93]) that use apparent kinetic constructions (including Hill or power-law type nonlinear induction/ repression models), Maria [89,90,93] used mechanistic VVWC models to adequately model GS-s of adjustable stationary and dynamic P.I.-s, including the switch efficiency (that is GS certainty, QSS stability, GS sensitivity to inducers, response rate, transition time to another QSS) by means of the no. and type of the included GERM effectors, TF-levels, reaction rate orders of self- and cross-repression schemes.

It is to be observed that a VVWC model requires that all cell species (individual or lumped) be considered in the model mass balance (including the lumped genome, proteome, and metabolome), because all these lumps contribute to the volume and dynamics via the isotonicity constraint (6-11). As proved by Maria [89,90,93] such an approach offers more realistic P.I.-s of the GS, in spite of a higher computational effort. The dynamic and stationary P.I.-s are easily adjusted by a suitable selection of the two GERM-s of the GS, and optimization of the GS model parameters such as the no. of buffering reactions, no. of TF per gene operator site, self-repression partial order, TF level and type, Hill-type induction/repression nonlinearity of some lumped steps.

To avoid complicating the GS model, [$G(P)$l] lumped GERM-s were adopted for the genome, and proteome replication in the cell. Approximate concentration of lumped genome and proteome can be easily estimated for a large number of micro-organisms by using –omics databanks such as KEGG [50]. The cell 'ballast' is accounted being included in the lumped metabolome. The scheme of such a GS model includes two [$G(P)$l] lumped GERM-s. To better represent the cross- and self-repression and fast induction of the two genes G2 and G3 expression, some of the rate expressions in the GERM models have been slightly modified that is [P1(P2)$_{-0.5}$(I2)$_{Hill}$;G2(P3)$_{Hill}$] + [P1(P3)$_{-0.5}$(I3)$_{Hill}$;G3(P2)$_{Hill}$], thus pointing the rate expression of the reactions involving each effector. The GS scheme is given in (Figure 15). The model includes only pseudo-elementary reactions, except those for the switch genes G2 & G3 and proteins P2 & P3 production, for which the apparent Hill-type kinetics of Voit, [35] has been adopted, that is:
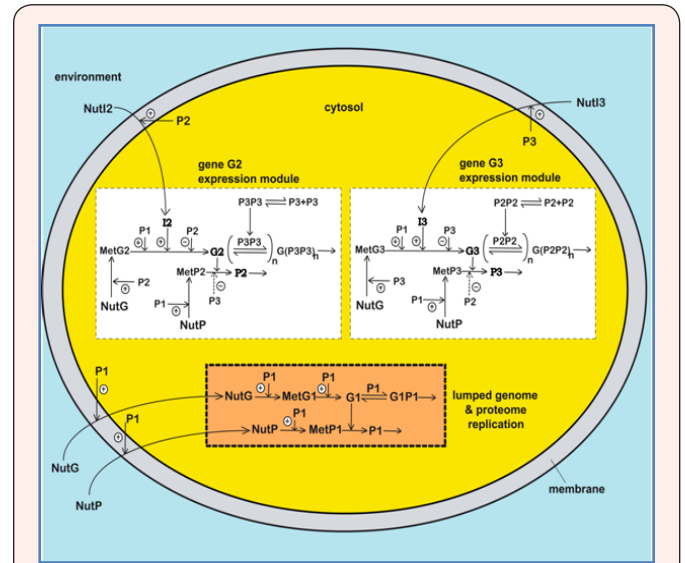

Figure 15

The genetic switch (GS) model including two gene expression (G2 and G3) in two self- and cross-regulated GERM-s. The GRC is placed in a growing *E. coli* cell, by mimicking the homeostasis and cell response to stationary and dynamic perturbations in the environmental species (inducers) NutI2 and NutI3 by using a VVWC model. The cell content ('ballast') influence is mimicked by including the lumped proteome P1 and genome G1 replication module [G1(P1)1]. Notations: G1/P1 = lumped genome/proteome; MetG1/MetP1 = lumped metabolome used for G1/P1 synthesis; P2, P3 = GS target proteins; MetG2, MetG3, MetP2, MetP3 = metabolites used for synthesis of G2, G3, P2, P3, respectively; I2, I3 = internal inducers; NutI2, NutI3 = external stimuli; NutG, and NutP (lumped external nutrients) precursors of (DNA, mRNA), and aminoacids, respectively; ± = positive / negative regulatory loops. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; dash vertical arrows indicate effects included in only some of the tested GS models in chapter 6 [adapted from Maria [93] by courtesy of CABEQ Jl.].

$$\frac{dc_{G2}}{dt} = k_{G2}c_{MetG2}c_{P1}\left(\frac{1+Bc_{I2}^4}{C_1+c_{I2}^4}\right)c_{P2}^{-0.5} - D\,c_{G2} \;\; ; \;\; \frac{dc_{P2}}{dt} = k_{P2}c_{MetP2}c_{G2}\left(\frac{1}{C_2+c_{P3}^2}\right) - D\,c_{P2} \;\; ;$$

$$\frac{dc_{G3}}{dt} = k_{G3}c_{MetG3}c_{P1}\left(\frac{1+Bc_{I3}^4}{C_3+c_{I3}^4}\right)c_{P3}^{-0.5} - D\,c_{G3} \;\; ; \;\; \frac{dc_{P3}}{dt} = k_{P3}c_{MetP3}c_{G3}\left(\frac{1}{C_4+c_{P2}^2}\right) - D\,c_{P3} \;\; ;$$

....(34)

Where G2 & G3 are the expressed genes, P2 & P3 are the corresponding encoding proteins, while I2 & I3 are the inducers activating the expression of G2 & G3 genes. The GS model of best performances includes 4 cross-buffering reactions G2(P3P3)4 and G3(P2P2)4. The derived GS kinetic model accounts for 19 individual and lumped species and includes 21 rate constants identified from solving the stationary model equations (26) with including P.I. optimization.

By considering all cell individual or lumped species in the VVWC model, the dilution rate is uniform for all species, while the degradative steps have been neglected. The gene G2/G3

expression activation accounts for four molecules (n=4) of inducer, allosterically binding to the promoter site, while a slow self-repression with the product is considered of an (adjustable) -0.5 apparent reaction order. The cross-repression of protein P2/P3 synthesis, of Hill type, accounts for only dimeric repressors (n=2) allosterically binding to the catalytic gene, even if a higher control (with tetramers) has been reported Salis & Kaznessis [8]; Kaznessis [9].

A typical simulation of the individual / lumped GS species response in E. coli is plotted in (Figure 16) after a "step" perturbation in the environmental stimulus NutI2 from 0 to 1nM. Predictions are generated by simulation over tenths of cell cycles by using the model with four cross-repressing reactions

of (G2,G3) expression, with optimised nR = 3 self-repression exponent and [TF] = 5nM.

As expected, species present in large amounts (of order 10^5-10^8 nM) display a negligible response to the NutI2 small perturbation (of 1nM), while the cellular species directly connected to the NutI2/I2 inducer pathway are very strongly affected. Even if the plots of (Figure 16) are represented for a large time-scale (thousands of minutes), the transient times for recovering the homeostasis are from the order of minutes up to few cell-cycles (hundreds of minutes), also revealed by Elowitz & Leibler [61] discussion on the transmission of the effect of certain perturbations from cell generation to generation as an expression of cell adaptation to the environment.
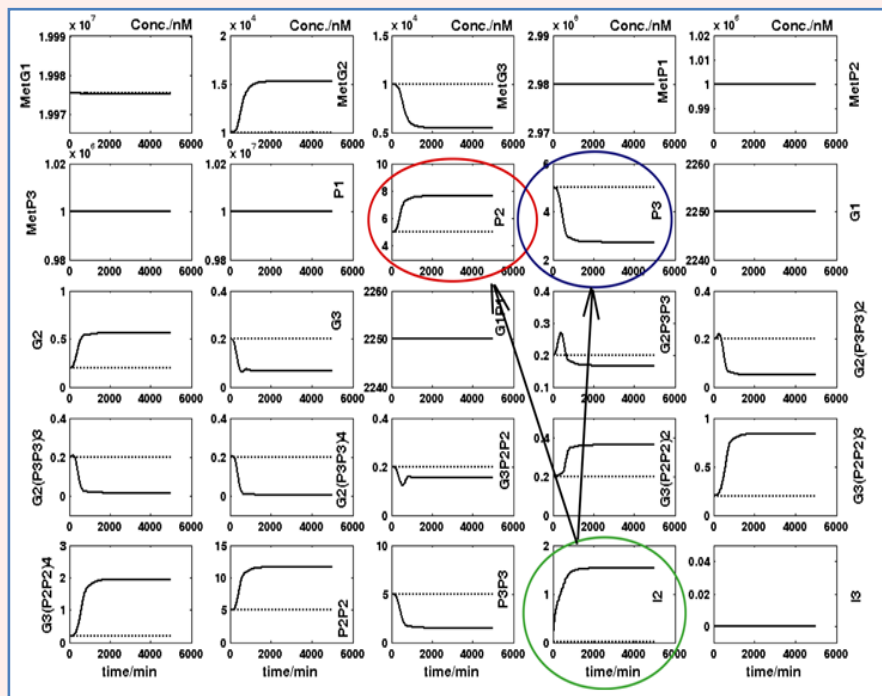

Figure 16

Individual or lumped cellular species dynamics after a "step"-like stationary perturbation in the external stimulus [NutI2] from 0 to 1 nM (applied at time t = 0) leading to a step-like response of the internal inducer I2. Predictions generated by simulations with the genetic switch GS model of Maria [89]. The optimized TF level is of [P2P2]s = [P3P3]s = 5 nM; the model includes 4 buffering reactions [G2(P3P3)4] and [G3(P2P2)4] (see Figure 15) Plots adapted from Maria [93] by courtesy of CABEQ Jl. The step-increase in the I2 inducer quickly leads to a step-like over-expression of P2 protein and repression of P3 protein. Evolution of the involved species in the regulatory loops is similar.

As discussed by Maria [89,93] such rather mechanistic VVWC models with explicit buffering reactions seem to be more flexible and versatile, offering better predictions of the regulatory system P.I.-s.

7. A whole-cell Model to Simulate Mercuric ion Reduction by *E. coli* Under Stationary or Perturbed Conditions

One worthy example of applying VVWC models to adequately represent complex modular GRC-s, is the structured model proposed by Maria [84,91,92] that reproduces the dynamics of the mer-operon expression in Gram-negative bacteria (*E. coli, Pseudomonas sp.*) to uptake the mercury ions from wastewaters under various environmental conditions. The model was constructed and validated by using the Philippidis, et al. [77,78,79] experimental data, and the Barkay, et al. [80] information on *mer*-operon characteristics.

Bacteria resistance to mercury is one of the most studied metallic-ion uptake and release process Barkay et al. [80] due to

its immediate large-scale application for mercury removal from industrial wastewaters Wagner-Döbler, et al. [81]; Deckwer, et al. [82]. The bacteria response to the presence of toxic mercuric ions in the environment is apparently surprising; instead of building carbon- and energy-intensive disposal "devices" into the cell (like chelate-compounds) to "neutralize" the cytosolic mercury and thus maintaining a tolerable level, a simpler and more efficient defending system is used. The metallic ions are catalytically reduced to the volatile metal, less toxic and easily removable from the cell by simple membranar diffusion Allen et al. [83]; Maria [92].
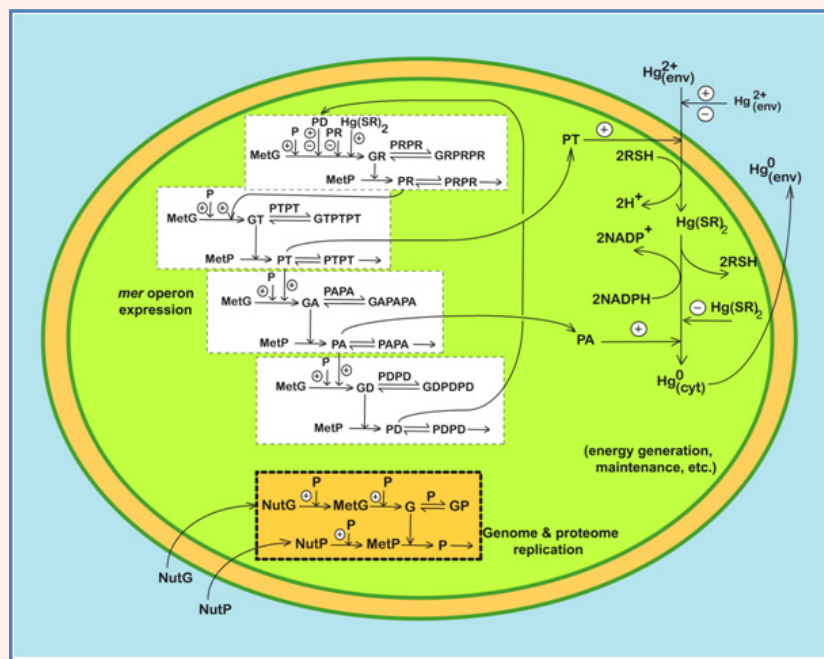
The whole-cell model of Maria [84,92] in the VVWC approach used to simulate the reduction of Hg2+ ions from environment to volatile Hg0 in *E. coli* bacteria. The simplified reaction path includes: Two modules for mediated transport of Hg2+ into cytosol (catalysed by PT) and its reduction (catalysed by PA); Five regulatory modules of *mer*-operon expression including successive synthesis of PR (the transcriptional activator of other protein synthesis), lumped PT permease, PA reductase, and of the control protein PD; One module for the lumped proteome P and genome G replication of [G(P)1] type. The regulatory system is placed in a growing cell, by mimicking the homeostasis and cell response to stationary and dynamic perturbations in the environmental [Hg2+]. The reductant NADPH and RSH are considered in excess into the cell. Figure adapted from Maria [91,92]. Notations: P= lumped proteome; G= lumped genome; NutG, NutP = lumped nutrients used for gene and protein synthesis; MetG/MetP = lumped metabolome (DNA or protein precursor); P• = proteins; G• = genes; RSH= low molecular mass cytosolic thiol redox buffer (such as glutathione); perpendicular arrows on the reaction path indicate the catalytic activation, repressing or inhibition actions; absence of a substrate or product indicates an assumed concentration invariance of these species; positive or negative feedback regulatory loops.

Such a process involves less cell resources and is favoured by the large content (millimolar concentrations) of low molecular-mass thiol redox buffers (RSH) able to bond and transport Hg^{2+} in cytosol, and of renewable NAD(P)H reductants able to convert it into neutral metal. A genetic regulatory circuit responsible for the *mer*-operon expression controls the whole process, by including 4 lumped genes (denoted by GR,GT,GA,GD in Figure 17) of individual expression levels induced and adjusted according to the level of mercury and other metabolites into cytosol. The whole process is tightly cross- and self-regulated to hinder the import of large amounts of mercury into the cell, which eventually might lead to the blockage of cell resources (RSH,

NADPH, metabolites, proteins), thus compromising the whole cell metabolism. The GRC model includes four GERM-s of simple but effective type as follows:

A. GERM to regulate the Hg^{2+} transport across the cellular membrane, mediated by three proteins (PmerP, PmerT, and PmerC) from the periplasmatic space, considered as a lumped permease PT in the model. Phillippidis et al. [78,80] found this transport step as being energy-dependent and the rate-determining step for the whole mercury uptake process. Once the mercuric ion complex arrives in the cytosol, thiol redox buffers (such as glutathione of millimolar concentrations) form

a dithiol derivative Hg $(SR)_2$. The GT lumped gene expression to get PT is induced by the regulatory protein PR and 'smoothed' by the 'ballast' effect of proteome P.

B. GERM to control the expression of the PR protein that induces and controls the whole *mer*-operon expression in the presence of cytosolic Hg^{2+} (even in nM concentrations). This GERM acts as an amplifier of the *mer*-expression leading to a quick (over ca. 30 s) cell response and mer-enzyme production. The GR gene expression to get PR is controlled by the protein PD, present in small amounts in the cell.

C. GERM to control the expression of PA enzyme responsible for the Hg $(SR)_2$ reduction to metallic mercury (relatively non-toxic for the cell), easily removable through membranar diffusion to be later removed from the bulk liquid phase by the continuously sparged air. The encoding GA gene expression is induced and controlled by the PT protein.

D. GERM controlling the protein PD synthesis. This protein has a complex role, by maintaining a certain level of GR expression even when mercury is absent in the cytosol Barkay, et al. [80].

E. GERM controlling the replication of the lumped cell proteome (P) and genome (G) (of concentrations 107 nM and 4500nM respectively in immobilized *E. coli* cells), thus mimicking the cell 'ballast' effect on the other cell expressions and reactions. The advantage of including the cell content in the VVWC model is the possibility to reproduce the smoothing effect of perturbations leading to more realistic transient times (*compared* to a cell with a 'sparing' content), the synchronized response to certain inducers, and the 'secondary perturbation' effect transmitted via the cell volume to which all cell components contribute (see the discussion of chapter 5.2.2, 5.2.5).

In total, the *mer*-operon expression GRC model includes 26 individual or lumped cellular species and 33 reactions. All reactions are considered elementary, excepting some of them for which extended experimental information exists, that is Maria [84,91,92]:

a. A Michaelis-Menten rate expression for the mercuric ion membrane permeation into the cell.

b. A Michaelis-Menten rate expression for the mercuric ion reduction in cytosol.

c. A Hill type quick induction of the GR expression that can rapidly initiate the production of permease PT (through the control protein PR) when mercuric ions are present in large amounts.

Dimerization reactions of TF-s are considered to be much rapid than the enzyme synthesis, while equal concentrations of active (G) and inactive (GPP) forms of the generic gene G are considered at homeostasis to maximize the GERM efficiency (eq.26e, cap. 5.1.2,4.3.1).

Lumped proteome P, present in a large amount, is included in all gene expression rates, thus leading to more realistic GERM efficiency indices (Maria [89,90]).The model rate constants are estimated from solving the cell stationary mass balances for nominal concentrations of observable species, but also from optimizing the GERM regulatory indices (e.g. adjust the optimum TF level of gene expression to get the minimum recovering times after a 10% dynamic perturbation in the key species, and smallest sensitivities of the homeostatic levels vs. external perturbations). Exceptions are the Michaelis-Menten rate constants for the mercury transport and reduction in cytosol adopted from the Phillipidis et al. [78-79]. kinetic data for various amount of Gmer plasmids in the cloned *E. coli* cells.

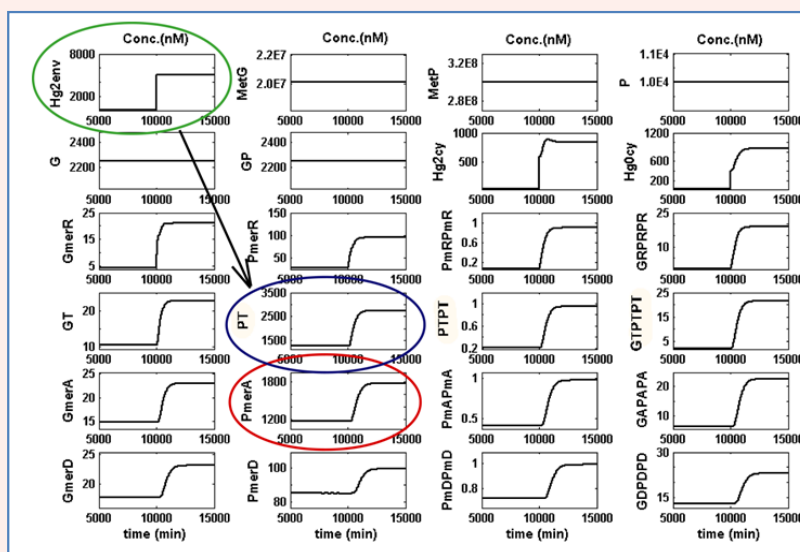This structured VVWC model displays a large number of advantages, being able to:

Figure 18

Typical evolution of cell species concentrations predicted by the cell model of Figure 17 after a step-perturbation of [Hgenv2+] is in the environment from = 0.1 μM to 5μM. Simulations correspond to a cloned *E. coli* cell with [plGmer]= 140nM, [TF]s = 1nM. Plots adapted from Maria [92]. The step-increase in the environmental [Hgenv2+]$_s$ and cytosolic [Hgcy2+]s inducer quickly leads to a step-like over-expression of the lumped PT permease and PmerA (PA) reductase.

I. Simulate the cell metabolism adaptation when the environmental mercury level changes. Such a reconfiguration of the levels of *mer*-genes and *mer*-proteins is presented in (Figure 18) as a step response of a cloned *E. coli* cell with *mer*-plasmids of [plGmer]= 140nM, after a step-perturbation in the environmental mercury level from $[Hg_{env}^{2+}]_s$ = 0.1 µM to 5µ M. The transient state toward the cell new homeostasis of adapted *mer*-gene/protein levels stretches over 15-20 cell cycles (of ca. 0.5 h each) as long as the environment perturbation is maintained.

II. Because the Hg^{2+} reduction rate constants are dependent on the *mer*-plasmid level, the cell model can predict the maximum level of *mer*-plasmids that can be added to the cell for improving its mercury up taking rate without exhausting the internal cell resources thus putting in danger the cell survival.

III. By coupling the structured cell model with the three-phase continuous bioreactor model (with immobilized *E. coli* cells on alginate beads), Maria & Luta [84] were able to determine optimal operating policies of the bioreactor in relationship to the culture of cloned cells.

8. Conclusions

As revealed by examples discussed in this study, there are important issues to be considered when developing modular VVWC models of GRC-s including a variable number of GERM-s.

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

In fact, such cell process models 'translate' from the 'language' of molecular biology to that of mechanistic chemistry and mathematics/computing languages, trying to preserve the structural, functional, and timing hierarchy of the cell components and functions. To avoid much extended ODE cell kinetic models, difficult to be identified, and to be used, the model structure should ensure a satisfactory trade-off between model simplicity and its predictive quality. The derived performance indices P.I.-s of the GERM-s and GRC-s under a VVWC formulation present more realistic estimates comparatively with those derived from the classical CVWC kinetic models which either over-estimate these P.I.-s or are unable to predict them accurately.

The VVWC cell model formulation also considers the genome and proteome replication (i.e. the cell "ballast"), of high influence on the cell metabolism, on the volume increase, and on the treatment of perturbations, in a simple way by mean of a lumped GERM. Consideration of the lumped proteome and genome replication in all VVWC cell models is mandatory in order to fulfil the isotonicity constraint. In such a VVWC model formulation, all cell species should be considered (individually or lumped), because all species net reaction rates contribute to the cell volume increase [eq. (9-10)]. Some examples presented in chapters 6-7 proved the feasibility of such modular constructions when modelling complex GRC-s (such as genetic switches, or genetic amplifiers).

The study also proves the importance of using a VVWC modelling environment instead of the classical CVWC to get more realistic simulation results, and more realistic estimates of the GRC, GERM quantitative regulatory indices, such as: dynamic regulatory efficiency, stationary regulatory efficiency, cascade control of the expression, the influence of the number and type of regulatory effectors, species connectivity, optimal level of TF species, Such VVWC cell simulators become more and more valuable tools in designing GMO with desirable characteristics, or for obtaining micro-organisms cloned with desirable plasmids with important applications in industry (new biotechnological processes, optimization of bioreactors, production of vaccines), or in medicine.

9. Acknowledgment

1.

10. References

1. Stelling J (2004) Mathematical models in microbial systems biology. Current Opinion in Microbiology 7(5): 513-518.

2. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420(6912): 190-193.

3. Crampin E J, Schnell S (2004) New approaches to modelling and analysis of biochemical reactions, pathways and networks. Progress in Biophysics & Molecular Biology 86(2004): 1-4.

4. Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. Eur J Biochem 267(6): 1583-1588.

5. Sewell C, Morgan J, Lindahl P (2002) Analysis of protein regulatory mechanisms in perturbed environments at steady state. J theor Biol 215(2): 151-167.

6. 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.

7. Heinemann M, Panke S (2006) Synthetic Biology-putting engineering into biology. Bioinformatics 22(22): 2790-2799.

8. Salis H, Kaznessis Y (2005) Numerical simulation of stochastic gene circuits. Computers & Chemical Engineering 29(3): 577-588.

9. Kaznessis Y N (2006) Multi-scale models for gene network engineering. Chemical Engineering Science 61: 940-953.

10. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli Cell 113(5): 597-607.

11. Klipp E, Nordlander B, Krüger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. Nature Biotechnology 23: 975-982.

12. Chen M T, Weiss R (2005) Artificial cell-cell communication in yeast Saccharomyces cerevisiae using signalling elements from Arabidopsis thaliana. Nat Biotechnol 23: 1551-1555.

13. Tian T, Burrage K (2006) Stochastic models for regulatory networks of the genetic toggle switch. Proc. Natl Acad Sci USA 103(22): 8372-8377.

14. Sotiropoulos V, Kaznessis YN (2007) Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. BMC Syst Biol 9 (2007): 1-7.

15. Tomshine J, Kaznessis YN (2006) Optimization of a stochastically simulated gene network model via simulated annealing. Biophys J 91(9): 3196-3205.

16. Zhu R, Ribeiro AS, Salahub D, Kauffman SA (2007) Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models. J Theor Biol 246(4): 725-45.

17. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, et al. (2000) Molecular cell biology. Freeman & Co, New York, USA.

18. Sauro HM, Kholodenko BN (2004) Quantitative analysis of signaling networks. Prog Biophys Mol Biol 86(1): 5-43.

19. Kobayashi H, Kaern M, Araki M, Chung K, Gardner, et al. (2004) Programmable cells: Interfacing natural and engineered gene networks. Proc. Natl Acad Sci 101(22): 8414-8419.

20. Bailey JE (1991) Toward a science of metabolic engineering. Science 252(5013): 1668-1675.

21. Nielsen J (1998) Metabolic engineering, techniques for analysis of targets for genetic manipulations. Biotechnol Bioeng 58(2-3): 125-132.

22. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Metabolic Engineering. Principles and Methodologies. Academic Press, San Diego, CA, pp.707.

23. Visser D, Schmid JW, Mauch K, Reuss M, Heijnen JJ (2004) Optimal re-design of primary metabolism in Escherichia coli using linlog kinetics. Metabolic Engineering 6(4): 378-390.

24. Kholodenko B N, Kiyatkin A, Bruggeman F J, Sontag E, Westerhoff H V, et al. (2002) Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. Proc Natl Acad Sci 99(20): 12841-12846.

25. Morgan JJ, Surovtsev IV, Lindahl PA (2004) A framework for whole-cell mathematical modelling. Jl. theor. Biology 231(4): 581-596.

26. Yang Q, Lindahl P, Morgan J (2003) Dynamic responses of protein homeostatic regulatory mechanisms to perturbations from steady state. J theor Biol 222(4): 407-423.

27. Tomita M, Hashimoto K, Takahashi K, Shimizu T, Matsuzaki Y, et al. (1999) E-Cell: Software environment for whole cell simulation. Bioinformatics 15(1): 72-84.

28. Tomita M. (2001) Whole-cell simulation: a grand challenge of the 21st century. Trends in Biotechnology 19: 205-210.

29. Bruggeman FJ, Westerhoff HV (2007) The nature of systems biology. Trends in Microbiology 15(1): 45-50.

30. Westerhoff HV (2006) Engineering life processes live: the Silicon cell. ESCAPE-16 Conference, Garmisch-Partenkirchen, Germany.

31. Styczynski MP, Stephanopoulos G (2005) Overview of computational methods for the inference of gene regulatory networks. Computers & Chemical Engineering 29(3): 519-534.

32. Shifton DC (2007) An introduction to metabolic control analysis. course at Univ. of Stellenbosch (South Africa). p.1-9.

33. Heinrich R, Schuster S (1996) The regulation of cellular systems. Chapman & Hall, London, UK.

34. Benner SA, Sismour AM (2005) Synthetic biology. Nature Reviews Genetics 6: 533-543.

35. Voit EO (2005) Smooth bistable S-systems. IEE Proc Syst Biol 152(4): 207-213.

36. Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. Nature 438(7067): 443-448.

37. Dueber J E, Yeh B J, Chak K, Lim W A (2003) Reprogramming control of an allosteric signaling switch through modular recombination. Science 301(5641): 1904-1908.

38. Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. Mol Syst Biol 1: 2005.0O18.

39. Lebar T, Bezeliak U, Golob A, Jerala M, Kadunc L et al. (2014) A bistable genetic switch based on designable DNA-binding domains. Nature communications 5007(2014).

40. Selvarasu S, Lee DY, Karimi IA (2007) Identifying synergistically switching pathways for multi-product strain improvement using multiobjective flux balance analysis, Proc. In: 17th European Symposium on Computer Aided Process Engineesing - ESCAPE17 (Plesu V, Agachi PS, Eds.), Elsevier, Amsterdam. Paper #T4-378. Also in: Computer Aided Chemical Engineering 24: 1007-1012.

41. Banga J (2008) Optimization in computational systems biology. BMC Systems Biology 2: 47.

42. Haraldsdottir H S, Thiele I, Fleming RMT (2012) Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. Biophys J 102(8): 1703–1711.

43. Zhu Y, Song J, Xu Z, Sun J, Zhang Y, et al. (2013) Development of thermodynamic optimum searching (TOS) to improve the prediction accuracy of flux balance analysis. Biotechnology and Bioengineering 110(3): 914-923.

44. Axe DD, Bailey JE (1994) Modeling the regulation of bacterial genes producing proteins that strongly influence growth Biotech. Bioengineering 43(3): 242-257.

45. Aris, R (1969) Elementary chemical reactor analysis, Prentice-Hall, New Jersey, USA.

46. Grainger J N R, Gaffney PE, West TT (1968) J Theor Biol 21: 123-130.

47. Kubitschek HE (1990) Cell volume increase in Escherichia coli after shifts to richer media. Journal of Bacteriology 172(1): 94-101.

48. Wall work SC, Grant DJW (1977) Physical Chemistry. Longman, London, UK.

49. Eco Cyc (2005) Encyclopedia of Escherichia coli K-12 genes and metabolism, SRI Intl., The Institute for Genomic Research, Univ. of California at San Diego, USA.

50. KEGG (2011) Kyoto encyclopedia of genes and genomes, Kanehisa Laboratories, Bioinformatics Center of Kyoto University, Japan.

51. Kinoshita A, Nakayama Y, Tomita M (2001) ICSB-2001, 2nd Int. Conf on Systems Biology. California Institute of Technology, Pasadena (Ca.), USA, 4-7.

52. Gibson MA, Bruck J (2001) A probabilistic model of a prokaryotic gene and its regulation, In: Computational modeling of genetic and biochemical networks, Cambridge (Mass.), MIT Press, USA.

53. Maria G (2004) A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. Chemical and Biochemical Engineering Quarterly 18(3): 195-222.

54. Savageau MA (2002) Alternatives designs for a genetic switch: Analysis of switching times using the piecewise power-law representation. Math Biosciences 180: 237-253.

55. Hlavacek WS, Savageau MA (1997) Completely uncoupled and perfectly coupled gene expression in repressible systems. J Mol Biol 266(3): 538-558.

56. Wall ME, Hlavacek WS, Savageau MA (2003) Design principles for regulator gene expression in a repressible gene circuit. J Mol Biol 332(4): 861-876.

57. Salvador A, Savageau MA (2003) Quantitative evolutionary design of glucose 6-phosphate dehydrogenase expression in human erythrocytes. Proc Natl Acad Sci 100(24): 14463-14468.

58. Sorribas A, Savageau MA (1989) A comparison of variant theories of intact biochemical systems. I. Enzyme-enzyme interactions and biochemical systems theory. Metabolic control theories. Math Biosci 94: 161-238.

59. Gabaldon T, Huynen MA (2004) Prediction of protein function and pathways in the genome era. Cell Mol Life Sci 61(7-8): 930-944.

60. Savageau MA, Voit EO (1987) Recasting nonlinear differential equations as S-systems: A canonical nonlinear form. Math Biosciences 87(1): 83-115.

61. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403: 335-338.

62. Speck C, Weigel C, Messer W. (1999) ATP-and ADP-dnaA protein, a molecular switch in gene regulation. EMBO J 18(21): 6169-6176.

63. Chen D, Feng J, Kruger R, Urh M, Inman R B, et al. (1998) Replication of R6K Gamma origin in vitro: Discrete start sites for DNA synthesis dependent on pi and its copy-up variants. J Mol Biol 282(4): 775-787.

64. Chen D, Feng J, Kruger R, Urh M, Inman R B, et al. (1998) Replication of R6K Gamma origin in vitro: Discrete start sites for DNA synthesis dependent on pi and its copy-up variants. J Mol Biol 282(4): 775-787.

65. Hargrove JL, Schmidt FH (1989) The role of mRNA and protein stability in gene expression, FASEB Jl 3(12): 2360-2370.

66. Kholodenko B N, Schuster S, Garcia J, Westerhoff HV, Cascante M (1998) Control analysis of metabolic systems involving quasi-equilibrium reactions. Biochimica Biophysica Acta 1379(3): 337-352.

67. Van Someren EP, Wessels LFA, Backer E, Reinders MJT (2003) Multi-criterion optimization for genetic network modeling. Signal Processing 83(2003): 763-775.

68. Kholodenko BN, Kiyatkin A, Bruggeman F J, Sontag E, Westerhoff H V, et al. (2002) Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. Proc Natl Acad Sci 99(20): 12841-12846.

69. Torres NV, Voit EO (2002) Pathway analysis and optimization in metabolic engineering, Cambridge University Press, Cambridge (MS), USA, pp. 299.

70. Brazhnik P, de la Fuente A, Mendes P (2002) Gene networks: how to put the function in genomics. Trends in Biotechnology 20(11): 467-472.

71. Cornish-Bowden A (2016) Biochemical evolution-The pursuit of perfection, Garland Science, New York.

72. Varma A, Morbidelli M, Wu H (1999) Parametric sensitivity in chemical systems. Cambridge University Press, Cambridge (MS), USA.

73. Liao JC, Lightfoot EN Jr (1987) Extending the quasi-steady state concept to analysis of metabolic networks. J Theoretical Biol 126(3): 253-273.

74. Hofmeyr J H S (2001) Metabolic control analysis in a nutshell, Proc. 2-nd Intl. Conf. on Systems Biology, Madison, WI.

75. Vance W, Arkin A, Ross J (2002) Determination of causal connectivities of species in reaction networks. Proceedings of the National Academy of Science 99(9): 5816-5821.

76. Ferrell JE (1997) How responses get more switch-like as you move down a protein kinase cascade. Trends Biochem Sci 22(8): 288-289.

77. Philippidis GP, Malmberg LH, WS HU, Schottel, JL (1991) Effect of gene amplification on mercuric ion reduction activity of Escherichia coli. Applied & Environmental Microbiology 57(12): 3558-3564.

78. Philippidis GP, Schottel JL, WS Hu (1991) Mathematical modelling and optimization of complex biocatalysis. In: A case study of mercuric reduction by Escherichia coli. Expression systems & Processes for DNA Products. pp. 35-49.

79. Philippidis GP, Schottel JL, Hu WS (1991) A model for mercuric ion reduction in recombinant Escherichia coli. Biotech Bioeng 37(1): 47-54.

80. Barkay T, Miller SM, Summers AO (2003) Bacterial mercury resistance from atoms to ecosystems. FEMS Microbiology Reviews 27(2-3): 355-384.

81. Wagner-Döbler I, von Canstein H, Li Y, Timmis K, Deckwer W.D. (2000) Removal of mercury from chemical wastewater by microorganisms in technical scale. Env Sci & Technol 34(21): 4628-4634.

82. Deckwer W D, Becker F U, Ledakowicz S, Wagner-Döbler I (2004) Microbial removal of ionic mercury in a three-phase fluidised bed reactor. Environ Sci Technol 38(6): 1858-1865.

83. Allen, George C , Kornberg A (1991) The priB Gene Encoding the Primosomal Replication n Protein of Escherichia coli* Jl. Biological Chemistry 266(18): 11610-11613.

84. 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.

85. 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.

86. Maria G (2005) Relations between apparent and intrinsic kinetics of programmable drug release in human plasma. Chemical Engineering Science 60(6): 1709-1723.

87. Maria G (2005) Modular-based modelling of protein synthesis regulation. Chemical and Biochemical Engineering Quarterly 19(3): 213-233.

88. Maria G, Szedlacsek S, Laslo AC, Zaharia E (2005) Modular-based simulation of regulated protein synthesis. 2nd Croatian-German Conference on Enzyme Reaction Engineering, 21-24 Sept, Dubrovnik (Croatia), Europe.

89. Maria G (2007) Modelling bistable genetic regulatory circuits under variable volume framework. Chemical and Biochemical Engineering Quarterly 21(4): 417-434.

90. Maria G (2009) lumped Dynamic models for a bistable genetic regulatory circuit within a whole-cell modelling framework. Asia-Pacific Journal of Chemical Engineering 4(6): 916-928.

91. 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.

92. Maria G (2010) A dynamic model to simulate the genetic regulatory circuit controlling the mercury ion uptake by E. coli cells. Revista de Chimie 61(2): 172-186.

93. Maria G (2014) Extended repression mechanisms in modelling bistable genetic switches of adjustable characteristics within a variable cell volume modelling framework. Chemical and Biochemical Engineering Quarterly 28(1): 35-51.

94. Maria G (2014) Insilico derivation of a reduced kinetic model for stationary or oscillating glycolysis in Escherichia coli bacterium. Chemical and Biochemical Engineering Quarterly, 28(4): 509-529.

95. Maria G, Xu Z, Sun J (2011) Investigating alternatives to in-silico find optimal fluxes and theoretical gene knockout strategies for E. coli cell. Chemical & Biochemical Engineering Quarterly 25(4): 403-424.

96. Ptashne M (1992) A genetic switch: Phage lambda and higher organisms. Cell Press & Blackwell Scientific Publs, Cambridge (Mass.), USA, pp. 82.

97. Myers CJ (2009) Engineering genetic circuits, Chapman & Hall/CRC Press, Boca Raton, USA.