



**Hybrid dynamic models linking cell-scale structured CCM pathways, genetic regulatory circuits GRC, and bioreactor state variables. Applications for solving bioengineering and bioinformatics problems**

**Hybrid dynamic models linking cell-scale structured CCM pathways, genetic regulatory circuits GRC, and bio-reactor state variables. Applications for solving bioengineering and bioinformatics problems**



**Gheorghe Maria\***

Department of Chemical & Biochemical Engineering Polytechnic University of Bucharest (Romania)

Corresponding member of the Romanian Academy (RA)

Head Chemical & Biochemical Eng. Commission of RA

[https://new.acad.ro/acad\\_membri/membri/Maria\\_Gheorghe.html](https://new.acad.ro/acad_membri/membri/Maria_Gheorghe.html)

**\*Corresponding Author**

**Gheorghe Maria**

Current Trends in Biomedical Engineering & Biosciences

Email: [gmaria99m@hotmail.com](mailto:gmaria99m@hotmail.com); [gmaria99m@yahoo.co.uk](mailto:gmaria99m@yahoo.co.uk);

[https://en.wikipedia.org/wiki/Gheorghe\\_Maria](https://en.wikipedia.org/wiki/Gheorghe_Maria)

**Published By**

Juniper publishers Inc.

United States

July 05, 2023

## List of Contents

1. Abstract
2. Keywords
3. Symbols used in the chap. 4
4. Introduction
- 1.1. General concepts
- 1.2. Modelling GRC under a WCVV framework
  - 1.2.1. The WCVV vs. WCCV when modelling the GRC dynamics in living cells
    - 1.2.1.1. WCCV formulation
    - 1.2.1.2. WCVV formulation and WCVV deterministic model hypotheses
    - 1.2.1.3. Advantages of using the WCVV kinetic modelling framework in living cells
    - 1.2.1.4. Lumping deterministic kinetic models
- 1.3. Modelling individual GERM-s under WCVV formulation. Rate constant estimation in WCVV models of GERM-s and GRC-s
- 1.4. GERM-s regulatory performance indices (P.I.)
- 1.5. Some rules to link GERM-s when modelling GRC-s
- 1.6. The book main objectives
5. Case study no. 1: The use of a hybrid WCVV-GRC structured kinetic model to optimize a SCR-TPFB bioreactor used for mercury uptake from wastewaters by immobilized *E. coli* cells cloned with *mer*-plasmids.
  - 1.1. Symbols used in the chap. 5
  - 1.2. The chapter purpose
  - 1.3. Mercury ion reduction in bacteria cells – the apparent kinetics
  - 1.4. The TPFB bioreactor model and its nominal operating conditions
  - 1.5. *E. coli* cell structured dynamic model for mercury uptake, and its use to optimize the TPFB bioreactor
    - 1.5.1. Generalities on the proposed extended HSMDM model
    - 1.5.2. Extended HSMDM model structure
    - 1.5.3. Fitting the extended HSMDM model parameters from experimental data
    - 1.5.4. Prediction of the bioreactor dynamics by using the extended HSMDM model compared to the unstructured model
    - 1.5.5. Optimization of the bioreactor dynamics by using a reduced unstructured hybrid model
    - 1.5.6. Parametric sensitivity analysis of the TPFB-2
5. Case study no. 2: the use of a hybrid modular CCM cell-scale structured kinetic model coupled with a FBR classical dynamic model (including macroscale state variables) to maximize the TRP production
  - 5.1. Symbols used in the chap. 6
  - 5.2. Bioprocess generalities
  - 5.3. The used *E. coli* GMO strain
  - 5.4. Experimental FBR bioreactor and the recorded kinetic data

- 5.5. The structured dynamic HSMDM model for the TRP production in a FBR
    - 5.5.1. The biomass [X] growth
    - 5.5.2. Module [A] – Glycolysis
    - 5.5.3. Module [B] - the ATP recovery system
    - 5.5.4. Module [C] - TRP operon expression system
    - 5.5.5. The FBR dynamic model
    - 5.5.6. Rate constants estimation for the HSMDM hybrid model
    - 5.5.7. Ways to intensify the TRP production in the FBR
  - 5.6. The fed-batch bioreactor (FBR) optimization problem
    - 5.6.1. Problem preliminary examination
    - 5.6.2. Formulation of the optimization problem
      - 5.6.2.1. Selection of the FBR control variables
      - 5.6.2.2. FBR optimization - objective function ( $\Omega$ ) choice
      - 5.6.2.3. Optimization problem constraints
      - 5.6.2.4. Selecting the number of time-arcs (Ndiv), and of the operating alternative
      - 5.6.2.5. The used numerical solvers
      - 5.6.2.6. The optimization problem solution particularities
  - 5.7. Optimization results and discussion
6. Case study no. 3: the use of a hybrid CCM cell-scale structured kinetic model coupled with a BR classical dynamic model (including macro-scale state variables) to maximize both biomass and succinate production by the *in-silico* design of GMO *E. coli* cells
    - 6.1. Symbols used in the chap. 7
    - 6.2. General remarks about CCM kinetic models, and ways used to solve the problem to *in-silico* design GMO-s of industrial interest by using a gene knock-out strategy
    - 6.3. The reduced CCM model of Edwards and Palsson [4] for the wild-strain of Escherichia coli
    - 6.4. Formulate the optimization problem to maximize the SUCC and biomass production in a BR
    - 6.5. Solving the FBA optimization problem to determine various gene knockout strategies to design optimized GMO-s
      - 6.5.1. Solving gene knockout LP, or MINLP problem by only accounting for the stoichiometric constraints
      - 6.5.2. Solving gene knockout LP or MINLP problem by accounting for the stoichiometric constraints, and some additional constraints
      - 6.5.3. Solving gene knockout MINLP problem with imposing a minimum limit for the succinate high production
      - 6.5.4. Conclusions about the use of HSMDM for the optimal design of GMO-s with desired characteristics
    - 6.6. Annex - Metabolic stoichiometric balance
  7. Conclusions and perspectives of using HSMDM for solving complex engineering problems.
  8. References

# 1. Abstract

This work is aiming to prove the feasibility and the advantages of using the novel concept to couple extended cell-scale **CCM**-based (central carbon metabolism) structured deterministic kinetic models with bioreactor classical dynamic models (including macro-scale state variables). This work presents a holistic 'closed loop' approach for the development of models of biological systems. The ever-increasing availability of experimental (qualitative and quantitative) information, at the cell metabolism level, but also on the bioreactors' operation necessitates the advancement of a systematic methodology to organize and utilize these data. The resulted *hybrid structured modular dynamic (kinetic) models* (**HSMDM**) were proved to successfully solving more accurately difficult bioengineering problems. In HSMDM, the cell-scale model part (including nano-level state variables) is linked to the biological reactor macro-scale state variables for improving the both model prediction quality and its validity range. The three approached examples here include development of **HSMDM**-s able: (case study *no.1*) to simulate the dynamics and to optimize the mercury uptake from wastewaters in a semi-continuous (**SCR**) three-phase fluidized bioreactor (**TPFB**), or (case study *no.2*) to simulate the dynamics of a fed-batch bioreactor (**FBR**) at both cell- and bulk-phase species levels, aiming to maximize the tryptophan (**TRP**) production, or (case study *no.3*) to optimize the both production of biomass and succinate (**SUCC**) in a batch bioreactor (**BR**). In all case studies, *in-silico* designed genetically modified (**GMO**) *E. coli*, or other bacteria cultures have been used. As proved, there are multiple advantages of using extended **HSMDM**-s. Thus, in the case study (*no.1*), a higher prediction detailing degree is reported, that is prediction of the dynamics of [26(cell species) + 3(bulk species)] vs. only [3 (bulk species)] by a classical macroscopic **SCR-TPFB** model, while covering a wider range of input [ $\text{Hg}^{2+}$ ] loads, with using cloned *E. coli* cells with various amounts of *mer*-plasmids [**Gmer**]. Also, the **HSMDM** offers the possibility to predict the bacteria metabolism adaptation to environmental changes over dozens of cell cycles, and the effect of cloning cells to modify their behavior under stationary or perturbed **SCR** operating conditions. In the case study (*no.2*) the **HSMDM** realizes a higher prediction detailing degree, by predicting the dynamics of [11(cell species) + 4(bulk species)] vs. only [3 (bulk species)] by a classical macro-level **FBR** model, while covering a wider range of control variables, and various **GMO** *E. coli* cells strains. Eventually this latter **HSMDM** model was used to derive the optimal operating policy of a **FBR** leading to the **TRP** production maximization. Besides, the **HSMDM** flexibility is high enough to consider a larger number of control variables for the studied bioreactors, that is: (i), biomass concentration, the inlet feed flow-rate, the inlet [ $\text{Hg}^{2+}$ ], and the [**Gmer**] concentration in the used cloned cells for the case study (*no.1*) and, (ii), the inlet feed flow-rate, and inlet substrate concentration [**GLC**] for the case study (*no.2*). (iii). In the case study (*no.3*), the **HSMDM** generates a Pareto-front of optimal operating alternatives of an **BR**, by using various *in-silico* design *E. coli* mutants. This approach uses the **HSMDM** model and a mixed-integer nonlinear programming (**MINLP**) rule, coupled with an effective adaptive random search to determine the optimal metabolic fluxes of a **GMO** in respect to multiple economic objectives associated to the gene knockout strategies. Exemplification is made for the case of designing an *E. coli* **GMO** that realizes maximization of both biomass and succinate production in an **BR** by using an extended structured **CCM** model from literature. Comparatively to the linear procedure **LP** that solves a combinatorial problem in a bi-level optimization approach, of dimensionality sharply increasing with the number of removed genes, the **MINLP** alternative using the **HSMDM** model includes the nonlinear influence of fluxes, and the number of checked knockout genes to the main goals. Besides, in case study (*no.3*), the **HSMDM** realizes a higher prediction detailing degree, by predicting the metabolic fluxes dynamics of [72(cell species), involved in 95 reactions + 1(bulk species, i.e. the biomass)] vs. only [1 (bulk species)] by a classical macroscopic **BR** model, while covering a wider range of control variables, and **GMO** *E. coli* strains.

**2. Keywords:** biochemical engineering concepts applied in bioinformatics; hybrid structured modular dynamic models; three-phase fluidized bioreactor for mercury uptake by cloned *E. coli* cells; fed-batch bioreactor for tryptophan production using genetically modified *E. coli* cells; mercury-operon expression regulation in modified *E. coli* cells; genetic regulatory circuits (GRC, GRN); individual gene expression regulatory module (GERM); cloned *E. coli* cells with *mer*-plasmids; WCCV (whole cell variable cell volume) modelling framework; gene knockout strategies to design optimized **GMO** for succinate production in **GMO** *E. coli* cells; Pareto optimal front to maximize both biomass and succinate production in the design **GMO** cells

**Abbreviations1:** CCM: central carbon metabolism; FBR: fed-batch bioreactor; GERM: Individual gene expression regulatory module; GRC: genetic regulatory circuits; HSMDM: hybrid structured modular dynamic (kinetic) model; P.I: GRC performance indices; SCR: semi-continuous reactor; TPFB: three-phase fluidized bioreactor; TRP: tryptophan; WCCV: whole cell constant cell volume modelling framework; WCVV: whole cell variable cell volume modelling framework

**Abbreviations2:** ADP, adp: adenosin-diphosphate; AMP, amp: adenosin-monophosphate; ATP, atp :adenosin-triphosphate; BCE: (bio) chemical engineering; CCM: Central carbon metabolism; G: The active Gene (DNA); GRC, GRN: Genetic regulatory circuits; GERM: Gene expression regulatory module; **GMO**: Genetically modified micro-organisms; GP: The inactive complex of G with the transcription factor P (its encoding protein in the reduced model here); GS: genetic switch; L: Species at which regulatory element acts; M: mRNA; Met (MetG, MetP): Metabolites (lumped DNA and protein precursor metabolites, respectively); **MINLP**: mixed-integer nonlinear programming; nM: Nano-moles/L, nano-molar (i.e.  $10^{-9}$  mol/L concentration); NG: Negligible; Nut (NutG, NutP): Nutrient (external nutrients imported to produce metabolites involved in the G and P synthesis respectively); ODE: Ordinary differential equations; P: Protein; PTS: Phosphotransferase; PPP: Pentose-phosphate pathway; Re(x): Real part of "x" variable; TF: Transcription factor; TCA: Citric acid cycle; QSS: Quasi steady-state; WC: Whole cell; WCCV: Whole cell of constant volume hypothesis; WCVV: Whole cell of variable volume hypothesis; [·]: Concentration".

### 3. Symbols used in the chap. 4

$C_j$  - species (lump, or 'pool') concentration

$D, D_s$  - cell content dilution rate (i.e. cell-volume logarithmic growing rate), under QSS cell growing conditions (balanced growth at homeostasis QSS = quasi steady state)

$\mathbf{g}, \mathbf{h}$  - kinetic model function vector

$\mathbf{J} = d\mathbf{g} / d\mathbf{C}$  - kinetic model Jacobian matrix

$\mathbf{k}$  - rate constant vector

$n_j$  - species " $j$ " number of moles, or number of effector

species binding the catalyst "L" in a **GERM**

$nr$  - number of reactions

$n_s$  - no. of species

$N_A$  - the Avogadro number

$r_j$  - species " $j$ " reaction rate

$R$  - universal gas constant

$t$  - time

$t_c$  - cell-cycle time

$T$  - temperature

$V$  - cell volume

**Greeks** -

$\lambda$  ( $\mathbf{J}$ ) - eigenvalues of the dynamic model Jacobian

$\nu_{ij}$  - stoichiometric coefficient of the species " $j$ " (individual or lumped) in the reaction " $i$ ";

$\pi$  - osmotic pressure

$\tau_j$  - species " $j$ " recovering time of the steady-state with an accepted tolerance (usually 1-5%)

**Index**

cyt - cytoplasm

env - environment

0 - initial

s - Stationary state

perturb - perturbed

## Acknowledgements

The author is deeply grateful to the following professors, who were kind enough to accept his scientific collaboration during their short (3 months) stays in their laboratories (summer 2006 to Prof. Deckwer, and summer 2009 to Prof. Zeng). In particular, the author is grateful for the generosity by which the two Professors provided him with basic information and experimental data in order to further elaborate structured cellular models for the case study no. 1 (Prof. Deckwer)[1,2], and for the case study no. 2 (Prof. Zeng) [3].



**Late Prof. Dr.-Ing. habil. Wolf-Dieter Deckwer (1941 – 2006)**  
Technischen Universität Carolo-Wilhelmina at Braunschweig  
also with:  
German Research Centre for Biotechnology  
(currently:  
Helmholtz Centre for Infection Research)



**Prof. Dr.-Ing. habil. An-Ping Zeng**  
Hamburg University of Technology  
Institute of Bioprocess and Biosystems Engineering

## Introduction

### General concepts

"As proved in the literature, the *in-silico* (math/kinetic model-based) numerical analysis of biochemical or biological processes are proved to be not only an essential but also an extremely beneficial tool for engineering evaluations aiming to determine the optimal operating policies of complex multi-enzymatic

reactors [5-10], or bioreactors [11-13].

In the **biochemical reactor case**, the trend is to use complex multi-enzymatic systems to successfully replace complex chemical syntheses, by using milder reaction conditions, and generating less waste.

Even if the multi-enzymatic systems are advantageous, the engineering part for optimizing such a complex process is not an easy task because it must account for the interacting enzymatic reactions, enzymes deactivation kinetics (if significant), multiple and often opposed optimization objectives, technological constraints, and uncertainties coming from multiple sources (model / constraints inaccuracies, disturbances in the control variables), and a highly nonlinear process dynamics [5,6,8,11,14,15]. All these parametric/model/data uncertainties require to update (with a certain frequency) the enzymatic process model, the optimal operating policies of the reactor being determined by using rather deterministic (model-based) optimization rules [16]. Multi-objective criteria, including economic benefits, operating and materials costs, product quality, etc., are used to off-line, or to on-line derive feasible optimal operating/control policies for various bioreactor types [15] by using specific numerical algorithms [6,13,17,18,19].

In the **biological reactor case**, development of extended cell-scale **CCM** structured kinetic models on a deterministic basis to adequately simulate *in detail* the cell metabolism self-regulation, cell growth, and its replication for such an astronomical cell metabolism complexity is practically impossible due to the lack of structured and comprehensive information, and computational limitations. A review of some trials is presented by Styczynski and Stephanopoulos [20], and by Maria [21-23]. That is because the cell metabolism 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 metabolism to the changing environment through complex genetic regulatory circuits (**GRC-s**), that includes individual or chains of „gene expression regulatory modules of reactions (**GERMs**), genetic switches, operon expression, etc. [21-23]. 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 reaction rates (fluxes) of the enzymatic reactions to get an optimized and balanced growth by using minimum resources (nutrients/substrates).

In spite of such tremendous modelling difficulties, the development of *structured reduced deterministic* (rather than stochastic) models [24] able to adequately reproduce the dynamics of some **CCM** complex metabolic syntheses [20,25,26], but also the dynamics of the **GRC-s** [21-23] tightly controlling the metabolic processes reported significant progresses over the last decades [27,28]. Even if they are rather based on sparse information from various sources, unconventional statistical identification, and lumping algorithms [21-23,29], such structured reduced deterministic kinetic models have been proved to be extremely useful for *in-silico* analyse and characterize the **CCM**, for designing novel **GRC-s** conferring new properties/functions to the mutant cells, or for engineering bioreactor evaluations. [3, 12, 21-23, 30].

The current (default) approach to solve the model-based design, optimization and control problems of industrial biological reactors is the use of *unstructured models* of Monod type (for cell culture reactors) or of Michaelis- Menten type (if only enzymatic reactions are retained) that ignores detailed representations of cell processes. The applied engineering rules are similar to those used for chemical processes and inspired from the nonlinear system control theory [13,15,31,32,33-38]. However, by accounting for only key process variables (biomass, substrate and product concentrations), these models do not properly

reflect the metabolic changes, being unsuitable to accurately predict the cell response to environmental perturbations by means of (self-) regulated cell metabolism [39,40].

The alternative is to use *structured* kinetic models, by accounting for cell metabolic reactions and component dynamics. Such deterministic models lead to a considerable improvement in the predictive power, with the expense of incorporating a larger



number of species mass balances including parameters (rate constants) difficult to be estimated from often incomplete data and, consequently, difficult to be used for industrial scale purposes. [40,41].

As a result, an impressive large number of valuable *structured deterministic* models (based on a mechanistic description of the metabolic enzymatic reactions taking place among individual or lumped species) have been proposed in the literature to simulate the cell **CCM** dynamics, with including tenths-to-hundreds of key species. Here, it is worth mentioning the *E. coli* model of Edwards and Palsson [4] used by [25,42-47] for various purposes, or the *S. cerevisiae* glycolysis model of Teusink et al. [48], or the **JWS** platform of Olivier and Snoep [49], and the **MPS** platform of Seressiotis and Bailey [50] to simulate cell metabolism (dynamics and/or stationary fluxes), to mention only few of them. Simulation platforms, such as **E-cell** of Tomita et al. [51,52], or **V-cell** of Slepchenko et al. [53], accounting for thousands of species and reactions, display extended capabilities to predict the dynamics of the cell metabolism under various conditions, based on EcoCyc, KEGG, Prodigic, Brenda and other -omics databanks (review of Maria [23]). A worthwhile **CCM**-based dynamic or stationary models were reported by Maria et al. [3,25,42] and schematically represented in (**Figure 4-1**).

Meritorious structured deterministic **CCM** kinetic models have been reviewed by Maria [22]. Deterministic kinetic models using continuous variables has been developed by Maria [25] for the glycolysis, and by [43,54-57] for the **CCM** in bacteria of industrial interest. Such models can adequately reproduce the cell response to continuous perturbations, the cell model structure and size being adapted based on the available -omics information. Even if such extended structured models are currently used only for research purposes, being difficult to be identified, it is a question of time until they will be adapted for industrial / engineering purposes in the form of *reduced hybrid structured modular dynamic(kinetic)models* **HSMDM**.

This work presents a holistic 'closed loop' approach for the development of models of biological systems [358]. The ever-increasing availability of experimental (qualitative and quantitative) information, at the cell metabolism level, but also about the bioreactors' operation necessitates the advancement of a systematic methodology to organise and utilise these data. The resulted **HSMDM** were proved to successfully solving more accurately difficult bioengineering problems. In such **HSMDM**, the cell-scale model part (including nano-level state variables) is linked to the biological reactor macro-scale state variables for improving the both model prediction quality and its validity range. The two case studies presented and discussed here proves this engineering aspect.

An alternative compromise is to use *hybrid* models that combine unstructured with structured process characteristics to generate more precise predictions (see the review of Maria [3]). Basically, hybrid models use a *two-level hierarchy*: the bioreactor macroscopic state variables linked with the nano-scale variables describing the cell key metabolic processes, and those of practical interest.

In fact, such a *hybrid structured* cell dynamic model must include only the essential parts of the central carbon metabolism (**CCM**), by incorporating the pathway responsible for the target metabolite synthesis, and the lumped modules of the cell core, that is the glycolysis, the **GLC** uptake system (i.e. the phosphotransferase (**PTS**), or an equivalent system), the **ATP**-recovery system (adenosin-triphosphate cycle), **TCA** (citric acid cycle), **PPP** (Pentose-phosphate pathway), and other reaction pathways modules if necessary in the **CCM** simulations. See for instance the discussion of Maria et al. [25,39,42,58].

A special interest was given to the accurate modelling of the glycolysis dynamics and its self-regulation [25,39,59-61] as long as most of the glycolysis intermediates are starting nodes for the internal production of lot of cell metabolites (e.g. amino-acids, or succinate (**SUCC**), citrate (**CIT**), amino-acids, like cysteine, lysine, phenyl-alanine, tryptophan (**TRP**), etc. [11,42,61,62].

This need to have good quality structured cell models to simulate the dynamics (and regulation) of the bacteria **CCM** became a subject of very high interest over the last decades, allowing *in-silico* design of **GMO-s** with desirable characteristics of various applications in the biosynthesis industry, civil engineering, and other fields [21-23].

**HSMDM advantages** Even if such a complex / extended model requires more experimental and computational efforts to be built-up, as proved by the two approached case studies of this work, the resulted *hybrid* (bi-level, that is cell species, and macro-level bioreactor state variables) dynamic models (**HSMDM**) present a large number of advantages compared to the classical (default) *unstructured* models of Monod type (for cell culture bioreactors), or of MichaelisMenten type (if only enzymatic reactions are retained) that ignores detailed representations of cell processes. Thus, among the multiple *advantages* of **HSMDM** models, are to be mentioned:

a) A higher prediction detailing degree. Thus, for the below case study (**no.1**) of chap. 5, a higher prediction detailing degree is reported, that is predicted dynamics of [26(cell species) + 3(bulk species)] vs. only [3 (bulk species)] by a classical macroscopic **SCR- TPFB** model, while covering a wider range of input [Hg<sup>2+</sup>] loads, with using cloned *E. coli* cells with various amounts of *mer*-plasmids [Gmer]. In the case study (**no.2**) of chap. 6, the **HSMDM** realizes a higher prediction detailing degree, by characterizing the dynamics of [11(cell species) + 4(bulk species)] vs. only [3 (bulk species)] by a classical macroscopic **FBR** model, while covering a wider range of control variables, and various **GMO** *E. coli* cells strains.

b) Prediction of the inner cell key reaction rates (different from the apparent rates observed in the bioreactor). See the case study (**no.1**), of [1,2,40], and the case study (**no.2**) of [3,12];

c) The model predictions can cover a wide range of input/control variables of the bioreactor. Thus, for the case study (**no.1**), the **HSMDM** of Maria and Luta [40] realizes a higher prediction detailing degree, that is [simulated 26(cell species) + 3(bulk species)] vs. only 3 (bulk) variable dynamics predicted by a classical macroscopic **SCR- TPFB** model, while covering a wide range of input [Hg<sup>2+</sup>] loads (0–100 mg/L), and cloned *E.coli* cells with various amounts of *mer*-plasmids [Gmer] of (3–140 nM).

d) The **HSMDM** model can predict the bacteria metabolism adaptation to environmental changes over several cell cycles, and the effect of cloning/GMO cells to modify their behavior under stationary or perturbed bioreactor operating conditions. See the case study (**no.1**) in chap. 5 for details.

e) The extended **HSMDM** can offer predictions of a higher accuracy when they are used to *in-silico* (model-based) engineering developments (bioreactor design, and its off-line optimization) compared to unstructured models (see the chap. 5-6 for the bioreactor cases). For instance, the model could better predict the optimal time stepwise feeding policy of a fed-batch reactor (**FBR**) to increase the tryptophan (**TRP**) production (the case study (**no.2**) in chap. 6) [3,12]. In the case study (**no.1**) of chap. 5, the extended **HSMDM** allows optimizing the operating policy of a **SCR- TPFB** regarding the biomass concentration, the inlet feed flow-rate, the inlet [Hg<sup>2+</sup>], the biomass support size, and the [Gmer] concentration in the used cloned cells [40, 63]. For a bi-enzymatic reactor case, Maria et al. [9] used a **HSMDM** to derive optimal operating policies of a **FBR** by accounting for multiple competing optimization objectives.

f) Complex **HSMDM**-s can be used for bioinformatics purposes, by evaluating the influence of the bioreactor operating conditions (that is the control macro-variables) on the dynamics of cell nano-scale key-intermediates and fluxes involved in the metabolite synthesis of interest (that is, for the case study **no.2** those belonging to glycolysis, **ATP**-recovery system, and **TRP**operon expression), thus directing the design of genetically modified cells with desirable ‘motifs’ [3].

g) Extended **HSMDM** can be used to obtain lumped dynamic models for rapid engineering calculations, by employing specific model reduction rules and additional kinetic data valid in *local* operating domains. See [29,64,65 ] for general models, or [66,67] for linear models. As a result of such an approach, the bioprocess complexity may be described by a succession of local reduced models *enfolding* on the real process. The local / reduced models include only the key metabolic pathways to obtain relevant (of interest) process state predictions;

h) The structured cell models (of **HSMDM** type) are also useful for understanding the cellular bioprocess in direct connection to the bioreactor operating mode. For instance, in the case study (**no.2**) of chap. 6, it can *insilico* be determined the conditions of occurrence of oscillations for the glycolysis [59], or oscillations in the **TRP**-operon expression [39,62], or those leading to a balanced cell growth (quasi-steady-state **QSS**, i.e. homeostasis) [58]. In the case study (**no.1**) in chap. 5, the extended **HSMDM** can predict the *mer*-enzymes expression levels and the cytosolic mercury reduction rate depending on the *mer*-plasmids level [Gmer] in the cloned *E. coli* cells.

i) Some other case studies supporting the use of complex **HSMDM** are given by: (**a**) [42] to design **GMO** *E. coli* strains for optimizing the succinate production (**SUCC**) in a batch reactor (**BR**); (**b**) reviews of Maria [11], and of Dorka [68] on **FBR** optimization for the monoclonal antibodies (**mAbs**) production, etc.

j) As documented by Maria [21-23], to overcome the cell process dynamics complexity, the metabolic pathway representation with continuous and/or stochastic individual or lumped reactions/variables remains the most adequate and preferred representation of cell processes, the adaptable-size and structure of the lumped model (species and/or reactions) depending on available information and the utilisation scope.

The main advantages of *deterministic / continuous variable* kinetic models are coming from the use of experience, concepts, math representation, rules, and algorithms of the '**biochemical reaction engineering**' (**Figure 4-2** and **Figure 4-3**) [21-23,29]. The reaction rate expressions in the deterministic models are the usual ones of biochemical reaction engineering, that is of Michaelis-Menten, or Hill type (see **Figure 4-4**). A large number of **CCM** kinetic models have been reported in the literature, such as those of [43,69,70]. A short discussion is given by Maria [3,25]. Such a **CCM**-based kinetic model (**Figure 4-1**) was used by [3,12] to optimize a **FBR**. In fact, the glycolysis together with the phosphotransferase (**PTS**)-system, or an equivalent one for GLC-uptake, and with the pentose-phosphate pathway (**PPP**), and with the tricarboxylic acid cycle (**TCA**), all these are part of the so-called central carbon metabolism (**CCM**) (**Figure 4-1**) [42].

The parameters (rate constants) of deterministic dynamic models are estimated by using the common (bio)chemical engineering rules of (**Figure 4-2** and **Figure 4-3**) [21-23,29], by using either dynamic (kinetic) data obtained in a chemostat under transient regime (e.g. pulse-like perturbations in the bioreactor influent [43], or using steady-state data (metabolites concentrations) obtained at homeostasis (i.e. balanced cell growth, [21-23]), by solving the math model stationary algebraic set of species mass balances, by accounting the all species from the reacting system, taken individually or lumped [see eqn.(9)]. Parameter estimate must fulfill physico-chemical meaning constraints related to metabolic reaction stoichiometry. Besides, the rate constants must be limited by the diffusional processes, and in agreement with the thermodynamic equilibrium steps.

Because the problem of estimating reaction rate constants for such hybrid structured kinetic models with continuous variables (of **ODE**-type, that is an ordinary differential-algebraic equations set, see below eq.(1) ) in the presence of constraints is one of **NLP** (nonlinear programming) type [29], with a convex search field and a multi-modal objective function, its solution is difficult, even if it uses high-performance numerical optimization algorithms included in commercial mathematical software (eg Matlab). Therefore, the feasible global solution for such estimation problems was found by applying a very efficient numerical algorithm, namely **MMA**'s optimization procedure of Maria [71,72].

Even if complicated and, often over-parameterized, the continuous variable dynamic deterministic **ODE** models (eq. 1A-B) of the **CCM** metabolic pathways, or of **GRC**-s present a significant number of advantages, being able to reproduce in detail molecular interactions, the cell slow or fast continuous response to exo/andogenous continuous perturbations [20,24]. 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 [73], by accounting for the regulatory system properties, that is: dynamics, feedback / feedforward, and optimality. And, most important, such **ODE** kinetic modelling approach allows using the strong tools of the classical (bio-)chemical engineering (**BCE**) modelling concepts summarized in (**Figure 4-2** and **Figure 4-3** ). The most important ones are the followings [21-23,60,74].

BCE-1. Fulfillment of the molecular species conservation law (analysis; species differential mass balance set)

BCE-2. Fulfillment of the atomic species conservation law (atomic species mass balance);

BCE-3. The thermodynamic analysis of reactions (that is quantitative assignment of reaction directionality) [75]

BCE-4. Set equilibrium reactions by using Gibbs free energy balance analysis; set cyclic reactions; find species at quasi-steady-state to replace its differential mass balance with an algebraic equation;

BCE-5. Extended **HSMDM**-s allow improved evaluation of steady-state flux distributions (i.e. stationary metabolic reaction rates) that provide important information for metabolic engineering [76]

BCE-6. Allow application of **ODE** model species and/or reaction lumping rules [29,64].

When developing deterministic models for the **CCM**, or for other cell metabolic processes of a micro-organism, to be further used for **GMO** design, an important aspect is to also include math (kinetic) models of individual **GERM**-s characterizing the gene expression control of the enzymes production. Also, by linking the interfering **GERM** modules, complex **GRC** regulatory chains can thus be obtained [21-23,77].

Development of dynamic models to adequately reproduce such complex synthesis related to the **CCM** [43,69,70], but also to the **GRC**-s 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 [21,24,29,64,78]. However, such structured models are extremely useful for *in-silico* design of novel **GRC**-s conferring new

properties/functions to the mutant cells, in response to external stimuli [42,79, 80-83, 85-90]. This topics belongs to the so-called '**computational systems biology**', or simply '**bioinformatics**'. In fact, the two emergent research/applicative fields are closely inter-connected, as depicted in (**Figure 4-59**).

**Systems Biology**, is defined as "the science of discovering, modelling, understanding and ultimately engineering at the molecular level the dynamic relationships between the biological molecules that define living organisms" (Leroy Hood, Head Inst. Systems Biology, Seattle, USA). **Systems Biology** is one of the modern tools, which uses advanced mathematical simulation models for *insilico* design of **GMOs** that possess specific and desired functions and characteristics. The works of Maria [21-23] presented a short review about **Systems Biology**, by including a short history, their modern concepts, and their (math/numerical-experimental) rules/tools, together with the (bio)chemical engineering (**BCE**) principles and deterministic modelling rules used by the Systems Biology for modelling cellular metabolic processes. This involves application of the classical **BCE** modelling techniques (mass balance, thermodynamic principles), algorithmic rules, nonlinear system control theory, and bioinformatics rules, briefly presented in (**Figure 4-2**, and **Figure 4-3**), and in the above (BCE1-BCE6) issues. The main principles, concepts, and rules of the **BCE** are shortly reviewed by [117-121]. The metabolic pathway representation with continuous and/or stochastic variables remains the most adequate and preferred representation of the cell processes, the adaptable-size and structure of the lumped model depending on available information and the utilisation scope [21-23].

When developing extended structured **HSMDM** models, besides **BCE**, and **Systems Biology** principles and rules (**Figure 4-2**, **Figure 4-3**, and **Figure 4-59**), the Bioinformatics concepts and rules play an essential role because they make the connection with the **bio-omics** data banks (see the below cell simulation platforms (**a-e**)), from which the most important information refers to the genome map and its correspondence to the proteome map for a certain micro-organism.

**Bioinformatics**. According to [122], Bioinformatics is an inter-disciplinary field that develops methods and software tools for understanding and interpret the biological data, in particular when the data sets are large and complex. As an interdisciplinary field of science, the bioinformatics combines several classic/modern disciplines, such as: biology, chemistry, physics, computer science, information engineering, mathematics, and statistics to analyze and interpret the biological data. Bioinformatics has been used rather for *in-silico* analyses of biological queries using a large number of computational and statistical techniques, aiming to design novel **GMO**-s of practical (industrial) use [123-133]. The most important rule of Bioinformatics refers to the rapid, and computer-rassisted genome sequentiation [122]. Some of the common rules used by bioinformatics are given in (**Figure 4-60**) [122].

Bioinformatics includes biological studies that use computer programming as part of their methodology, as well as specific analysis 'pipelines' that are repeatedly used, particularly in the field of genomics. Common uses of bioinformatics include identification of candidates genes and single nucleotide polymorphisms (SNPs). Often, such identification is made with the aim to better understand the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organizational principles within nucleic acid and protein sequences, called proteomics [122].

The main objective of the '**computational systems biology**' is to model the kinetics of entire living cells at the molecular level on a mechanistic base. Given the enormous complexity and unknown aspects of such systems, formulating reliable models with predictive ability remains only a dream. However, advances in genomics, transcriptomics, proteomics, metabolomics, and in the computing power provide hope that this objective might be realized within next couple of decades. Bioinformatic databases and software platforms are being constructed for modeling entire cells with massive amounts of data [51,52,81,92]. For example,

(a) **E-Cell** software allows simulating reaction pathways within compartment-based cells using the continuous-differential modeling approach [51,52,93,94] (**Figure 4-37-left**). The computing platform objects are [compartments, compounds, genes, reactions]. **E-cell** has been used in conjunction with the **EcoCyc** [95], and **KEGG** [96,97] databases to simulate the dynamics of 127 genes/protein found in *M. genitalium*.

(b) **v-Cell** [98,99]. The computing platform objects are (model, geometry & applications, biological interface).

(c) **CellML, JWS** (Silicon-cell) [49,100, 101,171,172] is a cell modelling and simulation framework (**Figure 4-37-right**) with compartments and membranes, each of which may include species, reactions and membrane fluxes.

(d) **M-Cell** simulation platform of [102] (**Figure 4-24-left**) allows simulating high-level complex cell sub-systems, such as neural communication networks, together with proteins and enzymes involved in exo-/endo-cytosis, synaptic transmission, transport and signal reception. The **M-Cell** simulator includes simulation of Brownian random walk, and Monte Carlo stochastic algorithms for modeling small numbers of diffusing ligands interacting with individual 3D binding sites in spatially complex environments.

(e) The **A-Cell** platform [103,104] (**Figure 4-24-right; Figure 4-20**) uses ('electrical circuit' like models) to simulate biochemical reaction schemes, neurons connections, and pathways. For other bio-modeling software packages the reader is referred to the reviews of [92,103,104,105].

To easier realize the numerical simulations of such complex cell math models, specific programming languages (SBML, [92,105]), or on-line simulation platforms (JWS, [49,100]) have been developed. By using such modern computing tools, simulation of various biological systems was possible, such as [21-23]:

- Single cell growth (e.g. *Escherichia coli*, *Haemophilus influenzae*, *Mycoplasma genitalium*, yeast, etc).
- Model metabolic oscillations (red-blood-cell synthesis, glycolysis, TCA cycle, oxidative phosphorylation, key species oscillations, etc.) [25,39,58,59,61,62,82, 106].
- Metabolic control of protein synthesis regulation (**GERM-s**, **GRC-s**) [2123,24,107-112].
- Modelling the central carbon metabolism (**CCM**) [26,43,70]. Some developed dynamic models are given by [25,42,69] (**Figure 4-38**).
- FBA (cell flux balance analysis) based design of **GMO-s** [77] (**Figure 4-39, Figure 4-40, Figure 4-41**).
- Modelling the cell cycle [113, 114].
- Modelling the drug release and cell-drug interactions [115,116].
- Modelling cellular communications, neuronal transmission
- Analysis of 'logical essence' of life (life minimal requirements)

At the same time, the exponential-like increase of the experimental biological information lead to development of valuable **bio- omics** databanks, such as (**Figure 4-42, Figure 4-43, Figure 4-44, Figure 4-45, Figure 4-46**):

- a) KEGG [96,97],
- b) JWS [49,100],
- c) EcoCyc [95],
- d) Roche [134], etc.

However, it is only over the last decades when *Systems Biology* reported notable successes due to a considerable increase in computing power of the modern computers. It is to mention here, for instance, the cell simulator platforms, and online model repository **JWS** of [49,100], or those developed by Rocha et al. [28], or by Tomita et al. [51,52], together with continuous expansion of the **bio-omics** databases JWS, KEGG, EcoCyc, Roche, etc. and reported advances in the numerical algorithms used by bioinformatics, (bio)chemical engineering, and nonlinear systems control theory [21-23].

Due to such favourable premises, related to the expansion of *bio- omics* databanks, and cell metabolism (CCM, GRC-s) dynamic models, novel works have been reported over the last decades. Among the milestone works in Systems Biology it is to mention the contributions in modelling / design of GRC-s, GERM-s, FBA, MCA of [73,135-138]. The number of published papers in the Systems Biology area increases with two orders of magnitude from 2000 to 2007, and it is still exponentially increasing, most of them being founded by programs of the European Science Foundation.

As stated in [139,140], tremendous applications of Systems Biology have been reported over the next decades in the below areas (Figure 4-47):

Designing mutant, cloned cells with desired 'motifs'

Cell biology

Genetics biology or genetics

Biotechnology, Bioengineering

Biomedical engineering

Biochemistry

Agricultural biology and Ecology

Biophysics

Food science

Immunology

Molecular biology

Biodiversity

Bioinformatics

And, before in a context of increasing calls for biology to be predictive, modelling and optimization are the only approaches biology has for making satisfactory predictions [141]. Due to the computing facilities offered by the algorithmic rules developed by the (bio)chemical engineering and nonlinear systems rules (Figure 4-2, Figure 4-3, Figure 4-4, Figure 4-8), the developed cell math models use a vectorial-matriceal approach (Figure 4-48), with a continuous model upgrading based on dynamic experimental data recorded in a chemostat (i.e. a continuously operated bioreactor), operated under steady-state, or in a dynamic regime following an input perturbation in the substrate/enzymes/biomass concentration in the solution fed in the bioreactor [43,77].

Given these developments, as well as the seemingly inexorable advances in computing power, it is tempting to believe that reliable whole-cell models (WC) with predictive power will be forthcoming once complete sets of 'bio-omic' information become available. However, a better theoretical understanding of cellular life, viewed *holistically*, may be required before we can understand how life *emerges* out of complex networks of molecular-level interactions between cellular components. This study represents a foyer into whole-cell kinetic modeling, in an attempt to understand holistic aspects of cell-system from a quantitative computational perspective.

The scope of this chapter is to examine some fundamental properties of living cells relevant to whole-cell modeling. Thus, a methodology to build-up regulatory kinetic schemes in a 'whole cell variable cell volume' (WCVV) modelling framework, and stationary/perturbed environmental conditions is developed for a generic protein/gene (P/G) pair synthesis process. The model elements of novelty consist in accounting variable cell-volume, and constant osmotic pressure conditions, and thus removing some drawbacks of classical (default) continuous-concentration-based simulators. Exemplification is made for several simple genetic regulatory schemes (GRC, or GRN) extracted from literature [21-24], and the impact of developed modeling approach on the prediction results is studied, such as: a more realistic prediction of regulatory scheme sensitivity to both stationary and dynamic perturbed conditions; increased species interconnectivity; effect of cell-ballast on cell-behavior vs. various types of perturbations; a more realistic way to design interconnected regulatory modules and links in GRC-s in a common volume growing environment. Other analysis aspects, not developed in this paper, can be easily approached under the present hypotheses such as system multiplicity, and characterization of periodic cell-phenomena.

### Modelling GRC under a WCVV framework

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. [21-23]

Various simple math models have been proposed to represent the elementary metabolic fluxes of a **CCM** [142], or of a **GRC** [20,81-85,106,138,143-162]. Eventually, such models allow a multi-criterion design and optimization of a target **GRC-s** [163].

Development of dynamic models to adequately reproduce such complex synthesis related to the **CCM** [43,69,70], but also to the **GRC-s** 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 [29,67,78]. However, such structured models are extremely useful for *in-silico* design of novel **GRC-s** conferring new properties/functions to the mutant cells, in response to external stimuli [42, 79-83, 86-90, 164-167].

A central part of cell metabolic 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 by which biological systems respond to signals, or environmental perturbations. The emergent field of such efforts is the so-called '*gene circuit engineering*' (**GCE**) 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. By using simulation of gene expression, the **GCE in-silico** design **GMO** 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 [23].

When the cell-model used to construct an extended **HSMDM** includes individual **GERM-s** (**Figure 4-7, Figure 4-9**), or complex **GRC-s** gathering chains of inter-connected **GERM-s**, genetic switches (**Figure 4-5**), or operon expression (**Figure 5-1**), genetic amplifiers, etc. [21-23,147,168,169,174], then the classical (default) 'whole-cell-constant-volume' (**WCCV**) kinetic modelling can not longer be applied because the regulatory properties of **GERM / GRC** are related to the cell-volume growth and lot of additional constraints derived from the cell holistic properties (e.g. isotonic constraint ensuring the cell membrane integrity). To also account for the cell growth, and the cell-holistic constraints, a novel holistic modelling framework of cell processes by accounting for the cell variable volume and other constraints was introduced by Maria [24,170], that is the so-called 'whole-cell-variable-volume' (**WCVV**), starting from the analogy with the chemical engineering concepts, that is, more specifically by the analogy with the kinetic models of chemical reactions conducted in variable-volume systems [173]. Implications of the novel **WCVV** concept when modelling metabolic cell processes (especially **GRC-s**) under variable cell-volume were systematically studied, compared to the **WCCV** models [60], positive results extrapolated, and widely promoted in a large number of applications by Maria [21-24,40,60,74,78,175,176,177,178] that is the so-called „whole-cell-variable-volume (**WCVV**) framework. The next chapter is aiming to briefly describe the characteristics of the **WCVV** approach, while Maria et al. [60] proved the superiority in the prediction accuracy offered by the **WCVV** kinetic modelling framework compared to the classical **WCCV**.

Generally, living cells are evolutionary, auto-catalytic, self-adjustable structures able to convert raw materials from environment into additional copies of themselves. Living cells are hierarchically organized, self-replicating, evolvable, and responsive biological systems to environmental stimuli. The structural and functional cell organization, including components and reactions, is extremely complex, 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 [21-23]. Relationships between structure, function and regulation in complex cellular networks are better understood at a low (component) level rather than at the highest-level [179,180].

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 [181]. 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 [21-23,182].

The *in-silico* re-design of the cell metabolism is an up-to-date subject in *Synthetic Biology*. But in this effort, *Synthetic Biology* is closely assisted by the *Systems Biology* focus on the cell organization, the former being one of the main tools in the *in-silico* design of genetically modified micro-organisms (**GMO**) with desired characteristics, and with applications in medicine, such as therapy of diseases (gene therapy), production of new devices based on cell-cell communicators, biosensors, production of vaccines, etc. The *Systems Biology* aims at understanding the dynamic interaction between components of a living system or between living systems. (<http://www.erasysbio.net/>). To realize these ambitious objectives, *Systems Biology* uses a wide range of tools, but mainly complex mathematical simulation models linked to -omics databanks [21-23].

In this context, the adequate modelling of the genetic regulatory circuits (**GRC**), made from linked **GERM**-s, together with modelling the cell central carbon metabolism (**CCM**) remain subjects of tremendous importance on which researches have been focus over the last decades, as long as **GRC**-s are the essential metabolic components used to re-design the whole cell metabolism, and in regulating the whole cell syntheses [21]. **GRC**-s, also denominated as '*genetic regulatory networks*' **GRN**-s, is a combination of **GERM**-s ensuring some functions into the cell. Due to the gene location in the **GRC** nodes, a more sophisticated definition was given by Das et al. [183], by using the graph-theory:

**Definition – Gene Regulatory Network (GRN):** a Gene Regulatory Network is a mixed graph  $G := (V, U, D)$  over a set  $V$  of nodes, corresponding to gene-activities, with unordered pairs  $U$ , the undirected edges, and ordered pairs  $D$ , the directed edges. A directed edge  $d_{ij}$  from  $v_i$  to  $v_j$  is present iff a causal effect runs from node  $v_i$  to  $v_j$  and there exist no nodes or subsets of nodes in  $V$  that are intermediating the causal influence (it may be mediated by hidden variables, i.e. variables not in  $V$ ). An undirected edge  $u_{ij}$  between nodes  $v_i$  and  $v_j$  is present iff gene-activities  $v_i$  and  $v_j$  are associated by other means than a direct causal influence, and there exist no nodes or subsets of nodes in  $V$  that explain that association (it is caused by a variable hidden to  $V$ ).

In fact, any lumped representation of a **GERM**, or a **GRC** should include, in one form or another, the main '*actors*' of such regulatory circuits, that is: metabolites ( $Met^*$ ) as substrates for genes and protein synthesis, genes ( $G^*$ ) encoding proteins ( $P^*$ ), as depicted in (**Figure 4-10**). This is a very simplified representation of the biochemistry in living cells (hiding hundreds of enzymatic reactions) conceptually decomposed in three '*spaces*'. Influences between gene activities, without explicitly accounting for the proteins and metabolites, result from a projection of all regulatory processes on the 'gene space' [137].

A central part of any **CCM** model concerns self-regulation of the metabolic processes via protein (enzymes) synthesis in **GERM**-s. Consequently, one particular application of the dynamic deterministic cell models is the study of **GRC**-s, in order to predict ways by which biological systems respond to signals, or environmental perturbations. The emergent field of such efforts is the so-called '*gene circuit engineering*' and a large number of examples have been reported with *in-silico* re-creation / design of **GRC**-s conferring new properties/functions to the mutant cells (i.e. desired '*motifs*' in response to external stimuli) [21,184]. Such an effort is facilitated by the use of **GERM** simulation models [21,24,185]. As mentioned by the pioneers of this field, 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* [186-188]. 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 [26]. In this context, **GERM** and **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 [21-23]. Examples of such **GRC** modulated functions include: toggle-switches, hysteretic **GRC** behaviour, **GRC** oscillator, specific treatment of external signals, **GRC** signalling circuits and cell-cell communicators [107]. The development of dynamic models on a deterministic basis to adequately simulate *in detail* the cell metabolism self-regulation, cell growth, and replication for such an astronomical cell metabolism complexity is practical impossible due to lack of structured information and computational limitations. A review of some trials was presented by Styczynski and Stephanopoulos [20]. In spite of such tremendous modelling difficulties, development of *reduced* dynamic models to adequately reproduce the cell complex syntheses related to the **CCM** [20,25,26], but also the **GRC** tightly controlling the metabolic processes [107] reported significant progresses over the last decades in spite of the lack of structured experimental kinetic information. Even of being rather based on sparse information from various sources, and on unconventional identification / lumping algorithms [21,24,29,64], such structured deterministic kinetic models have been proved to be extremely useful for *in-silico* design of novel **GRC**-s [21-23,147,168,169,178,184].

Examples of such **GRC** modulated functions include [21,107,174,189]:

**a) toggle-switch**, i.e. mutual repression control in two gene expression modules, and creation of decision-making branch points between on/off states according to the presence of certain inducers (**Figure 4-5**);



- b) **hysteretic GRC behaviour**, that is a bio-device able to behave in a history dependent fashion, in accordance to the presence of a certain inducer in the environment;
- c) **GRC oscillator** producing regular fluctuations in network elements and reporter proteins, and making the **GRC** to evolve among two or several quasisteady-states;
- d) **external signals treatment** by controlled expression such as amplitude filters, noise filters or signal / stimuli amplifiers;
- e) **GRC signalling circuits** and cell-cell communicators, acting as ‘programmable’ memory units. [23]

As discussed by Maria [21-24,60,74], the classical (default) modelling tools of metabolic cell processes are based on the ‘**Constant Volume Whole-Cell (WCCV)**’ continuous variable **ODE** dynamic models which, do not explicitly consider the cell volume exponential increase during the cell growth. As proved by Maria [60], such an approach may lead to biased and distorted conclusions on the **GERM**’s performances, thus making difficult the modular constructions of **GRC**-s by linking individual **GERM**-s. By contrast, the holistic ‘**whole-cell of variable-volume (WCVV)**’ modelling framework introduced, extended, and widely promoted by Maria [21-24,40,60,74,78,175,176,177,178] has been proved to be more realistic and robust, by explicitly including in the model relationships the cell-volume growth, with preserving the cell-osmotic pressure (that is the cell membrane integrity). The added isotonicity constraint by Maria [21-24,170] was proved to be essential for predicting more adequate performance regulatory indices of **GERM**-s and **GRC**-s.”

### The WCVV vs. WCCV when modelling the GRC dynamics in living cells

This chapter is aiming to exemplify, in a simple way, the importance of using a **WCVV** modelling framework compared to the classical (default) **WCCV** models when simulating some regulatory properties of **GERM**-s or **GRC**-s, by explicitly accounting for the cell-volume growth, and system thermodynamic isotonicity (constant osmotic pressure). Exemplification is made for the case of a simple generic **GERM** model with characteristics taken from *E. coli* cells [24,95,174-176,190,191], by mimicking the cell homeostasis and its response to dynamic perturbations. This work subject importance is very high, as long as a large number of cell simulators are developed and used for practical applications in the biosynthesis industry, and in medicine. The isotonicity constraint is proving to be a natural way to preserve the homeostatic properties of the cell system [21-24,60,78], instead of imposing others constraints, such as ‘the total enzyme activity’ and ‘total enzyme concentration’ constraints [73,110].

A comparison of model prediction quality in the case of a **GERM** of [G(P)1] type (**Figure 4-7**, and **Figure 4-9**) modelled under **WCCV** or **WCVV**, clearly indicate that **WCCV** can lead to biased and distorted conclusions on **GERM** regulatory performances (under both stationary or as response to dynamic pulse-like perturbations), thus making difficult the modular construction of **GRC**-s by linking individual **GERM**-s (**Figure 4-12**, and **Figure 4-13**) [60,175,179].

### WCCV formulation

For a system of chemical or biochemical reactions conducted in a cellular closed volume ‘**V**’ (assumed an open system of uniform content), the classical (default) 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 including the considered system states (biological/chemical species index ‘**j**’, taken individual or lumped), in the following **WCCV** formulation (with referring to the whole system volume):

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

The above formulation assumes a homogeneous constant volume with no inner gradients or species diffusion resistance into the cell. When continuous variable ODE dynamic models are used to model cell enzymatic/metabolic processes, the default-modelling framework Eq. (1A-B) is that of a constant volume and, implicitly, of a constant osmotic pressure ( $\pi$ ) in isothermal systems, according to the assumed fulfilled Pfeffer’s law in diluted solutions (i.e. the cytosol system) [24,113,170]:

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

Eventually some **WCCV** models, accounts for the cell-growing rate as a pseudo ' $D_s$  cay' rate of key-species (often lumped with the degrading rate) in a so-called 'diluting' rate (denoted here by an average  $D_s$  see below for its significance). In fact, by ignoring the direct influence of the cell volume increase, the **WCCV** dynamic model cannot ensure the system isotonicity constraint fulfilment because the sum of species number of moles doubles over the cell cycle. Such a **WCCV** dynamic model might be satisfactory for modelling many cell subsystems, but not for an accurate modelling of cell regulatory / metabolic processes under perturbed conditions, or for division of cells [113], distorting the prediction quality, as reviewed by Maria [21-24,60,175,176]. Other researchers [110] tried to preserve the homeostatic properties of the cell system, not by imposing the isotonic constraint Eq. (2), but 'the total enzyme activity' and 'total enzyme concentration' constraints (see [73])."

### WCVV formulation

At this point, it is to strongly emphasize that living cells are systems of variable volume. They double their volume during the cell cycle. For chemical or biochemical systems of variable-volume, another formulation is more appropriate, being given by Aris [173], and later developed and promoted by Maria [21-24,170] in modelling cell subsystems by also including the cell isotonicity constraint in the so-called **WCVV** modelling framework [21-24,40,60,74,78,175,176,177,178]. In mathematical terms, the species mass balance Eq.(1) should be re-written in the following form:

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

$$c(t=0) = c_s \tag{3A}$$

Where:  $C$  = cell species concentration vector;  $t$  = time;  $k$  = model constants; 's' index = at steady-state;  $h$  = cell kinetic model functions. Also:

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

Here, it is to remark two possibilities to calculate the cell dilution ' $D$ '

necessary for solving the model Eq. (3A). The simplest, but not the accurate one, is to use a value averaged over the cell cycle, that is:

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

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

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

The second alternative to evaluate the cell dilution ' $D$ ' is those to impose a constraint accounting for the cell-volume growth while preserving a constant osmotic pressure and membrane integrity. Thus, by derivation of the Pfeffer's law Eq. (2) in respect to  $V$ , and by division to  $V$ , one obtains the '*isotonic*' dilution rate  $D_i$  [21,24,60]:

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

It is to observe in Eq. (5) that the cell content dilution rate  $D_i$  is linked to the all species (taken individually or lumped) reaction rates via the isotonicity constraint. As species reaction rates varies during the cell cycle, it clearly results that formulation Eq. (5) offers a more accurate estimation of the (variable) cell dilution at any time. Such a system isotonicity constraint is more 'natural' and eventually includes, 'the total enzyme activity' and 'total enzyme concentration' constraints suggested by Komasilovs et al. [110].

In the above relationships eqns.(2, 5), the following notations have been used: T = absolute temperature, R = universal gas constant, V = cell (cytosol) volume. As revealed by the Pfeffer's law eqn.(2) in diluted solutions [192], and by the eq.(5), the volume dynamics is directly linked to the molecular species dynamics under isotonic and isothermal conditions. Consequently, the cell dilution 'D' results as a sum of reacting rates of all cell species (individual or lumped). The ( RT/ π ) term can be easily deduced in an isotonic cell system, from the fulfilment of the following invariance relationship derived from eqn.(2):

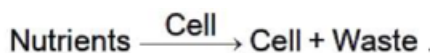
$$V(t) = \frac{RT}{\pi} \sum_{j=1}^{ns} n_j(t) \Rightarrow \frac{RT}{\pi} = \frac{V(t)}{\sum_{j=1}^{ns} n_j(t)} = \frac{1}{\sum_{j=1}^{ns} C_j} = \frac{1}{\sum_{j=1}^{ns} C_{j0}} = \text{constant} \quad (6)$$

The basic hypotheses of the WCVV dynamic models of type Eqs. (3-6) are briefly presented in the (Table 4-1), and in the below sub-paragraph. These formulations are valid over ca. 80% of the cell cycle representing the balanced cell growth before its division [113]."

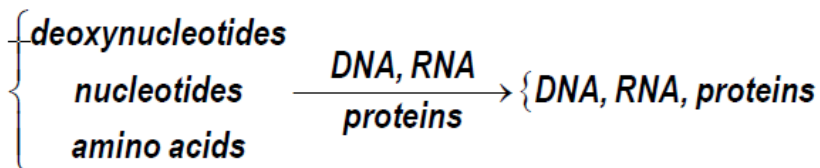
### WCVV deterministic model hypotheses

"Life at its simplest level involves two major divisions of interacting molecular species called the *cell* and the *environment*. The *environment* consists of molecules dissolved in water and largely separated from the cell. In their simplest form, cells consist of hydrophilic molecules in aqueous volumes (*cytosol*), encapsulated by semi-permeable hydrophobic *membranes* composed of phospholipids and proteins.

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



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



DNA and protein are co-catalysts for RNA synthesis from ribonucleotides. RNA and protein are co-catalysts for the synthesis of proteins from amino acids. DNA, RNA, and proteins are co-catalysts for the synthesis of DNA from deoxyribonucleotides. The substrates for these processes (deoxyribonucleotides, nucleotides, amino acids etc.) are *metabolites*, synthesized from imported environmental *nutrients* through complex metabolic pathways [193].

The whole chemical/biochemical cell processes are called '*cell metabolism*', defined as: Metabolism is the set of life-sustaining chemical transformations within the cells of living organisms. The three main purposes of metabolism are the conversion of food/fuel to energy to run cellular processes, the conversion of food/fuel to building blocks for proteins, lipids, nucleic acids, and some carbohydrates, and the elimination of nitrogenous wastes. These enzyme-catalyzed reactions allow organisms to grow and reproduce, maintain their structures, and respond to their environments [193].

The basic equations and *hypotheses* of a *deterministic* WCVV simplified cell model (with continuous variables) presented in this work, also called a '*mechanical cell*', are the followings [21-24,60] (Figure 4-11):

a) 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*.

b) The membrane, of negligible volume, presents a negligible resistance to nutrient diffusion; the membrane dynamics being neglected in the cell model, being assumed to follow the cell growing dynamics. The cell is placed in an environment with which it interacts through a semi-permeable membrane.

c) Changing nutrient levels in the environment. Genes (generically denoted by G), and proteins (generically denoted by P) are in a mutually autocatalytic relationship. Variable cell volume and constant osmotic pressure. Regulatory mechanisms to achieve internal homeostasis to be further explained and detailed (chap. 4.2, 4.3, 4.4, 4.5).

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

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

f) 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' [21-24,175,176].

g) The inner osmotic pressure ( $\pi_{\text{cyt}}$ ) is constant, and all time equal with the environmental pressure ( $\pi_{\text{env}}$ ), thus ensuring the membrane integrity ( $\pi_{\text{cyt}} = \pi_{\text{env}} = \text{constant}$ ). As a consequence, the isotonic osmolarity under isothermal conditions leads to the equality  $RT/\pi_{\text{cyt}} = RT/\pi_{\text{env}}$ . Otherwise, the osmosis will eventually lead to an equal osmotic pressure ( $\pi_{\text{cyt}} = \pi_{\text{env}}$ ). Even if, in a real cell, such equality is approximately fulfilled due to perturbations and transport gradients, and in spite of migrating nutrients from environment into the cell, the overall environment concentration is considered to remain unchanged. On the other hand, species inside the cell transform the nutrients into metabolites and react to make more cell components. In turn, increased amounts of polymerases are then used to import increasing amounts of nutrients. The net result is an exponential increase of cellular components in time, which translates, through isotonic osmolarity assumption, into an exponential increase in volume with time [  $V = V_0 \exp(+Ds \cdot t)$  ] [21-24,60,175,176].

h) The cell content reports a continuous dilution, that is a species concentration decline due to the continuous increase of the denominator of the expression  $C_j = n_j(t)/V(t)$ . In spite of that, concentrations of key species remain constant because the numerator (copynumbers) increases at the same rate with the denominator. So, the overall concentration of cellular components is time-invariant at the homeostasis.

i) Species concentrations at the cell level are usually expressed in nanomoles, being computed with the relationship of equation 7:

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

j) where  $N_A$  is the Avogadro number. For instance, for an *E. coli* cell with an approximate volume of  $V_{\text{cyt},0} = 1.66 \cdot 10^{-15}$  L [191], concentration of one generic gene G copynumber is:  $[G]_s = (1/(6.022 \times 10^{23}))(1.66 \times 10^{-15}) = 1 \text{ nM}$  (that is  $10^{-9}$  mol/L).

k) 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 definition,  $D = d(\ln(V))/dt$ , in eq.(5) ]. Under stationary growing conditions, that is a constant D over the cell cycle.

l) Integration of this relationship indicates an exponential increase of the cell volume, that is [  $V = V_0 \exp(+Ds \cdot t)$  ].

m) Under stationary growing conditions, species synthesis rates ( $r_j$ ) must equal to first-order dilution rates ( $D_s C_{j,s}$ ), leading to time-invariant (index 's') species concentrations  $C_{j,s}$ , i.e. the homeostatic conditions (corresponding to a balanced steady-state growth). Under QSS cell growing conditions, the ODE model mass balance eq. (3A) is leading to the following nonlinear algebraic mass balance set:

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

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

n)  $j = 1, \dots, ns$  (no. of species). This QSS mass balance eq.(8) is used to estimate the rate constants 'k' from the known (from experiments) stationary species concentration vector  $C_s$ , with also imposing some optimal properties of the cell system. Some examples are given by [1,2,21-24,40,60,78,107,174-176,189,194].

o) 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 averages 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 [21,23].

p) A metabolic kinetic model in a WCVV approach should be written in the form eq. (3-6). In such a formulation, all cell species should be considered (individually or lumped), because all species net reaction rates contribute to the cell volume increase (eq. (6)). As the cell volume is doubling during the cell cycle, this continuous volume variation cannot be neglected.

The simplest representation of the core of such a 'mechanical cell' is shown in (Figure 4-12-down-right). It exists in an environment consisting of two nutrients  $Nut_g$  and  $Nut_p$ . The cell contains one gene (lumped genome), and protein (lumped proteome) [40] in a mutually autocatalytic relationship, two lumped metabolites ( $Met_g$  and  $Met_p$ ) used in the synthesis of the G/P pair, and various regulatory elements promoting internal homeostasis. A membrane is presumed to demarcate the cell from its environment but is not an explicit component of the system."

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

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

Relationships (5-6) are important constraints imposed to the **WCVV** cell model (3A-B), eventually leading to different simulation results compared to the **WCCV** kinetic models that neglect the cell volume growth and isotonic effects (see some examples given by Maria [21-24,60,175,176]). On the contrary, application of the default classical **WCCV-ODE** kinetic models of eqns. (1A-B) type with neglecting the isotonicity constraints presents a large number of inconveniences, related to ignoring lots of cell properties (discussed in detail in [21-24,60,175,176]), that is:

- a) the influence of the cell ballast in smoothing the homeostasis perturbations;
- b) the secondary perturbations transmitted via cell volume following a primary perturbation;
- c) the more realistic evaluation of **GERM-s** regulatory performance indices ( **PI-s**, see below chap. 4.4);
- d) the more realistic evaluation of the recovering/transient times after perturbations;
- e) loss of the intrinsic model stability;
- f) loss of the self-regulatory properties after a dynamic perturbation, etc.

"When applied to model **GRC-s** (see chap. 4.3, 4.4, and 5), the **WCVV** modelling hypotheses described in (Table 4-1) must include some constrains referring to the optimality of cell metabolic processes,that is:

- a) Reaction rates must be maximal, but with rate constants limited by the diffusional processes;
- b) The total enzyme (proteine) content of the cell is limited by the isotonicity condition (i.e. constant osmotic pressure under Pfeiffers'law for diluted solutions, eq. 2);
- c) As a corollary of the above constraint, the species differential mass balances must be written under the variable cell volume constraint (eq.3A-B);
- d) Also the total cell energy (ATP) and reducing agent (NADH) resources are limited (see for instance the **HSMDM** model of chap. 6);
- e) The reaction intermediate level must be minimum; The cell model at homeostasis must be stable, that is will reach the steady-state after termination of a perturbation, see [21-24,60,73,78].

f) In math terms, for an **ODE** cell model, of the form eqn.(3A-B), the above cell model stability constraint at homeostasis translates in the condition that the real part of all  $\lambda(i)$  (i) must be negative, that is:  $Re(\lambda(i) (i)) < 0$  for all 'i'. Here, the

eigenvalues  $\lambda(i)$  (i) of the Jacobian matrix  $J_c = (\partial h(C, k) / \partial C)$ , are evaluated at a checked quasi-steady-state (**QSS**) of cell species concentrations (**Cs**). The Jacobian matrix elements refers to the **WCVV** model eq.(3A), that is:  $J(i,k) = \partial h(i)(C, k) / \partial C(k)$ . where  $h(i)$  are the right-side functions of the **ODE** cell kinetic model eqn.(3A-B), detailed as:

g)

$$\frac{1}{V(t)} \frac{dn_j}{dt} = \sum_{i=1}^{nr} s_{ij} r_i(n/V, k, t) = h_j(C, k, t) \quad (9)$$

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

where notations are the followings: **C(j)** = (cell-)species **j** concentration; **V** = system (cell cytosol) volume; **n(j)** = species **j** number of moles; **r(j)** = **j**-th reaction rate; **s(i,j)** = stoichiometric coefficient of the species '**j**' (individual or lumped) in the '**i**'-th reaction; **t** = time; **k** = rate constant vector; **i** = 1, ..., nr (no. of reactions).

- i) It is self-understood that, as  $\text{Max}(\text{Re}(\lambda(i))) < 0$  is smaller as this cell **QSS** is more stable.
- ii) The key-species concentrations must be constant at **QSS** (homeostasis).

In fact, the cell metabolism optimality derives from the requirement to get a maximum growth / quick replication over a defined / limited cell cycle, by using minimum resources from the environment. These requirements are better illustrated by the 5 main characteristics of the cell systems, below underlined in the outline 4-1.

Amazing, but the first pioneers in dynamic modelling of biological systems were not the (bio)chemical engineers which are better trained in 'translating' from the 'language' of molecular biology to that of mechanistic (bio)chemistry, by preserving the structural hierarchy and component functions. The first dynamic models of some cell processes have been reported by the electronists [195,196]. Later, such 'electronic circuits-like' models have been extensively used to understand intermediate levels of regulation [103], but they failed to reproduce in detail molecular interactions with slow and continuous responses to perturbations and, eventually, they have been abandoned. However, the electronists underlined the main characteristics of the cell systems, which must be included in any simulation model (Figure 4-20, and Outline 4-1):

**Outline 4-1.** The main characteristics of the cell systems.

- a) The **dynamic character** of species interactions and processes [140].
- b) The **feedback/feedforward** character of processes ensuring their selfregulation [197].
- c) **Optimal regulation** of cell syntheses with fastest reaction rates, smallest amounts of intermediates, and best P.I.-s (chap. 4.4) [140], with (d),
- d) **Consuming minimum of resources** (nutrients/substrates), and cell energy (ATP, NADH, etc.), but with (e),
- e) Ensuring **maximum reaction rates** [73].

All these cell metabolic characteristics will be accounted in all the subsequent cell *in-silico* WCVV simulators based on extended/reduced mathematical models. All these characteristics are in fully agreement to the Darwin theory "Living organisms have evolved to maximize their chances for survival". It explains structures, behaviors of living organisms. (Figure 4-21). From such very incipient efforts to model-based design GMO-s, 40 years latter [196] pointed-out the tremendous advanced in the Systems Biology and *in-silico* design of GMO-s, or even tissues, by means of computational systems biology [198,199].

A review of *mathematical model types (including the WCVV models)* used to describe metabolic processes is presented by [20-24,188]. Each model type presents advantages but also limitations. Roughly, to model the complex metabolic regulatory mechanisms at a molecular level, two main approaches have been developed over decades: a structure-oriented (*topological*) analysis, and the dynamic (*kinetic*) models [21-23,179,200]. Each theory presents strengths and shortcomings in providing an integrated predictive description of the cellular regulatory network, as briefly reviewed by Maria [21-24] .

Structure-oriented analyses or *topological models* ignore some mechanistic details and the process kinetics, and use the only *network topology* to quantitatively characterize to what extent the metabolic reactions determine the fluxes and metabolic concentrations [73]. The so-called ‘metabolic control analysis’ (MCA) is focus on using various types of sensitivity coefficients (the so-called ‘response coefficients’), which are quantitative measures of how much a perturbation (an influential variable) affects the cell-system states [e.g. reaction rates, metabolic fluxes (stationary reaction rates), species concentrations] around the steady-state (QSS). The systemic response of fluxes or concentrations to perturbation parameters (i.e. the ‘control coefficients’), or of reaction rates to perturbations (i.e. the ‘elasticity coefficients’) have to fulfil the ‘summation theorems’, which reflect the network structural properties, and the ‘connectivity theorems’ related to the properties of single enzymes in connection to the system behaviour.

MCA methods are able to efficiently characterize the metabolic network robustness and functionality, linked with the cell phenotype and gene regulation. MCA allows a rapid evaluation of the system response to perturbations (especially of the enzymatic activity), possibilities of control and self-regulation for the whole path or some subunits. Functional subunits are metabolic subsystems, called ‘modules’, such as amino acid or protein synthesis, protein degradation, mitochondria metabolic path, etc. [182]. Because the living cells are self-evolutionary systems, new reactions recruited by cells together with enzyme adaptations can lead to an increase in the cell biological organisation and to optimal performance indices. When constructing methods to optimize evolutionary metabolic systems, MCA concepts and appropriate performance criteria have been used, leading to: maximize reaction rates and steady-state fluxes; minimize metabolic intermediate concentrations; minimize transient times; optimise the reaction stoichiometry (network topology); maximize thermodynamic efficiency. All these objectives are subjected to various mass balance, thermodynamic, and biological constraints [73]. However, by not accounting for the system dynamics, and grounding the analysis on the linear system theory, topological methods presents inherent limitations (see for instance some violations of stoichiometric constraints discussed by Atauri et al. [201], or the use of modified control coefficients [202]).

From the mathematical point of view, various structured (mechanism-based) dynamic models have been proposed to simulate the metabolic processes and their regulation, accounting for continuous, discrete, and/or stochastic variables, in a modular construction, ‘circuit-like’ network, or compartmented simulation platforms [24,181,200]. Briefly, the math models used by Systems Biology are of the following types [21-24]

a) *Deterministic continuous variable* dynamic models can perfectly represent the cell response to continuous perturbations, and their structure and size can be easily adapted based on the available *bio- omics* information [21-24,73,108,188,200]. *Deterministic continuous* variable kinetic models present a large number of advantages, as previously mentioned. Besides, it is to underlined the huge advantages coming from the used concepts, rules, and algorithms of (Bio)chemical engineering and nonlinear system control theory, as discussed by [21-23,60,74,177,178], and briefly represented in (Figure 4-2, and Figure 4-3), and by issues (a-j) at the end of chap. 4.1.

Classical approach to develop *deterministic dynamic models* is based on a hypothetical reaction mechanism, kinetic equations, and known stoichiometry. This route meets difficulties when the analysis is expanded to large-scale metabolic networks of the CCM (Figure 4-23) because the necessary mechanistic details and standard kinetic data to derive the rate constants are difficult to be obtained. However, advances in genomics, transcriptomics, proteomics, and metabolomics, lead to a continuous expansion of bioinformatic databases, while advanced numerical techniques, non-conventional estimation procedures, and massive software platforms reported progresses in formulating such reliable cell models. Valuable *structured dynamic models*, based on cell biochemical mechanisms, have been developed for simulating various (sub)systems [21-24].

**Boolean (discrete) variable models.** Such models works with a topological structures of the GRC-s. An example is displayed in (Figure 4-22, and Figure 4-5) [180,200]. Due to the very large number of states  $O(10^3 - 10^4)$ , and  $O(10^3)$  of transcriptional factors (TF) involved in the gene expression, such GRC models are organized in clusters, modules, of a multi-layer structures (Figure 4-22) [77,200].

In the *Boolean modelling approach*, variables can take only discrete values (usually 0, and 1). Even if less realistic, such an approach is computationally tractable, involving networks of genes that are either ‘on’(1), or ‘off’(0) (e.g. a gene is either fully expressed or not expressed at all; Figure 4-22) according to simple Boolean relationships, in a finite space. Such a coarse representation is used to obtain a first model for a complex bio-system including a large number of components, until more detailed data on process dynamics become available. ‘Electronic circuits’ structures (see an example in (Figure 4-24), from

[103,104]) have been extensively used to understand intermediate levels of regulation, but they cannot reproduce in detail molecular interactions with slow and continuous responses to perturbations, and they offer any information on the process / species concentration dynamics.

**b) Stochastic variable models** [91,203,204]. *Stochastic models* replace the ‘average’ solution of continuous-variable (deterministic) ODE kinetics (e.g. species concentrations) by a detailed random-based simulator accounting for the exact number of molecules present in the system. Because the small number of molecules for a certain species (present in traces in a cell) it is more sensitive to stochasticity of a metabolic process than the species present in larger amounts, simulation via continuous models sometimes can lack of enough accuracy for random process representation (as cell signalling, gene mutation, etc.). Monte Carlo simulators are used to predict individual species molecular interactions, while rate equations are replaced by individual reaction probabilities, and the model output is stochastic in nature. Even if the required computational effort is extremely high, stochastic representation can be useful sometime to simulate the cell system dynamics by accounting for a large number of species of which spatial location is important, or for a large distribution of concentrations [91,203,204] (Figure 4-24).

**c) Mixed variable** models [149,150,200]. Such models try to take advantages from each of the model type (a-c) above mentioned.

The multiple advantages of the WCVV modelling framework are discussed, and exemplified by Maria [21-24,60,74,107,175,176]. In short, the novel modelling concept/framework WCVV proposed, extrapolated, and widely promoted in a large number of applications by Maria [21-24,40,60,74,78,170,175,176,177,178] to derive cell kinetic models, in a holistic approach, ensures cell processes homeostasis, and the individual/holistic GRC regulatory properties, by including in a natural way constraints related to the cell *system isotonicity*, and the *variable-volume* in relationship to the species reaction rates, and the *lumped proteome/genome* replication [21-24,40,60,175,176]. Such an *isotonicity constraint* is required to ensure the cell membrane integrity, but also to preserve the homeostatic properties of the cell system, not by imposing ‘the total enzyme activity’, or the ‘total enzyme concentration’ constraints used by the classical (default) constant-volume cell modelling approach (WCCV). As proved by Maria et al.[60], compared to the classical WCCV models, the WCVV novel modelling framework is leading to more accurate simulate of cell metabolic effects, such as: relationships between the external conditions, species net synthesis reactions, osmotic pressure, cell content (ballast) influence on smoothing the continuous perturbations in external nutrient concentrations, the more realistic representation of GERM regulatory modules, etc.

Besides, as shortly presented in the (Table 4-2) the WCVV holistic modelling framework proposed, extrapolated, and widely promoted in a large number of applications in bioengineering and bioinformatics by Maria [21-24,40,60,74,78,170,175,176,177,178] is proved to be more accurate and present a large number of advantages.

Table 4-2. Some advantages when using the holistic WCVV framework when modelling GRC-s [21-24,175,176].

## Lumping deterministic kinetic models

“To not complicate too much the **WCVV** dynamic model (**Figure 4-14**), usually a reduction in the number of cell species and reactions by common lumping rules of (bio)chemical engineering [29,64,67,78] is usually performed. Such a model reduction strategy of the metabolic kinetic models present a series of disadvantages, such as: (a) a loss in model adequacy; (b) a loss in the simulated system flexibility (due to the reduced number of considered intermediates and species interactions); (c) an increased possibility to get multiple (rival) reduced models of proximate characteristics for the same cell system, difficult to be delimited; (d) a loss in the model prediction capabilities; (e) lumped model parameters can lack of physical meaning; (f) a loss / alteration of systemic / holistic properties (e.g. cell system stability, multiplicity, sensitivity, regulatory characteristics). In spite of that, lumped deterministic models can successfully simulate a broad range of metabolic processes [21-24]. The (**Figure 4-19**) summarizes the main reasons to reduce an extended cell kinetic model.

To avoid large deterministic models, difficult to be identified (due to the lack of experimental structured kinetic information), and difficult to be used, a *lumping procedure* should be applied (**Figure 4-15**).

Reduction in the model structure (via lumping of species, and/or reactions) is necessary due to the following reasons [21-24, 29,64,67,78]



- a) the high complexity of cell metabolic processes vs. available data
- b) large number of species, reactions, transport parameters, and interactions
- c) low data observability & reproducibility
- d) metabolic process variability
- e) interpretable representation of cell complexity
- f) requirement to get quick simulations of cell behavior under various environmental conditions
- g) computational tractability and easier application of algorithmic lumping rules of extended kinetic models imported from the (bio)chemical engineering numerical rules, and from the nonlinear systems numerical treatment.

However, a tradeoff between model complexity and adequacy must be maintained [29,78], when such models are used for the *in-silico* design of **GMO-s**, by *in-silico* re-programming the cell metabolism [3,12,42,205,] or by optimal cell cloning [1,2,40]. 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 [206]. The **ODE** model Jacobian matrix is defined as the derivatives of model functions in respect to model state-variables, that is species concentrations the in cell metabolic models; see its calculation relationships before eq.(9)]

This classic approach to develop dynamic reduced models is based on the biochemical reaction engineering rules, that is: propose a hypothetical reaction mechanism, formulate kinetic equations, and reaction stoichiometry, and try to validate them on an experimental basis. This route meets difficulties when the analysis is expanded to large-scale metabolic networks, because the necessary mechanistic details and standard structured kinetic data to derive the rate constants are difficult to be obtained. However, advances in genomics, transcriptomics, proteomics, and metabolomics, lead to a continuous expansion of bioinformatic databases, while advanced numerical techniques, non-conventional estimation procedures, and massive software platforms, and lumping algorithms [24,29,64,78], reported progresses in formulating lumped (reduced) adequate cell dynamic models. Valuable *structured dynamic* models, based on cell biochemical mechanisms, have been developed for simulating various (sub) systems (see examples of Maria [21-24].

However, here it is to mention, that the work with *reduced kinetic models* of cell **CCM**-metabolic syntheses, and of **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).

Here can be mentioned only a few of the classical chemical engineering rules used for *reducing an extended kinetic model* [29,64,67,78](**Figure 4-15**):

- a) 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.)
- b) Reduce the list of species, by eliminating unimportant components and/or lumping some species, by using various measures [29,207], e.g. small values for the product of the target '*i*'-th species lifetime  $LT(i) = -1/J(i,i)$ , and its production rate  $r(i)$ , where the Jacobian matrix elements  $J(i,i)$  refers to the **WCVV** model eq.(3A), that is:  $J(i,k) = \partial h(i)(C,k) / \partial C(k)$  where  $h(i)$  are the right-side functions of the extended **ODE** cell kinetic model eqn.(3A-B), detailed in eqn.(9). Decompose the kinetics into fast and slow 'parts' allowing application of the quasi-steady-state-approximation (**QSSA**) to reduce its dimensionality [29,64,67].
- c) When the **ODE** kinetic model is linear in parameters, then the reduction procedure of Maria [29,64,67,78], can be successfully applied by preserving the **ODE** model invariants (that is the eigenvalues and the eigenvectors of the model Jacobian) [67].

d) Use of mixed integer **NLP** estimation rules (**MINLP**), to concomitantly estimate the rate constants and reduce the **ODE** model structure, by replacing the rate constants  $[k(j)]$  with the terms  $[k(j)*y(j)]$ , where the binary variables  $y(j) \in [0,1]$  have also to be estimated in order to retain the significant kinetic terms, by eliminating the rate constants corresponding to  $y(j) = 0$  [29,65]. In such a way, the obtained reduced **ODE** model will also ensure a satisfactory model adequacy vs. the considered experimental data.

e) If the model structure is too extended vs. the available information, besides computational methods, various experimental techniques can be applied for reducing the structure/size of the ODE kinetic model [29]. Experimental rules can point-out or '**mask**' intermediate species and/or steps. Computational rules imply kinetic model estimation, study of the effect of parameter changes on the model solution (sensitivity analysis), and identification of the redundant parts of the model or variables [29,64,67,78].

The current trend in kinetic modelling of cell processes is to use more structured and complex strategies, by taking into account constraint representation, algorithms for model development and pathways synthesis, by using all types of biological information, conventional or not, and assembling suitable reactions and kinetic modules from databanks. Complex software is now able to realize integrated platforms for model synthesis, parameter estimation, model reduction and discrimination, by using the (bio)chemical engineering concepts and tools [21-23,74]. The modular approach and '**automatic**' generation of **ODE**, differential-algebraic (**DAE**), or stochastic models, allow simulation of complex chemical systems [208,209,210,211-217] or biochemical systems by using modern software such as (MPS, MetaModel, GEPASI, ESSYNS, METASIM, ProMoT/DIVA, BioSpice, Cellerator, Dbsolve, Jarnac, StochSim), see the review of Hucka et al. [218].

Oriented and unified programming languages have been developed (CellML of Hedley et al. [172]; SBML of Hucka et al. [92,218]) to include the bio-system organization and complexity in integrated platforms for cellular sub-systems simulation. These platforms include representation of cells, neurons, bioinformatic sequences, bio-polymer sequences, complex molecular structures, gene expression, gene-finding (E-Cell of Tomita et al. [51,52]; V-Cell of Moraru et al. [99], and of Slepchenko et al. [53]; M-Cell of Bartol and Stiles [102]; A-Cell of Ichikawa [103,104]. Integrated modelling and simulation platforms tend to use a large variety of chemical and biological **bio-omics** databanks, including physico-chemical properties, species biodegradability [219,220], reactions in diluted solutions [221], properties of enzymes, proteins, genes, metabolic reactions, etc. (**CRGM**-database [105]; **NIH**-database [222]).

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 [78]). Such a rule allows disassembling the whole system in parts (model functional 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.[21,67]

The math modelling efforts have intensified a lot after 2000 when the human genome has been deciphered (**Figure 4-17**), being proved that the difficult task to model and design complex biological circuits with a *building blocks strategy* can be accomplished by properly defining the cell basic components, functions, and structural organisation (**Figure 4-18**). Because many cell regulatory systems are organized as '**modules**' [223], it is natural to model **GRC**-s and other metabolic processes by using a **modular** approach [21-24]. Further analyses including engineered **GRC**-s can lead to simulate and design of **GMO**-s, of desirable characteristics, that is [40,81,107]: 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**' by the modified **GRC**-s in **GMO**-s, one may create potent applications in industrial, environmental, and medical fields (e.g. biosensors, gene therapy). Valuable implementation tools of the design **GRC**-s in real cells have been reported over the last years [79,184,224].

The **WWVC** modelling framework will be further exemplified when developing the **HSMDM** for the case study (**no.1**). including a complex **GRC** with 7 linked **GERM**-s of the *mer*-operon in the *E. coli* cell.

By using the **WCVV** modelling framework, structured reduced dynamic models of various complexity have been developed to simulate individual **GERMs**, [21-24,175,176] but also linked **GERM**-s in **GRC** of modulated functions (e.g. toggle-switch, amplitude filters, modified operons, etc.), used to design **GMOs** of practical interest, such as genetic switches [107], or optimally *in-silico* cloned *E. coli* cells for a better uptake of mercury from waste waters [40].”

### Modelling individual GERM-s under WCVV formulation

In order not to overly complicate the **HSMDM** models that also include **GRC**-s, it is necessary to have a “library” of kinetic models to represent individual **GERM**-s, to be used for build-up **GRC**-s of desirable properties (e.g. genetic switches, operon expression, etc.). See the **case study no. 1** (chap. 5) as an example. Obviously, the selection of the most suitable **GERM** to be included in the **GRC** chain depends on its regulation performances (that is, the so-called self-regulation performance indices (**P.I.**, chap. 4.4), related to the **GERM** type.

This chapter briefly presents the main **GERM**-s models proposed by Maria [21-24, 185] used in the construction of **HSMDM**-s, in terms of the reduced reaction scheme, kinetic model, and their associated **P.I.**-s.

As experimentally proved in the literature (reviews [21-24]), the **GRC**-s (or **GRN**-s) „that control the synthesis of all proteins (enzymes) in the cell, present a modular construction, every operon (a cluster of genes under the control of a single promoter) including a variable number of interacting **GERM**-s. However, it is wellknown that one **GERM** interacts with no more than other 23-25 **GERM**-s [225], while most of **GERM** structures are repeatable. Consequently, when developing the **GRC** analysis and reaction schemes / kinetic model, the modular approach is preferred due to several advantages: **(i)** A separate analysis of the constitutive **GERM**-s in conditions that mimic the stationary or perturbed cell growth; **(ii)** The **GERM** modules are then *in-silico* linked to construct the target **GRC** of an optimized regulatory efficiency that ensures key-species homeostasis and cell network holistic properties (**Figure 4-16**, **Figure 4-25**, and **Figure 4-5**). **(iii)** *In-silico* investigations of **GERM**-s and **GRC**-s characteristics focus on the tight control of gene expression, the quick dynamic response, the high sensitivity to specific inducers, and the **GRC** robustness (i.e. a low sensitivity vs. undesired inducers). Such advanced regulatory structures must ensure the homeostasis (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 metabolites or nutrients [21-24, 185] (**Figure 4-26**, and **Figure 4-27**).

To model complex **GERM**-s, intensive efforts have been invested over the last decades, and various types of dynamic models have been proposed, both in a deterministic [20,25,26,107,173,186-188,192], or stochastic approach [78,95,173,192,225]. See also the reviews [25,174,190,226] concerning structured deterministic models with using continuous variables, built-up from time-series experiments [191,227].

However, to not complicate the resulted simulation model when coupling **GERM** chains in complex **GRC**-s, simple **GERM** dynamic models have been proposed and investigated by various researchers [21-24,26,107,187,188,190,223], as displayed in (**Figure 4-9**) , with Hill-type [107,174,189,196], or pseudo-Hilltype [107,174] activation (see **Figure 4-4**).

To make this rule easier, Maria [21-24] elaborated a library with reduced representations of **GERM**-s (**Figure 4-7**) to be used for every particular case. Of course, these individual **GERM** modules differ by the regulatory performance indices (**P.I.**) to be further defined, in response to stationary or dynamic perturbations into the cell, or in the environment.

These simplified deterministic models of lumped **GERM** and structures have been proved to adequately represent complex **GRC**-s (**Figure 4-5**, or **Figure 5-1**). The simplest **GERM** structure with one regulatory element is those denoted by **G(P)1**, or of better regulatory efficiency **G(PP)1** (**Figure 4-9**).

The generic [**G(P)1**] regulatory module (schematically represented in (**Figure 4-28**-down-right, and **Figure 4-7**, the row-up) refers to the synthesis of a generic protein **P** and the simultaneous replication of its encoding gene **G**. The lumped **G(P)1** model includes only one regulatory element (a so-called “effector”, that is a fast ‘buffer’ reversible reaction  $G + P \rightleftharpoons GP(\text{inactive})$ ) (**Figure 4-7**), aiming at controlling the **P** synthesis rate and its homeostatic **QSS** level. The following notations have been used: **G** = active part of the gene encoding protein **P**; **GP** = inactive part of the gene encoding protein **P**; **MetG**, **MetP** = lumped DNA and protein precursor metabolites, respectively.

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

The module nomenclature used in (Figure 4-7) for such **GERM** models, proposed by Yang et al. [228], and by Maria [24] is those of  $[L_i(O_i)n_i ; \dots ; L_i(O_i)n_i]$  (Figure 4-29). It includes the assembled regulatory units  $L_i(O_i)n_i$ . One unit 'i' is formed by the component **L(i)** (e.g. enzymes or even genes **G**, **P**, **M**, etc.) at which regulatory element acts, and **n(i)** = 0,1,2,... number of 'effectors'/**TF**, generically denoted by species **O(i)** (that is 'effectors' like **P**, **PP**, **PPPP**, **R**, **RR**, **RRRR**, etc.) binding the 'catalyst' **L**. For instance, a **[G(P)2]** unit of (Figure 4-7) includes two successive binding steps of **G** with the product **P**, that is  $G + P \rightleftharpoons GP + P \rightleftharpoons GPP$ , all intermediate species **GP**, **GPP**, being inactive catalytically, while the mass conservation law is all time fulfilled, i.e.  $\sum_{i=0}^{\infty} [G(P)] = \text{constant}$ . Such a representation accounts for the protein concentration diminishment due to the cell-growth dilution effect, but could also include protein degradation by proteolysis. It is also to observe that such **GERM** lumped models try to account essential properties of the gene expression, that is a highly self- / cross- regulated and mutually catalyzed process by means of the produced enzymes / effectors. As depicted in (Figure 4-7) for the **[G(P)1]** module case, the protein **P** synthesis is formally catalysed by its encoding gene **G**. In turn, **P** protein formally catalyse the **G** synthesis, but also modulate the **G** catalyst activity (via the fast buffering reaction  $G + P \rightleftharpoons GP$ ).

Even if such a generic **[G(P)1]** regulatory module is more complex, by including a large number of reactions involved in the regulation of the gene expression (schematically represented in Figure 4-29), it was proved (see the case study no. 1 of chap. 5) to satisfactory reproduce the dynamics of complex **GRC**-s in **HSMDM** models. As proved by Maria [21-24,78] this simplified formulation of **GERM**-s in (Figure 4-7) implicitly ensures the *homeostatic regulation* of the gene expression and **G/P** mutual auto-catalysis of their synthesis.

Cells are regulated such that their components are maintained at relatively *invariant* concentrations despite the presence of inherent external/internal perturbations. Recently, the effectiveness of various simplistic regulatory mechanisms in maintaining protein homeostasis in the presence of perturbations have been evaluated by [21-24,170,228,229]. Some of these representative **GERM** mechanisms, representatives of which are shown in (Figure 4-49), reflect synthesis (transcription and translation) and decay of a generic protein **P**. Both processes are required to maintain protein [**P**] concentration at a *nominal* steadystate [**P**]**ns** (**ns** = nominal steady-state).

Protein synthesis involves many components, but in our mechanisms, all except for the generic key-gene **G** encoding **P** (and in some cases the corresponding mRNA) are ignored and assumed to be present at constant concentrations. In the '**control**' mechanism, denoted with **[G(P)0]**, **P** is synthesized in a single reaction catalyzed by **G** (Figure 4-49, mechanism no. 1). In a constant-volume **WCCV** model, [185,229] considered a **P** degradation by a first-order reaction in **P**. In real cells, proteins are generally stable, and so these *fictitious* decay terms actually reflect the cell dilution ('*D*') caused by volume expansion as cells grow, that is eq.(3a-b), eq.(4a-b), eq. (5).

Other mechanisms included negative feedback regulatory elements in various combinations. In mechanism **[G(P)1]**, one **P** binds **G** reversibly (Figure 449, mechanism no. 2). The resulting **GP** form is catalytically inactive, and so this relationship serves to regulate protein synthesis. The dissociation equilibrium constant is set to equal [**P**]**ns**, ensuring that **[G]**ns** = **[GP]**ns**** at [**P**]**ns**. Thus, when [**P**] > [**P**]**ns**, then **P** tends to bind more **G** and attenuate protein synthesis. Conversely, when [**P**] < [**P**]**ns**, then **P** tends to dissociate from **GP**, thereby increasing the rate of protein synthesis. This arrangement leads to optimal regulatory effectiveness. Mechanism **[G(P)1;M(P)1]** distinguishes between transcription and translation (Figure 4-49, mechanism no. 3). In it, **G** catalyzes the synthesis of **M** (i.e. mRNA), and **M** catalyses the synthesis of **P**. Also included is the reversible binding of **P** to **M**, stimulating the degradation of **M** to form **M'**. In the mechanism **[G(PP)2]**, two copies of a **PP** dimer reversibly bind **G** (Figure 449, mechanism no. 4), mimicking the binding of transcription factors which often bind promoters as oligomers and in multiple copies.**

As revealed by (Figure 4-50), the efficiency of all these four (no. 1-4) **GERM** mechanisms is very good when coping with a

dynamic perturbation, that is a 10% negative perturbation (impulse-like) of **[P]** from **[P]ns = 1000 nM**, to **[P]ns = 900 nM**. It is to remark that the species recovering trajectories are as faster as the GERM efficiency schema is better, that is in the relative order: **[G(PP)2] > [G(P)1;M(P)1] > [G(P)1] > [G(P)0]** A more detailed discussion on this subject is offered by Maria [21,24,185], by highlighting the role of the number of effectors on the **GERM P.I.-s** (chap.4.4)."

### Rate constant estimation in WCVV models of GERM-s and GRC-s

In the **WCVV** differential models, the large number of rate constants are estimated by using several methods. If the stationary cell species concentration vector  $C_s$  is known from the experimental data (for the all individual or lumped components considered in the kinetic model), then the rate constant vector  $k$  (and even unknown  $C_s$ ) of the kinetic model eq.(3A-B) results by solving the nonlinear algebraic set eq.(8), for every cell subsystem (e.g. a **GERM** or a **GRC**), by using an effective procedure [29,71,72,239], including those of the Maple™ symbolic computing platform. „As the (pi) term is known from the initial condition, and the number of model parameters is usually higher than the number of observed cell species, supplementary optimization rules can be applied to determine some rate constants, by imposing optimum regulatory criteria (see below **P.I.-s** [175,176]) for the **GERM-s**, such as minimum recovering time of the stationary concentrations (homeostasis) after a dynamic ('impulse'-like) perturbation in a key-species [24], and using effective solvers [[29,71,72,239]:

$$\left[ \hat{k}, \hat{C}_s \right] = \arg \text{Min}(\tau_p), \text{ subjected to fulfilment of eq. (8), } \left[ \hat{k}, \hat{C}_s \right] > 0, \text{ and.} \tag{10}$$

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

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

With the following notations: superscript  $[U^{\wedge}]$ = estimated value;  $\tau_p$  = the recovering time of  $[P]_s$  with a tolerance of 1% $[P]_s$  after applying a 10% $[P]_s$  impulse perturbation of the  $[P]_s$ ;  $L_i$  (e.g. enzymes or even genes G, P, M, etc.) is a **GERM** component at which regulatory element O, TF, R acts (**Figure 4-29**). To estimate  $[\hat{k}, \hat{C}_s]$ , other regulatory global properties can also be used together with the constraints eqns. (6,8,10) [24,163]. The reverse reaction rate constants in the rapid buffer reactions of **GERM-s**, of type **G+P**  $\leftrightarrow$  **GP**, are adopted at values five to seven orders of magnitude higher than **D** in eq. (3A-B) (see the proof of Maria [107]). That is because fast buffering reactions are close to equilibrium and have little effect on metabolic control coefficients. As a consequence, rate constants of such rapid reactions are much higher than those of the core synthesis and dilution rates."

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

"As proved in previous works [21-24], the performances indices (**P.I.-s**) of **GERM-s** of **[G(P)n]** type in (**Figure 4-7**), are as better as the number 'n' of buffer reactions increases (**Figure 4-30**). Also, Maria [21-24] proved that when **P** is acting as a **TF**, its efficiency is better if it is present in a dimeric form (**PP**), in **GERM-s** of **[G(PP)n]** type in (**Figure 4-7**, and **Figure 4-30**). Maria [21-24] also proved that the **GERM** regulatory efficiency is better if **TF** is a dimmer **PP** acting at both **G** and **M** levels of the expression (middle and down-rows of **Figure 4-7**), thus developing a cascade control scheme of the expression where transcription and translation regulatory steps are separately considered, that is **GERM-s** of **[G(PP)n;M(PP)n']** type.

It clearly appears (**Figure 4-26**) that, as the number of effectors increases in **GERM-s** as their **P.I.-s** are better. Perturbations of the species steady-state (homeostatic) concentrations are caused by environmental processes. In a **GERM** case, these processes tend to increase or decrease the key-protein stationary level  $[P]_s$ . These processes occur in addition to those of the "core" system (**G/P** pair replication over the cell cycle).

**GERM** or **GRC** regulatory performance indices **P.I.-s** are of two types [21,24] (**Figure 4-31**, and **Figure 4-32**): *stationary* and *dynamic*. Briefly they are presented in the (**Table 4-3**), together with the associated optimization objective (goal), for a general nonlinear dynamic cell model described by eq. (3A-B). See also an intuitive display in (**Figure 4-31**, and **Figure 4-32**) Detailed information is given by Maria [21-24]. The monodromy matrix **A**, necessary to express the species **QSS-level** stability 'strength' is e<sup>t</sup>  $dC / dt = h(C, k); C(0) = C_s; dA / dt = J_C A; A(0) = I; J_C = (\partial h(C, k) / \partial C)$ , differential relationship:

$$\tag{11}$$

**Table 4-3** . The regulatory efficiency performance indices **PI**-s proposed to evaluate the perturbation treatment efficiency by a generic **GERM** of (**Figure 4-9**, or **Figure 4-7**) type, following the definitions of Maria [24]. Abbreviations: Min = to be minimized; Max = to be maximized. Note:  $k(\text{syn})$  and  $k(\text{decline})$  refer to the  $\rightarrow P \rightarrow$  overall reaction. Notations: 'n' = nominal value; 's' = stationary value; (\*) see eq. (11) and [185] for the monodromy matrix **A** calculation;  $\lambda(i)$  = i-th eigenvalues of the Jacobian matrix  $J_c = (\partial h(C, k) / \partial C)_s$  defined before eq.(9); **A** = monodromy matrix, defined by eq. (11);  $\tau(j)$  = species 'j' recovering time of its **QSS**-level, with an accepted tolerance (usually 1-5%); Nut = nutrient;  $\text{Re}(\lambda)$  = real part of '  $\lambda$  '; **AVG** = average; **STD** = standard deviation (st. dev.);  $C_j$  = species 'j' concentration;  $R_p$  = dynamic regulatory (recovering) index (equivalent with the recovering rate after a dynamic perturbation); **QSS** = quasi-steady-state; P denotes the key-protein expressed in the analysed generic **GERM**;  $S_{NutP}^i$  = the sensitivity of NutP(j) vs- concentration C(i) of species 'i'. Adapted after [23].

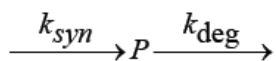
In short, explanations of **PI**-s presented in (**Table 4-3**) are the followings. **A) Stationary PI** are defined as response to a stationary perturbation (**Figure 4-32**), that is transition from a **QSS** concentration to another **QSS** following a 'step'-like perturbation of one cell component concentration.

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 generic **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 also **Figure 4-32**). Equivalently, the **P**-synthesis regulatory module will tend to maintain  $[P]_{ns}$  within certain limits,  $[P]_{min} \leq [P]_{ns} \leq [P]_{max}$ . In this respect, a relative  $R_{ss} = \pm 10\%$  maximum deviation has been proposed by [228,229], (see **Table 4-3**) to get an effective **GERM**. A measure of the species 'i' steady-state concentration ( $C_i, s$ ) 'resistance' to various stationary perturbations { in the rate constants,  $k(j)$ , or in nutrient concentrations,  $[Nut, j]$  } is given by the magnitude of the relative sensitivity coefficients at **QSS**, i.e.  $S(C_i; k_j)$  and  $S(C_i, s; [Nut, j])$  defined in (**Table 4-3**). In general terms, a system state sensitivity vs. A perturbation is defined (in absolute terms) as  $S(\text{state}; \text{perturb.}) = \partial(\text{state}) / \partial(\text{perturbation})$  [231]. Other stationary **PI**-s are discussed by Maria [21-24], that is (**Figure 4-32**):

**A-i).- Transition time** necessary to each **GERM** component to return to their stationary concentration (**QSS**) after a step-like perturbation in one component of the regulatory module (individual or lumped component);

**A-ii).- Responsiveness** to exo/endogeneous signaling species of the analysed **GERM** or **GRC** can be represented by the small transient times necessary for a species 'j' **QSS**-level to reach a new **QSS** (with a certain tolerance, usually 15%) after applying a stationary external stimulus [189]. Consequently, the **PI** measure of the **GERM** efficiency to move fast to a new **QSS** is given by the duration of the *transition time*  $\tau_p$  (in the case of **P**-species in **Figure 4-32**) necessary to a certain component to reach the new steady-state concentration.

Another regulatory **PI**, that is  $A_{unsync} = k_{syn} \times k_{decline}$  (**Table 4-3**), 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 [232]. For instance, in the case of the **P**-species, these rate constants belong to the synthesis and degradation lumped reactions:



**A-iii).- Stationary efficiency.** This stationary **PI** is related to the small sensitivities  $S(C(i); NutP(j))$  of the key-species levels  $C(i)$  vs. changes in the external nutrient levels  $NutP(j)$ , and is simply denoted by  $S_{NutP}^i$ . These sensitivities are computed from solving a sensitivity nonlinear algebraic set obtained by assuming **QSS** conditions eqn. (8) of the **ODE** kinetic model eqn. (3A-B), and known nominal species stationary concentrations  $C_s$ . Then, differentiation of the steady-state conditions eqn. (8) leads to the evaluation of the state sensitivity vs. nutrient levels, i.e.  $S(C(i); Nut(j)) = \partial C_i / \partial Nut(j)$ , (see details of Maria [21-24]).

**A-iv).- The steady-state Cs stability strength** . This **GERM** property is related to the strong capacity of the regulatory systems 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 **PI** -s, this **GERM** property is related to the **GERM** system characteristics. Basically, as  $\text{Max}(\text{Re}(\lambda(i))) < 0$  is smaller as this **QSS** is more stable. Here, the eigenvalues  $\lambda(i)$  of the Jacobian matrix  $J_c = (\partial h(C, k) / \partial C)_s$  (see its definition before eqn.(9)) are evaluated at a checked **QSS** of the species concentration vector (**Cs**).

In a more systematic approach, the steady-state **Cs** 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 [185], by using eq.(11). For a stable **Cs**, i.e.  $|\lambda_{A_i}| < 1$ , as  $|\lambda_{A_i}|$  are smaller,

as the stability of the **Cs** state is stronger and that **QSS** recovers faster after a small dynamic perturbation. Here,  $\lambda_{A_i}$  denotes the eigenvalues of the

$A(T)$  matrix, while  $I$  = identity matrix. In other words, **QSS** stability strength involves:  $\text{Min}(\text{Max}(\text{Re}(l(i))))$ , with  $\text{Re}(\lambda_{(i)}) < 0$  for all 'i', and  $\text{Min}|\lambda_{A_i}| < 1$ .

**B) Dynamic P.I.** are defined in relation to the **GRC** response to a dynamic perturbation (**Figure 4-32**), that is the recover of the **QSS** following an 'impulse'like perturbation in the stationary concentration of one of the cell component.

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 4-32**, and **Figure 4-27** for a generic **P**-protein case). The computed recovering time  $\tau$  (**rec,j**) necessary to each component 'j' to reach-back their stationary concentration (with a tolerance of 15% proposed by Maria, 2005b) 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 [73]. In terms of the evolution and stability of component **QSS** concentrations included in a dynamic cell system expressed by an **ODE** model (1) or (2), these properties can be evaluated from the analysis of the eigenvalues  $\lambda_{(i)}$  (i = no. of species) of the linearized kinetic model Jacobian matrix  $J_c = (\partial h(C,k)/\partial C)_s$ , of elements  $J(i,k) = \partial h(i)(C,k) / \partial C(k)$  defined before eqn.(9). 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, that is  $\text{Re}(\lambda_{(i)}) < 0$ , for all 'i' [73,233,234]. 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 (stationary) perturbation.

Here it is to mention the works of Maria [21-24,185], and of Sewell et al.[229], proving 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  $[G] = [GP]$ ,  $[Gi]=[GiPjPj]$ , of  $[M] = [MP]$ ,  $[Mi]=[MiPjPj]$ , etc. (for the **GERM**-s types displayed in **Figure 4-7**).

The main dynamic **P.I.**-s discussed by Maria [21-24] are the followings (**Figure 4-32**):

**B-a).- recovering time** ( $\tau$  (rec,j), or simply  $\tau$  (j)) necessary to each **GERM**'s component to return to their stationary concentration (**QSS**) after an impulse-like perturbation in one component. This **P.I.** measure of the **GERM** efficiency to fast recover the key-species stationary concentrations is given by the time  $\tau$  (j) necessary to the species 'j' to recover its steady-state concentration (with an assumed tolerance of 1%, as proposed by Maria, 2005b)). As an example, in **Figure 4-27**, and **Figure 4-33** is presented how a simple generic **GERM** of **[G(P)1]** type is better working, by presenting a better dynamic efficiency compared to the simplest  $[G(P)]_0$  gene-expression module. Thus, the stationary **[P]s** and **[G]s** are recovered faster after an impulse perturbation in the **[P]s**, that is a -10% decline of **[P]s** at an arbitrary time  $t=0$  [175,176]. As another example, in (**Figure 4-34**) is presented how the dynamic regulatory efficiency (that is the **QSS** recovering time) depends on the **GERM**'s structure, and its number of effectors (**TF**). Thus, the regulatory efficiency increases in the order: **[G(P)]0** ('1') < **[G(P)1]** ('2') < **[G(P)1;M(P)1]** ('3') < **[G(PP)2]** ('4').

**B-b).- recovering rate**  $R_p$  (**Table 4-3**) necessary to each **GERM** component to return to their stationary concentration (**QSS**) after an impulse-like perturbation in one component. As an example, in (**Figure 4-26**) is presented how  $R_p$  depends on the **GERM** structure, and its number of effectors (**TF**). The recovering rate  $R_p$  reflects the recovering properties of the regulated key-**P** synthesis by the **GERM** modular system. In a simpler way, the species 'j' recovering times  $\tau$  (j)  $\sim 1/R_p$  and trajectories  $C_j(t)$  can be obtained by *simulation*, that is by simulating the **GERM** system dynamics [ with using the **GERM ODE** model eqn. (1A-B), or eqn.(3A-B), or eqn.(8) ] 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 [24]. Species recovering trajectory and amplitude are both very important (**Figure 4-26**, and **Figure 4-34**). As proved by Maria [185], **GERM**s display very different recovering trajectories and amplitudes according to their reaction pathway structure (**Figure 4-26**). The most effective are the **GERM** types ensuring the smallest amplitude of the recovering pathway, thus not disturbing the other cell metabolic processes. As underlined by Maria [185], the recovering trajectories in the **G/P** phase plane is more 'linear'-like for the efficient **GERM**-s, by

presenting a lower amplitude, thus not disturbing other cell reactions and regulatory circuits.

**B-c).- Regulatory robustness.** The regulatory robustness of a **GERM** model is defined by Maria [24] as being the property to realize (Min)  $(\partial R_D / \partial k)$ , where  $R_D$  denotes the key-species recovering rates, while ' $k$ ' is rate constant vector of the **GERM** model (depending on the micro-organism type). 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 (depending on the each individual expression).

**B-d).- Species interconnectivity** in a **GERM** modular regulatory schema of reactions can be viewed as a degree to which they 'assist' each other, and 'cooperate' during the **GERM** system recovering to realize optimal regulatory performances. 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. (3A-B, 4A-B, 5, 6), and **WCVV** model hypotheses of the (Table 4-2). Vance et al. [235] reviewed and proposed several quick experimental/computational rules to check a reaction schema via species interconnectivities. By inducing experimental perturbations to a (bio)chemical system, by means of concentration tracers, or by fluctuating the inputs of the system, one can measure the perturbation propagation through the consecutive/parallel reaction pathway. Then, various techniques can determine the 'distance' among observed species, and specific rules can be used to include this information in elaborating a reaction schema [2,24,40,78,107,174-176,189]. Maria [24,78,107,174,189] proposed an approximate measure of species interconnectivity related to the species recovering-times after a dynamic perturbation, that is:  $AVG(\tau(j))$  and  $STD(\tau(j))$ , i.e. the average and the standard deviation of the species individual recovering times  $\tau(j)$ , respectively. As **AVG** and **STD** are larger, as the cell dynamic regulatory effectiveness is lower, species less interconnected, and components recover more disparately (scattered recover times). The higher the number of effectors and buffering reactions, the better these dynamic regulatory indices of the **GERM** are [2,21-24,25,40,78,107,174-176,185,189].

**B-e) Cell sub-system QSS stability** (and, in particular, of a **GERM**) refers to the system's capacity to recover a **QSS** after cessation of a dynamic perturbation. Such a property can be highlighted by analysing the **QSS** of the dynamic **GERM ODE** model eqn.(3A-B) or eqn.(8). The stability property can be evaluated from the analysis of the eigenvalues  $\lambda(i)$  ( $i$  = no. of species) of the linearized model Jacobian matrix of elements  $J(i,k) = \partial h(i)(C,k) / \partial C(k)$  defined before eqn.(9). The **QSS** is asymptotically stable if the real parts of the Jacobian eigenvalues  $\lambda(i)$  are all negative, that is  $Re(\lambda(i)) < 0$  for all ' $i$ ' [73,231,233,234]. 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:

a) A characteristic of the **WCVV** models including the Pfeiffers' constraint eq.(2), translated in eqn.(3A-B, 4A-B, 5, 6), is that they are always stable (intrinsic stability), because, as proved by Morgan et al. [113], always  $Max(Re(\lambda(i))) = -D$  (see eq.(3A-B, 5)).

b) By contrast, one fundamental deficiency of the classical **WCCV** model formulation is the lack of the intrinsic stability of the cell system model, because these models do not include neither the Pfeiffers' constraint eq. (2), nor an equivalent constraint. Consequently, the **GERM** regulatory mechanism for recovering the system homeostasis (illustrated in **Figure 4-7**, **Figure 4-27**, **Figure 4-34**), is not longer efficiently working under a classical **WCCV** modelling framework, and it becomes invalid, ineffective, and not applicable."

### Some rules to link GERM-s when modelling GRC-s

The **HSMDM** model of the **case study no. 1** of this book (chap. 5), includes a complex **GRC** obtained by linking 7 individual **GERM**-s. To ensure the optimal efficiency of the resulted **GRC**, and their functions into the cell, Maria [21-24] elaborated a couple of rules to be followed, and accounted for, when linking **GERM**-s to form a **GRC**. This chapter is aiming to briefly review the main linking rules, by following the reported results of Maria [21-24,175,176].

When modelling a **GRC** consisting of a network of **GERM**-s, „they are two problems to be considered: (i) what types of **GERM**-s have to be chosen from the library of (**Figure 4-7**) to match the individual gene expression characteristics, and (ii) what rules are to applied for linking **GERM**-s to obtain the holistic regulatory properties of the **GRC** in the context of the cell balanced growth.

When develop a math (kinetic) cell structured model linking **GERM**-s to construct a certain **GRC** reproducing a certain function of the cell, there are two contrary trends in the *in-silico* analysis: (a) on one hand is the natural approach of using simple **GERM** structures to reduce the large computational effort to estimate rate constants of a **HSMDM** model, and for its



simulation; **(b)** on the other hand, it is important to use effective and flexible **GERM**-s able to reproduce individual enzyme-synthesis, but also holistic properties of the **GRC** (of **Table 4-3**; some complex examples are provided by Maria [23]). Below, there are reviewed some rules to be applied for linking **GERM**-s by thus adjusting their regulatory properties to finally obtain the **GRC** with the observed regulatory properties.

**Rule 1) The effect of the no. of regulatory effectors (n).** By definition, **GERM** models (see **Figure 4-7**, and the nomenclature given in chap. 4.3) include an adjustable number of 'regulatory effectors', that is: 'n' for the **[G(P)n]** type, or **[G(PP)n]** series; 'n' and 'n1' for the **[G(P)n; M(P)n1]** series. As proved by Maria [24,78,185], and by Yang et al. [228], a quasi-linear relationship of the **GERM** module **P.I.**-s function of no. of regulatory effectors 'n' can be derived for every **GERM** type, of the form  $P.I. = a_0 + \sum_i a_i n_i$  (Notations: a(o),a(i) denotes the correlation constants related to a certain **P.I.** and module type). Here, **P.I.** denotes the regulatory performance index (chap. 4.4), such as *RD*,  $AVG(\tau(j))$ ,  $STD(\tau(j))$ , stability strength, etc. Also, **n(i)** = number of effectors (**P**, **PP**, **O**, etc.) acting in the 'i-th' allosteric regulatory unit **[Li(Oi)n(i)]**. Such a dependence can also be observed in (**Figure 4-30**, **Figure 4-26**). In short, Maria [24,78,107,174] proved that (**Figure 4-30**, **Figure 4-26**):

**a)** **P.I.** improves ca. 1.3-2 times (or even more) for every added regulatory unit to the **GERM** model. Multiple regulatory units lead to an average recovering time  $AVG(\tau(j))$  of the all **GERM** species much lower than the cell cycle duration 'tc', under a constant logarithmic volume growing rate,  $D = \ln(2)/tc$ . (see eqn. 4B)

**b)** Combinations of **GERM** regulatory schemes (with different effectors) can improve the regulatory **P.I.**-s.

**c)** Certain **GERM** 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)n1]**, or of **[PP]**s in modules **[G(PP)n]** of (**Figure 4-7**). 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 (see the example given by Maria [189]).

**Rule 2) Ranging the number of transcription factors TF and buffering reactions.** To select the suitable **GERM** structure that fits the available experimental (kinetic) data of the cell metabolism, the first problem to be solved is related to the number of buffering reactions of type  $G + P \rightleftharpoons GP$  or  $M + P \rightleftharpoons MP$  necessary to be included in the model to obtain the desired optimal **P.I.**-s (**Figure 4-7**, **Figure 4-34**, **Figure 4-35**, **Figure 4-36**). Evaluation of **P.I.**-s for a large number of **GERM** types (**Figure 4-7**) [2,21-24,40,74,78,107,174,189,194] indicated that the dynamic regulatory efficiency of **[G(P)n]**, or of **[G(PP)n]** modules is nearly linearly increasing with the number 'n' of buffering reactions (following a quasi-linear correlation  $P.I. = a_0 + \sum_i a_i n_i$ ; see also **Figure 4-30**, and **Figure 4-26**). Moreover, the plots of (**Figure 4-26**) reveal that this **P.I.**-s increase is more pronounced in the case of **GERM**-s with **[G(PP)n]** model structures, that use dimeric **TF**-s (that is **PP** instead of simple **P**). Also **GERM**-s of type **[G(P)n; M(P)n1]** models that use a control scheme in cascade of the gene expression reported superior regulatory **P.I.**-s.

Such a **GERM** module efficiency ranking concerns not only the dynamic efficiency, but also most of **P.I.**-s, as 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 [21-24,78,107,174,189]:

**a)** Modules reporting high stationary-regulation **P.I.**-s usually, also reported high dynamic-regulation **P.I.**-s.

**b)** The catalyst activity control at a single enzyme level, that is structures **[G(P)0]**, **[G(PP)0]**, **[G(P)n;M(P)0]** of (**Figure 4-26**, and **Figure 4-34**), which are lacking of buffering reactions able to modulate the gene **G** and **M** catalytic activity appear to be of lowest regulatory efficiency.

**c)** Multiple copies of effector molecules (i.e. **O**, **R**, **P** in **Figure 4-28**, **Figure 4-7**, **Figure 4-5**, **Figure 4-36**), which reversibly and sequentially (allosterically) bind the catalyst (**G**, **M**) in negative feedbacks, improve the regulation effectiveness.

A structured cascade control of the 'catalyst' activity, with negative feedback loops at each level as in the **[G(P)n;M(P)n1]** model series, improves the **GERM** regulatory **P.I.**-s 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 (see the example of chap. 4.3, before the subchap. "Rate constant estimation in **WCVV** models of **GERM**-s or **GRC**-s). As an example in (**Figure 4-28**, **Figure 4-7**, **Figure 4-5**, **Figure 4-36**), by placing regulatory elements (**R**, **P**, **PP**,...) at the level of mRNA (i.e. the species denoted by **M**), and at the level of DNA (i.e. species denoted by **G**) in the **[G(P)n;M(P)n1]** model is highly effective.

d) It was proven [21,23,24] that the **PI** -s of **GERM**-s increase nearly linearly with the number **n(i)** of effectors (**P**, **PP**, **O**, **R**, etc. in **Figure 4-51**) acting in the **i**th allosteric unit [**L(i)(O(i))n(i)**] of buffering reactions applied at various level of control of the gene expression. Such an observation is valid for both dynamic and stationary **PI**-s of the (**Table 4-3**), and chap. 4.4.

e) **PI** -s improves ca. 1.3-2 times (or even more) for every added regulatory unit to the same **GERM** type (**Figure 4-30**). 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,  $Ds = \ln(2)/t_c$  (eq.4B).

f) Combinations of regulatory schemes and units (with different effectors) can improve the regulatory **PI**-s [21-24,78,174].

g) 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)n1**] and of [**PP**]s in the 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. Thus, Maria [107,189] proved existence of an optimal [**TF**] concentration leading to optimal **PI**-s of a **GERM** (**Figure 4-51**).

**Rule 3) The effect of the mutual catalysis in the G/P synthesis.** One essential aspect of the [**G(P)n**], [**G(PP)n**], and [**G(P)n;M(P)n1**] kinetic models of **GERM**-s is the mutual catalysis of **G** and its encoding protein **P** synthesis. If one adds the **WCVV** modelling constraints eqn.(3A-B, 4A-B, 6, 8), 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, as discussed by Maria [21-24,78,174,185]. This direct and indirect link (via **GP**, **GPP**, etc., and the cell volume/osmotic pressure) of **G** and **P** syntheses ensures a quick recovering of both stationary [**G**]s and [**P**]s after any small perturbation. To prove this in a simple way, one considers the synthesis of a generic **G/P** pair in a **GERM** of [**G(P)1**] type (denoted by '2') or a [**G(P)0**] type (denoted by '1') (**Figure 4-7**, and **Figure 4-34**). After estimating the rate constants from solving the stationary model equations by using the homeostatic concentrations of (**Figure 4-52**, high ballast cell case), one determines the dynamic efficiency for each **GERM** by applying a negative 10% impulse perturbation in the [**P**]s = 1000 nM at an arbitrary time  $t=0$ . The obtained recovering trajectories of **P** and **G** obtained by model simulations are plotted in (**Figure 4-27**, and **Figure 4-34**). The plots reveal a very good regulatory efficiency of the [**G(P)1**] type of **GERM**, both **G**, and **P** species presenting relatively short recovering rates, and negligible for the other species. These plots reveal, in a simple way, the self-regulation of the **G/P** pair synthesis: after an impulse perturbation leading to the decline of [**P**]s from 1000 nM to 900 nM, the very fast buffering reaction  $G + P \rightleftharpoons GP$  leads to restore the active **G**, whose concentration quickly increases from 0.5 nM 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 the **G**-lump steady-state. For comparison, as revealed by the results displayed in the (**Figure 4-33**), 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**] case compared to [**G(P)0**], due to the reported smaller  $STD(\tau(j))$  (**Table 4-3**). Consequently, removal of the buffering reaction that automatically adjusts the 'catalytic activity' of **G**, will: (i) decrease the species inter-connectivity (by increasing the standard deviation of the recovering times); (ii) will increase the species recovering times; (iii) will increase the sensitivities of the species steady-state vs. external nutrients (see sensitivity coefficients vs. **NutG** in the **Figure 4-33**). As expected, the **PI**-s of the **GERM**-s depend not only on (a) the no. of effectors (buffering reactions), but also on (b) the **TF** types (**P**, or **PP**), and even more (c) on the used control scheme (i.e. simple buffer reaction of **G**-activity, or in a *cascade* of buffer reactions, like the [**G(P)n;M(P)n1**] reaction schemes).

To exemplify these issues, one considers the same generic **G/P** gene expression example with the species homeostatic stationary concentrations given by Maria [21-24,175,176] (the high ballast cell case). For comparison, one considers the gene encoding gene **G** expression by means of **GERM**-s of various structures given in the (**Figure 4-34**), 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 even [**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. Additionally, the requirement of getting a maximum dynamic responsiveness and efficiency, as discussed by Maria [21-24,185], 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 the (**Figure 4-34**). It is to remark that the incomplete [**G(P)0**] module reports the worst dynamic efficiency, with very slow recovering tendencies after a perturbation, as depicted in (**Figure 4-34**). Better performances are reported by [**G(P)1**] **GERM** module type. Even if a better regulatory efficiency is reported by the cascade control of separately considered transcription and translation of 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 **PI.**). Due to such reasons, **GERM**-s modules of type **[G(PP)2]** will be used to build-up the **GRC** for the *mer*-operon expression in the case study no. 1 (chap. 5) of this book.

All the above analyzed **GERM**-s have been modelled in a **WCVV** framework (eq. **3a-b, 4a-b, 5-8**). For such **WCVV** kinetic models, it is to remark the way by which the variable cell-volume plays an important role to species interconnectivity (direct or indirect via the cell volume) within a certain **GERM** regulatory module or among linked modules. Even if species connectivity can be expressed in several ways [24,235], this index is directly dependent on the manner by which species in a **GERM** or in a **GRC** recover more or less independently after a perturbation. When the species connectivity increases, then 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 the **Figure 4-33**). By contrary, when the species are more disconnected, they recover in a more disparate way, and the **GERM** presents weaker **PI.**-s. [reporting not only larger species recovering times  $\tau$  (rec)(j), but also larger state sensitivities to external nutrients, see **Table 4-3**].

Thus, the *mutual autocatalysis* **G/P** 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 reflect the holistic properties of the *in-silico* (model-based) analyzed **GRC**-s.

**Rule 4) The effect of cell system isotonicity.** The effect of the isotonicity constraint eqn.(**2, 3a-b, 4a-b, 5-8**) of a **WCVV** cell model can be easily proved [21-24]. By simulating the species dynamics which belong to a **GERM** of 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  $t=0$ , on the key-species (**G, P**) can be observed in (**Figure 4-27**), while species recovering times are given in (**Figure 4-33**). By contrast, in a **WCCV** cell model formulation, when the isotonicity constraint is missing from the model, the key-species do not recover after a dynamic perturbation (!). By contrast, as revealed by the performed simulations with the a **[G(P)1]** module, 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 Maria [21-24,175,176], the **WCVV** models, with 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.

**Corolary.** The effect of the isotonicity constraint eqn. (**2, 3a-b, 4a-b, 5-8**) in a **WCVV** cell model is also related to the more realistic prediction of **PI.**-s of **GERM**-s. As proved by Maria et al.[23,60], the **WCCV** dynamic models might be satisfactory for modelling many cell sub-systems, but not for an accurate modelling of cell **GRC**-s, and the cell holistic properties under perturbed conditions, or the division of cells [113], by distorting very much or even misrepresenting the predicted results, as exemplified by Maria [21-24,60] for both stationary and perturbed cell growing conditions (**Figure 4-53**, and **Figure 4-54**).

**Rule 5) The importance of the adjustable regulatory TF-s in a GERM.** As proved by the example of (**Figure 4-34**), and those of [24,185], dimeric **TF**-s, such as **PP** in **[G(PP)n]** instead of simple **P** in **[G(P)n]**, leads to several conclusions:

a) 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**, in a small quantity). Some **GERM** modules reported an increased **PI.** flexibility, due to 'adjustable' intermediate **TF** species levels. This is the case, for instance, of adjusting **[M]s** in the module **[G(P)n;M(P)n1]**, or of **[PP]s** in the modules **[G(PP)n]**. Optimal levels of these intermediate species can be determined from matching various optimization criteria, rendering complex **GRC**-s to be more flexible in optimally reproducing certain desired cell-synthesis regulatory properties.

b) The dynamic regulatory efficiency (defined in **Table 4-3**) decreases in the following order [21,23,24,185]:

$\text{Min}(\tau(\text{rec})P) : [G(PP)2] > [G(P)1;M(P)1] > [G(P)1] > [G(P)0]$   $\text{Min(STD): } G(PP)2 > [G(P)1;M(P)1] > [G(P)1] > [G(P)0]$

c).- The stationary regulatory efficiency of [P]s decreases in the same order:  $\text{Min } S([P] ; [\text{Nut}G]) : [G(PP)2] > [G(P)1;M(P)1] > [G(P)1] > [G(P)0]$

**Rule 6) The effect of the cell ballast on the GERM efficiency.** When constructing simplified WCVV cell models, it is important to know what is the minimum level of simplification to not essentially affect the holistic properties of the cell, and their key-functions (Outline 4-1). This paragraph proves why it is essential to include in a WCVV cellular 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 that all species (individually, or lumped) to be accounted in the cell model, because species concentrations and rates are linked through the common cell volume eq.(2, 4A, 6). As proved by Maria [21,170,175,176], and (Figure 4-52), in such WCVV cell model constructions, the recovery rates are properties of all interactions within the system rather than of the individual elements thereof [73].

However, another important question derived from the isotonicity constraint refers to the degree of importance of the cell content (ballast) concerning the cell reactions, and the species QSS-levels 'resistance' to perturbations. In other words, the P.I.-s of a GERM or of a GRC 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 prove that in a simple way, one considers a GERM of a generic [G(P)1] type placed in an *E. coli* cell with two different nominal conditions given in (Table 4-4) for a high-ballast cell, and a low-ballast cell. To not complicate these models, lumped gene and protein metabolites have been considered. Being present in a large amount (that is lumped [MetG] = 3E+6 nM, and lumped [MetP] = 3e+8 nM), these components also play the role of cell ballast, their concentrations being set to values much larger than those of the other cell species. Simulations of Maria and Scoban [175,176], allowed obtaining the species trajectories, and their recovering times 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 displayed in (Figure 4-52). The species recovering

times are presented in (Table 4-5) for the [G(P)1] model comparatively to a [G(P)0] model, both used under the WCVV approach. The results clearly indicate the regulatory superiority of the [G(P)1] model."

" Selection of appropriate lumped [MetG] and lumped [MetP] 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 recovering times  $\tau(\text{rec})(P)$  for the key-protein P. For instance, in the [G(P)1] module case, of low ballast case, with lumped [MetG] = 2000 nM, and lumped [MetP] = 3000 nM, and (all [Cj]) = 12001 nM, the resulted recovering times of the key-species G/P are  $\tau(\text{rec})(P) = 103$  min, and  $\tau(\text{rec})(G) = 223$  min after a -10% impulse perturbation in the [P]s of 1000 nM at an arbitrary t = 0 (Figure 4-52). Whereas for a high-ballast cell case with lumped [MetG] = 3e+6 nM, and lumped [MetP] = 3e+8 nM, and (all [Cj]) = 6.06e+8 nM, the resulted  $\tau(\text{rec})(P) = 127$  min, and  $\tau(\text{rec})(G) = 118$  min after a -10% impulse perturbation in the [P]s of 1000 nM at an arbitrary t=0 (Table 4-5, and Figure 4-52).

We refer to this as the 'Inertial Effect'. In an in-silico model estimation/simulation, it arises because the invariance relationships described above (chap. 4.2, 4.2.1.2) for the WCVV approach require that larger rate constants for P and G synthesis be used to counter-balance lower [MetP] and [MetG] content into the cell, and these constants are determinants for the key-species recovering rates ( $\tau(\text{rec})(j)$ ) after a dynamic perturbation. On the other hand, when metabolite concentrations were low (low-ballast cell case), perturbation of the cell volume (by the [P]s-perturbation) was greater than those when they cell-ballast was high (the volume dynamics plots not presented here; see [170,175,176]). The attenuation of the perturbation-induced volume changes by large metabolite concentrations is called the 'Ballast Effect'. Cell 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 (Table 4-5, and Figure 4-52). Thus, increasing metabolite concentrations attenuates the impact of perturbations on all cellular components but negatively influences their recovery times.

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.

**Rule 7) The effect of GERM complexity on the resulted GRC efficiency, when linking GERM-s. One important issue to be solved when linking GERM-s**

One important issue to be solved when linking **GERM-s** 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 [21-24,78] 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 WCVV 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 by using ODE kinetic models with continuous (or stochastic) variables, and difficult to be estimated, due to the very large number of unknown parameters (rate constants, diffusion coefficients, etc.), and unknown steady-state concentrations. Beside, cell model constructions with too complex cell modules lead to inoperable large models very difficult, if not impossible, to be used for GMO cell design purposes. The alternative is to use reduced ODE models with a number of lumped species and enough reactions (chap. 4.2.1.4.) 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 adequately 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 [78] compared two GERM's linking alternatives (Figure 4-35, and Figure 4-55):

Alternative A: [G1(P1)1] + [G2(P2)1] (10 individual and lumped components). Alternative B: [G1(P1)1;M1(P1)1] + [G2(P2)1;M2(P2)1] (14 individual and lumped components). The GERM mutual linking in the alternative A is as following: the expressed P1 in [G1(P1)1] is the metabolase that converts NutG 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 modules [G1(P1)1] + [G2(P2)1].

The GERM linking in the alternative B 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].

Simulations revealed that alternative B is superior, presenting better stationary P.I.-s (Figure 4-35), and dynamic P.I.-s (Figure 4-55). The 'cost' in increasing complexity is minimum, the GRC including 14 species into the model (alternative B) compared with only 10 species (alternative A). To conclude, in spite of a slightly more complex structure (14 vs. 10 individual and lumped components, and two more buffering reactions), the GRC of alternative 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 in a GERM of [G(P)n;M(P)n1] type (Figure 4-7) 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, and STD indices). The same positive conclusions also concern the GERM structures of [G(PP)n] type (Figure 4-7), as proved by (Figures 4-26, 4-30, 4-34, 4-49, 4-50).

**Rule 8) Cooperative vs. concurrent linking of GERM-s in GRC and species interconnectivity.**

When coupling two or more GERM modules in a GRC of a certain living cell, the lumped nutrients, and metabolites involved in the G/P syntheses are roughly the same (Figure 4-54). The modelling problem is what alternative should be chosen (see Figure 4-56)??? A competitive scheme (due to the common substrate, i.e. MetG1,2 and MetP1,2), or a cooperative scheme, the two GERM-s mutually supporting their regulatory efficiency??? 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] module

types, there are tested three alternatives of coupling GERM-s in a GRC, as illustrated in (Figure 4-56), that is:

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

**Alternative 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.

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

Tests were performed by using the **WCVV** modelling framework, and the nominal high-ballast cell condition of (Table 4-4). Simulations lead to very interesting conclusions, as followings [24]:

a) In the **Alternative A**, one links two modules  $[G1(P1)0] + [G2(P2)0]$ , both ensuring regulation of the two proteins (**P1**, **P2**) synthesis, in a *concurential* disconnected way (Figure 4-56). For this hypothetic system, synthesis of **P1/G1** and **P2/G2** from metabolites is realized with any interference between modules {the simulated case study in Figure 4-57 corresponds to the following steadystate:  $[P1]s = 1000$  nM,  $[P2]s = 100$  nM,  $[G1]s = 1$  nM,  $[G2]s = 1$  nM}. The only connection between the two GERM-s 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  $[P1]s$ , 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, see the details of [24].

b) In the **Alternative B**, the simple *cooperative* linking of  $[G1(P1)0] + [G2(P2)0]$  modules in (Figure 4-56, and Figure 4-57) ensures specific individual functions of each protein, i.e. **P1** lump plays both permease and metabolase functions, while **P2** is a polymerase.

c) In the **Alternative C**, the simple *cooperative* linking of  $[G1(P1)0] + [G2(P2)0]$  system of the **Alternative B** has been *consolidated* by adding simple effectors for the gene activity control, thus resulting the consolidated cooperative system  $[G1(P1)1] + [G2(P2)1]$ , of (Figure 4-56, and Figure 4-57), where the effectors **P1** and **P2** act in two buffering reactions,  $G1+P1 \rightleftharpoons G1P1$ , and  $G2+P2 \rightleftharpoons G2P2$ , respectively, with the stationary states  $[G1]s = [G1P1]s = 1/2$  nM, and  $[G2]s = [G2P2]s = 1/2$  nM, thus ensuring maximum dynamic P.I. -s..

The species steady-state (QSS) concentrations are given in (Table 4-4).

The same *cooperative* linking rule of GERM-s can be repeatedly applied, by using the same strategy. Of course, other GERM types can be used as well. For instance, when  $[G(PP)n]$  modules are used, the more effective effectors are the dimers **PP**, acting in 'n' buffering reactions of the type,  $G+PP \rightleftharpoons GPP \rightleftharpoons \dots \rightleftharpoons GPn$ , with the stationary states  $[G]s = [GP1]s = [GP2]s = \dots = [GPn]s = 1/(n+1)$  nM. The **WCVV** model rate constants should be estimated from the species stationary concentration vector **Cs**, and by imposing regulatory optimal characteristics discussed by Maria [21-24], and presented in the chap. 4.3 (eq.(10)). From the same reasons, stationary levels of active and inactive forms of catalyst should be adopted,  $[L]s = [TF1]s = [TF2]s = \dots = [TFn]s = 1/(n+1)$ . Besides, the dissociation constant of the  $(L:TFn)$  complex in the buffering reactions  $k(\text{diss}) \gg D$  has been adopted, e.g.  $k(\text{diss}) \approx (1e+5 \text{ to } 1e+7)D$ , being much higher than other rate constants of the GERM reactions [Maria, 2005b]. In subsequent works, Maria [1,2,21-24,25,40,60,78,107,189,174-176,194] also proved that optimization of the GERM's P.I.-s with the multiobjective criteria summarized by Maria [21,24,189] (eqn.(10)) leads to small values for the intermediate  $[PP]s$  (the active parts of dimeric TF-s).

The stability and the dynamic regulatory characteristics of all three GRC systems (**A-C**) (above defined) have been determined by studying their QSSrecover after a +/-10% $[P1]s$  impulse perturbation. The results, presented by Maria [24,175,176], reveal the following aspects concerning the alternative GRC systems [ **A**, **B**, **C** ] :

a) All three systems are stable, that is  $\text{Max}(\text{Re}(\lambda(j))) = -D < 0$  (where  $\lambda(j)$  are the eigenvalues of the ODE model Jacobian matrix  $J_c$ , defined above eqn. (9)). The GRC systems [ **B** ] and [ **C** ] recover faster after a dynamic perturbation in  $[P1]s$ . It results that the *cooperative* module linking is superior to the competitive GERM's linking alternative [ **A** ], being only one *viable* linking alternative that ensures the system homeostasis and the cell equilibrated metabolism. The [ **B** ] and [ **C** ] GERM's linking alternatives are superior because they preserve specific functions of each protein inside the cell. The GERM's linking alternative [ **C** ] presents the best P.I.-s from the all three checked linking alternatives due to the additional regulatory effector of type  $G+P \rightleftharpoons GP$ .

- b) The **GRC** system is as better regulated as the effector is more effective (e.g. the use of multiple buffering reactions, with dimeric **TF**-s, and a cascade control of the expression, of type  $[G(PP)n ; M(PP)n]$  (not investigated here here).
- c) The use of efficient effectors and multiple regulation units can improve very much the dynamic **PI**, in the following relative order:  $[G(P)n] < [G(PP)n] < [G(P)n ; M(P)n1] < [G(PP)n ; M(PP)n1]$ .
- d) Dynamic perturbations affect rather species present in small amounts inside the cell, while recovering times for the other species (e.g. lumped metabolites **MetP**, **MetG** of large concentrations) are negligible.

The design procedure of a regulatory network **GRC** can be continued in the same way, by accounting for the expression of additional genes (and their suitable **GERM**-s, according to the experimental information about the regulation dynamics). For instance, in the simplified representation of Maria [24], a 3-rd **GERM** for the **P3** synthesis can be added to the **Alternative [C]**, by allocating specific functions to the **P1**, **P2**, and **P3** lumped proteins, as follows: **P1** and **P3** are permease and metabolase enzymes, which ensure nutrient import inside the cell, and their transformation in gene-metabolites (**MetG1-MetG2-MetG3**) and protein-metabolites (**MetP1-MetP2-MetP3**) respectively. Protein **P2** lumps polymerases able to catalyze the genes **G1**, **G2**, and **G3** production. If one considers the simplest effector case, the resulted **GRC** includes the following three modules  $[G1(P1)1] + [G2(P2)1] + [G3(P3)1]$ , which regulate the synthesis of **P1**, **P2** and **P3**, in a *cooperatively interconnected* way, which preserves the protein functions.

**Rule 9) The optimal value of TF.** It is self-understood that, in a realistic **WCVV** model, the holistic properties the whole cell and of 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 PI**-s. Such optimal **[TF]**s are obtained by solving a multi-objective optimization problem [21,175,176,189], see chap. 6, eqn.(10). An example was provided by Maria [189] in the case of a genetic switch (**GS**) in *E. coli* cell, modelled under the **WCVV** approach. The two considered self- and cross-repressing gene expression modules of the **GS** are of the type  $[G2(P2P2)1(P3P3)1] + [G3(P3P3)1G3(P2P2)1]$  (**Figure 4-58**). These well chosen **GERM**-s models are very flexible, allowing adjusting the regulatory properties of the **GS** (that is the switch certainty, its good responsivity to inducers, and its good dynamic and stationary efficiency). Besides, based an adequate **WCVV** math model, and by applying an optimization procedure, Maria [189] proved in a **GS** case that it exists an optimal level of the **TF**-s (that is  $[P2P2]$ s, or  $[P3P3]$ s in the present case, see **Figure 4-52**) 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 experimental data [189].

**Rule 10) Additional aspects to be considered 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 pathways, accounting for its structural and functional organization, seems to be a promising route to be followed, as it is proved by the case study no. 1 (chap. 5) of this work.. Because a limited number of **GERM** types exist (see the library of **Figure 4-7**), individual **GERM**-s can be separately analysed, and checked for efficiency in conditions that mimic the stationary and perturbed cell growing conditions, as above exemplified with some simple **GERM**-s, and **GRC**-s. Efficient **GERM**-s (of regulatory indices of **Table 4-3**) are then linked accordingly to certain rules to mimic the real metabolic process, by ensuring the overall **GRC** efficiency, cell system homeostasis, and protein individual functions. Module linking rules are not fully established, but some and above mentioned principles (i.e. rules 1-9) reviewed by Maria [21-23], should be fulfilled. The hierarchically organised genetic regulatory network (**GRN**, **GRC**) 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 [24] advanced some supplementary rules to be accounted for, that is:

10-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 strength of linking reactions among modules would have to decline as the number of linked modules increases.

10-b) When linking **GERM**-s, the main questions arise on the connectivity mechanism and on the *cooperative* vs. *uncooperative* way by which proteins interact over the parallel/consecutive metabolic pathways inside or between **GERM**-s [24,82,146,147,185]. 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 the above 'rule no. 8'). 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 internal/external perturbations.

10-c) Protein (enzyme) interactions are very complex, being part of the cell metabolism and distributed over the **GRN** 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 (**Figure 4-10**, and **Figure 4-18**). Specific proteins are involved in nutrient permeation (permeases), in metabolite synthesis (metabolases), or in the 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 (see chap. 7). 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, thus avoiding the slower diffusional processes [21-24].

10-d) By applying the former rule, it results that, when building-up a kinetic model of a **GRC**, comprising a couple of **GERM-s**, 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 if some **GERM** modules do not include reactions in common, the modules are anyway linked through the cell volume (to which all cell species contribute, eqns.(2,6)) and due to some intermediates controlling module interactions in the **GRC**. Only the **WCVV** modelling framework is able to account for such cell regulatory characteristics.

10-e) A natural strategy for building complex and realistic cell dynamic models is to analyse independent functional reaction pathway modules or groups of closely interacting cellular components, and then link them. The **WCVV** approach may facilitate this strategy. Each module (such as **GERM-s**, **GRC-s**, or parts of the **CCM**) 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 cell. 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 investigated module. In such a way, each tested cell carrying a certain defined **GERM**, or **GRC** 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. This is an important **WCVV** approach, allowing to independently simulate some cell reaction pathway modules, by placing them into a 'virtual cell' of which characteristics and properties are reproduced by a lumped dynamic model.

10-f) Application of the **WCVV** modelling approach demonstrates that each cell component affects, and is affected by, all other cellular components [21-24,78,174]. Indirect inter-connectivities among species / reactions arise because all components in a cell contribute to cell volume [eqn.(2, 6)], and cell volume influences component concentrations [eqn.(3A-B)]. Thus, perturbations in one component reverberate throughout the all cell species. The importance of these indirect relationships will vary with the diversity and complexity of cellular components. Increasing numbers and concentrations add '*ballast*' to the cell, strengthening these indirect relationships, while increasing diversity allows individual metabolites to be present at lower concentrations, thus improving the dynamic responses of **GERM-s** and of **GRC-s** to perturbations. Another issue, thus far unexamined, is how specific types of interconnectivities affect the regulatory behaviour of cells. This could be probed by using the experimental methods developed by Vance et al.[235] to deduce connectivities in biochemical pathways from the effects of an impulse perturbation in one of the species.

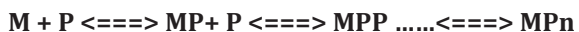
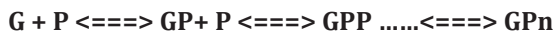
10-g) When modelling complex operon structures [see the chap. 5 of this book], simple, but effective **GERM** structures (also in agreement to the experimental observations on the cell processes dynamics) should be adopted to not complicate too much the **WCVV** kinetic model. The default **GERM** is the **[G(P)1]**. But, according to the experimental data and interactions among genes and proteins, more complicated, and effective **GERM** constructions, such as **[G(PP)1]** can be elaborated, as those described in the application presented in the below chap. 5 of this book

**Rule 11) The effect of cascade control on the GERM efficiency.** Among **GERMs** reviewed and tested under a **WCVV** modelling framework, the most significant are those of **[G(P)n]** type, whose effectiveness nearly linearly increases with the number (**n**)



of buffering reactions (**Figure 4-30**, and **Figure 4-26**). Due to their simple structure such **GERM**-s are the most suitable to construct very complex **GRC**-s. 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**) in an optimized small amount. As previously discussed (Rules 1, 2, 3, 5,7), the most effective are the **GERM**-s with a cascade control of the expression, by means of buffering reactions applied at both key gene **G**, and mRNA (**M**) catalyst level, that is of type **[G(P)n;M(P)n1]** (**Figure 4-7**). [21-24,78,185] *in-silico* proved the superiority of the **[G(P)n;M(P)n1]** gene expression regulatory modules. The conclusions are the followings:

(I).- The very rapid buffering reactions, such as:



have been proved to be very effective regulatory elements, by quickly adjusting the active/inactive **G/GP/GPP/GPn** or **M/MP/MPP/MPn** species level ratios, thus efficiently coping with the perturbations.

(II).- Numerical tests revealed that the **P.I.**-s of some simple **GERM**-s increase in the approximate order:

**[G(P)0]** (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 **P.I.**-s (**Table 4-3**) per regulatory element is of ca. 1.3 (under **WCVV** modelling framework), while the same improvement is of ca. 2.5 under the classical (inadequate, and unlikely) **WCCV** modelling framework [21,24,78,107,174-176,189]. It clearly appears that the **WCVV** modelling framework is more realistic, the default **WCCV** approach tending to over-estimate the **P.I.**-s, and to distort the **GERM**-s regulatory properties (see chap. 4.2.1, chap. 4.2.1.3, and **Figure 4-12**, and **Figure 4-13**)."

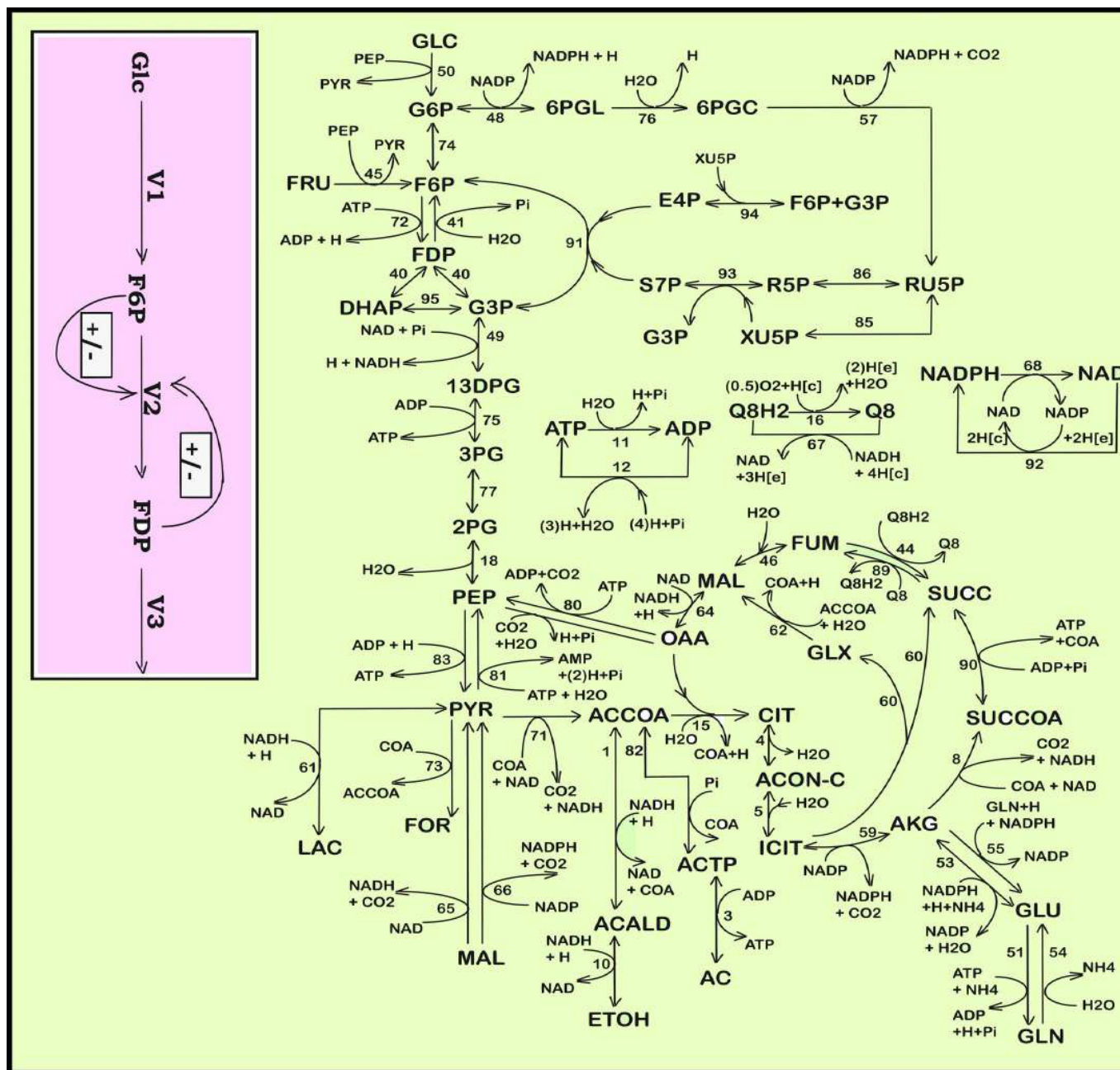
## The book main objectives

To elaborate such a **HSMDM** dynamic model, the engineering analysis cycle of a biological process includes several steps as schematically represented in (**Figure 4-8**), that is: experimental lab-scale investigations, followed by the math (kinetic) modelling, and the numerical analysis of the biological process. Once validated, such a dynamic (structured or apparent/global) model will be used to optimize/control the **FBR** or **SCR** industrial bioreactor operation. If inconsistencies occur between the recorded data vs. model predictions, then an intermediate step is required to update (with a certain frequency) the biological process model.

Although complicated and often over-parameterized, dynamic deterministic **ODE** models with continuous variables of **CCM**-s or **GRC**-s have a significant number of advantages, being able to reproduce in detail, by numerical simulations, molecular interactions, continuous slow or rapid response of the cell to continuous exo / endogenous disturbances [20,24]. In addition, the use of **ODE** kinetic models has the advantage of being easily approachable from a numerical point of view, being flexible, easy to extend and suitable to be characterized using nonlinear systems theory tools [73], by taking into account properties of the control system, that is: dynamics, direct and reverse control loops, and optimality. And, most importantly, such a **CCM-GRC WCVV** modelling approach by means of **ODE** kinetic models by also accounting the cell variable volume under the isotonic constraint, allows the use of powerful tools, namely numerical algorithms and the classical (bio)chemical engineering modelling concepts summarized in (**Figure 4-3**, and **Figure 4-4**). The most important of these {issues (**BCE1-BCE6**), of chap. 4.1) used in the kinetic modeling of (bio)chemical, or biological processes are mentioned above and by Maria [21-24,60,178].

An extended discussion of the (bio-)chemical engineering concepts, principles, and tools is given by [119,121,236].

This work is aiming to prove the feasibility and advantage of using the relatively novel **HSMDM** concept by coupling extended **CCM**-based cell structured deterministic nano-scale models with the macro-scale state-variables of the analyzed bioreactor models. The resulted hybrid dynamic model was successfully used for engineering evaluations. Exemplifications are made for two case studies. Thus, the case study (**no.1**) in chap. 5 describes the way by which Maria and Luta [40] derived a lumped **HSMDM** able to simulate the dynamics of the *mer*-operon expression and self-regulation together with the dynamics of a **SCR-TPFB bioreactor**, using cloned *E. coli* cells (with variable *mer*-plasmid concentrations), aiming to optimize the performances of a bioreactor used for mercury uptake from wastewaters. The case study (**no.2**) in chap. 6 describes the way by which Maria [3,12] derived an extended **HSMDM** able to simulate the dynamics of the **CCM** key-species, and of the **TRP**-operon expression and self-



**Figure. 4-1:** Simplified representation of the CCM pathway in *E. coli* of [4,42] (the “wild” cell including the PTS-system). “Fluxes characterizing the membranar transport [*Metabolite*(e) « *Metabolite*(c)] and the exchange with environment have been omitted from the plot. See [42] for details and explanations regarding the numbered reactions. Notations: [e]= environment; [c]=cytosol. Adapted from [42] with the courtesy of CABEQ J]. The considered 72 metabolites, the stoichiometry of the 95 numbered reactions, and the net fluxes for specified conditions are given by Maria et al. [42]. The pink rectangle indicates the chemical node inducing glycolytic oscillations [39,59]. Notations + and - denotes the feedback positive or negative regulatory loops respectively. **GLC** = glucose; **F6P**= fructose-6-phosphate; **FDP** = fructose-1,6-biphosphate; see the abbreviation list for species names; **V1-V6** = lumped reaction rates indicated by Maria [3].” Species notations are explained in the abbreviation list of [3].

Chemical engineering modelling rules and concepts applied in Systems Biology and Bioinformatics, after [21-23].

## Chemical engineering modeling rules and essential tools

- **Mass conservation law:**
  - **molecular species conservation law**
    - stoichiometry analysis;
    - formulate differential mass balance equations in kinetic models
  - **atomic species conservation law**- atomic species mass balance
- **A large variety of numerical algorithms for estimating / solving nonlinear math dynamic models in the presence of complex constraints** (Levenspiel, O., Wiley, 1999; Rasmuson et al., *Mathematical Modelling in Chemical Engineering*, Cambridge Univ. Press, 2014)
- **Thermodynamic analysis of reactions**
  - quantitative assignment of reaction directionality (Zhu et al., 2013)
  - set equilibrium reactions; Gibbs free energy balance analysis
  - set cyclic reactions; find species at quasi-steady-state
  - improve calculating steady-state flux distributions that provide important information for metabolic engineering (Zhu et al., *B&B*, 110, 914 (2013))
- **Some kinetic modelling rules**
  - **lumped reactions** according to their time constant,  $\tau = 1/k$  for 1-st order;  $1/k/(\text{coreactant})$  for 2-nd order; Maria, G., *Chem. Biochem. Eng. Q.* 18(3), 195(2004).
  - **lumped species** according to their life time  $LT = -1/J(i, i)$ ;  $J =$  system Jacobian [Maria, G., *Chem. Eng. Sci.* 60, 1709-1723 (2005)].

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

Figure 4-2. Chemical engineering modelling rules and concepts applied in Systems Biology and Bioinformatics, after [21-23].

## Biochemical engineering modeling principles

- **Modular approach:**
  - by applying the building blocks rules of **Synthetic Biology**
  - **math modelling** is the most comprehensive mean for a rational design of regulatory **GRC**
- **Interferring Regulatory GRC-s** ensure production of enzymes of optimal structure & properties
- **Enzymes ensure an optimized cell metabolism:**
  - **GRC maximum regulatory efficiency**
  - Cell regulatory and adaptive properties are based on an **Optimized cell metabolism**, that is :
    - **homeostatic** mechanisms, that is quasi-constant key-species concentrations and output levels, realized:
      - by **adjusting** the synthesis rates,
      - by **switching** between alternative substrates, or development pathways,
      - by maintaining **key-species homeostasis**,
      - with using **minimum of resources** (substrates, energy),
      - with producing **minimum amount of intermediates**,
      - with **maximum reaction rates, optimized fluxes**,
      - with **minimum recovery and transition times** after perturbations

Maria, G *A review of some novel concepts applied to modular modelling of genetic regulatory circuits*, Juniper publ., Newbury Park, CA, 2017, ISBN 978-1-946628-07-7(USA). <https://juniperpublishers.com/ebook-info.php>

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

Figure 4-3: Biochemical engineering modelling rules and some concepts to be applied in Systems Biology and Bioinformatics, after [21-23].

## Mathematical representations



- Boolean / Discrete variables ('E-circuits', coarse GRC representation)
- Continuous variables, continuous processes, stationary and dynamic perturbations (Michaelis-Menten-type, Hill-type, or power-law S-system representation)

$$\frac{dX}{dt} = r_m \frac{(Act)}{K_a + (Act)} \left( 1 - \frac{(Rep)}{K_r} \right) - bX$$

$$\frac{dX}{dt} = r_m \left( d_1 + \frac{d_2 I^n}{d_3^n + I^n} \right) - bX$$

$$\frac{dX_j}{dt} = a_j \prod_i X_i^{g_{ij}} - b_j X_j^{h_j}$$

- Stochastic models for representing: gene mutations, cell signalling, faulty switches, multimodal cell population distribution; p=probability

$$\frac{dp(X;t / X_0;t_0)}{dt} = \sum \{ a_j(X-\delta;t) p(X-\delta;t / X_0;t_0) - a_j(X;t) p(X;t / X_0;t_0) \}$$

- Topological models (GRC clusters, modules, multi-layer structure)
- Large numbers of states  $O(10^3 - 10^4)$ , gene TF  $O(10^3)$ , and parameters
- Adaptable lumped structure vs. available -omics information and utilization scope

- Maria, G. *Chem. Biochem. Eng. Q.* 19, 213-233 (2005); *CABEQ* 20, 353-373 (2006); *CABEQ* 17, 99 (2003)

Figure 4-4: Some kinetic expressions used by the ODE type dynamic models of the cell genetic circuits (GRC, GERM), or of the cell metabolic biochemical reactions [21-24,107]. Adaptation after [23].

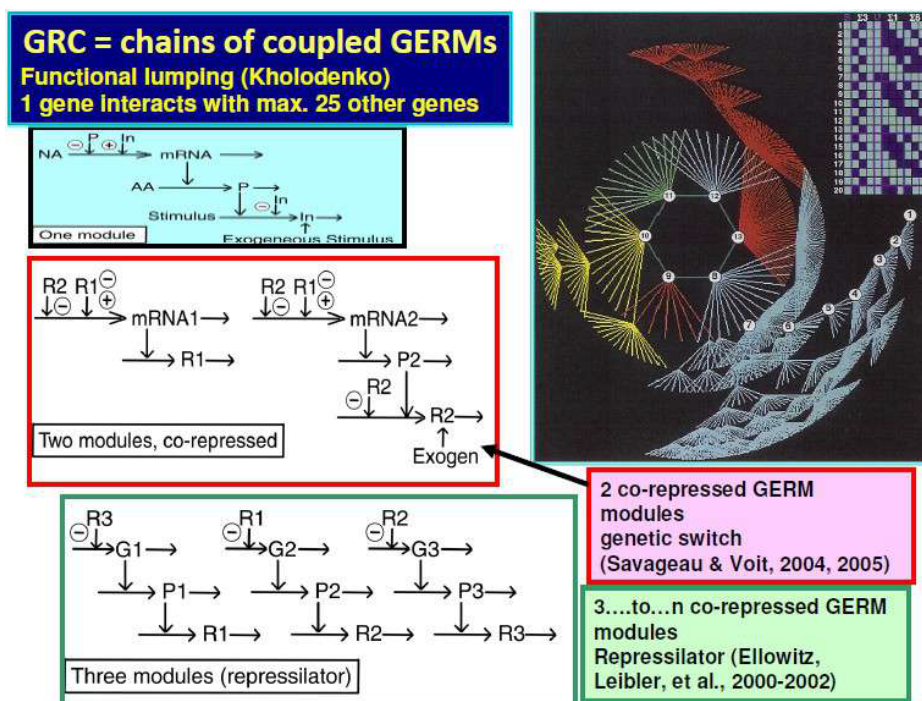
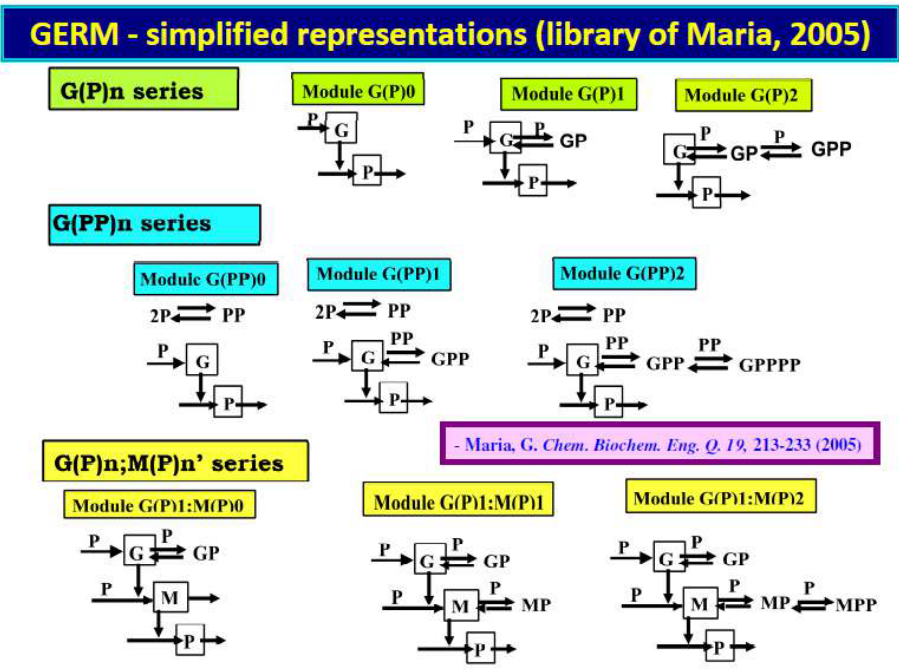
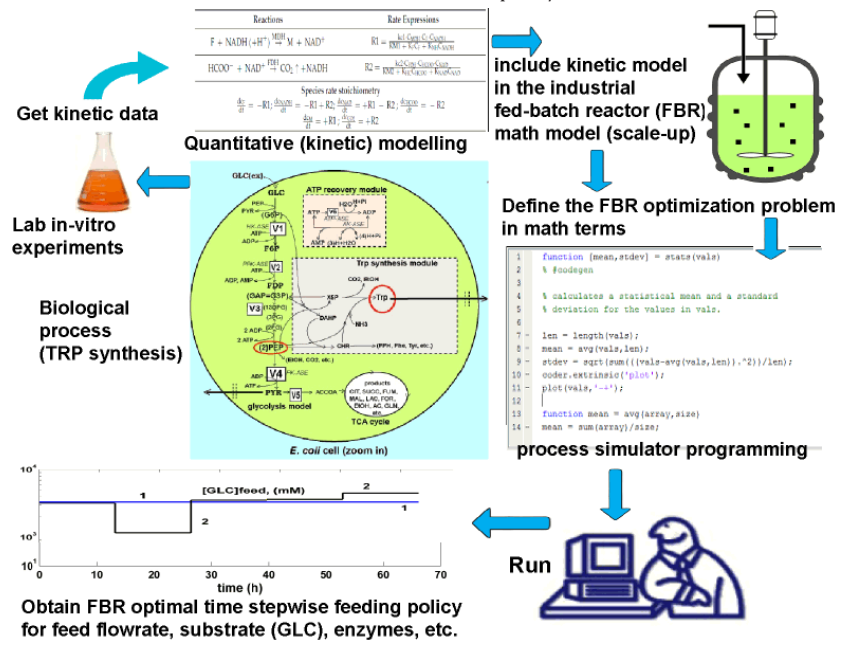


Figure 4-5: Example of linked GERM-s to form GRC-s (genetic switches here). See the reviews of Maria [21-24,60,74,107,178].

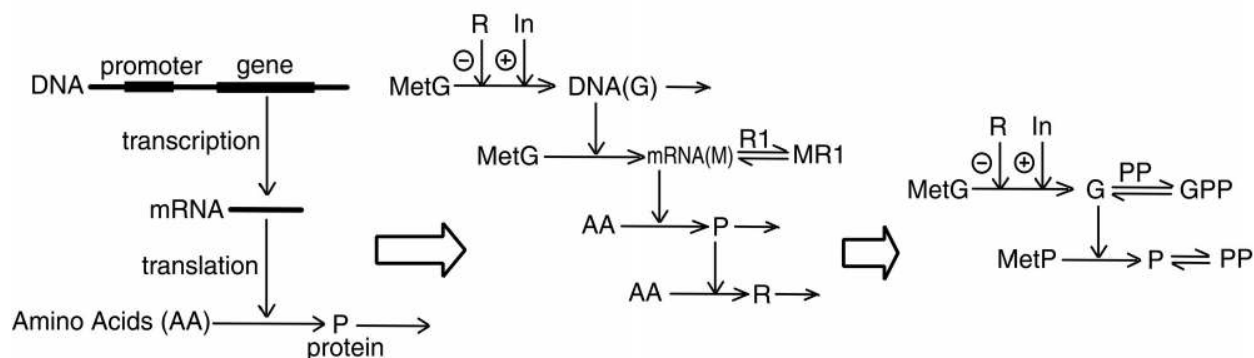


**Figure 4-7:** The library of Maria [21-24] with lumped modular models to represent the GERM-s dynamics. Adapted from [23-24] with the courtesy of CABEQ J1., Simplified representations of some generic gene expression G/P regulatory modules (GERM) types following [24]. The horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicates an assumed concentration invariance of these species; **Up-row**: simplified representation of the gene expression model corresponding to [ G(P)n ] regulatory module types. The transcriptional factor is the protein P itself, the selfregulation over the transcription and translation steps being lumped together. To improve the system homeostasis stability and self-regulation, despite of perturbations in nutrients Nut\*, and metabolites Met\*, or of internal cell changes, a very rapid buffering reaction  $G + P \rightleftharpoons GP$  (inactive) has been added. **Middle-row**: simplified representation of the gene expression model corresponding to a [ G(PP)n ] regulatory module types. The transcriptional factor is the dimmer PP. **Down -row**: simplified representation of the gene expression model corresponding to [ G(P)1; M(PP)n ] regulatory module types. The models account for the cascade control of the expression via the separate transcription and translation steps. Notations: G\* = DNA gene encoding P\*; M = mRNA; P, PP = allosteric effectors of the transcription/translation."

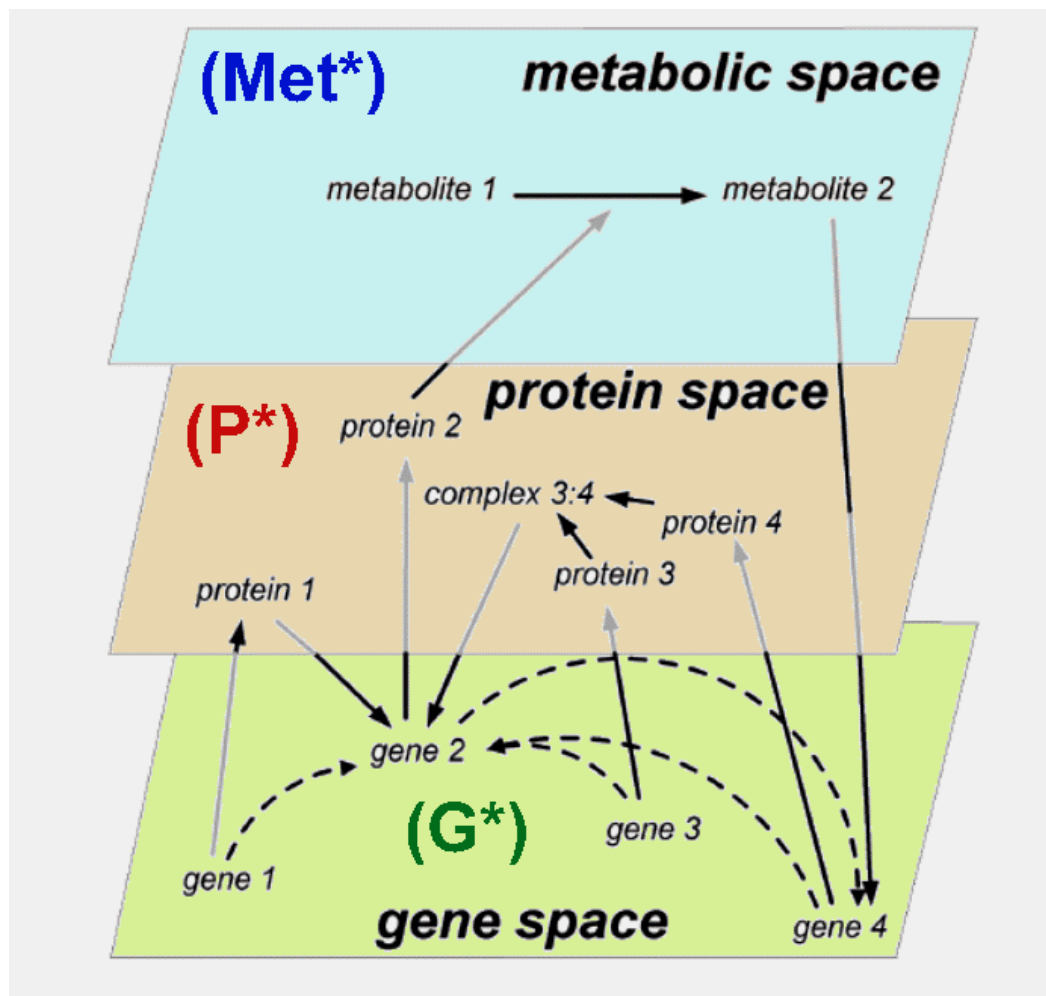


**Figure 4-8:** The engineering analysis cycle of a biological process: experimental labscale investigations, coupled with the math (kinetic) modelling, and the numericalanalysis of the biological process aiming to optimize the FBR industrial bioreactor operation. [The green cell] Simplified CCM reaction pathway in E. coli used in the hybridstructured model, adapted after [12,42,25]. It includes four linked reaction modules:[a] glycolysis; [b] ATP recovery system (the pink rectangle, including the synthesis of adenosin co-metabolites ATP, ADP, AMP); [c] the TRP synthesis (the gray area), and the biomass [X] growth model. This reaction pathway has been used by Maria [3] to derive a hybrid FBR dynamic model for the TRP synthesis. Connection of the

TRP synthesis to glycolysis is realized through the PEP node. **Notations:** GLC(ex)= glucose in the cell environment. Species abbreviations are given in the abbreviations list of [3]. Species in parenthesis are not explicitly included in the glycolysis model. Italic letters denote the enzymes. Squares include notations of enzymatic reactions V1-V6 included in the glycolysis model. Adapted from [25,42] with the courtesy of CABEQ JI, and completed according to the kinetic model of [42]. [Right-up] The scheme of a FBR. [Left-down] The optimal feeding-policy with GLC solution in the FBR, after [12]."



**Figure 4-9:** Protein P synthesis – simplified representations of a generic GERM regulatory module (horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; G = gene encoding P; M = mRNA; R = repressor; In = inducer; Met= metabolites). See reviews of Maria [21-24,60,74,107,178].



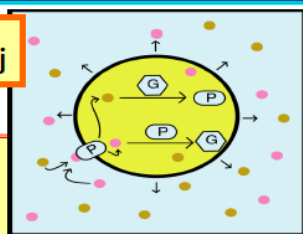
**Figure 4-10:** Abstract depiction of cellular physiology. Adapted from [183].

**Classical approach (CVWC)= variable cell volume is neglected; cell dilution is included in the species decay terms.**  
**Variable Volume Whole-Cell (VVWC) modelling background (promoted by Maria, G., 2005-2017)**

$$dC_j/dt = r_j - DC_j ; D = d \ln(V)/dt = [RT / \pi(\text{cyt})] \sum r_j$$

$$RT / \pi = 1 / \sum C_j = \text{const.}$$

- Hypotheses:**
- Isothermal, isotonic, open system, negligible gradients
  - Quasi-constant osmotic pressure (inner and outer cell)
  - Semi-permeable membrane, following the cell growing dynamics
  - Include all cell components at a certain degree of detail
  - Account for holistic cell properties (ballast effect, constant osmotic pressure, variable volume, cell cycle periodicity)



- Identification of model parameters (k) and unobservables species conc. (C<sub>i</sub>):**
- Max. QSS-recovering rate of the cell system after dynamic perturbations
  - Min. QSS-sensitivity to stationary perturbations
  - Min. transition times between steady-states after a stationary perturbation
  - Max. regulatory effectiveness of GRC (with maximizing the stability strength)

- Maria, G. *Chem. Biochem. Eng. Q.* 19, 213-233 (2005); *CABEQ* 20, 353-373 (2006); *CABEQ* 17, 99 (2003)

Figure 4-11: The novel modelling framework of WCVV promoted by Maria [21-24,60,74,107].

improve the system homeostasis stability and self-regulation, despite of perturbations in nutrients **Nut\***, and metabolites **Met\***, or of internal cell changes, a very rapid buffering reaction  $G + P \rightleftharpoons GP(\text{inactive})$  has been added. **Middle-row**. simplified representation of the gene expression model corresponding to a **[G(PP)n]** regulatory module types. The transcriptional factor is the dimmer PP. **Down -row**. simplified representation of the gene expression model corresponding to **[G(P)1; M(PP)n]** regulatory module types. The models account for the cascade control of the expression via the separate transcription and translation steps. Notations: **G\*** = DNA gene encoding **P\***; **M** = mRNA; **P, PP** = allosteric effectors of the transcription/translation.

### Comparison constant-vs.-variable cell volume approach

Isotonic system constraint induces secondary perturbations to any change in the cell content  
Pfeiffers' law

$$\pi V = RT \sum n_j$$

$$RT / \pi = V_0 / \sum n_{j,0} = 1 / \sum C_j = \text{const.}$$

$$D = d \ln(V)/dt = [RT / \pi(\text{cyt})] \sum r_j$$

$$\begin{cases} \frac{d C_{MetG}}{dt} = r_1 - r_3 \\ \frac{d C_{MetP}}{dt} = r_2 - r_4 \\ \frac{d C_P}{dt} = r_4 - r_5 + r_6 \\ \frac{d C_G}{dt} = r_3 - r_5 + r_6 \\ \frac{d C_{GP}}{dt} = r_5 - r_6 \end{cases}$$

**default CVWC**

$$\begin{cases} \frac{d C_{MetG}}{dt} = r_1 - r_3 - D \cdot C_{MetG} \\ \frac{d C_{MetP}}{dt} = r_2 - r_4 - D \cdot C_{MetP} \\ \frac{d C_P}{dt} = r_4 - r_5 + r_6 - D \cdot C_P \\ \frac{d C_G}{dt} = r_3 - r_5 + r_6 - D \cdot C_G \\ \frac{d C_{GP}}{dt} = r_5 - r_6 - D \cdot C_{GP} \end{cases}$$

**VVWC Promoted by Maria 2005-17**

Figure 4-12. Comparison of WCCV vs. WCVV modelling approach in the case of a simple **[G(P)1] - GERM** [23,60,175,176].

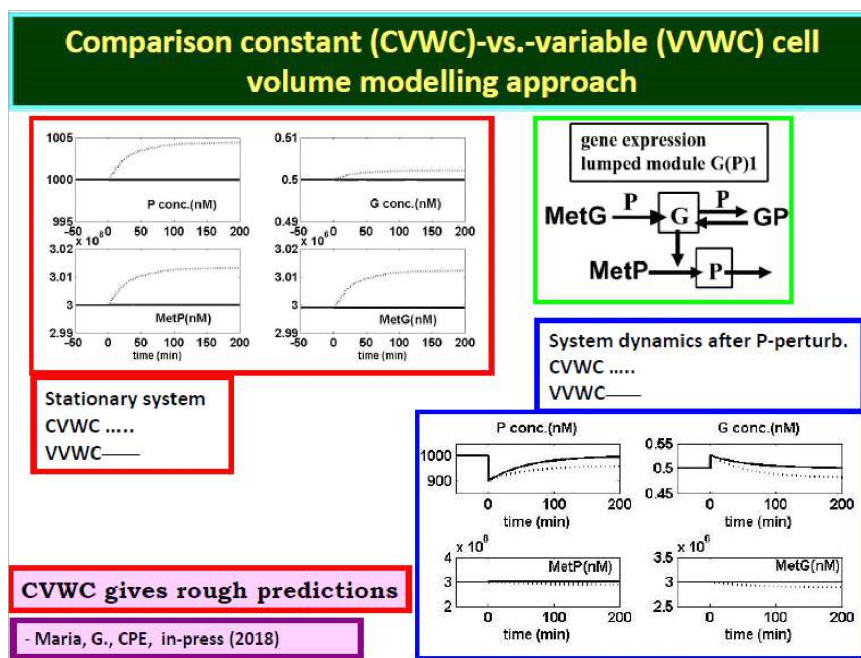


Figure 4-13. Comparison of the WCCV and WCVV modelling approach in the case of a simple G(P)1 gene expression regulatory module [23,60,175,176].

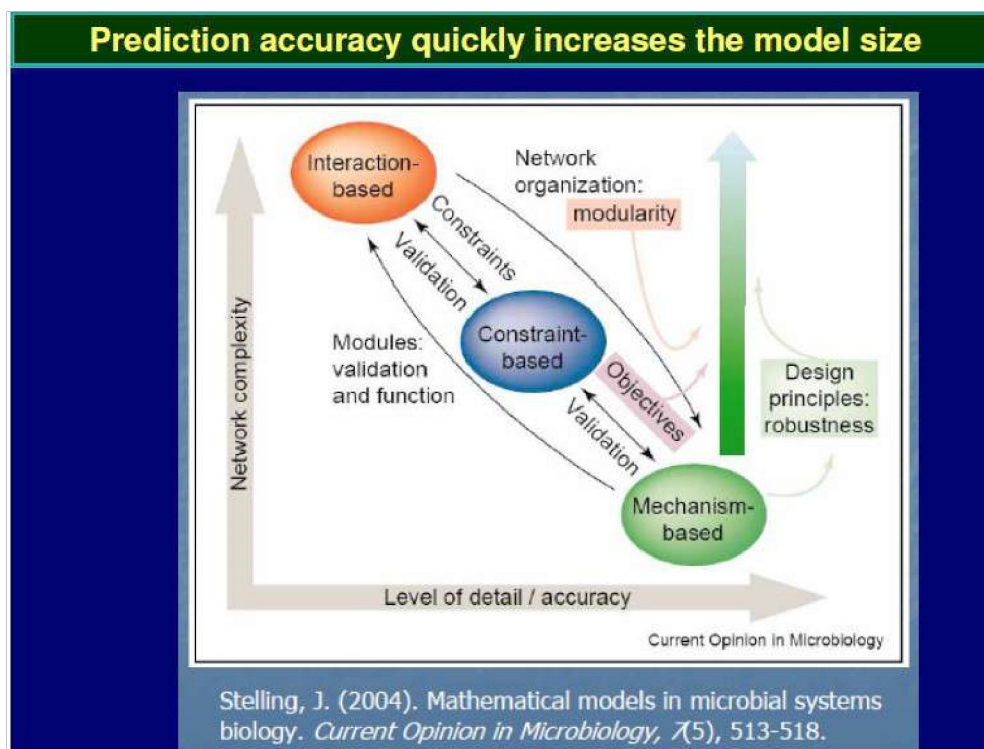


Figure 4-14: The exponential-like increase of the model size with the prediction accuracy [180].



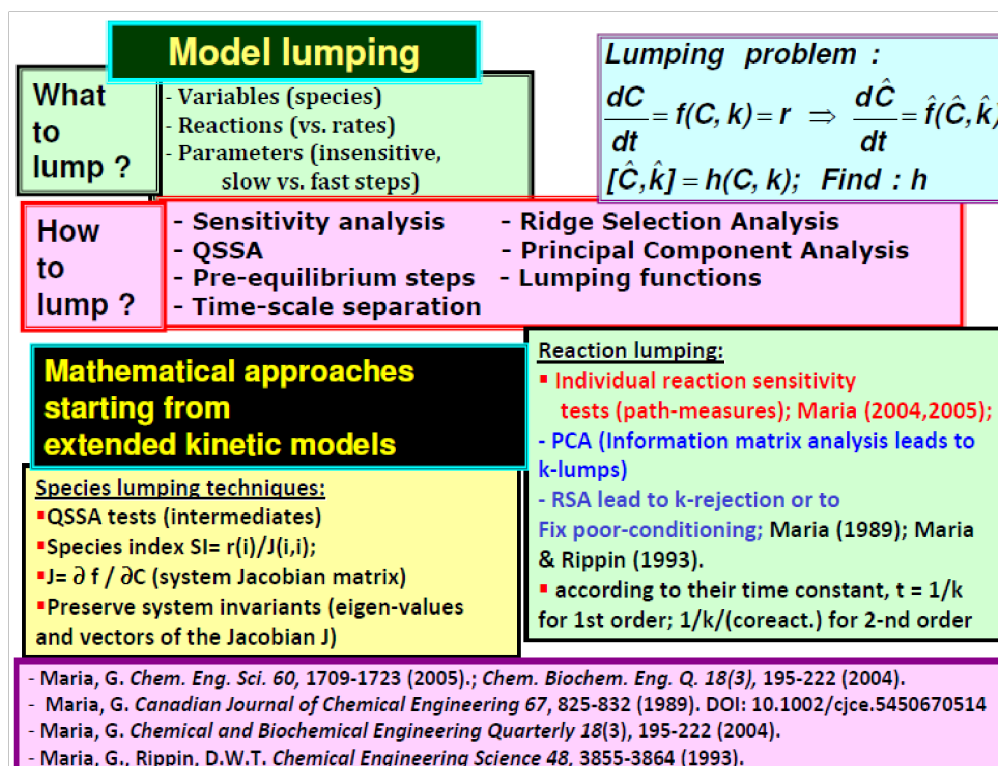


Figure 4-15: Some guide-lines used in obtaining reduced dynamic models [29,64,67,78].

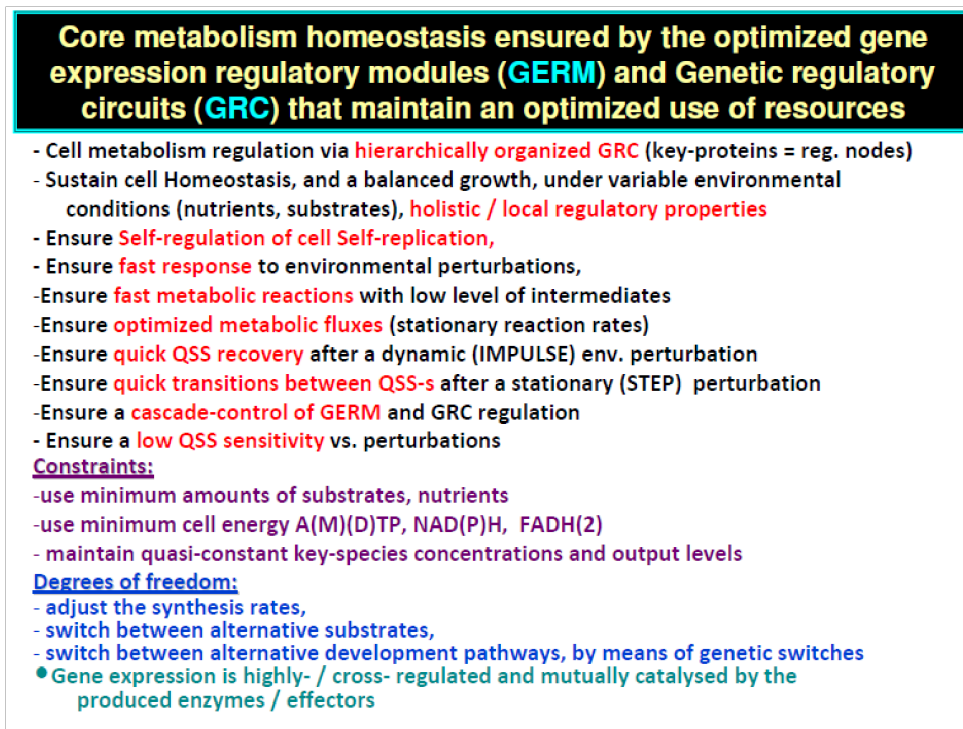


Figure 4-16: The role of gene expression regulatory modules (GERM) and of the genetic regulatory circuits (GRC) [21-24].

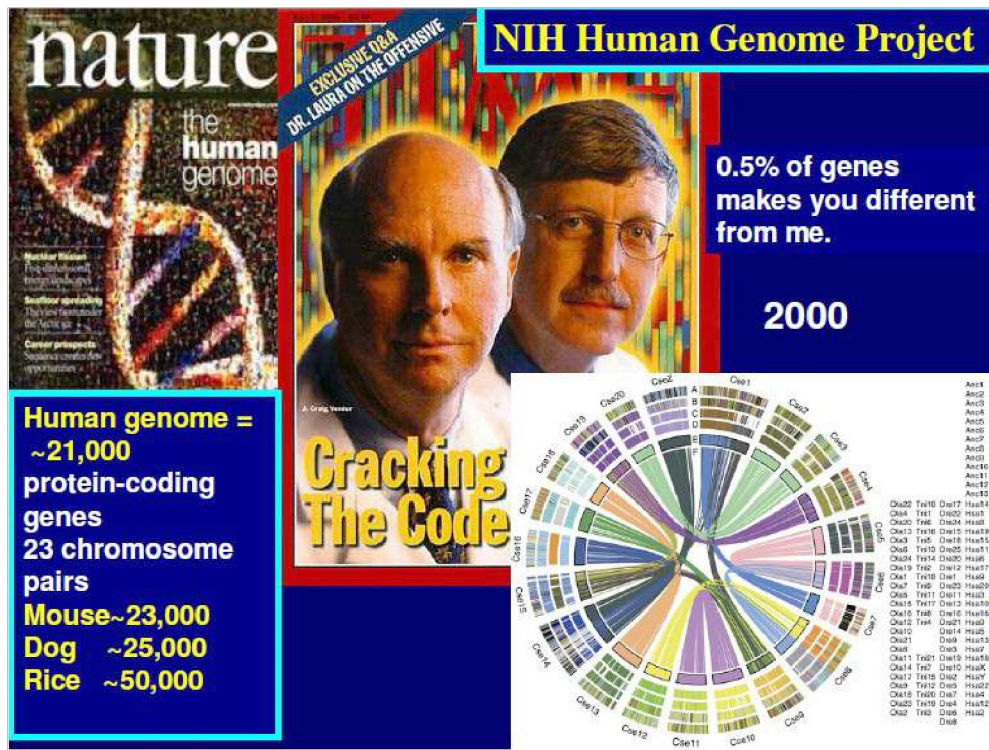


Figure4-17: Around 2000 the human genome has been deciphered.

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

Figure 4-18: The hierarchical organization of a living cells [21-23].

## Lumping in modelling CELL metabolism – Why ?

**Reasons:**

- too complex cell mechanisms vs. available data
- large number of species, reactions, transport parameters, interactions
- low data observability & reproducibility
- metabolic process variability
- interpretable representation of cell complexity
- quick simulations of cell behavior under various environmental conditions
- computational tractability and easier application of algorithmic rules
- design **GMO, in-silico** re-program cell metabolism, design of biosensors, drug target release, bioprocess optimization & control, gene therapy, optimal cell cloning, etc..

**Disadvantages:** Multiple reduced structures of different characteristics

- Loss of information on certain species and reaction steps
- Loss in system flexibility (no. of intermediates and species interactions)
- Multiple reduced structures of proximate characteristics
- Loss in prediction capabilities
- Lumped model parameters can lack of physical meaning
- Loss / alteration of systemic / holistic properties (stability, multiplicity, sensitivity, regulatory characteristics)

Figure 4-19: Reasons to use lumped models in description of cell metabolic processes.

Pioneers in kinetic modelling Biological systems

1952

Key issues in Systems Biology

- Dynamics
- Feedback
- Optimality

Hodgkin, A. L. and Huxley, A. F. (1952) A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. J. Physiol. (Lond.), 117:500-544.

Figure 4-20: First attempts to model the cell CCM, GRC-s are inspired from the electric circuit's theory (the so-called A-cell [103,104]).

## Mathematical modelling of gene networks starts around 1990-2000

**Metabolic space**

Metabolite 1 → Metabolite 2

**Protein space**

Protein 1, Protein 2, Protein 3, Protein 4, Complex 3:4

**Gene space**

Gene 1, Gene 2, Gene 3, Gene 4

Reinhart Heinrich (1946–2006)

**Reinhart Heinrich 1946- 2006**

**The regulation of Cellular Systems**

1996

THE REGULATION OF CELLULAR SYSTEMS  
REINHART HEINRICH AND STEFAN SCHUSTER

### Optimization in biology

**Darwin:** “Living organisms have evolved to maximize their chances for survival; It explains structures, behaviors of living organisms”

Stefan Schuster 1961-

Figure 4-21: Pioneers in modelling genetic regulatory circuits [73].

## State of the art in modelling kinetics of the Central Carbon Metabolism CCM

**Boolean topological representation of GRC (Bower & Bolouri, 2001)**

**VWC dynamic model of iron metabolism in mitochondria (Hudder, Maria, et al., 2002, NIST Conf., Gaithersburg (USA), Sept. 17-18, 2002 )**

Figure 4-22: Some cell simulators: one of Boolean type: (left) [200]; (right) a mixed Boolean-continuous model concerning the iron metabolism in

mitochondria [194].

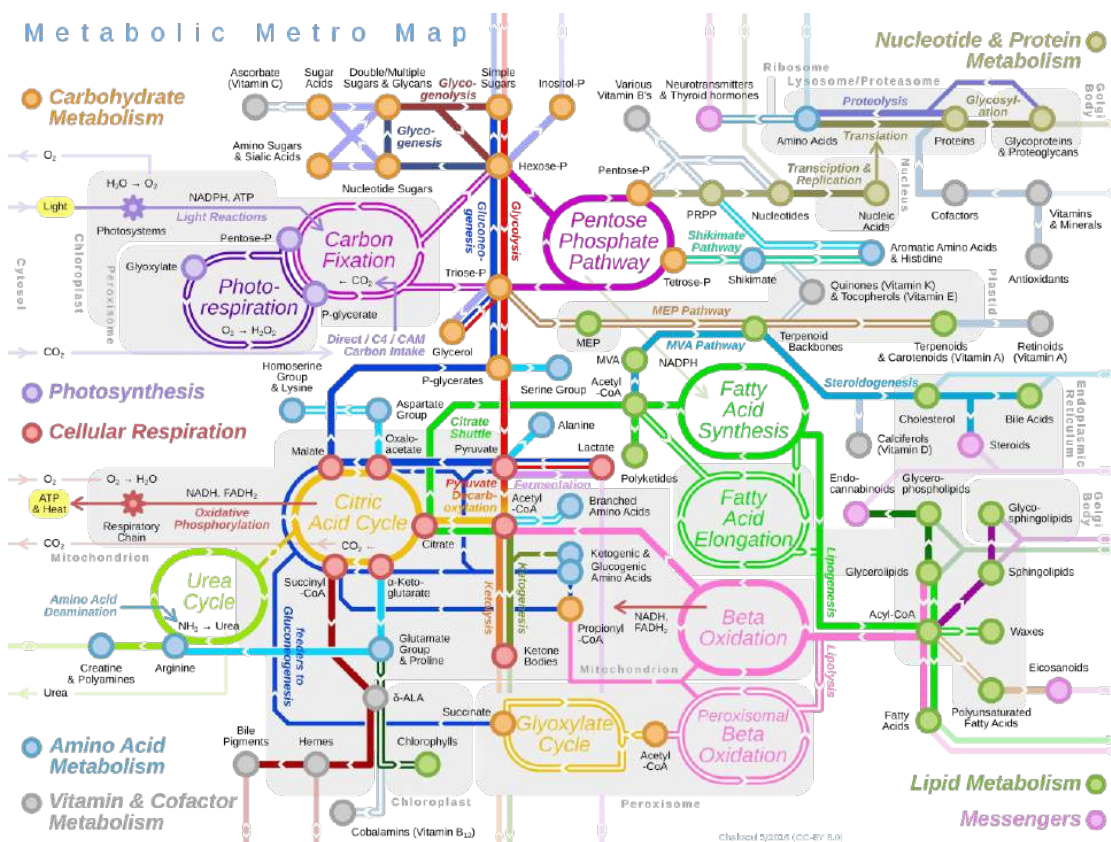


Figure 4-23: Part of the CCM in an eukaryotic cell. Source =

[https://en.wikipedia.org/wiki/File:Metabolic\\_Metro\\_Map.svg](https://en.wikipedia.org/wiki/File:Metabolic_Metro_Map.svg)

### Lumped modular approach in modelling GRC

- **Modular Representation** of cell complexity vs. few structured data (low observ.)
  - **modules** = semi-autonomous groups of reactions & species, or functional units generating an identifiable cell function (individually characterized)
  - **disassemble** the system in modules (**Reverse Engineering**)
  - **recreate** the same system from scratch – module linking in GRC (**Integrative Understanding**), by preserving individual and holistic regulatory properties (efficiency, functionality, cell structural-functional-temporal internal hierarchy)
  - **improve identifiability and estimability** by using **lumping** of species / reactions, **aggregate pools**, slow / fast manifold separation
- **Building-blocks strategy** allows designing new cellular systems
  - Genetic circuit engineering used to design specific cell functions
  - Genetic network GERM components may be extracted, replaced, replicated, altered, etc.
- **Application of similarity analysis to develop biological inspired engineering systems (**Synthetic Biology**)**
  - Engineering inspired artificial operations (switching, oscillating, spatial patterns)
  - Reprogram enzyme functions
  - Reprogram more complex cell functions (cell-cell communicators, biosynthesis)
- **Computational tractability**
  - easy numerical analysis and application of algorithms for system characterization (multiplicity, stability, flexibility, robustness, similarity analysis)
  - development of cell simulation platforms using modularization algorithms, quick simulations of cell behaviour under variable environmental conditions

Figure 4-24: Some cell simulators of stochastic type (left)[102], and (right) of electric-circuit type [103,104].

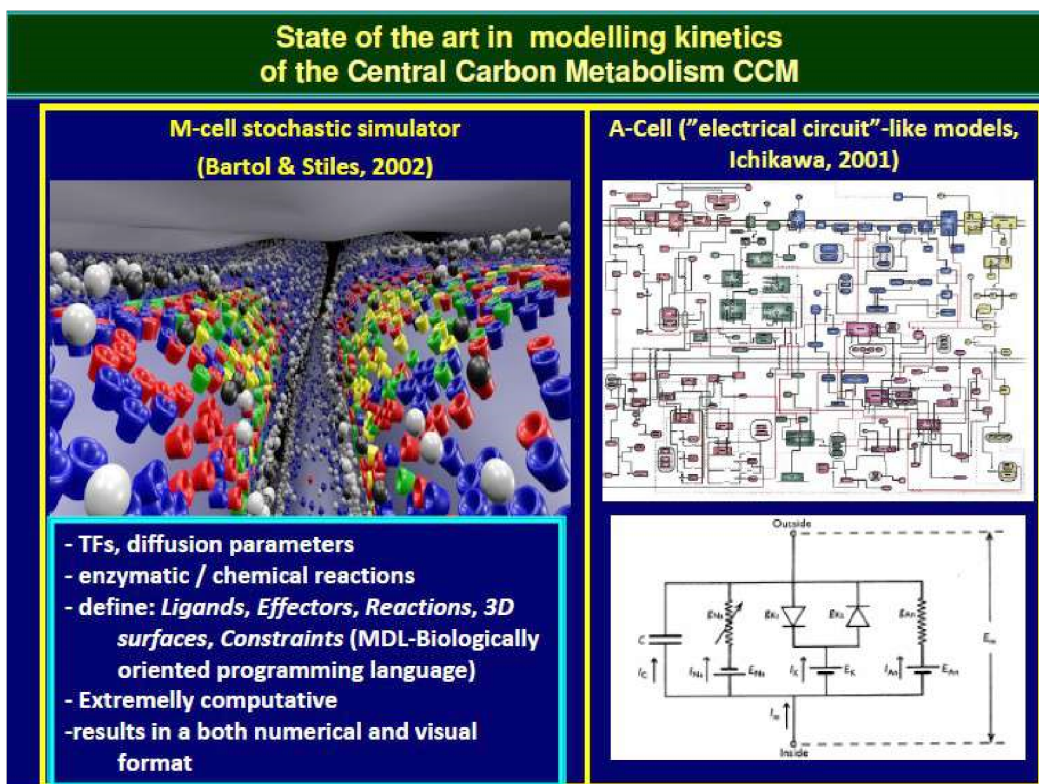


Figure 4-25: Importance of the lumped modular math modelling for the *in-silico* design of GMO-s by using HSMDM models [21-23].

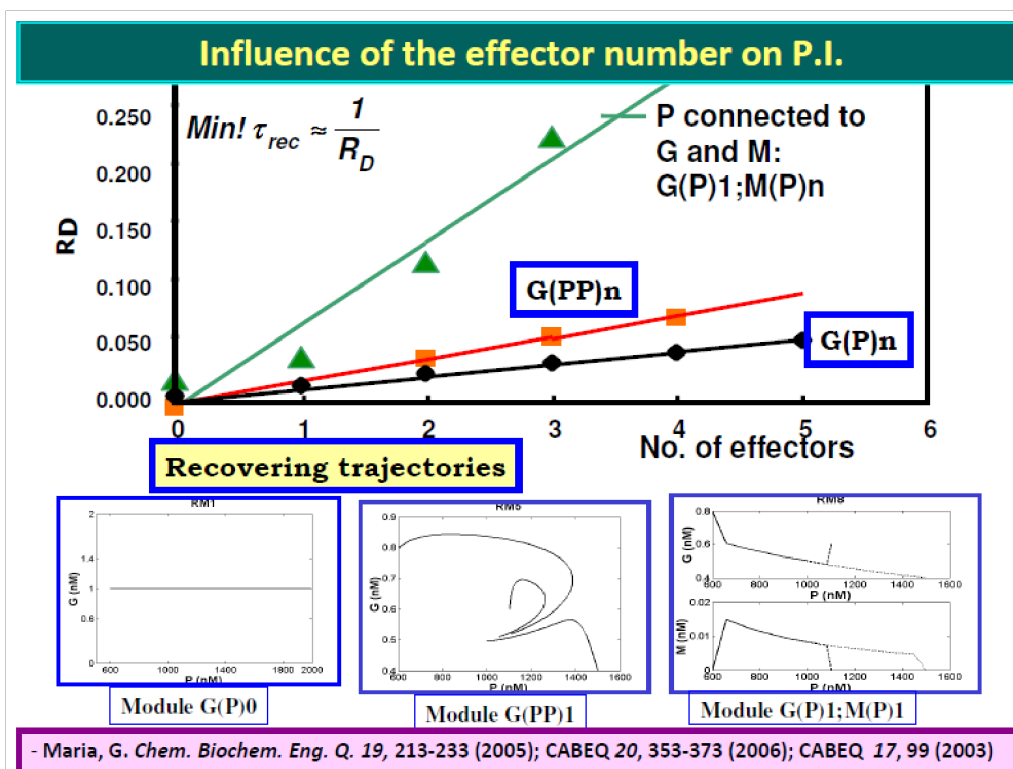


Figure 4-26: Influence of the GERM's number of effectors on some of their properties, i.e. the QSS recovering rate (top), and the amplitude / species recovering trajectory after a 10% dynamic perturbation in the [P]s [21-24,185].

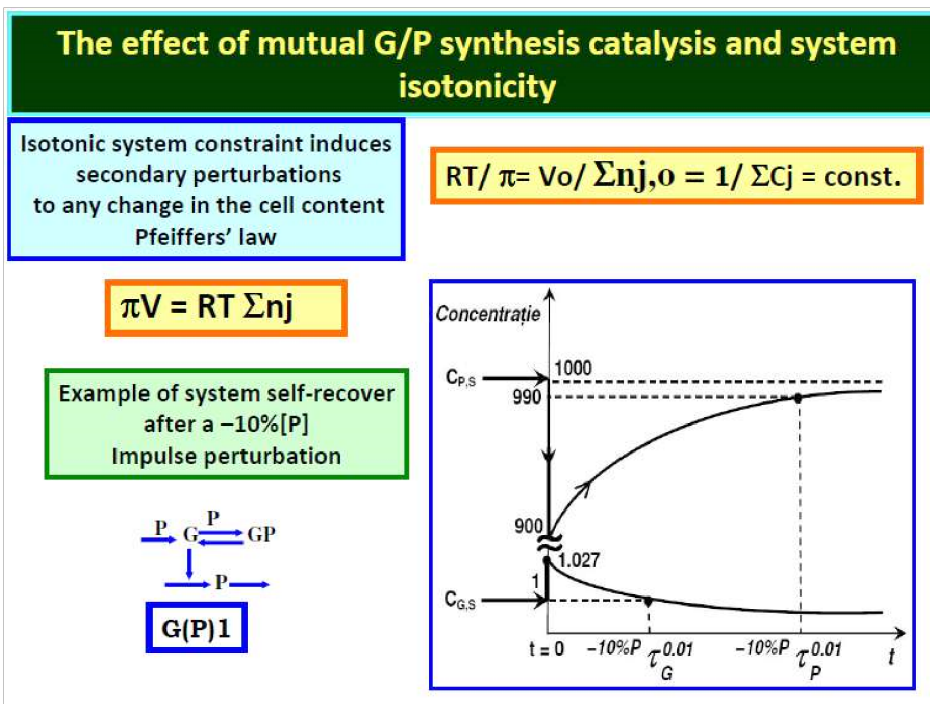


Figure 4-27: The effect of the mutual G/P mutual self-catalysis, and of the isotonicity in the case of a simple [G(P)1] gene expression regulatory module GERM [21-24,175,176]. Species QSS recovery after a 10% dynamic (impulselike) perturbation in [P]s. Notations:  $\pi$  = cell osmotic pressure;  $V$  = cell volume;  $T$  = temperature;  $G$  = generic gene;  $P$  = the protein encoded by  $G$ ;  $C(j)$  = cell species concentrations; „s” index = at QSS; „o” index = initial; = species  $P$  recovering time of its QSS,  $[P]_s$ , with a 1% precision (i.e. 0.01).

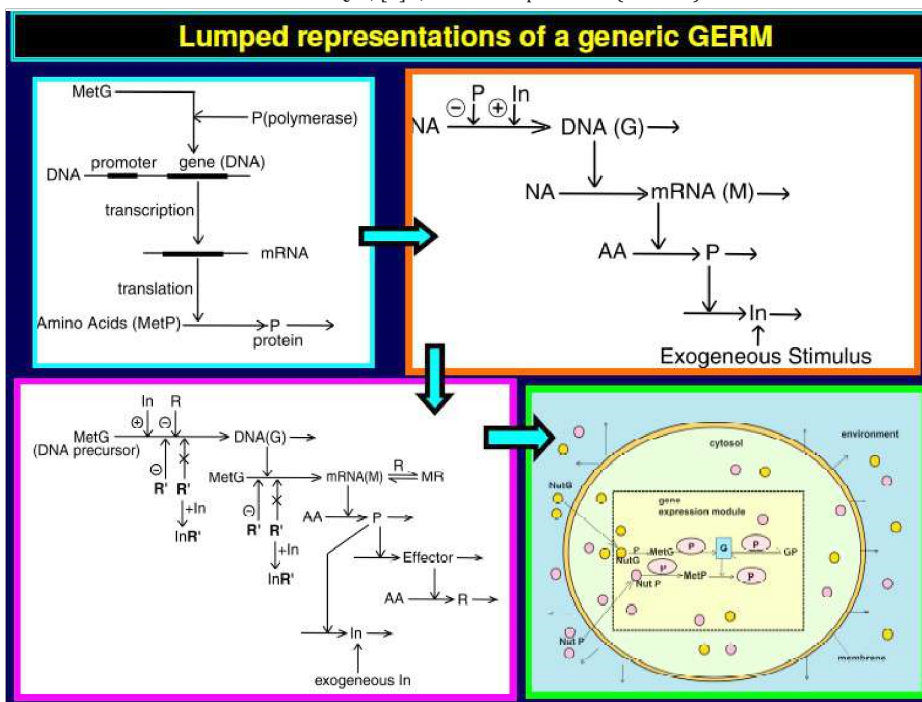
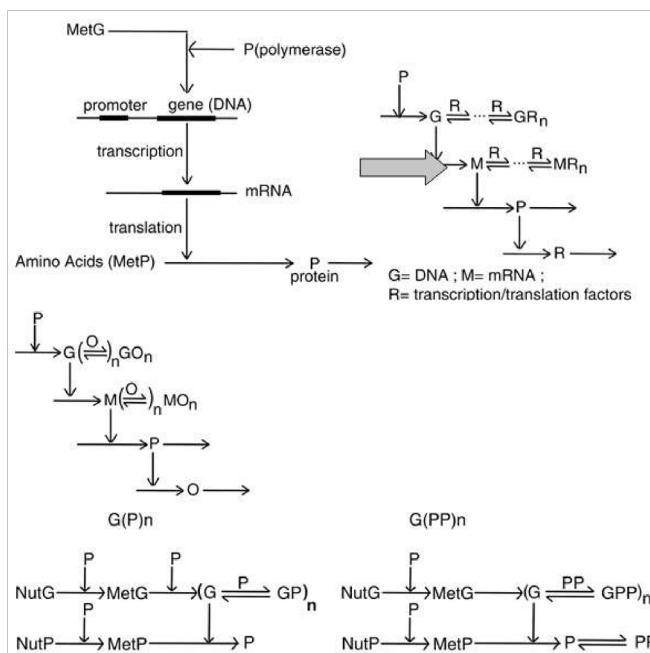
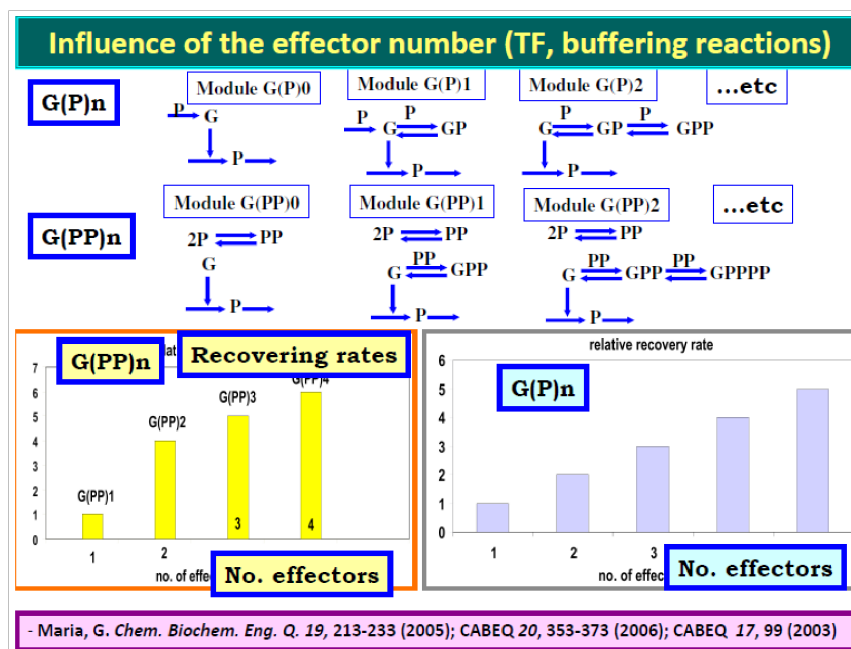


Figure 4-28: Several „simplified representations of a gene expression regulatory modules (GERM) [21-24]. Adapted from [24,107] with the courtesy of CABEQ JI. Down-right. Simplified reaction scheme of a generic gene  $G$  expression, by using a regulatory module of [G(P)1] type. The model was used to exemplify the synthesis of a generic  $P$  protein in the *E. coli* cell by [24]. 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  $G(\text{active}) + P \rightleftharpoons GP(\text{inactive})$  has been added. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions;  $G$  = active part of the gene encoding protein  $P$ ;  $GP$  = inactive part of the gene encoding protein  $P$ ;  $MetG$ ,  $MetP$  = lumped DNA and protein precursor metabolites, respectively.”



**Figure 4-29:** Protein P synthesis - simplified representations of a generic gene expression regulatory module (GERM). The horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicate an assumed concentration invariance of these species; **G** = gene encoding P; **M** = mRNA). The right structure corresponds to a  $[G(R)n; M(R)n]$  module type. **Uprow:** the simplified representation of the gene expression two-steps: transcription and translation (left), and (right) a simplified reaction schema of the gene expression [21,24]. The right model corresponds to a  $[G(R)n; M(R)n]$  GERM type. **Center:** (adapted from [24]): Protein P synthesis - simplified representation with a GERM self-regulated expression module of type  $[G(O)n; M(O)n]$ . Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; absence of a substrate or product indicate an assumed concentration invariance of these species; **G** = DNA gene encoding P; **M** = mRNA; **O** = allosteric effectors). **Down-row:** two types of GERM simplified representations for protein synthesis:  $[G(P)n]$  (left) and  $[G(PP)n]$  (right) [24].



**Figure 4-30:** Influence of the number of effectors in a GERM on their regulatory performance indices PI.-s[21-23].



### GERM or GRC regulatory performance indices translated in mathematical terms

- **Performance indices** (QSS= quasi-steady-state):
  - **SELECTIVITY**: to inner/external stimuli
  - **SENSITIVITY**: QSS levels vs. endo-/exogeneous stimuli
  - **ROBUSTNESS**: small sensitivity (States vs. Parameters)
  - **EFFICIENCY**: large margins of QSS stability vs. small changes, quick recovering of QSS
  - **RESPONSIVENESS**: small transition time to the new QSS after a step perturbation
  - **STABILITY STRENGTH**: high QSS recovery rates (RD) after an impulse perturbation;
  - **SPECIES CONNECTIVITY** during transitions: small St.Dev. of recovering / transition times
- **Apply unconventional estimators**:
  - Optimise individual module properties
  - Reproduce GRC holistic properties
- **Impose physical-biological constraints**:
  - Maximum regulatory efficiency
  - Structural, functional, temporal cell hierarchy (systemic properties)
  - Key species homeostasis
  - Cell system flexibility vs. environmental changes
  - Holistic properties of GRC (gene connectivity and sharing functionality, minim basic-levels of effectors / intermediates, system robustness)

Maria, G A review of some novel concepts applied to modular modelling of genetic regulatory circuits , Juniper publ., Newbury Park, CA, 2017, ISBN 978-1-946628-07-7(USA). <https://juniperpublishers.com/ebook-info.php>

- Maria, G. *Chem. Biochem. Eng. Q.* 19, 213-233 (2005); *CABEQ* 20, 353-373 (2006); *CABEQ* 17, 99 (2003)

Figure 4-31: Seeking for modelling the GRC regulatory properties (performance indices, P.I.-s [21].

### GERM or GRC regulatory performance indices (P.I.) translated in mathematical terms

**Stationary indices:**

- Stability : (Min) change of homeostase vs. stationary perturbations
- Responsiveness : (Min) transition times toward a new steady - state
- Sensitivity : (Min) S(steadystate vs. stationary perturbations)
- Robustness : (Min) sensitivity of steady - state to rate constants

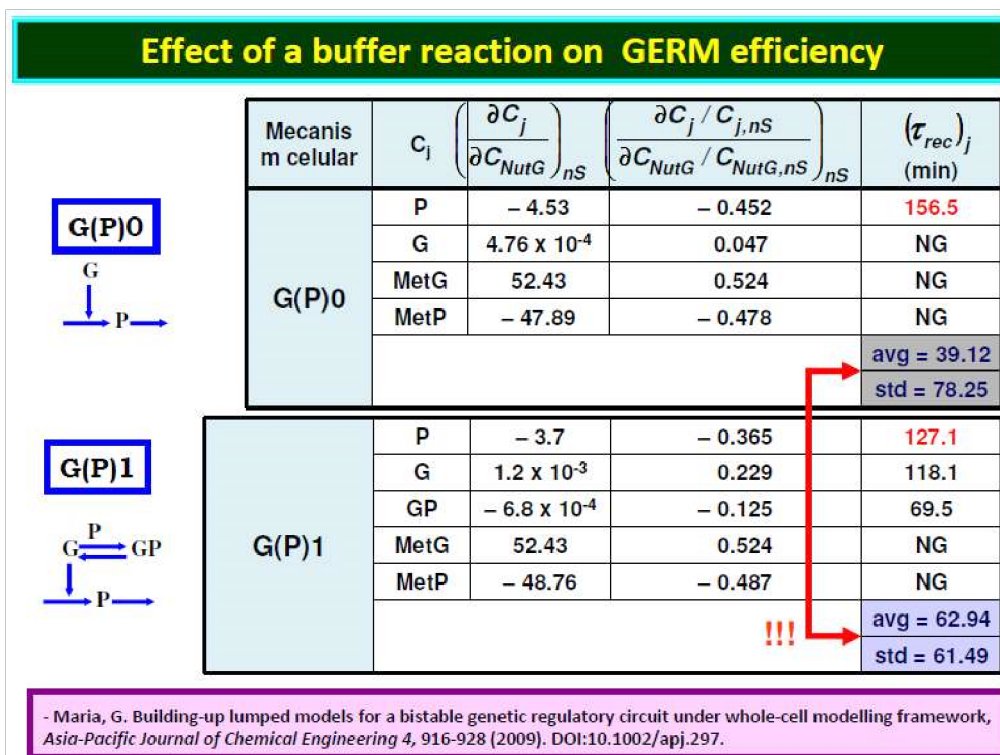
**Dynamic indices:**

- Stability strength: (Min)Max (Re(L(i)) < 0), L(i)= Jacobian eigenval.
- Dynamic efficiency: (Min) species recovering times after Delta perturb.
- Species interconnectivity: (Min){Avg.St.Dev.}(sp. Rec. times)
- Regulatory robustness: (Min)(sp. Rec. rates vs. rate constants)

Maria, G A review of some novel concepts applied to modular modelling of genetic regulatory circuits , Juniper publ., Newbury Park, CA, 2017, ISBN 978-1-946628-07-7(USA). <https://juniperpublishers.com/ebook-info.php>

- Maria, G. *Chem. Biochem. Eng. Q.* 19, 213-233 (2005); *CABEQ* 20, 353-373 (2006); *CABEQ* 17, 99 (2003)

Figure 4-32: Define some of the GERM regulatory performance indices P.I.-s [21].



- Maria, G. Building-up lumped models for a bistable genetic regulatory circuit under whole-cell modelling framework, *Asia-Pacific Journal of Chemical Engineering* 4, 916-928 (2009). DOI:10.1002/apj.297.

Figure 4-33: The effect of a buffer reaction effector (G+P ↔ GP) on the [G(P)1] gene expression regulatory module dynamic efficiency [175,176,189].

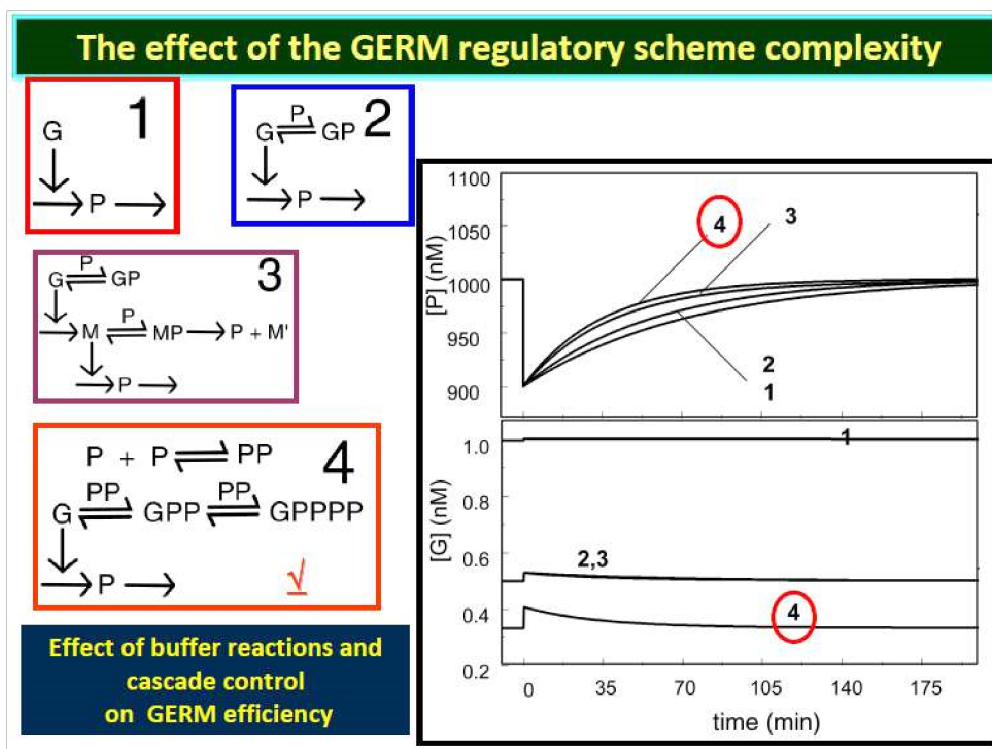


Figure 4-34: GERM model complexity reflected on its dynamic regulation efficiency [21,175,176].

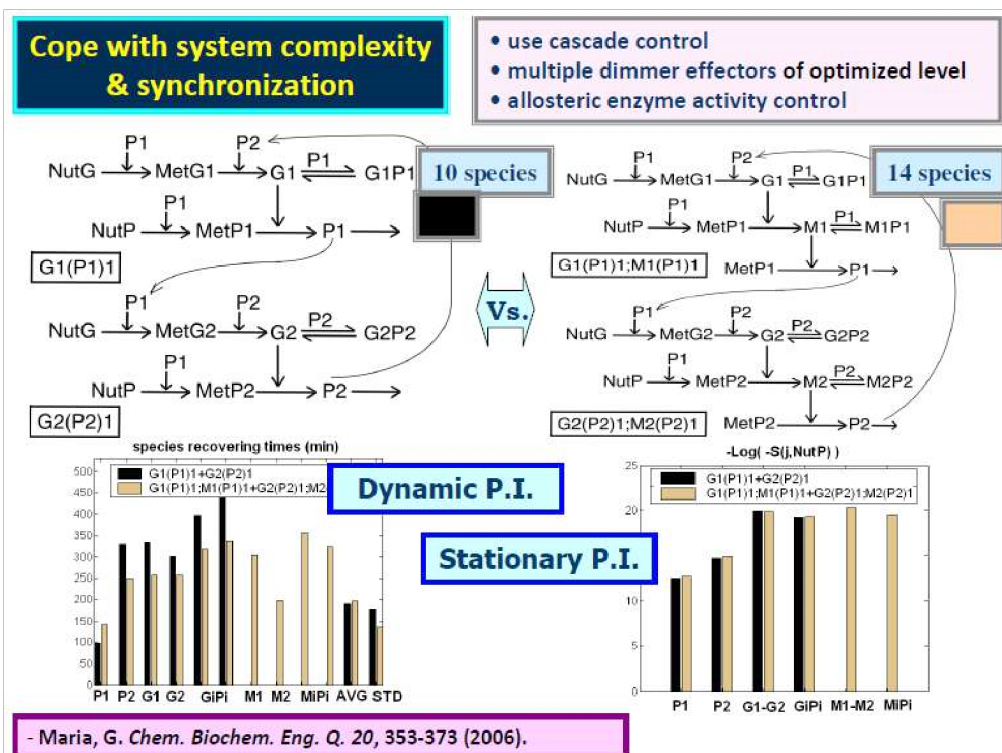


Figure 4-35: Effect of the GERM model complexity on the GRC stationary performances (P.I.) of Table 1-3. The GRC includes 2 linked (through P1) gene expression modules, i.e. (structure "1") [G1(P1)1] + [G2(P2)1], vs. [G1(P1)1;M1(P1)1] + [G2(P2)1; M2(P2)1] (structure "2"). The GRC structure no."2", of a higher complexity, reported better stationary and dynamic P.I.-s [78]

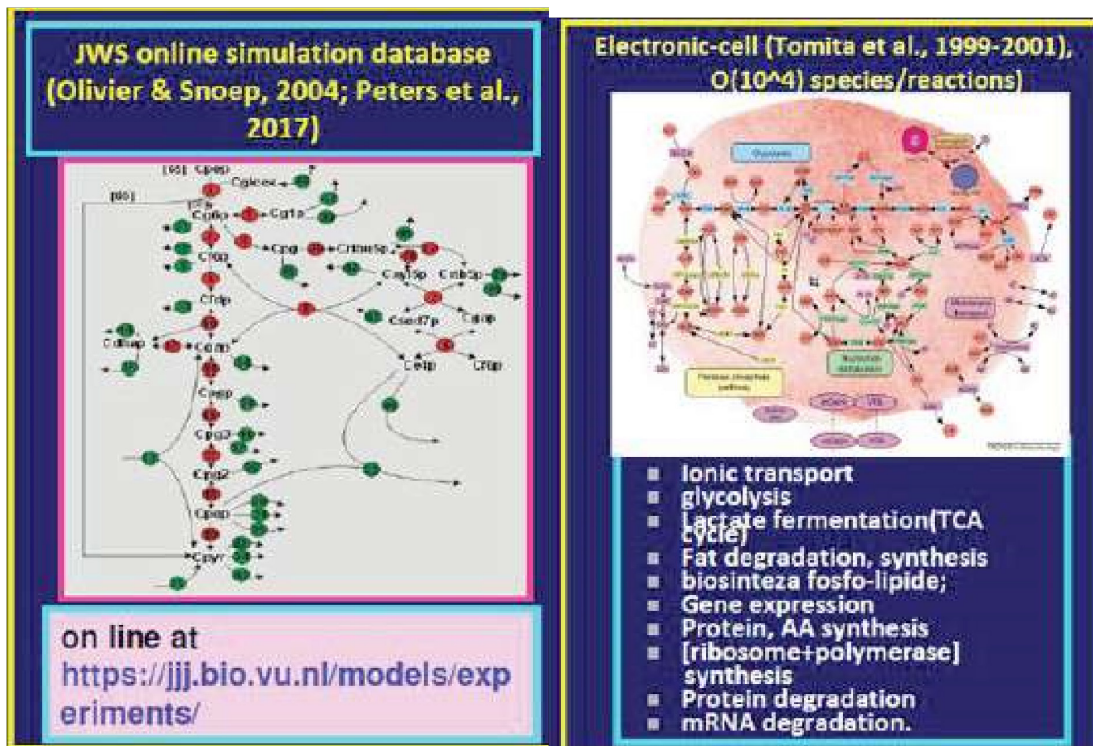


Figure 4-36: Effect of the TF level on the GERM efficiency [174,189]. Downright= the minimum (optimum) level of TF = P2P2 = P3P3 in a genetic switch case [189].

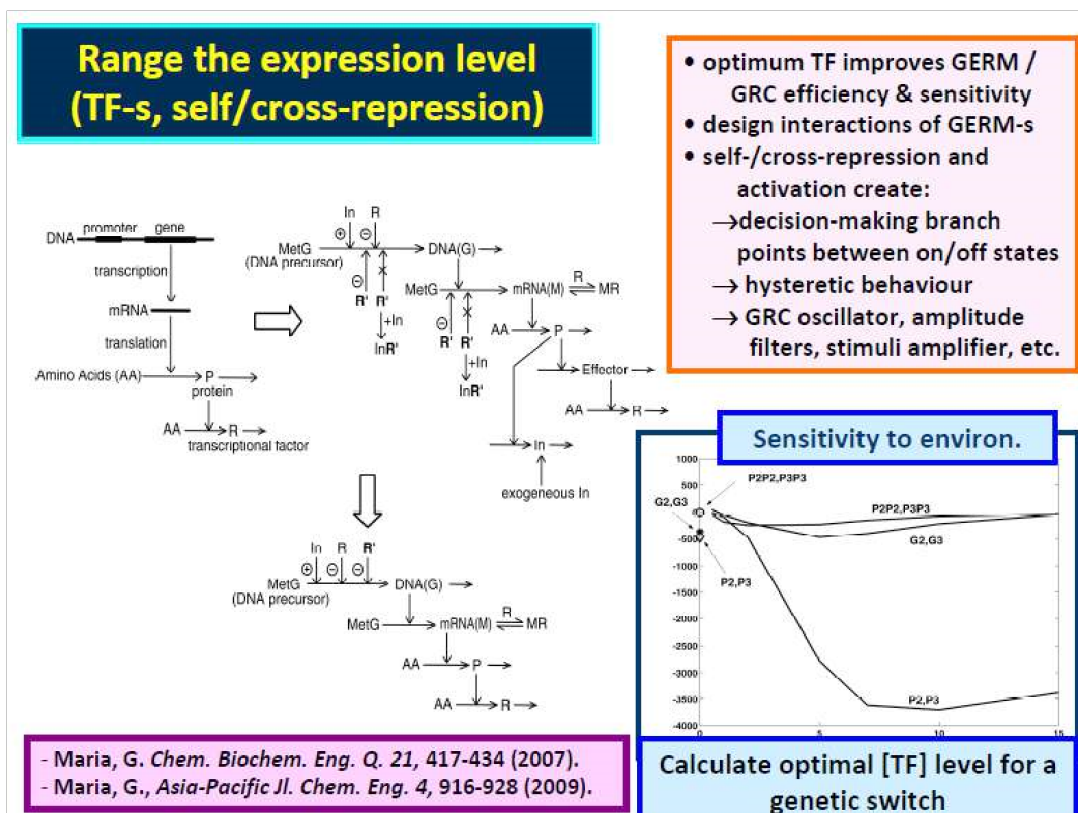


Figure 4-37: (left) E-CELL simulator of the CCM Tomita et al., 1999]; (right) CellML (JWS) cell metabolism dynamic or stationary simulation [49,100].

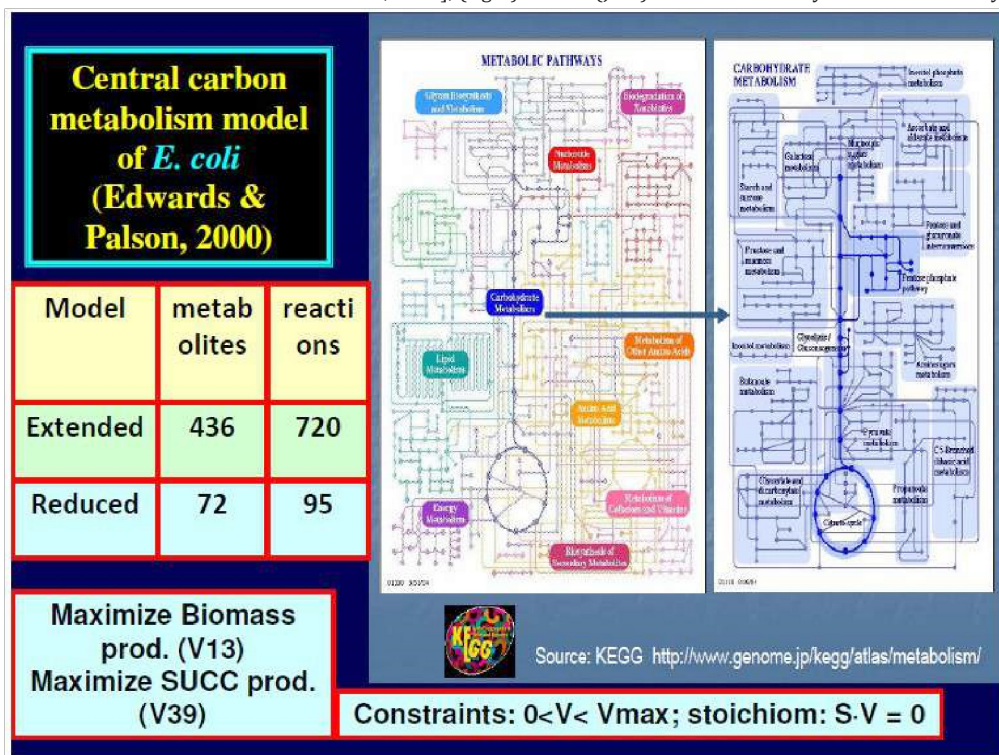


Figure 4-38: Carbohydrate metabolism in *E. coli* [96,97]. The extended kinetic model of Edwards and Palson [4] was reduced by Maria et al. [42] to only 72 key-metabolites, over 95 key-reactions.

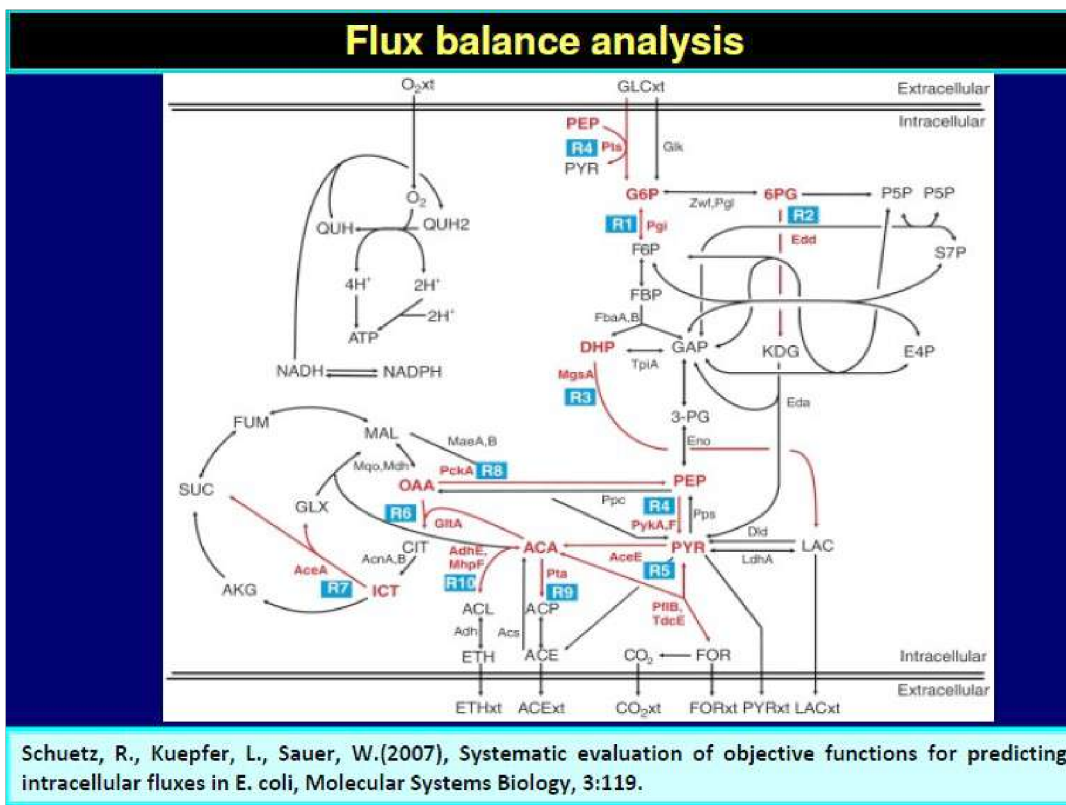


Figure 4-39: Flux balance analysis – the main objective of metabolic engineering, and an essential preliminary step in design GMO-s. [237].

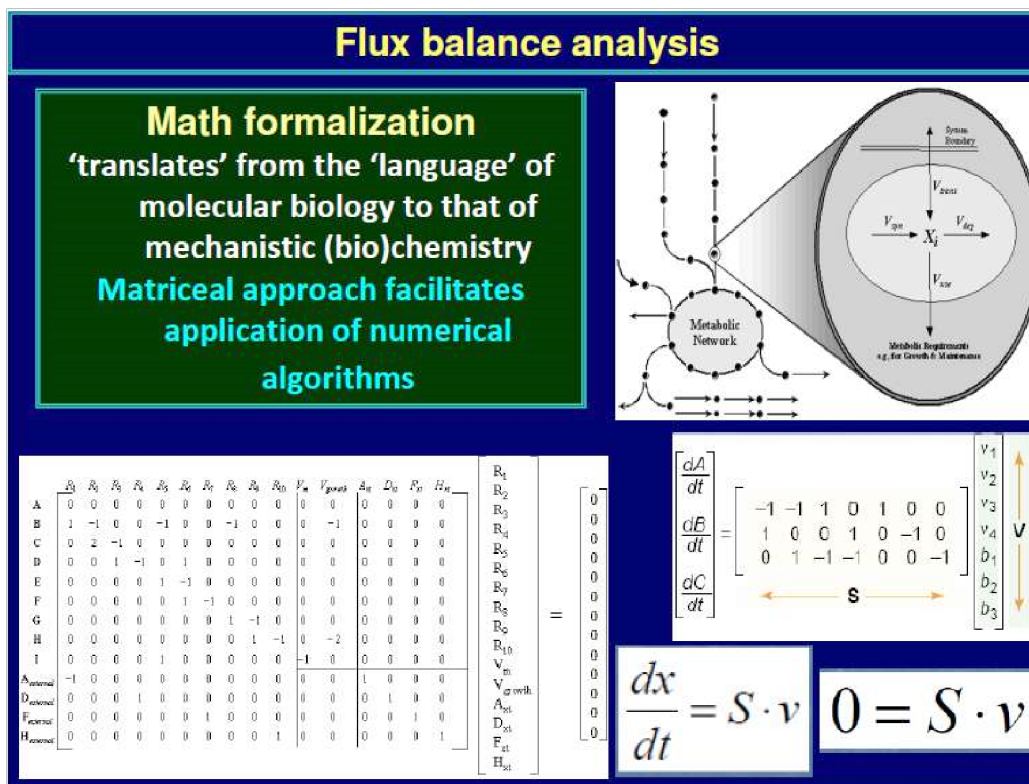


Figure 4-40: (continuation) Flux balance analysis is working with matrix math models [188].

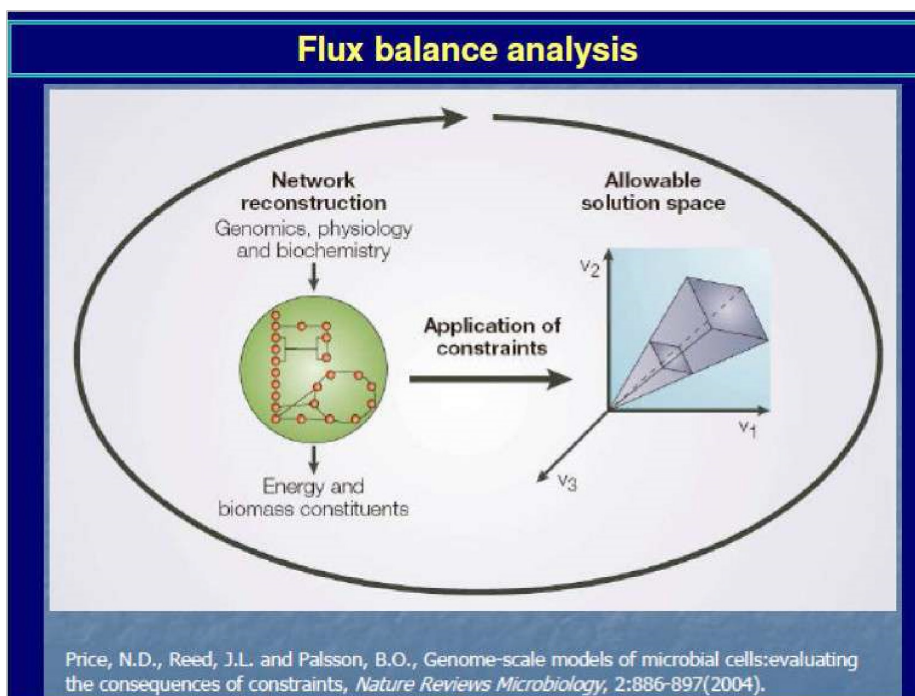


Figure 4-41: (continuation) Solving a Flux balance analysis problem translates in (non)linear programming problem [238].

### KEGG

### Omics databanks Simulation platforms

**JWS online simulation database (Olivier & Snoep, 2004; Peters et al. 2017)**

on line at <https://jjj.bio.vu.nl/models/experiments/>

Figure 4-42: Some developed -omics databanks: KEGG [96,97]; JWS [49,100].

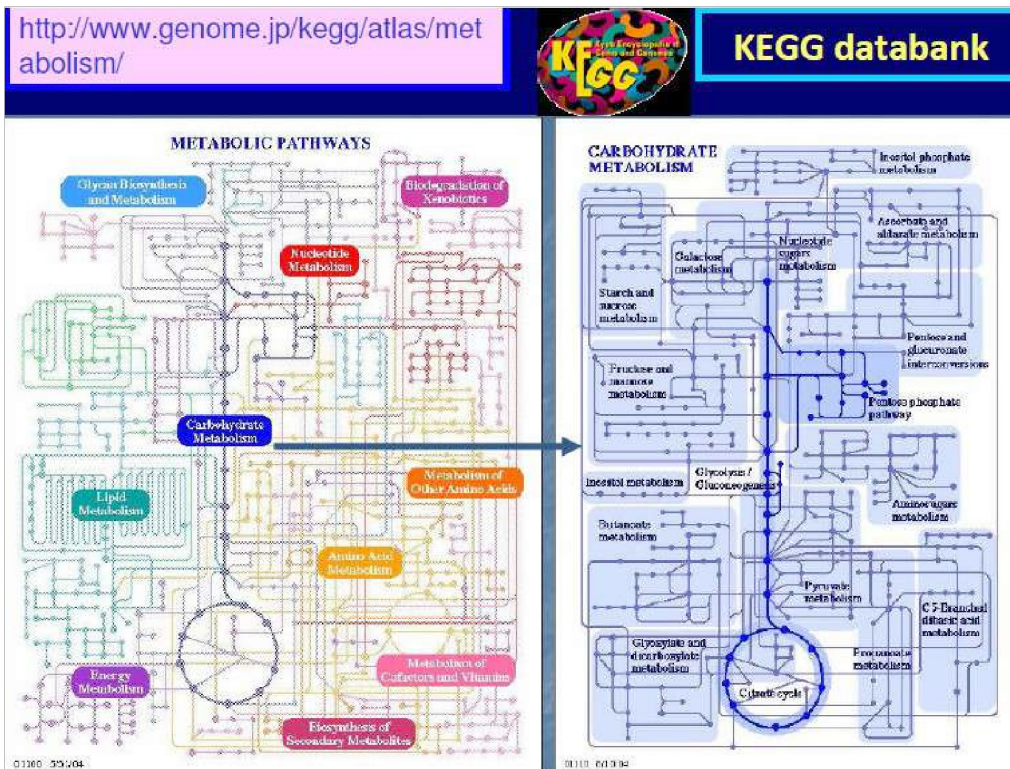


Figure 4-43: KEGG databank [96,97].

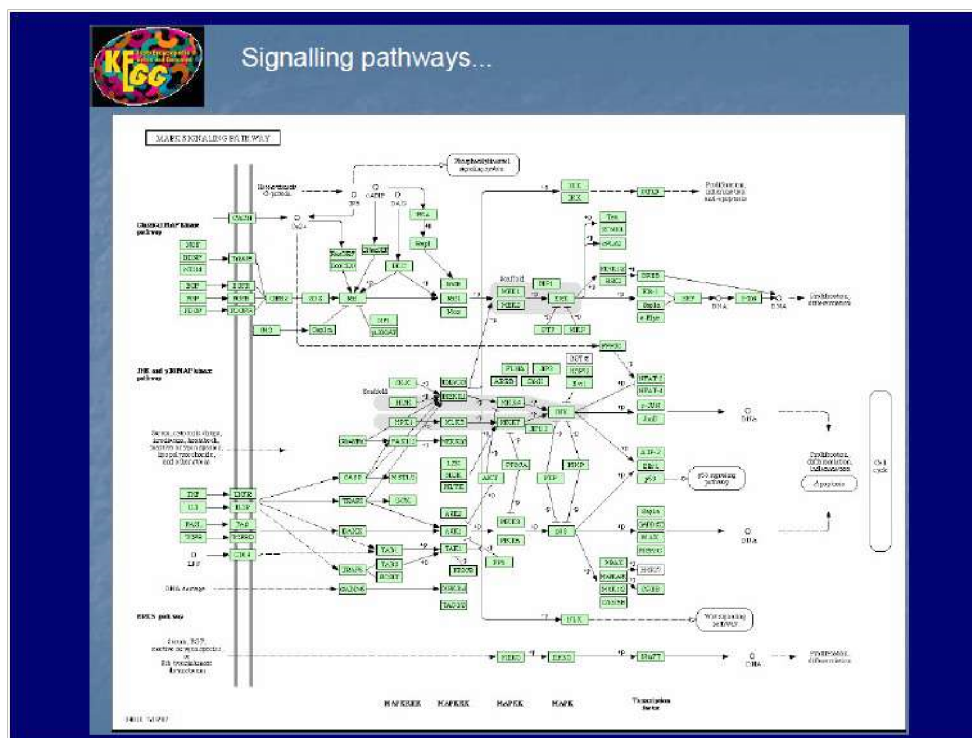


Figure 4-44: (continuation) KEGG databank [96,97].

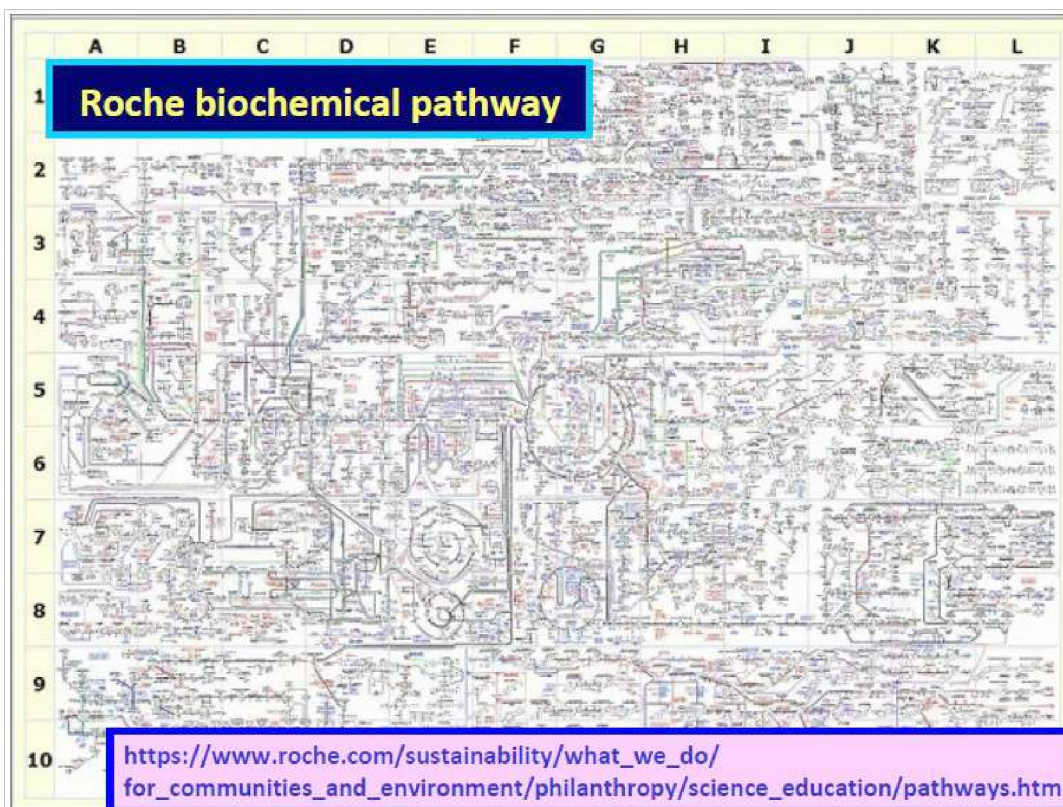


Figure 4-45: Roche biochemical databank [134].

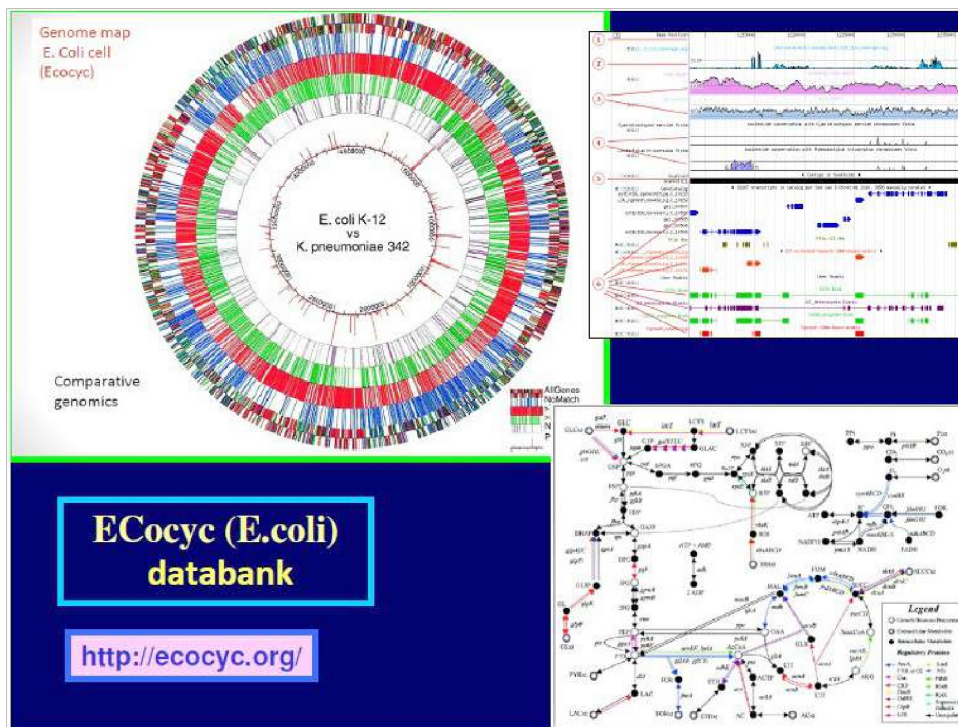


Figure 4-46: EcoCyc biochemical databank [95].



### Where systems biology (*in-silico* design of GMO) has been applied ???

- Design mutant, cloned cells with desired 'motifs'
- Genetics biology or genetics
- Biotechnology, Bioengineering, Industrial Biosynthesis
- Biomedical engineering
- Biochemistry
- Agricultural biology Ecology
- Biophysics
- Bioinformatics
- Cell biology
- Food science
- Immunology
- Molecular biology
- Biodiversity

“...In a context of increasing calls for biology to be predictive, modeling and optimization are the only approaches biology has for making predictions from first principles...”

Source: Current Trends in Biomedical Eng & Biosciences 3(2): CTBEB.MS.ID.5555606 (2017)

Figure 4-47: Area where *Systems Biology* has been applied.

### Experiments ↔ Mathematical modelling cycle in (bio-)chemical / molecular engineering

**Math formalization:**

**vectors, matrices**

**(species, conc., rates, fluxes, diff. mass balances)**

**Copynumbers (n-moles)**  
[n1, n2,...nj,..., ns]

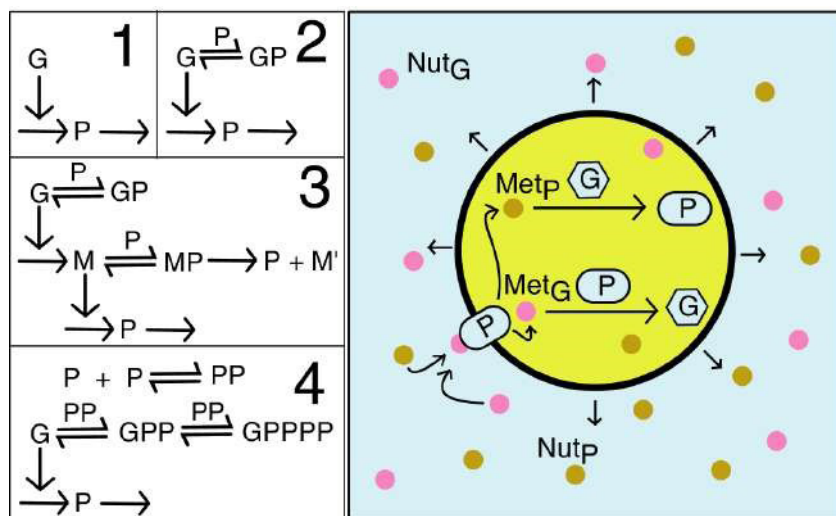
**Species diff. mass balances**

$$r = \left( \frac{dC_1}{dt} \quad \frac{dC_2}{dt} \quad \dots \quad \frac{dC_j}{dt} \quad \dots \quad \frac{dC_s}{dt} \right) = (f_1 \quad f_2 \dots f_j \dots f_s)$$

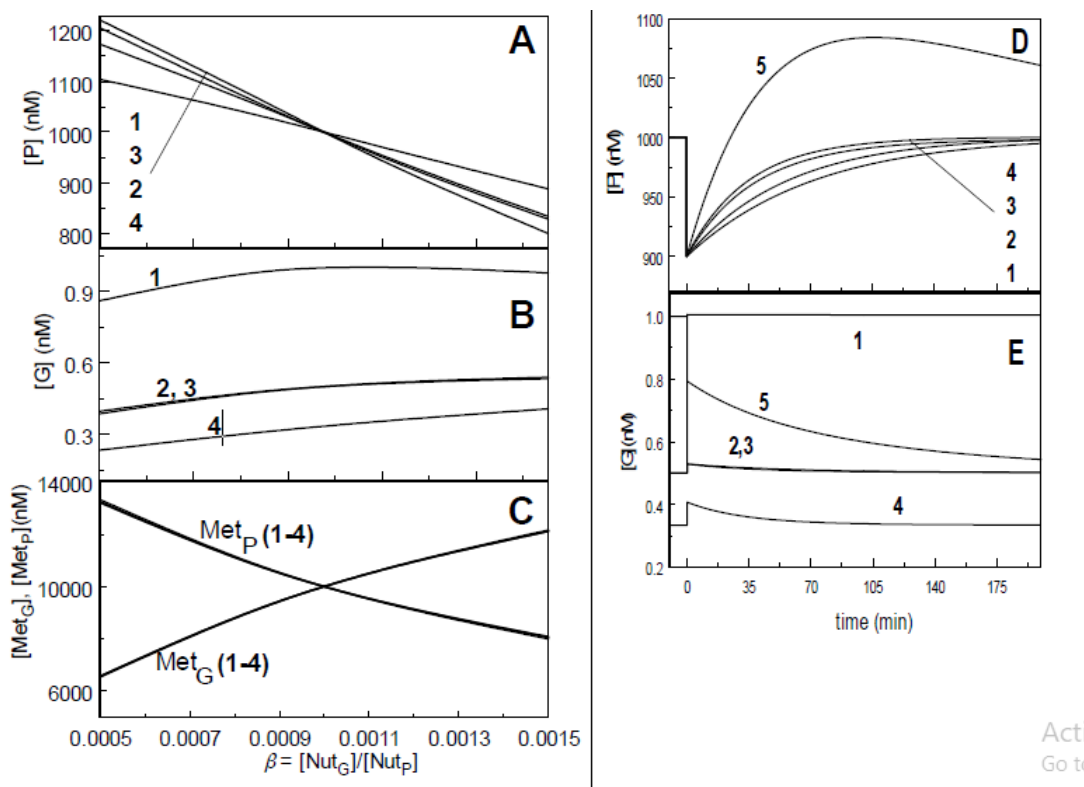
**Concentrations, (nM)**  
 $C_j = n_j/V_{cyt}$

$$J = \left[ \frac{df_i}{dC_j} \right], \text{ system Jacobian (invariant over lumping)}$$

Figure 4-48: Experiment-modelling cycle to obtain math models in biochemical / metabolic engineering. Math formalization includes working with vectors and matrices.” [239].



**Figure 4-49:** Protein homeostatic regulatory mechanisms no. 1-4, of type  $[G(P)0]$ ,  $[G(P)1]$ ,  $[G(P)1;M(P)1]$ , and  $[G(PP)2]$ . Horizontal arrows indicate reactions. Vertical arrows indicate that the component catalyzes the designated Components above horizontal arrows indicate substrates. Adapted after [21-24].



**Figure 4-50:** [A-C]: Steady-State trajectories of  $P$ ,  $G$ ,  $MetG$  and  $MetP$  vs. environmental  $b = NutG/NutP$  for the following GERM representations of (Figure 4-49): (1)-  $[G(P)0]$ ; (2)-  $[G(P)1]$ ; (3)-  $[G(P)1;M(P)1]$ ; and (4)-  $[G(PP)2]$ . See the nomenclature in (Figure 4-7). [D-E]: Dynamic Trajectories (2-5) of  $P$  and  $G$  after a 10% negative perturbation of  $[P]$  from  $[P]_{ns} = 1000$  nM, for GERM-s of type no. (1-4). At  $[G(P)1]$  curve no.1,  $[G]$  was set to 1 nM at the moment of perturbation. Nominal QSS concentrations before the Pperturbation at an arbitrary  $t=100$  min, are  $[MetG]_{ns} = [MetP]_{ns} = 10,000$  nM,  $[P]_{ns}=1000$  nM,  $[G]_{ns} = [GP]_{ns} = 1/2$  nM,  $[M]_{ns} = [MP]_{ns} = 1/2$  nM.. Adapted after [21-24].

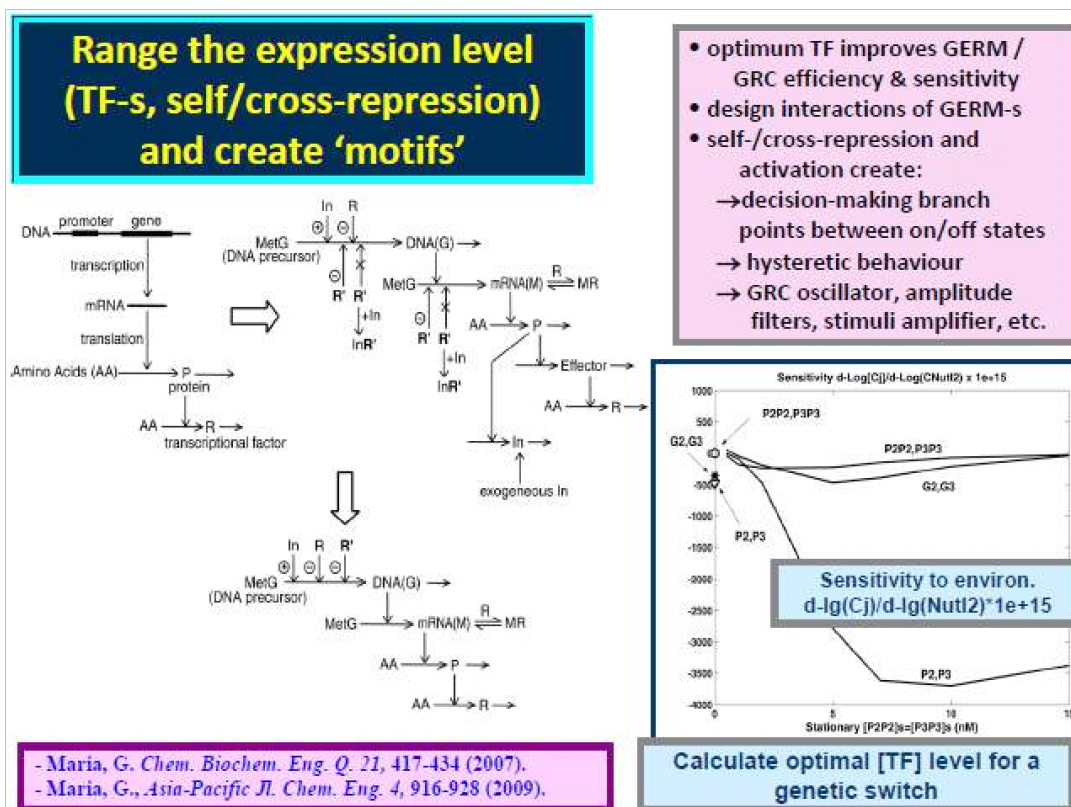


Figure 4-51: Effect of the TF level (expressed proteins P2, P3 here) on the GERM efficiency [174,189].

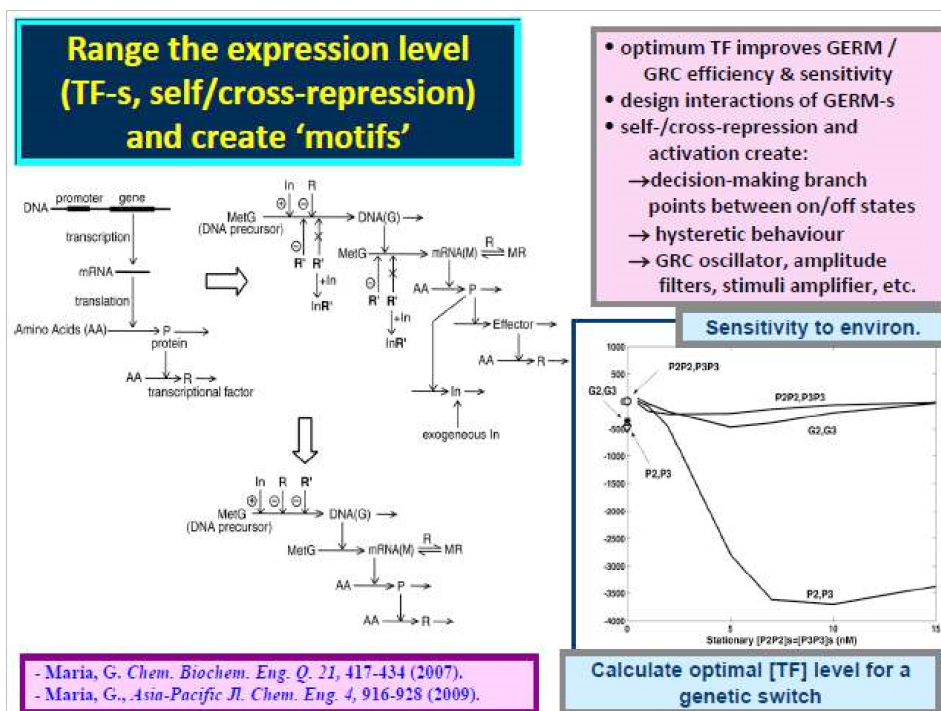


Figure 4-52: Exemplification of the cell content ballast effect on the species recovering times to homeostasis, in the case of a  $[G(P)1]$  gene expression regulatory module. Gene  $G$  (top-right) and its encoding protein  $P$  (down-left) recovery trajectories after a  $-10\%$  impulse perturbation in the  $[P]_s = 1000$  nM at  $t=0$ . Solid line trajectories correspond to a high ballast cell, while the dash line trajectories to a low ballast cell. The species concentrations in nM are given in the Table 2 of [176].

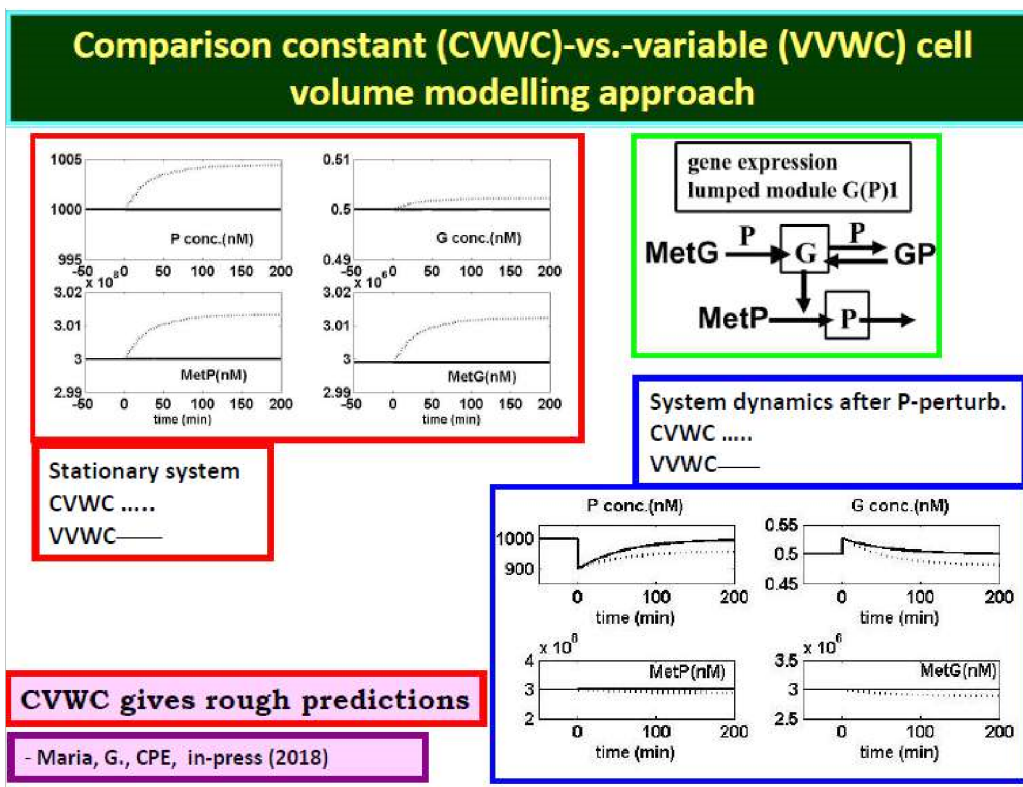


Figure 4-53: (continued) Comparison of the WCCV vs. WCVV modelling approach in the case of a simple [G(P)1] gene expression regulatory module [60].

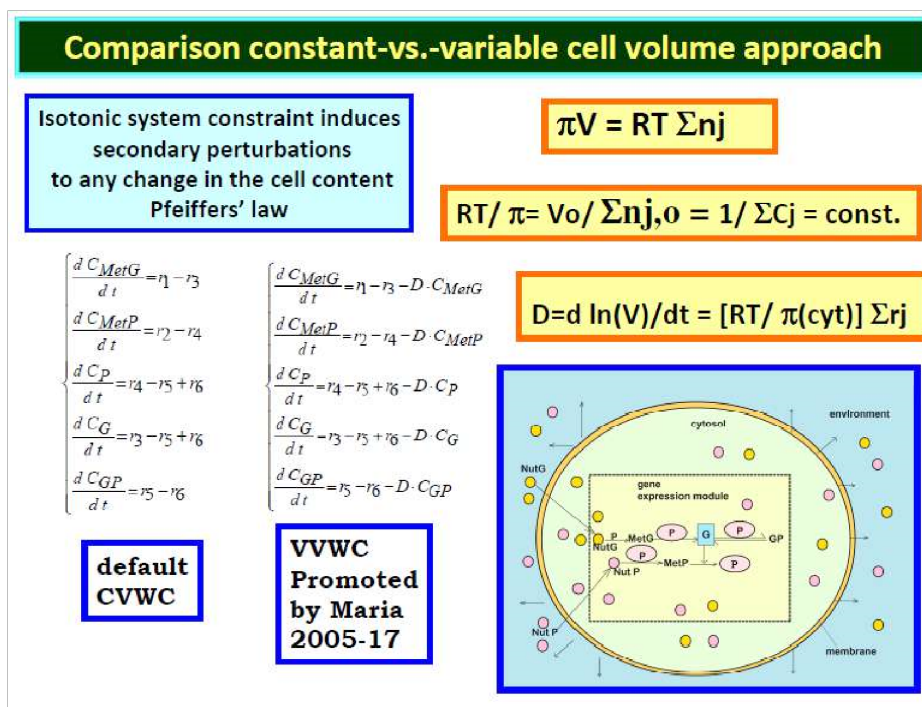
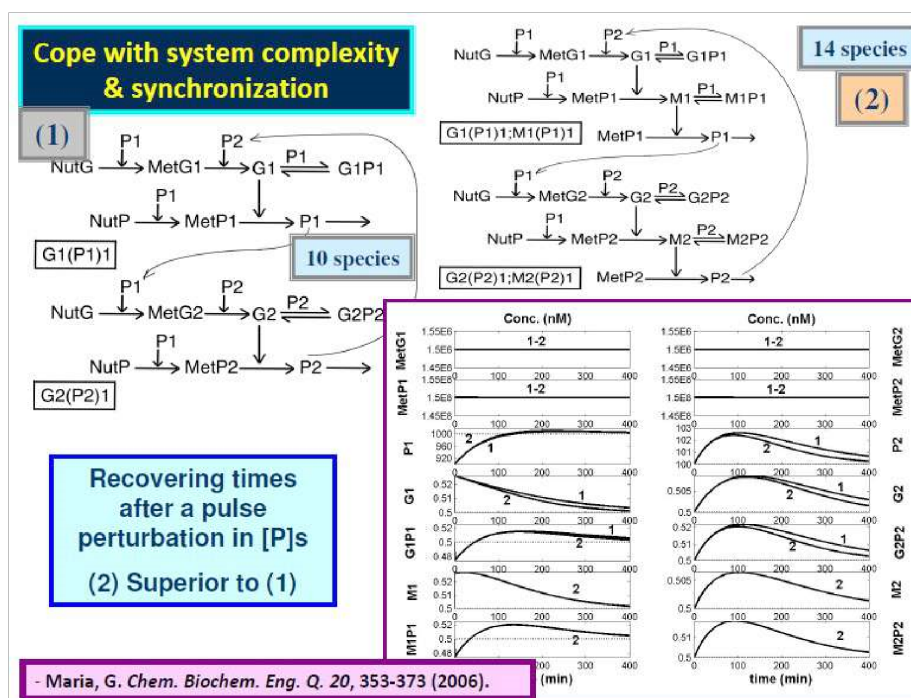
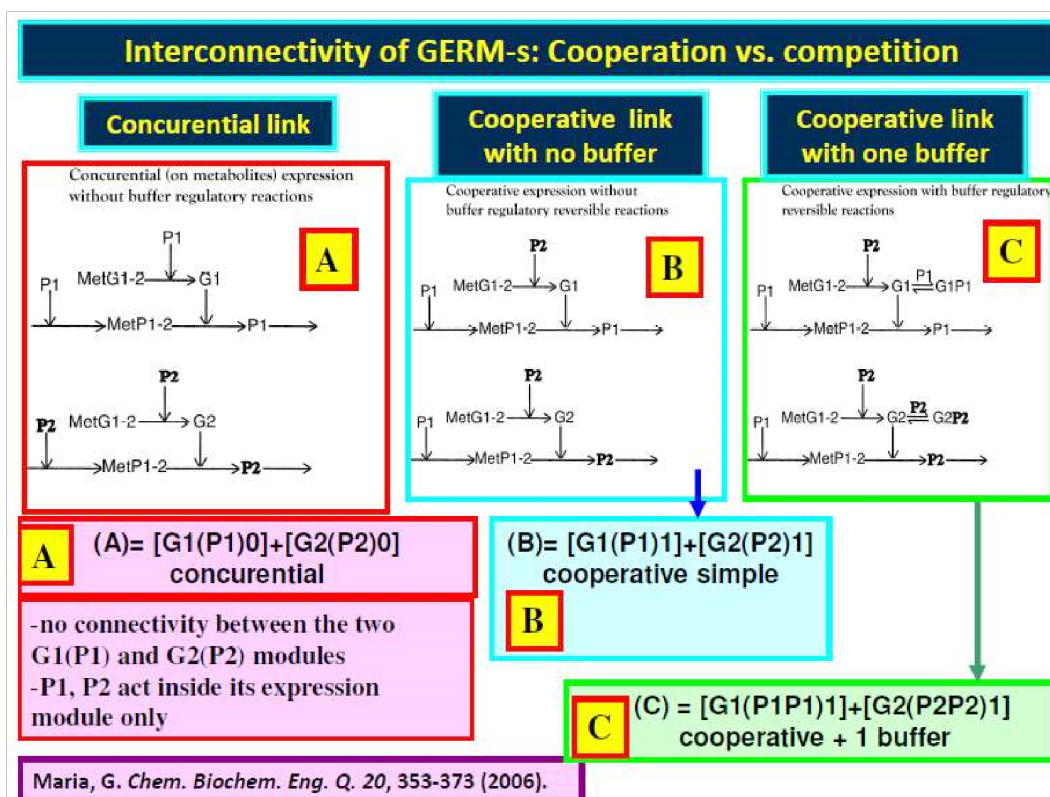


Figure 4-54: Comparison of the WCCV vs. WCVV modelling approach in the case of a simple [G(P)1] gene expression regulatory module [60].



**Figure 4-55:** Effect of the GERM model complexity on the GRC dynamic performances (P.I.). The GRC includes 2 linked (through P1) gene expression modules, i.e. (structure "1")  $[G1(P1)1] + [G2(P2)1]$ , vs.  $[G1(P1)1; M1(P1)1] + [G2(P2)1; M2(P2)1]$  (structure "2"). The GRC structure no."2", of a higher complexity, reported better P.I.-s. After [78].



**Figure 4-56:** Effect of the GERM inter-connectivity, and of individual species functions into the cell on the GRC efficiency [78].

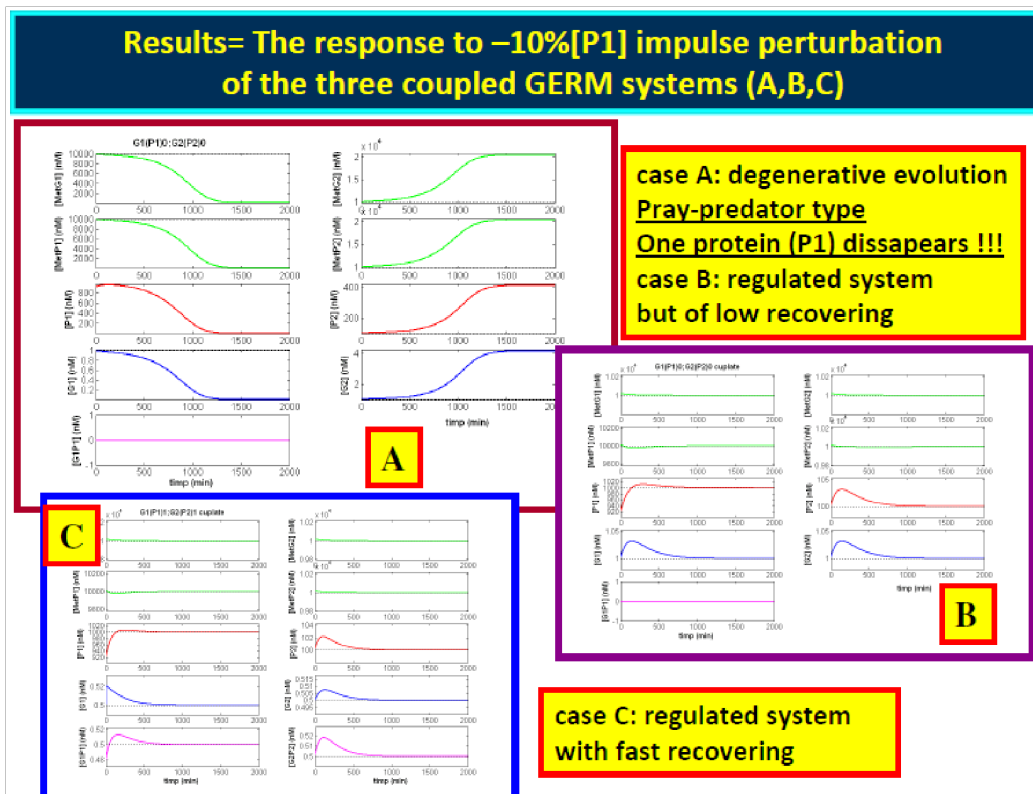


Figure 4-57: (continued) Effect of the GERM inter-connectivity, and of individual species functions into the cell on the GRC efficiency [78].

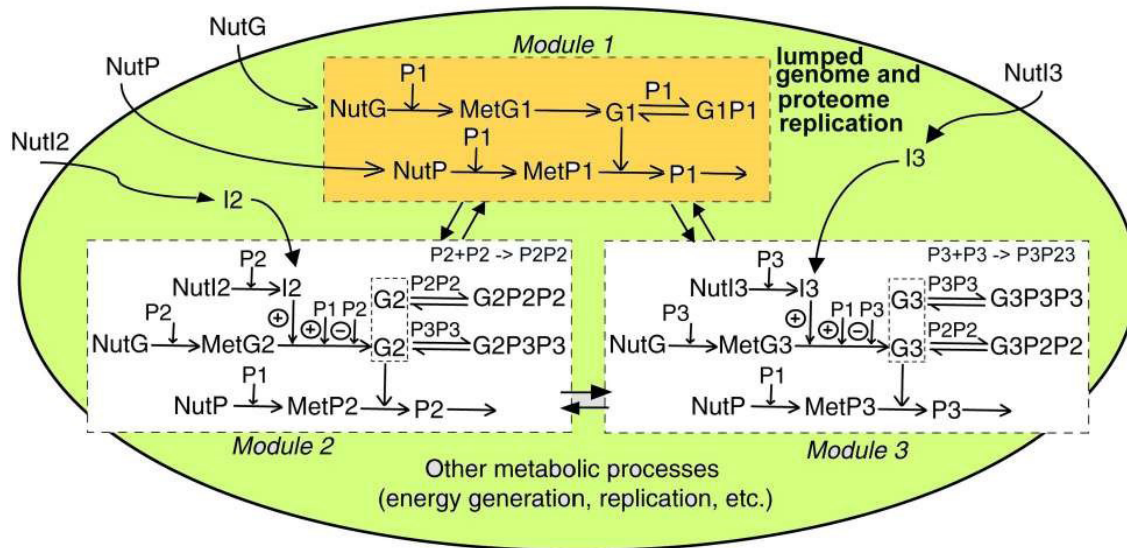


Figure 4-58: "Schema of a design genetic switch (GS): GERM modules 2 and 3, for G2 and G3 expression, are placed in a WCVV of a cell of equilibrated growth; the G1 expression (module 1) mimics the whole cell content replication; inducers I2 and I3 activate the G2 and G3 synthesis, respectively. (Notations: G1 = lumped genome; P1 = lumped proteome; MetG1, MetP1 = lumped metabolome used for the synthesis of G1 and P1, respectively; NutP and NutG = lumped external nutrients; P2 and P3 = two individual proteins involved in the GS of the type=[G2(P2P2)1(P3P3)1] + [G3(P3P3)1G3(P2P2)1]G2 and G3 = two individual genes involved in the GS; MetG2, MetG3, MetP2, MetP3 = individual metabolites involved in the GS; I2 and I3 = internal inducers for the G2 and G3 expression, respectively; NutI2 and NutI3 = external stimuli responsible for the I2 and I3 synthesis;  $\wedge$  /  $\ominus$  = positive / negative regulatory loops; NutG and MetG are precursors of DNA (and mRNA); NutP and MetP are precursors for the production of amino-acids used in the protein synthesis (P1, P2, P3, P2P2, P3P3). The TF-s, of type P2P2 and P3P3 are kept at a minimum level."



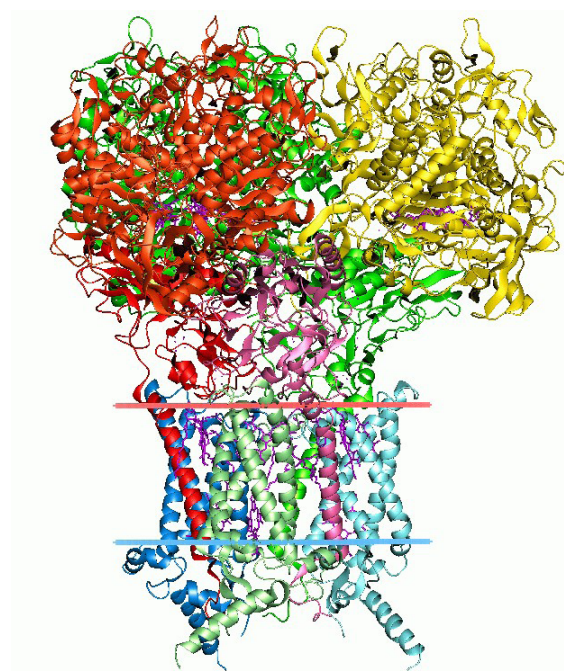


Figure 4-60: Some of the common rules used by the Bioinformatics [122].

**Top:** Example of a Multiple Sequence Alignment (MUSCLE format) analysis on human head louse species. These are sequences being compared in a MUSCLE multiple sequence alignment (MSA). Each sequence name (leftmost column) is from various louse species, while the sequences themselves are in the second column. Source = [https://en.wikipedia.org/wiki/Bioinformatics#/media/File:Muscle\\_alignment\\_view.png](https://en.wikipedia.org/wiki/Bioinformatics#/media/File:Muscle_alignment_view.png)

**Down-left:** Interactions between proteins are frequently visualized and analyzed using networks. This network is made up of protein-protein interactions from *Treponema pallidum*, the causative agent of syphilis and other diseases. Source = [https://en.wikipedia.org/wiki/Bioinformatics#/media/File:The\\_protein\\_interaction\\_network\\_of\\_Treponema\\_pallidum.png](https://en.wikipedia.org/wiki/Bioinformatics#/media/File:The_protein_interaction_network_of_Treponema_pallidum.png)

**Down-right:** 3-dimensional protein structures such as this one are common subjects in bioinformatic analyses. Source = [https://en.wikipedia.org/wiki/Bioinformatics#/media/File:1kqf\\_opm.png](https://en.wikipedia.org/wiki/Bioinformatics#/media/File:1kqf_opm.png)

**Applications of GRC models:**

→

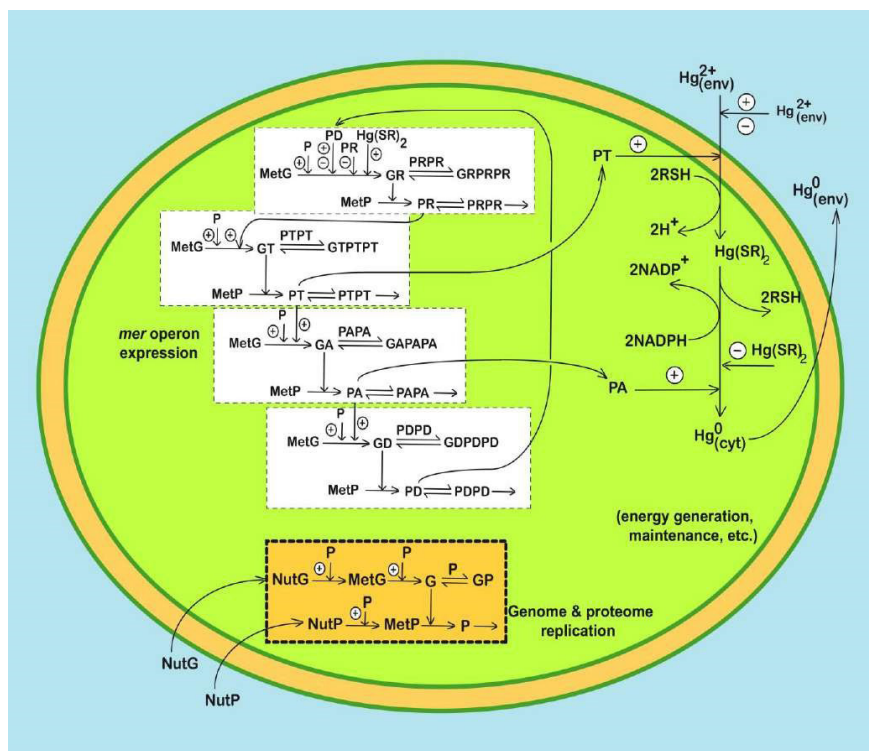
- Dynamic simulation of GRC properties (cell response to perturbations)
- *In-silico* design of GMO with desirable motifs

- **Design systems of inter-connected genes with specific functions, ('motifs' of various mini-functions, e.g. biosensors):**
  - **switches** (mutual repression control for two gene expressions)
  - **oscillators** (regular fluctuations, GRC evolving among QSSs)
  - **signal amplifiers, amplitude filters** (treatment of external signals)
  - **memory storage** (part of cell-cell communications)
- Use **modular approach** ('building-blocks' concept) to functionally link individual genes, by using inter-connected regulatory loops for controlling the expression
- **Desired properties of a genetic switch:**
  - high sensitivity to specific external inducers
  - quick and robust response to inducers (low sensitivity to noise)
  - tight control of gene expression (depending on inducers), avoiding overshoots in produced enzymes

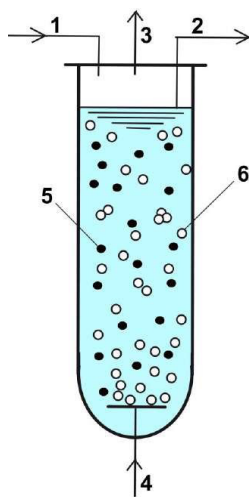
- Maria, G. *Chem. Biochem. Eng. Q.* 17(2), 99-117 (2003); Maria, G. *Chem. Biochem. Eng. Q.* 19, 213-233 (2005).  
 - Maria, G., *Chem. Biochem. Eng. Q.* 20, 333 (2006); Maria, G., Xu, Z., Sun, J., *Chem. Biochem. Eng. Q.* 25, 403, 2011.

Figure 5-1: Some applications of GRC models [21,24].





**Figure 5-2.** Time-dependent mercury transfer between the *E. coli* cell and the three-phase fluidized bed bioreactor (TPFB), according to the proposed hybrid dynamic model HSMDM of Maria [1,2,40,63]. Enzymes concentrations in *E. coli* determines the apparent reaction rates in the bioreactor model (especially PT lumped permease, and PA lumped reductase), while the reactor state-variables (e.g.  $[Hg^{2+}]$ , nutrients) determine the cell metabolism, and *mer*-operon expression adaptation. *E. coli* cell model notations: The simplified GRC pathway of the *mer*-operon expression for the mercury ion uptake in *E. coli* cells includes 7 gene expression regulatory modules (GERM-s): 2 modules for mediated transport of  $Hg^{2+}$  into cytosol (catalysed by PT) and its reduction (catalysed by PA); 5 regulatory modules of *mer* operon expression including successive synthesis of PR (the transcriptional activator of other protein synthesis; its synthesis being triggered by the import of mercury ions, linked as  $Hg(SR)_2$  into the cell), lumped PT permease, PA reductase, and of the control protein PD. One additional regulatory module deals with the lumped proteome P and genome G replication into the cell (also mimicking the cell "ballast"). The GRC is placed in a growing cell, by mimicking the homeostasis and cell response to stationary and dynamic perturbations in  $[Hg^{2+}]_{env}$ . The reductant NADPH and RSH are considered in excess in the cell. Notations: P = lumped proteome; G = lumped genome; NutG, NutP = lumped nutrients used for gene and protein synthesis; 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; A / Q positive or negative feedback regulatory loops.



**Figure 5-3:** Simplified scheme of the three-phase fluidized bedreactor (TPFB). Notations: 1 – mercury ion fed solution including nutrients (C/saccharides, N/ureea,P/salts, mineral sources, pH-bufferadditives, anti-bodies, etc.); 2 –liquid outlet; 3 – Hg vapor in airflow; 4 – sterile air input; 5 –immobilized bacteria; 6 – air bubbles.The bioreactor performances depends on: suspended biomass concentration

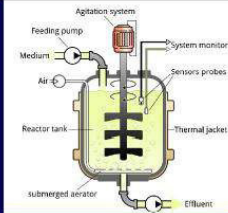

and efficiency, and its porous support size, feed flowrate, inlet  $[Hg^{2+}]$ , nutrients, additives, aeration rate, pH, etc. In turn, the cell efficiency (mercury reduction rate) depends on the PT, PA enzymes concentration (dependent on the cell resources and environmental  $[Hg^{2+}]$ ).

**Ex.2. *In-silico* design of a cloned *E. coli* to maximize the mercury uptake from wastewaters**

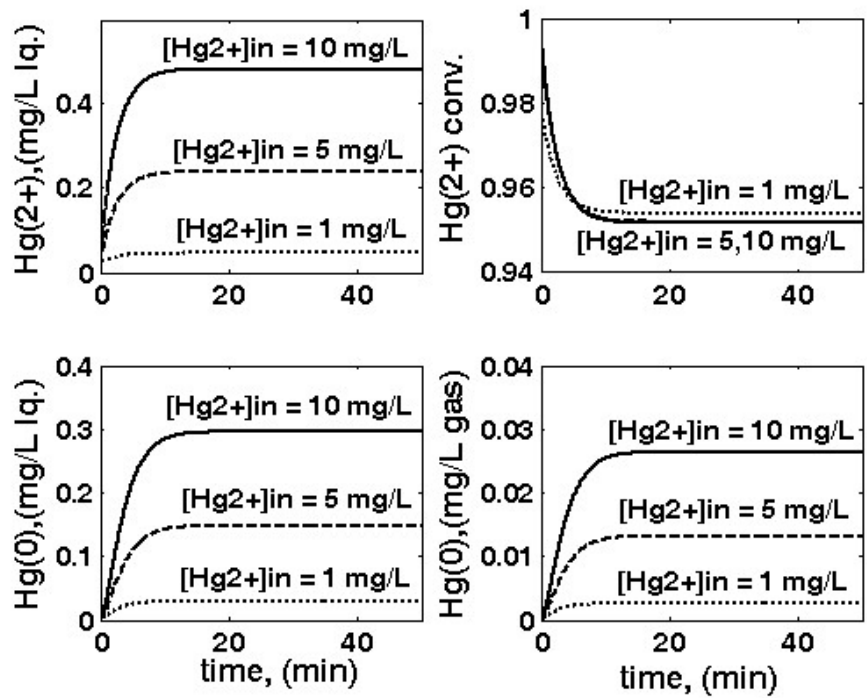
- Optimization of the three-phase fluidized bed-reactor for mercury uptake by immobilized cloned *E. coli* on pumice beads
- Coupling dynamic cell simulator with the reactor model
- Simulate bacteria response and reactor performance to environmental perturbations

**In collaboration with (late) Prof. Wolf Deckwer from TU Braunschweig (Germany), & DFG SFB-578 / 2006**

- Maria, G., Luta, I., Maria, C., *Chemical Papers* 66, 67, 1364 (2013).  
- Maria, G., Luta, I., *Computers & Chemical Engineering*, 58, 98 (2013).

**Figure 5-4:** Case study of chap. 5 - *In-silico* design of a cloned *E. coli* with a maximized capacity of mercury uptake from wastewaters [1,2,40,63,241,242,243]. [Down-right] Prof. G. Maria and late Prof. W. Deckwer at the 2nd Croatian- German Conference on Enzyme Reaction Engineering, 21-24 Sept. 2005, Dubrovnik (Croatia), sharing opinions about the bioprocess of mercury removal from wastewaters by using cultures of cloned *E. coli* cells.



**Figure 5-5:** Three-phase fluidized bed reactor sensitivity related to the variations in the inlet  $[+2L Hg]_{in}$ , leading to variations of the following operating variables: outlet  $[+2L Hg]$  (up-left); outlet  $[0L Hg]$  (down-left);  $+2L Hg$  conversion (up-right); and outlet  $[0G Hg]$  (down-right). Nominal conditions:  $[G_{mer}] = 3 \text{ nM}$ ;  $FL = 0.02 \text{ L min}^{-1}$ ; biomass  $cX = 1 \text{ g L}^{-1}$ ; particle size  $d_p = 1 \text{ mm}$ . Notation „in” refers to the bioreactor inlet.

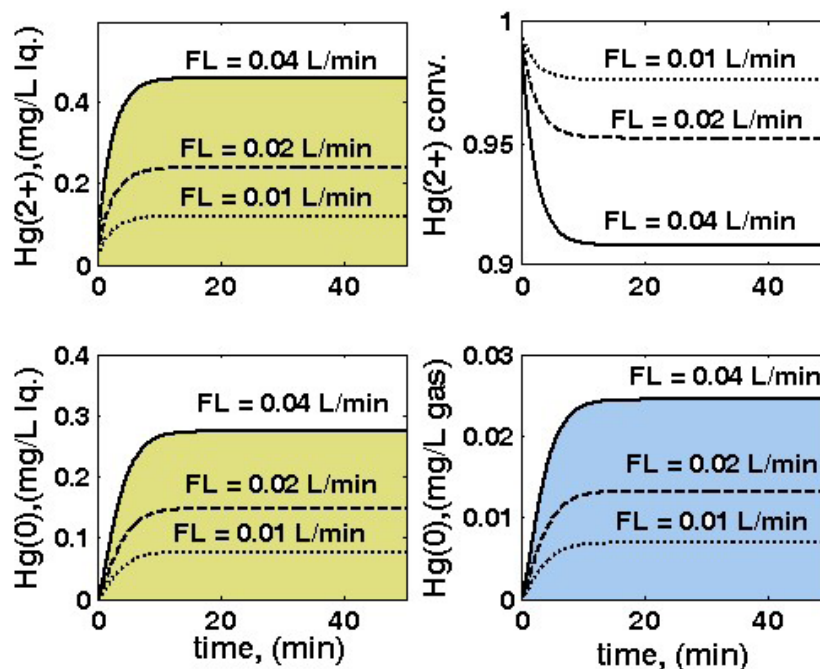


Figure 5-6: Three-phase fluidized bed reactor sensitivity related to inlet liquid flow rate  $FL$  variations, leading to variations of: outlet [ +2L Hg ](up-left); outlet [ 0L Hg ](down-left); +2L Hg conversion (up-right); outlet [ 0G Hg ](down-right). Nominal conditions:  $[G_{mer}] = 3 \text{ nM}$ ;  $[ +2L \text{ Hg } ]_{in} = 20 \text{ mg L}^{-1}$ ; biomass  $Xc = 1 \text{ g L}^{-1}$ ; particle size  $p d = 1 \text{ mm}$ .

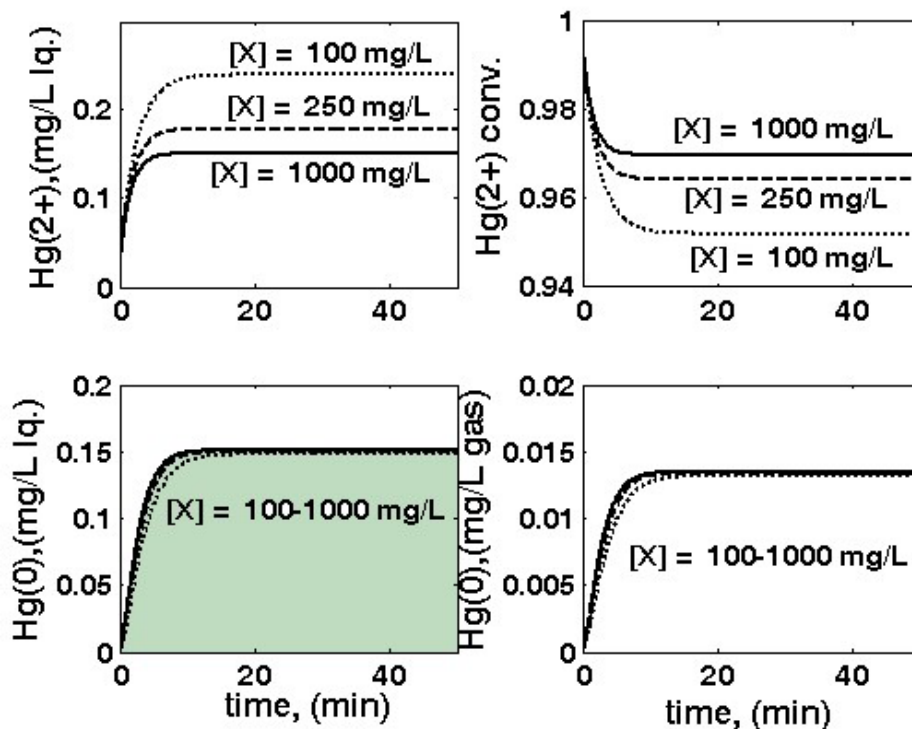
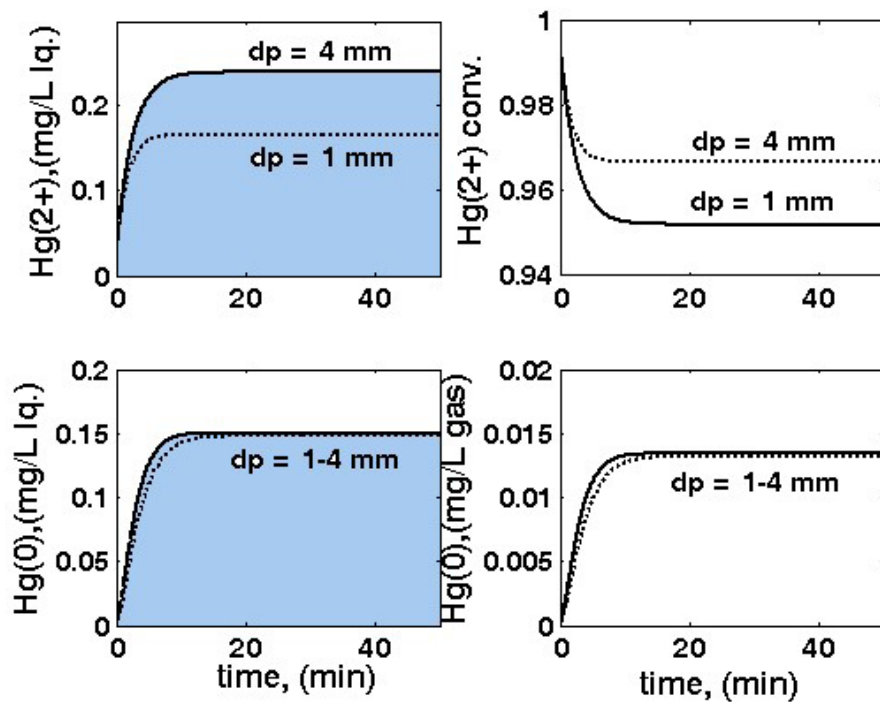
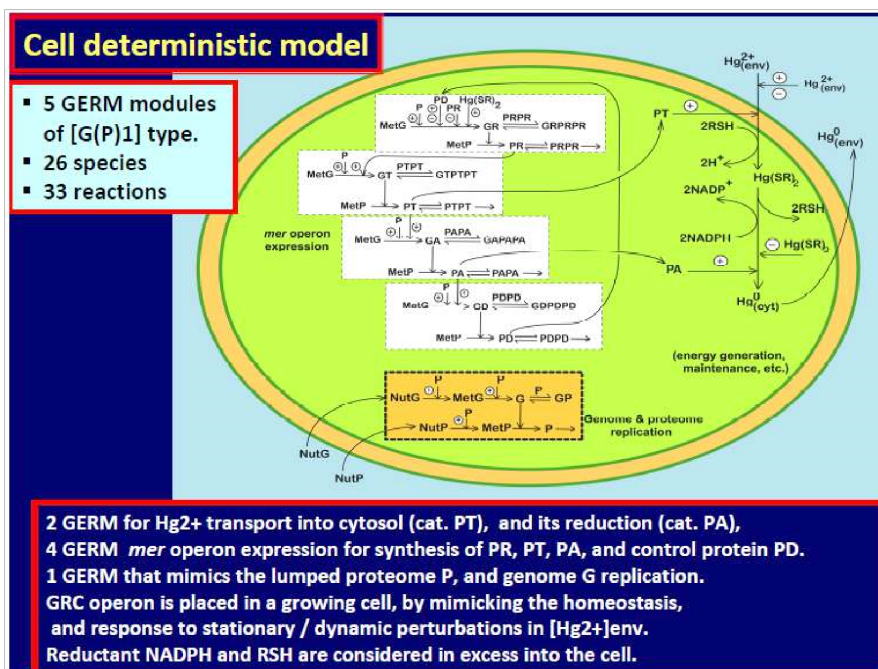


Figure 5-7: Three-phase fluidized bed reactor sensitivity related to biomass concentration  $cX$  variations, leading to variations of: outlet [ +2L Hg ](up-left); outlet [ 0L Hg ](down-left); +2L Hg conversion (up-right); outlet [ 0G Hg ](down-right). Nominal conditions:  $[G_{mer}] = 3 \text{ nM}$ ;  $FL = 0.02 \text{ L min}^{-1}$ ;  $[ +2L \text{ Hg } ]_{in} = 20 \text{ mg L}^{-1}$ ; particle size  $p d = 1 \text{ mm}$ .



**Figure 5-8.** Three-phase fluidized bed reactor sensitivity related to particle size  $dp$  variations, leading to variations of: outlet [ +2 L Hg ](up-left); outlet [ 0 L Hg ](down-left); +2 L Hg conversion (up-right); outlet mercury in the outlet-gas [ 0 G Hg ](down-right). Nominal conditions: [Gmer] = 3 nM;  $FL = 0.02 \text{ L min}^{-1}$  ; [ 2+ L Hg ]in = 20 mg L<sup>-1</sup>; biomass  $cX = 1 \text{ g L}^{-1}$ .



**Figure 5-9:**The reduced cell model, accounting for the *mer*-operon expression (4 GERM-s), coupled with 2 kinetic modules referring to the main enzymatic reactions. Adapted from [1,2,23,40].

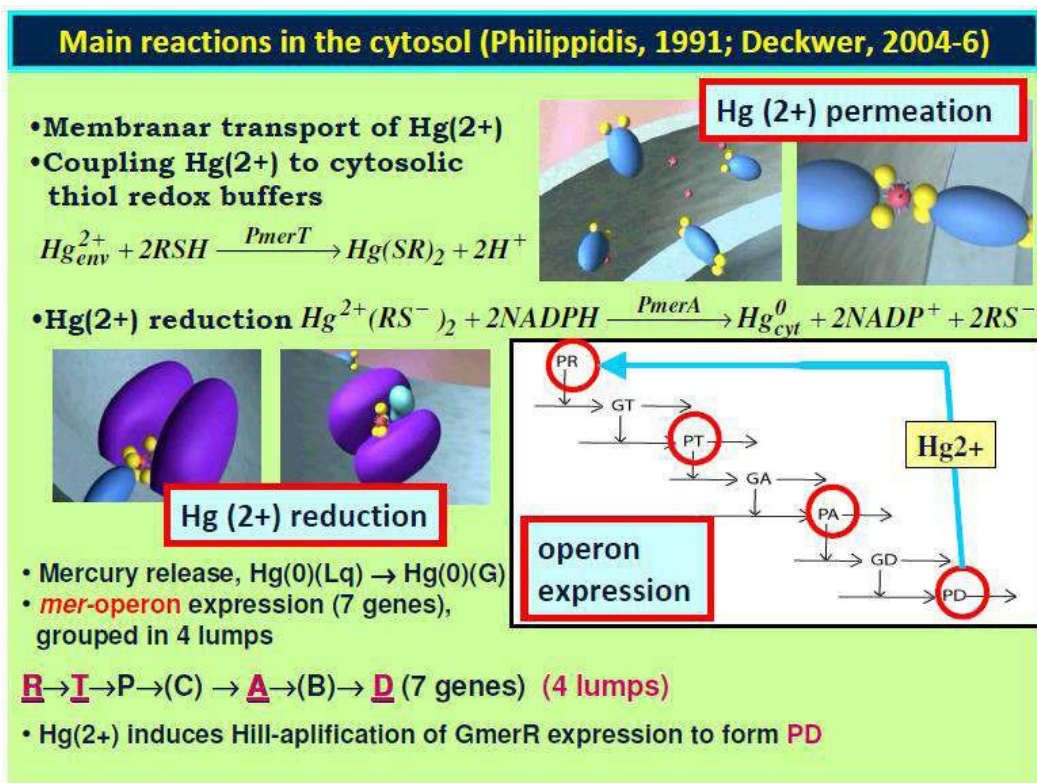


Figure 5-10: The nano-scale cellular enzymatic process in the immobilized *E. coli* bacteria: *mer*-operon (4 gene lumps) expression (Figure 5-9), and the main enzymatic reactions (mercury permeation, and its reduction). Adapted from [1,2,23].

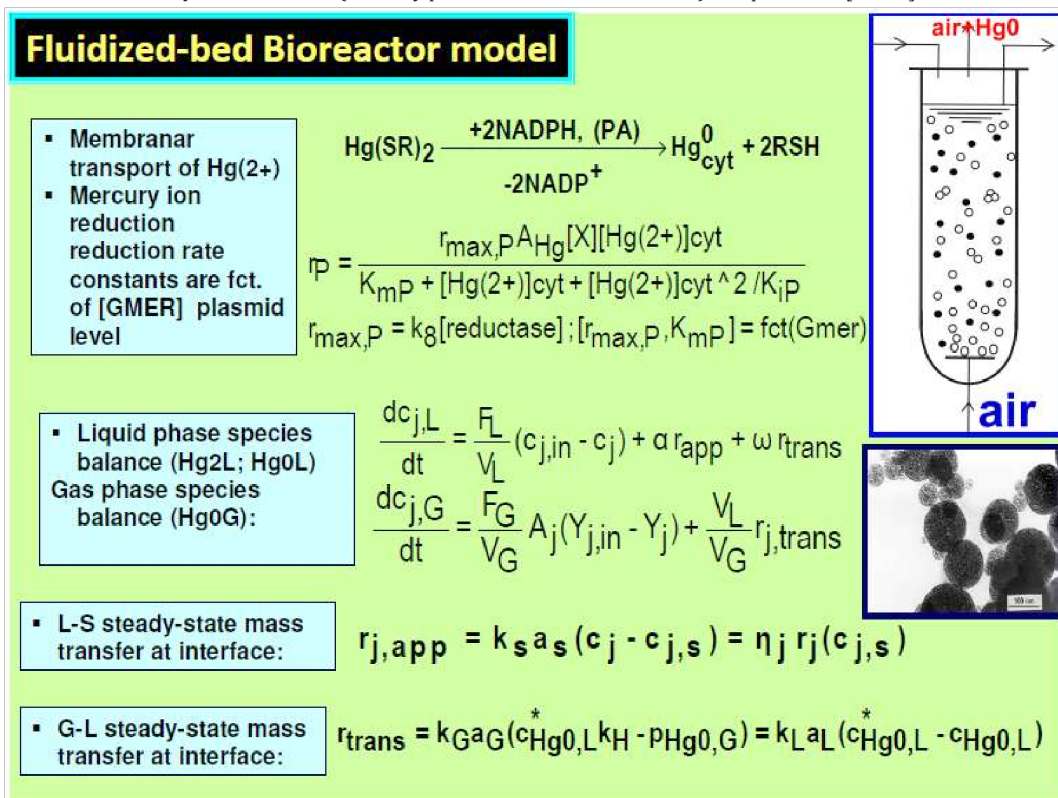


Figure 5-11: The macroscopic model of the three-phase fluidized-bed bioreactor (TPFB) with suspended immobilized *E. coli* on pumice beads. The Michaelis-Menten rate constants depend on the *mer*-plasmid level [1,2,40].

## E. Coli cell characteristics (K-12 strain, EcoCyc, 2005)

Basic idea in design =

- 1) *in-silico* modulate the GS characteristics with a suitable model;
- 2) choose or clone *E. coli* with 2 suitable **GERM** of desired characteristics

- ca. 1'000 ribosomal proteins of 1'000-10'000 copies
  - ca. 3'500 non-ribosomal proteins of avg. 100 copies
  - ca. 4'500 polypeptides of avg. 100 copies.
- In total, the proteome concentration is [P1]= 1e+7 nM
- born cell volume = 1.66e-15 L

● ca. 4500 genes (of one copy) ; overall [G1](total) = 4500 nM

● cell cycle = 100 min; dilution rate  $D = \ln(2) / 100, 1/\text{min}$

● equilibrated growth conditions (isotonic constraints):

$$[\text{NutP}] = 3e+8 \text{ nM} ; [\text{NutG}] = 3e+7 \text{ nM} ; \Sigma[\text{MetP}] = 3e+8 \text{ nM}$$

$$\Sigma[\text{C}]_{\text{env}} = \Sigma[\text{C}]_{\text{cyt}} = \Sigma[\text{MetG}] + \Sigma[\text{MetP}]_{\text{cyt}} - \Sigma[\text{P}]_{\text{cyt}} - \Sigma[\text{G}]_{\text{cyt}}$$

$$\pi[\text{cyt}] = \pi[\text{env}] \text{ (osmotic pressure; isotonic system)}$$

Maria, G. *Chem. Biochem. Eng. Q.* 28, 35-51 (2014).

**E. Coli case study data**

Figure 5-12: The main characteristics of the *E. coli* (K-12 strain) used in the case study no. 1. Data from [95].

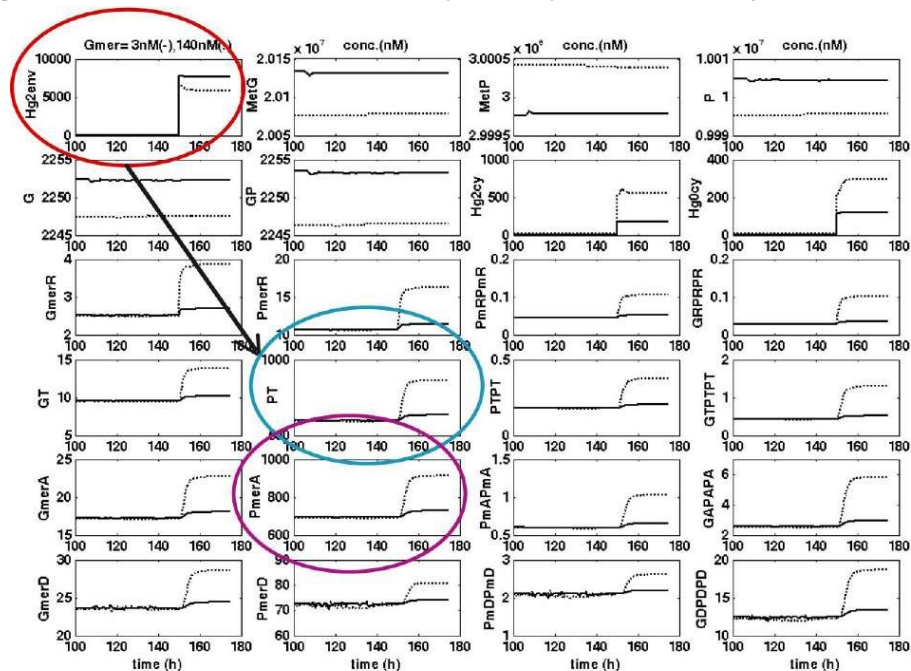
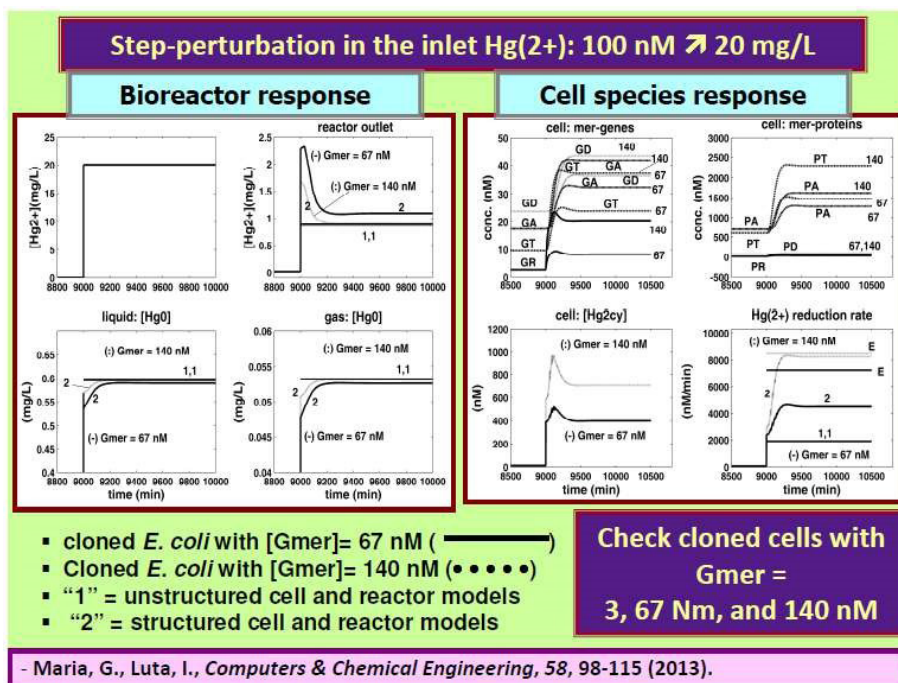
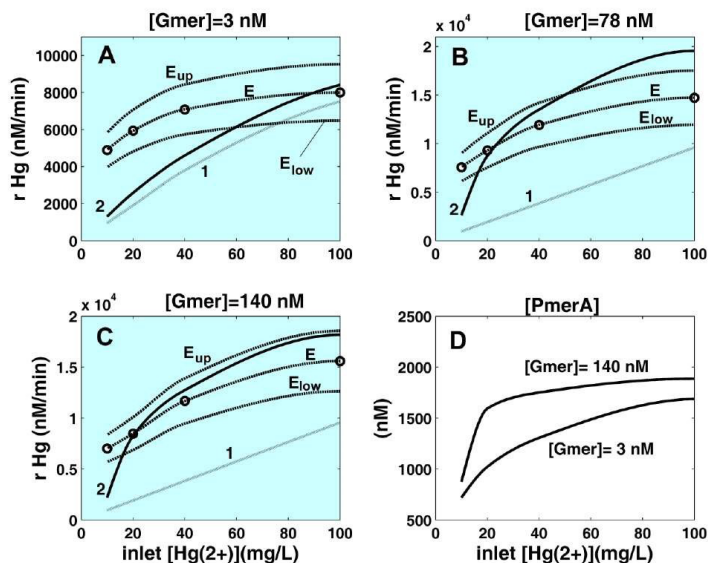


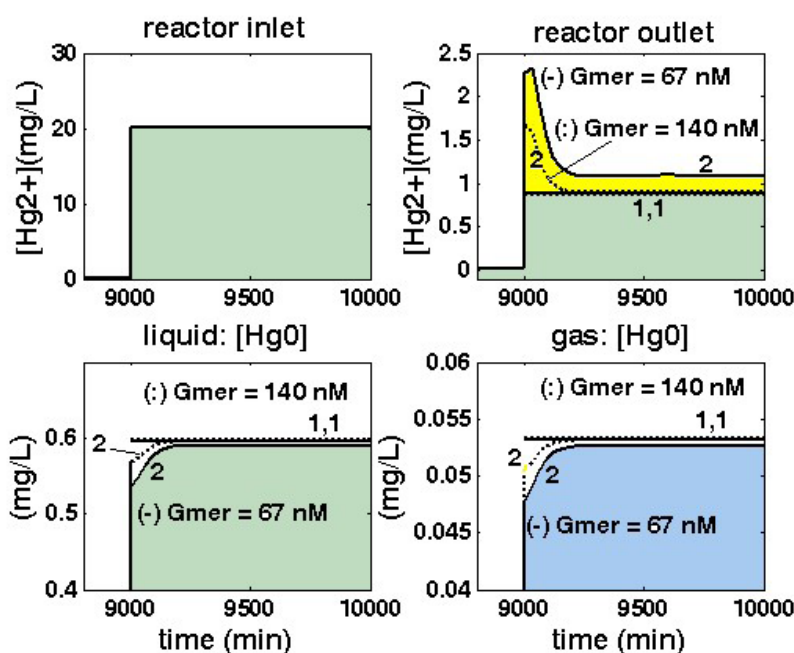
Figure 5-13: Typical evolution of relevant species concentrations predicted by the *E. coli* cell model, after a “step” perturbation in the bioreactor inlet from [ + 2 env Hg ]s = 0.1 to 10 mM (ca. 2 mg/L), for the case of cell cloned with [Gmer]= 3 nM (—), or with [Gmer]= 140 nM (•••••). The arrow indicates the quick and vigorous response of the two key *mer*-enzymes: PT = the mercury ions permease, and PA = the mercury ions reductase.



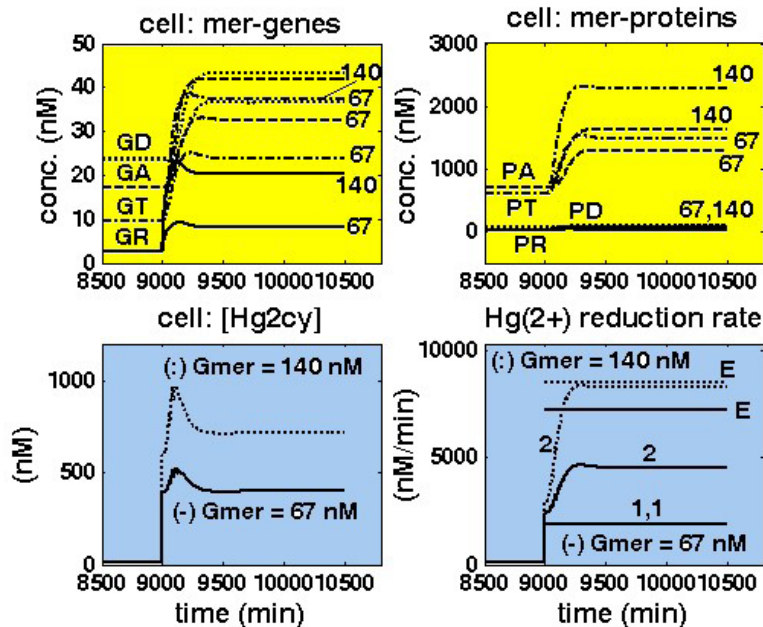
**Figure. 5-14:** [Left]. Dynamics of metallic mercury concentrations in the liquid and gas phases, and in the bioreactor outlet after a “step” perturbation in the reactor inlet from  $[+2L Hg] = 0.1 \text{ mM}$  to  $100 \text{ mM}$  (ca.  $20 \text{ mg L}^{-1}$ ), for immobilized cells on pumice granules under nominal conditions of **Table 5-2**. Comparison is made for the cases of cells cloned with  $[Gmer] = 67 \text{ nM}$  (—), or with  $[Gmer] = 140 \text{ nM}$  (•••••). The curves indexed by “1” denote predictions of the unstructured (apparent) reactor model (**Table 5-1**, **Table 5-3**), while curves indexed by “2” denote the predictions of the structured reactor model (including the cell model of **Table 5-4**). [Right]. Concentration dynamics of relevant *mer*-species inside the cell, and the mercury reduction rate following the same “step” perturbation in the reactor inlet from  $[+2L Hg] = 0.1 \text{ mM}$  to  $100 \text{ mM}$  (ca.  $20 \text{ mg L}^{-1}$ ), for immobilized *E. coli* cells on pumice granules under nominal conditions of **Table 5-2**. Notation “E” denotes the experimental curves of Philippidis et al.[244]. Adapted from [40].



**Figure. 5-15:** The stationary reduction rate ( $r_{Hg}$ ) of mercuric ions by *E. coli* cells for various inlet  $[Hg_{2L}]$  + ]. Comparison includes predictions of the structured extended **HSMDM** model (“2 —”) vs. the Philippidis et al.[244-246] experimental curves (“E -----”) (diffusion free). The upper / lower (“Eup”, “Elow”) denote the confidence bounds of the experimental curves. Curves denoted by (“1 •••••”) are the predictions of the unstructured bioreactor model (cap.5.5.5, **Table 5-1** plus **Table 5-3**), with including the mass transport terms. Comparison is made for the *E. coli* cells cloned with *mer*-plasmids in the amount of  $[Gmer] = 3 \text{ nM}$  (**A**),  $[Gmer] = 78 \text{ nM}$  (**B**), and  $[Gmer] = 140 \text{ nM}$  (**C**). Plots (**D**) display the predicted cytosolic concentration of *mer*-reductase  $[PmerA] = [PA]$  for  $[Gmer] = 3 \text{ nM}$ , and  $[Gmer] = 140 \text{ nM}$ .



**Figure 5-16.** Dynamics of mercury concentrations in the liquid and gas phases after a “step” perturbation in the reactor inlet from  $[+2 \text{ L Hg}] = 0.1 \text{ mM}$  to  $100 \text{ mM}$  (ca.  $20 \text{ mg/L}$ ), for immobilized cells on pumice granules under nominal conditions of (Table 5-2). Comparison is made for the cases of cells cloned with  $[\text{Gmer}] = 67 \text{ nM}$  (—), or with  $[\text{Gmer}] = 140 \text{ nM}$  (•••••). The curves indexed by “1” denote predictions of the unstructured (apparent) reactor model (Table 5-1, together with Table 5-3), while curves indexed by “2” denote the predictions of the structured HSMMDM reactor model (including the cell model of Table 5-4).



**Figure 5-17:** Concentration dynamics of relevant species, and mercury reduction rate following a “step” perturbation in the reactor inlet from  $[+2 \text{ L Hg}] = 0.1 \text{ mM}$  to  $100 \text{ mM}$  (ca.  $20 \text{ mg L}^{-1}$ ), for immobilized *E. coli* cells on pumice granules under nominal conditions of (Table 5-2). Comparison is made for cloned cells with  $[\text{Gmer}] = 67 \text{ nM}$  (67, —), or with  $[\text{Gmer}] = 140 \text{ nM}$  (140, •••••) mer-plasmids. The curves indexed by “1” denote predictions of the unstructured (apparent) reactor model (Table 5-1, together with Table 5-3), while the curves indexed by “2” denote the predictions of the structured HSMMDM reactor model (by also including the cell model of Table 5-4). Notation “E” denotes the experimental curves of Philippidis et al. [244-246].



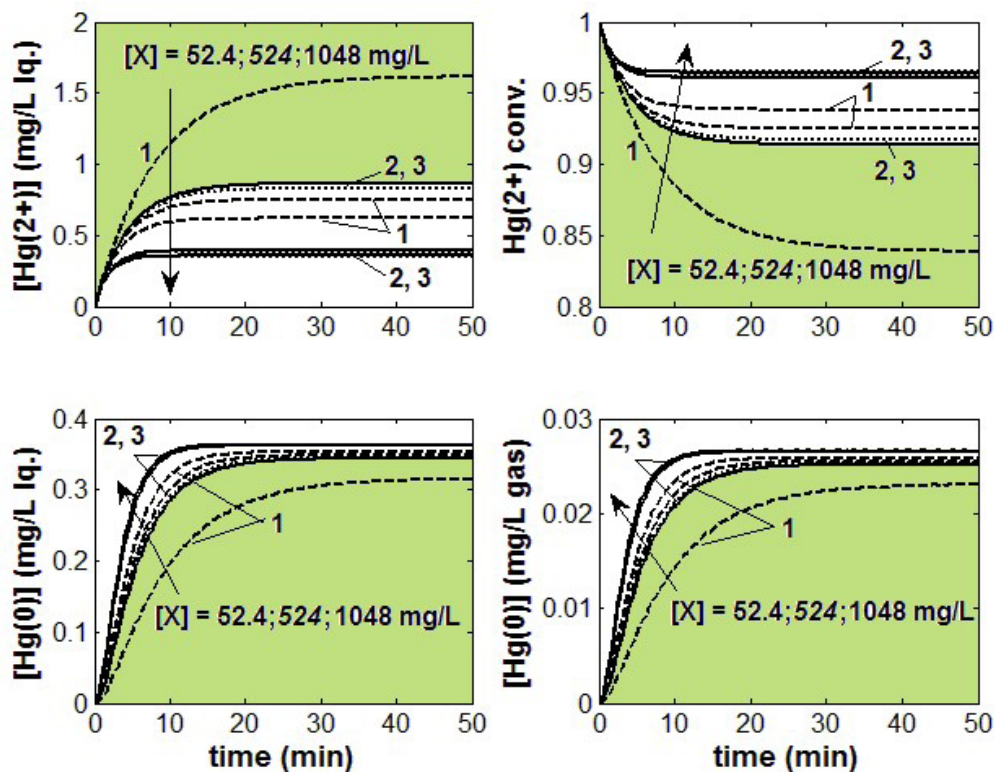


Figure 5-18: TPFB-2 bioreactor sensitivity to biomass  $[X]$  variations (under nominal conditions of Table 5-6), for the outlet  $[+2L Hg]$  (up-left), outlet  $[0L Hg]$  (down-left),  $+2L Hg$  conversion (up-right), and outlet  $[0G Hg]$  (down-right). Predictions of 1(---) PFOM, 2(·) MM1, and 3(—) PHM models.

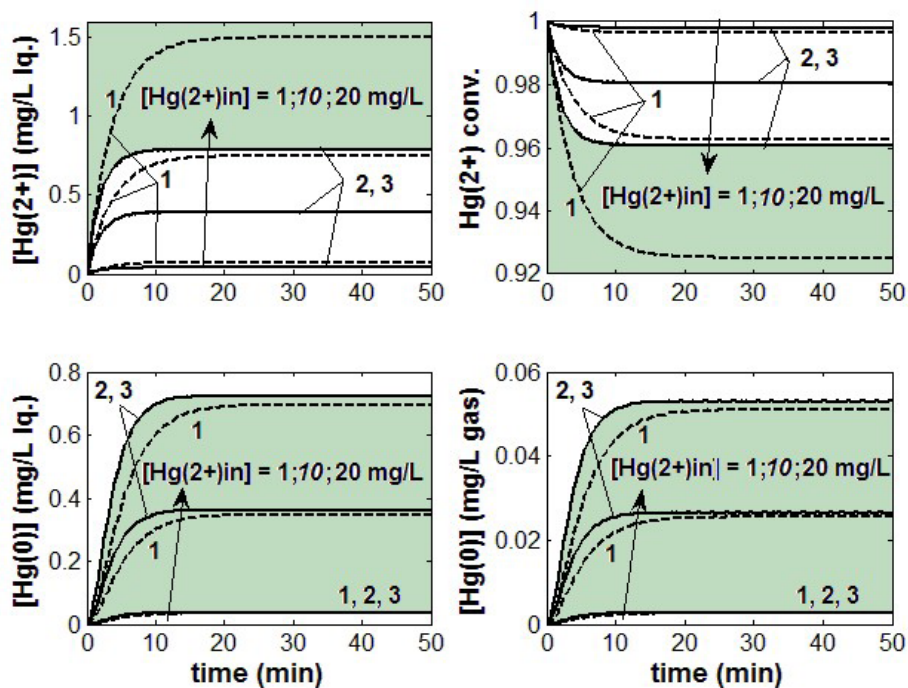


Figure 5-19: TPFB-2 bioreactor sensitivity to  $[+2in Hg]$  variations (under nominal conditions of Table 5-6), for the outlet  $[+2L Hg]$  (up-left), outlet  $[0L Hg]$  (down-left),  $+2L Hg$  conversion (up-right), and outlet  $[0G Hg]$  (down-right). Predictions of 1(---) PFOM, 2(·) MM1, and 3(—) PHM models.

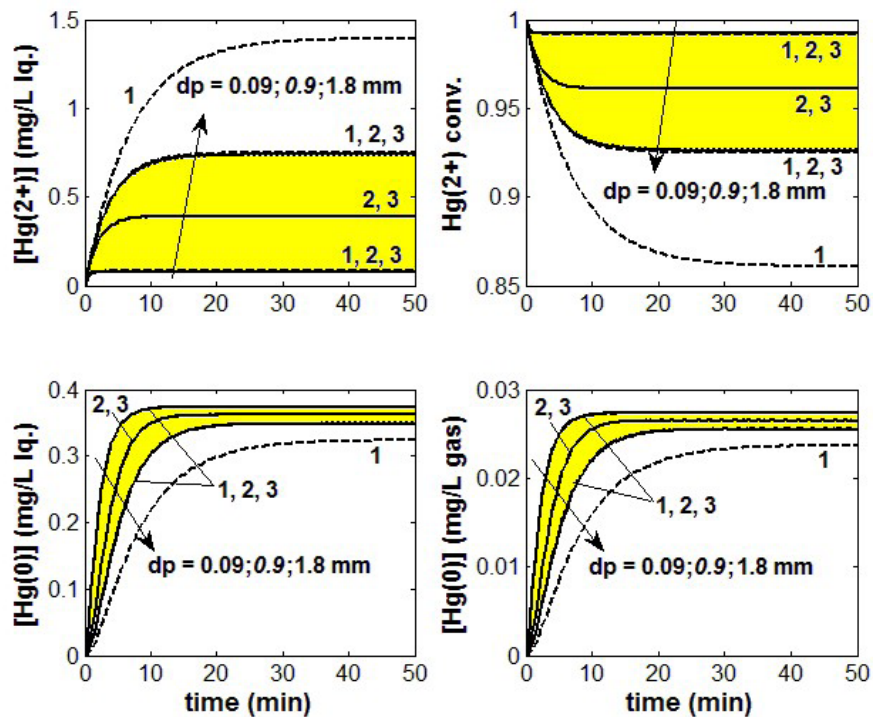


Figure 5-20: TPFB-2 bioreactor sensitivity to particle size  $dp$  variations (under nominal conditions of Table 5-6), for the outlet [ +2L Hg ](up-left), outlet [ 0L Hg ](down-left), +2L Hg conversion (up-right), and outlet [ 0G Hg ](down-right). Predictions of 1(---) PFOM, 2(:) MM1, and 3(—) PHM models.

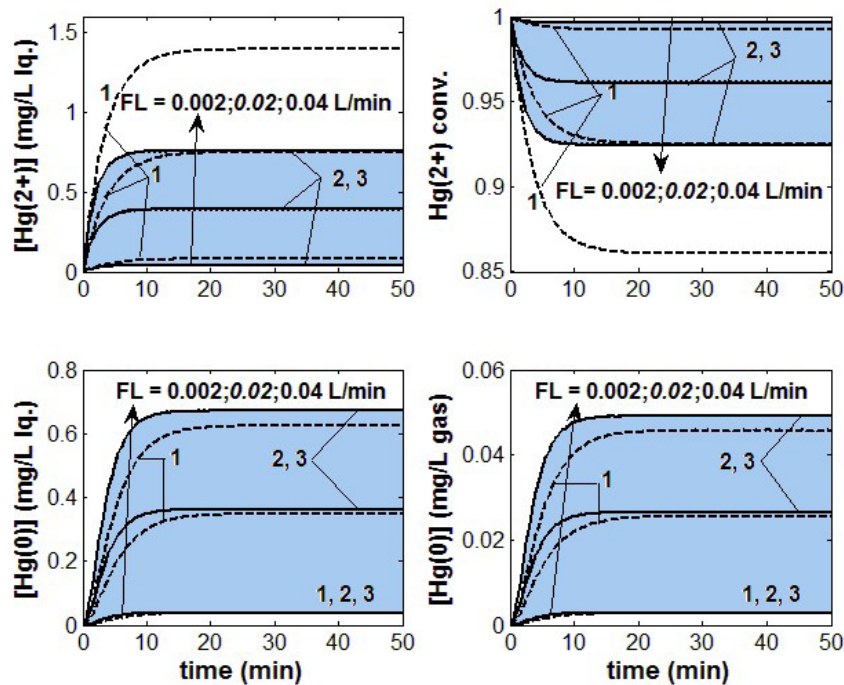


Figure 5-21:TPFB-2 reactor sensitivity to inlet flow rate  $LF$  variations (under nominal conditions of Table 5-6), for the outlet [ +2L Hg ](up-left), outlet[ 0L Hg ](down-left), +2L Hg conversion (up-right), and outlet [ 0G Hg ](down-right). Predictions of 1(---) PFOM, 2(:) MM1, and 3(—) PHM models.

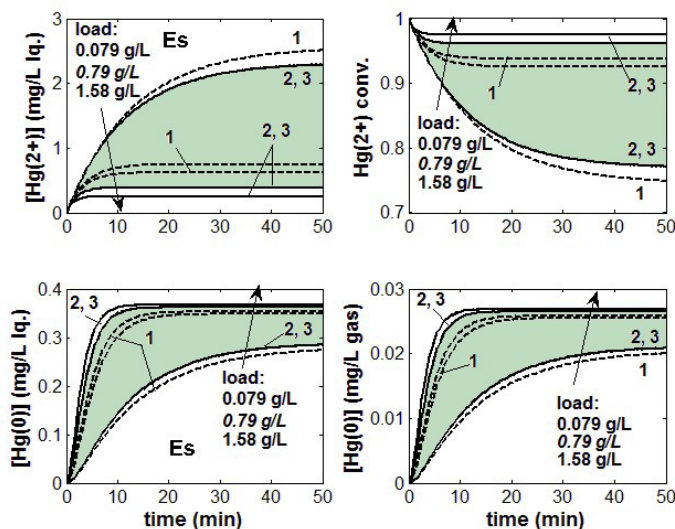


Figure 5-22: TPFB-2 reactor sensitivity to solid fraction  $\epsilon_s$  variations (under nominal conditions of Table 5-6), for the outlet [ +2L Hg ](up-left), outlet[ 0L Hg ](down-left), + 2L Hg conversion (up-right), and outlet [ 0G Hg ](down-right). Predictions of 1(---) PFOM, 2(·) MM1, and 3(—) PHM models

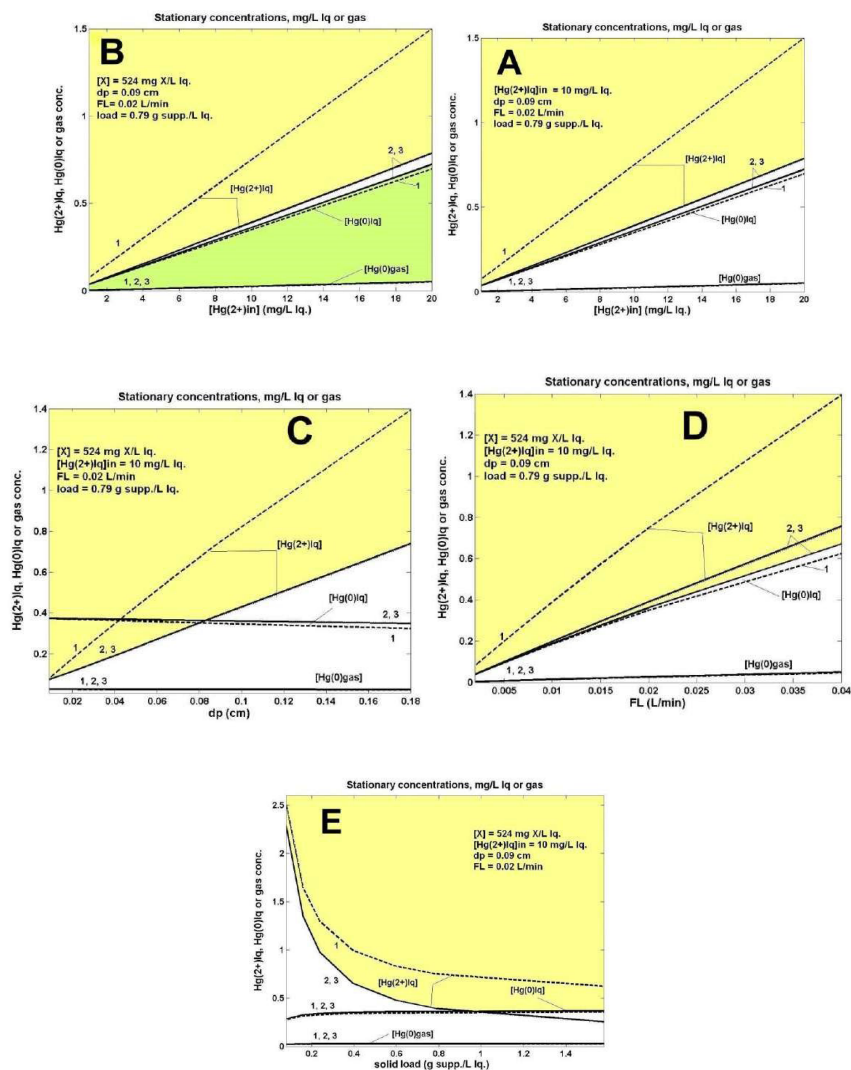
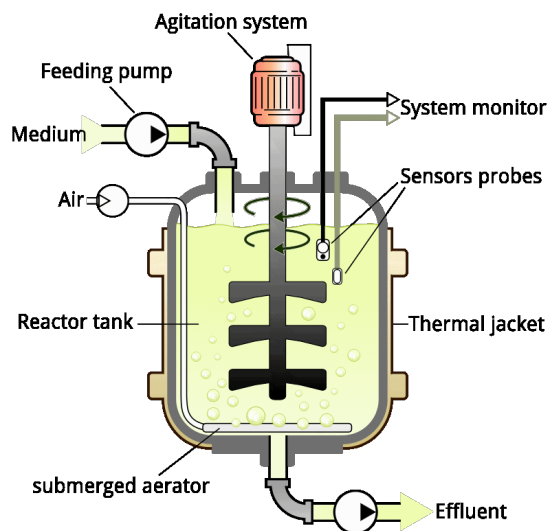
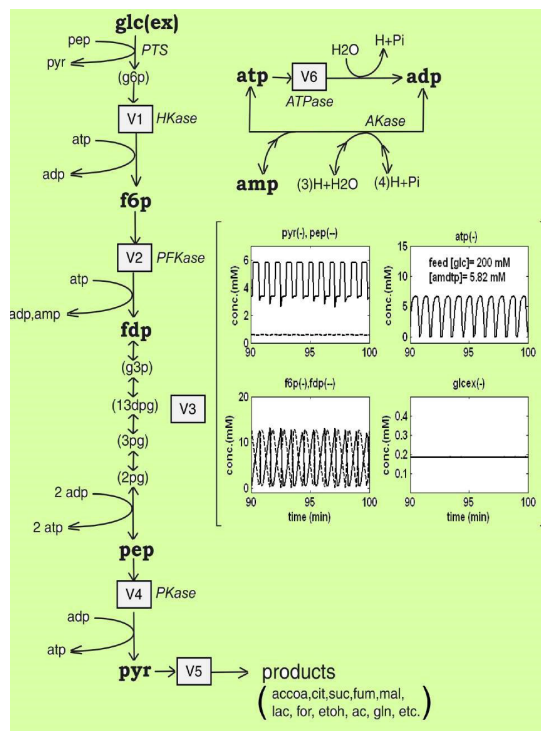


Figure 5-23: Predicted stationary mercury concentrations (at the QSS of the TPFB-2), as ions in lq., metal in liquid, and as metal vapours in gas, that is: [Hg2L+ ], [ Hg0L],and [ Hg0G], respectively. Predictions are made for the control parameter values of (Outline 5-1), that is for: various

biomass loads of biomass [X] (A); various inlet[ 2 HgL+ ]in (B); various particle sizes d p (C); various inlet liquid flow rate FL (D); or initial load of solid εs (E). Predictions are made by using the following unstructured models of the bioprocess kinetics: 1(---) PFOM; 2(:) MM1, and 3( ) PHM.

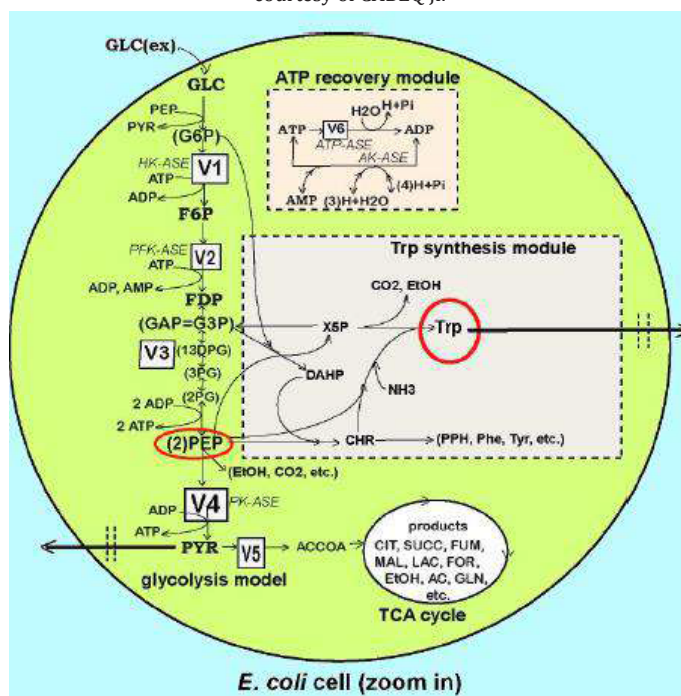


**Figure 6-1:** Simplified scheme of a BR or a FBR used to conduct enzymatic or biological processes. In the BR operating mode, substrate(s), biocatalyst, and additives are initially loaded in the recommended amounts (concentrations). In the FBR operating mode, the substrate(s)/ biomass (immobilized or not), and additives (nutrients, pH-control substances) are continuously fed, following a certain (optimal) policy. The vigorous medium oxygenation with sparged air or pure oxygen ensures the optimal growing conditions for the biomass (see details of Chen [268]). Source = [https://en.wikipedia.org/wiki/Bioreactor#/media/File:Bioreactor\\_principle.svg](https://en.wikipedia.org/wiki/Bioreactor#/media/File:Bioreactor_principle.svg)

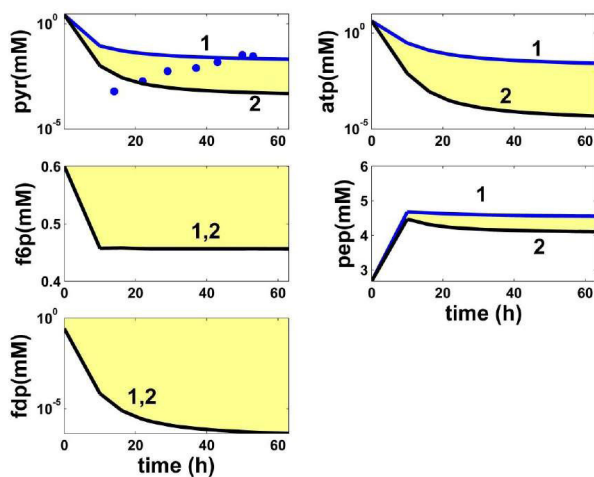


**Figure 6-2:** The simplified reaction schemes of glycolysis in *E. coli* used by Maria [25] to develop a kinetic model with including 9 (individual, or lumped species), participating to only 7 lumped reactions. The kinetic model also include the adenosin co-metabolites ATP, ADP, AMP reactions involved in the ATP recovering system (the top rectangle in the figure). Squares include notations of

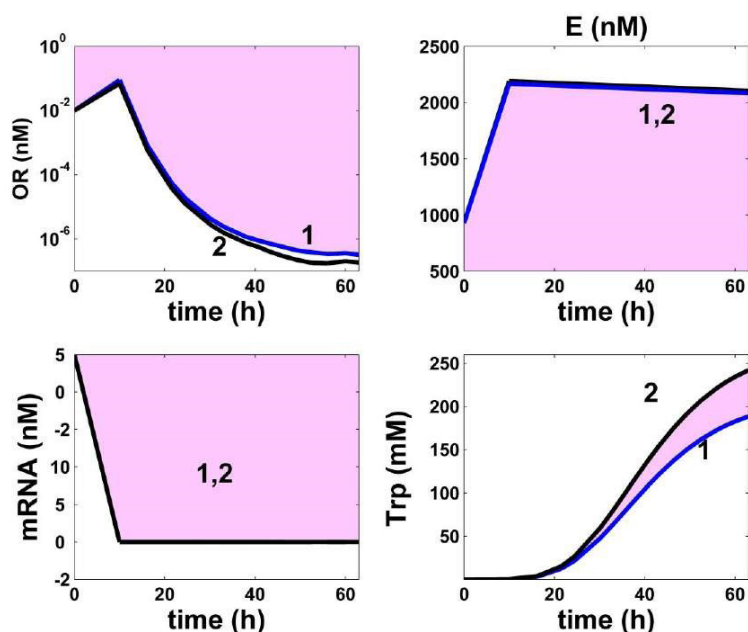
enzymatic reactions of this model. Species in parenthesis are not explicitly included in the kinetic model. Italic letters denote the enzymes. Species oscillating concentrations in the white figures, correspond to the oscillating glycolysis conditions discussed by Maria [25,39,59,61,306]. Adapted from [25] with the courtesy of CABEQ JI.



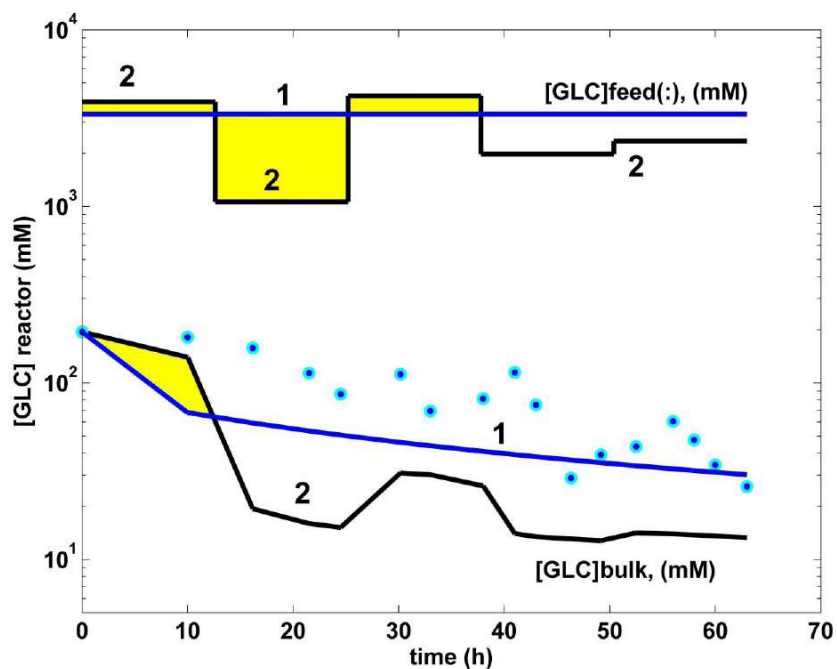
**Figure 6-3:** Simplified structured reaction pathway in *E. coli* for the glycolysis (after Maria [25]), and for the **TRP** synthesis (the gray area) (after Maria et al. [62]). This reaction pathway has been used by Maria et al. [58,62] to derive a **TRP** synthesis kinetic model. Connection of the **TRP** synthesis to glycolysis is realized through the **PEP** node [39,62]. The modular model structure also includes the synthesis of adenosin cometabolites **ATP**, **ADP**, **AMP**, as part of the **ATP** recovery system (the pink rectangle in the figure). **Notations:** **GLC(ex)**= glucose in the cell environment. Species abbreviations are given in the abbreviations list. Species in parenthesis are not explicitly included in the glycolysis model. Italic letters denote the enzymes. Squares include notations of enzymatic reactions **V1-V6** included in the glycolysis model presented by Maria [3,12]. Adapted from Maria [25] with the courtesy of CABEQ JI, and completed according to the Maria [3] kinetic model."



**Figure 6-4:** "Model-based simulated trajectories (\_\_\_) for the glycolytic key species (**PYR**, **F6P**, **FDP**, **ATP**, **PEP**) in the modified *E. coli* T5 strain for the **FBR** operated in two alternatives: (i) (2, black) optimal operation derived in this paper (variable fed [GLC], and feed flow-rate), and (ii). (1, blue), and the experimental data (•, blue) of Chen [268] recorded under nominal, not-optimal operating conditions of (Table 6-1), that is a constant fed [GLC], and feed flowrate. Species abbreviations are given in the abbreviations list." Adapted from Maria and Renea [12].



**Figure 6-5:** Model-based simulated trajectories (—) for the key-species involved in the TRP-operon expression module (TRP, OR, MRNA, E) in the modified *E. coli* T5 strain for the FBR operated in two alternatives: (i) (2, black) optimal operation derived in this paper (variable fed [GLC], and feed flow-rate), and (ii) (1, blue), under nominal, not-optimal operating conditions of (Table 6-1), that is a constant fed [GLC], and feed flow-rate. Species abbreviations are given in the abbreviations list.” Adapted from Maria and Renea [12].



**Figure 6-6: Top-curves.** “The time step-wise optimal feeding policy (2, black) of the GLC concentration in the bioreactor,  $feed_{cglc}(j) (j=1, \dots, 5 \text{ time arcs})$ , derived by Maria and Renea [12] (variable fed [GLC], and feed flow-rate). Comparison is made with the experimental FBR (1, blue) operated under the nominal (not-optimal) operating conditions of (Table 6-1), that is with a constant feed flow-rate, and with a constant GLC concentration in the feed. Both cases are using the same modified *E. coli* T5 strain. **Down-curves.** Model-based simulated trajectories (—) of glucose (GLC) in the bioreactor bulk, for the FBR operated in two alternatives: (i) (2, black) optimal operation derived in this paper (variable fed [GLC], and variable feed flow-rate); (ii) (1, blue) trajectories, and the experimental data (•, blue) of Chen [268] derived under nominal, not-optimal operating conditions of (Table 6-1), that is a constant fed [GLC], and a constant feed flow-rate [268].” Adapted from Maria and Renea [12].

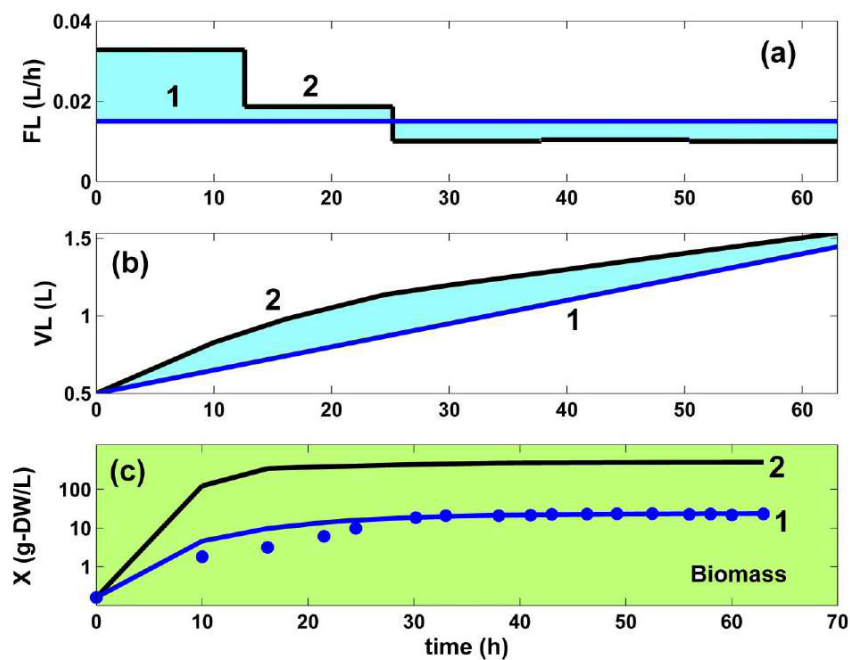


Figure 6-7:

(a) "The time step-wise optimal policy of the feed flow-rate (FL),  $FL_j$  ( $j=1, \dots, 5$  time-arcs) in the bioreactor (\_\_\_) for the FBR operated in two alternatives: (i) (2, black) optimal operation derived by Maria and Renea [12] (variable fed [GLC], and feed flow-rate); (ii) (1, blue) trajectories under nominal, not-optimal operation of (Table 6-1), that is a constant fed [GLC], and feed flowrate [268]. Both cases are using the same modified *E. coli* T5 strain of Chen [268].

(b) The liquid volume (VL) dynamics in two alternatives: (i) of using the optimal policy of the feed flow-rate (FL) in the bioreactor (2, black) derived by Maria and Renea [12], or (ii) of using (1, blue) the non-optimally operated FBR under the nominal conditions of (Table 6-1), that is with a constant fed [GLC] and feed flow-rate [268].

(c) The model-based predictions of the biomass (X) concentration in the same FBR with using the modified *E. coli* T5 strain of Chen [268], but operated in two alternatives: (i) (2, black) optimal operation derived by Maria and Renea [12] (i.e. variable fed [GLC], and feed flow-rate), or (ii) (1, blue) simulations, and the experimental data ( $\bullet$ , blue) of Chen [268] under nominal, not-optimal operating conditions of (Table 6-1), that is a constant fed [GLC], and feed flow-rate [268]." Adapted from Maria and Renea [12].

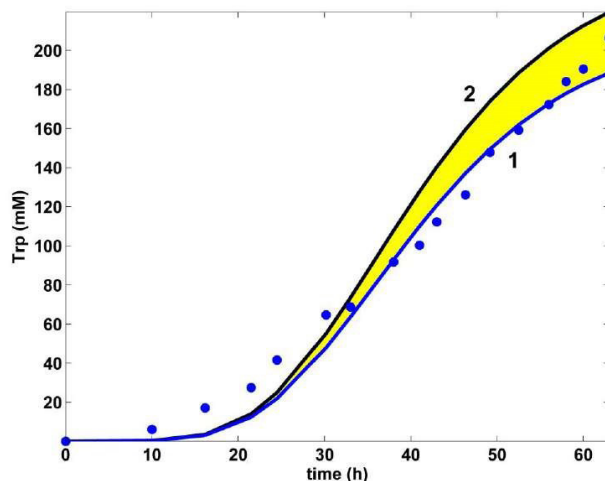
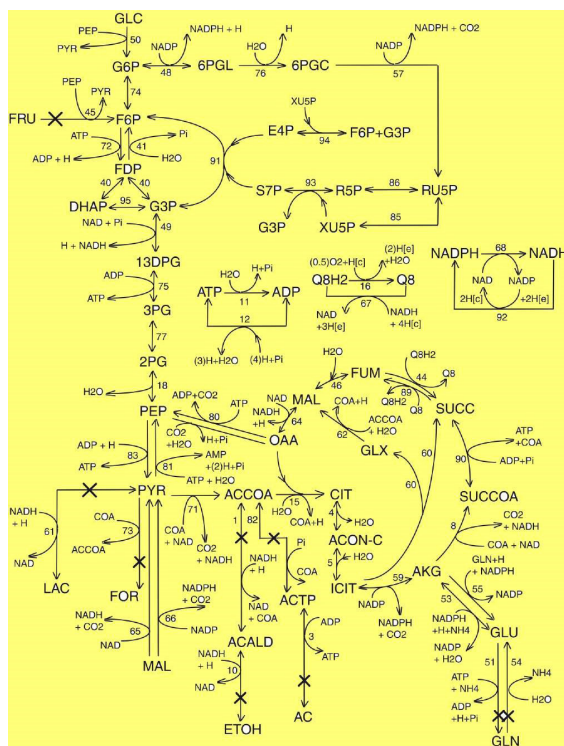
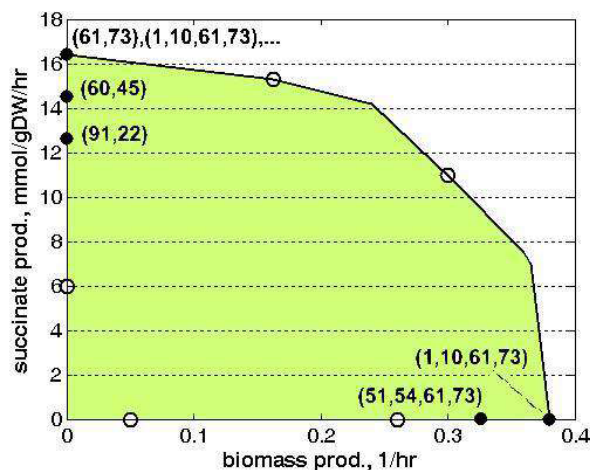


Figure 6-8: Model-based predictions of the tryptophan (Trp) concentration dynamics in the same FBR of Chen [268] with using the modified *E. coli* T5 strain, but operated in two alternatives: (i) (2, black) optimal operation derived by Maria and Renea [12] (i.e. variable fed [GLC], and feed flow-rate), or (ii) (1, blue) simulations [3], and the experimental data ( $\bullet$ , blue) of Chen [268] for the nominal, not-optimal operating conditions of (Table 6-1), that is a constant fed [GLC], and feed flow-rate." Adapted from Maria and Renea [12].



**Figure 7-1:** The CCM reduced pathway of *Escherichia coli* after the model of Edwards and Palsson [4]. This 'wild' strain includes the PTS-system for GLC import from the environment. "Fluxes characterizing the membranar transport ( ) and the exchange with environment ( ) have been omitted from the plot ([e]= environment; [c]= cytosol). This is the case of fluxes number: 2(ACALD), 6(AC,H), 9(AKG,H), 14(CO<sub>2</sub>), 17(LAC,H), 19(ETOH,H), 20(AC), 21(ACALD), 22(AKG), 23(CO<sub>2</sub>), 24(ETOH), 25(FOR), 26(FRU), 27(FUM), 28(GLC), 29(GLN), 30(GLU), 31(H), 32(H<sub>2</sub>O), 33(LAC), 34(MAL), 35(NH<sub>4</sub>), 36(O<sub>2</sub>), 37(Pi), 38(PYR), 39(SUCC), 42(FOR,H), 43(FOR), 47(FUM,H), 52(GLN,ATP), 56(GLU,H), 58(H<sub>2</sub>O), 63(MAL,H), 69(NH<sub>4</sub>), 70(O<sub>2</sub>), 78(Pi,H), 84(PYR,H), 87(SUCC,2H), 88(SUCC,H). Notations: (Met) = diffusional transport of metabolite Met; (Met,H) =transport of metabolite Met via proton symport; (Met,ATP) = transport of metabolite Met via ABC system. Species abbreviations are explained in the **Table 7-1**. Adapted from [42] with the courtesy of CABEQ JI. The considered 72 metabolites, the stoichiometry of the 95 numbered reactions, and the net fluxes for specified conditions are given by Maria et al. [42]."

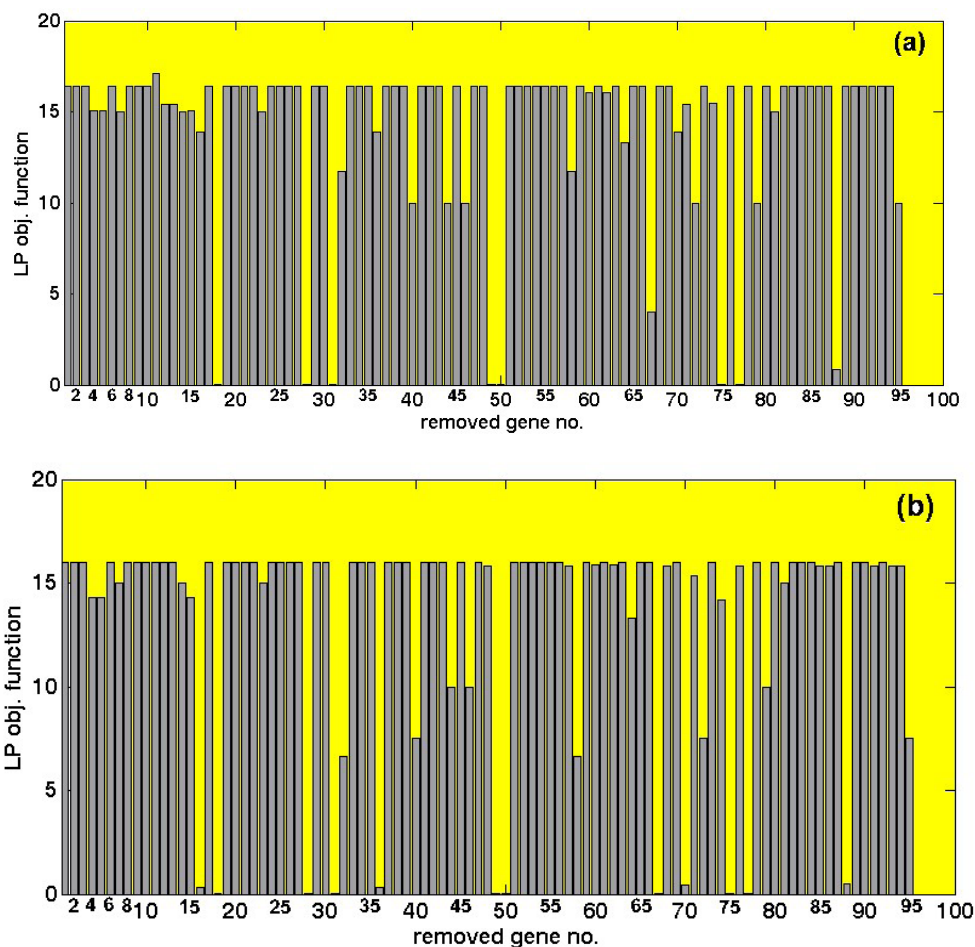


**Figure 7-2:** The Pareto-optimal front of succinate and biomass simultaneous maximum production in the **GMO** *E. coli* cells (□), and location of some restricted gene-knockout mutants predicted by Burgard et al. [310] (○, under anaerobic conditions), and by the present study for (●, under aerobic conditions). Numbers in the parenthesis denote the removed genes from the wild-type CCM *E. coli* cell. Adapted from [42] with the courtesy of CABEQ JI.

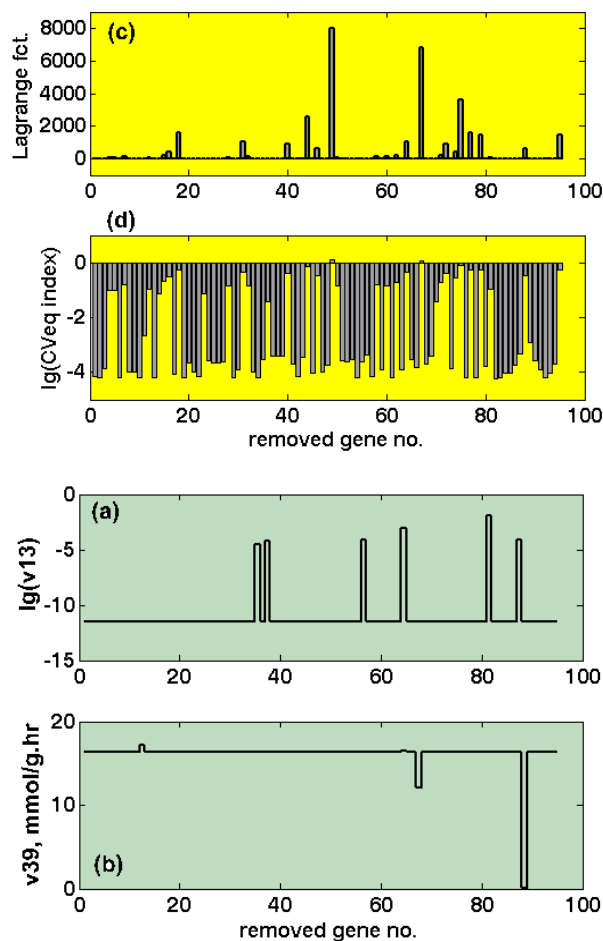




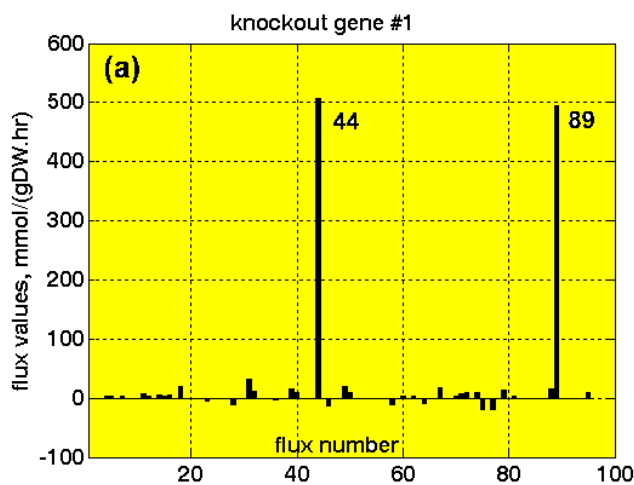
**Figure 7-3:** "Flux distribution (absolute values in mmol/gDW/hr) in the wild-type *E. coli* for succinate and biomass production maximization (with only the basic CONSTR constraints of eq.(5)): (a) LP solution ( $\tau = 2.90 \cdot 10^{-12}$ ,  $\mu = 16.3840$ ,  $L = 1.36 \cdot 10^{-3}$ ); (b) NLP solution ( $\tau = 2.97 \cdot 10^{-12}$ ,  $\mu = 16.3842$ ,  $L = 1.07 \cdot 10^{-5}$ ,  $DF=94$ ). The reversible succinate-to-fumarate transformation (fluxes#44 and #89) is the main reaction responsible for succinate production maximization. Adapted from [42] with the courtesy of CABEQ JI."

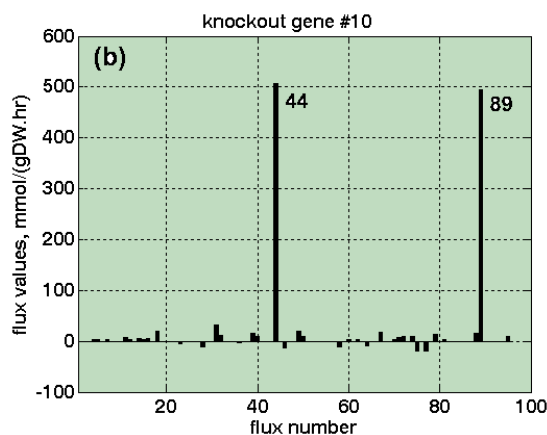


**Figure 7-4:** LP objective function ( $\tau$ ) for succinate and biomass production maximization of various mutants of *E. coli* when successively removing one single gene (from gene #1 to gene #95). (a) Imposed basic constraints CONSTR of eq.(5); (b) Imposed basic constraints CONSTR of eq.(5), and the supplementary constraints  $v_{28} = v_{50}$  and  $v_{51} = v_{52}$ . Adapted from [42] with the courtesy of CABEQ JI."

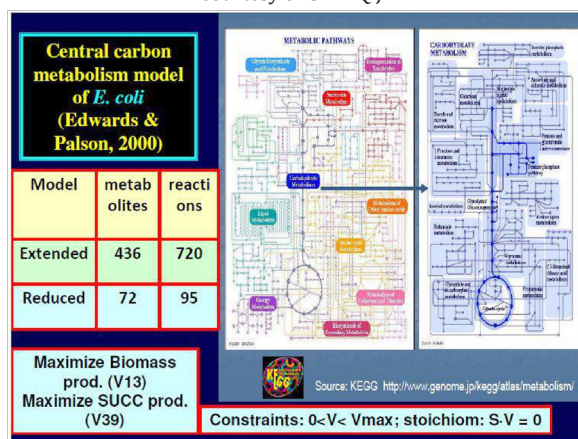


**Figure 7-5:** "Example of optimal solutions with the MINLP for succinate ( )*(b)*, and for biomass ( )*(a)* production maximization of various mutants of *E. coli* when successively removing one single gene (from gene #1 to gene #95;  $DF=95$ , under the basic CONSTR of eq.(5)). The MINLP Lagrange objective function (L)*(c)* is displayed together with the logarithm of base ten of constraint index *(d)*. All inequality constraints are met ( ). The reversible succinate-to-fumarate transformation (fluxes #44 and #89) is the main reaction responsible for succinate production maximization. Adapted from [42] with the courtesy of CABEQJ!"

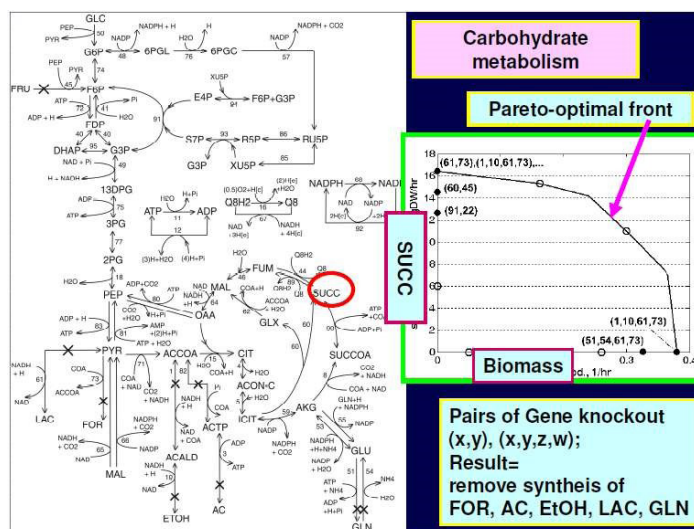




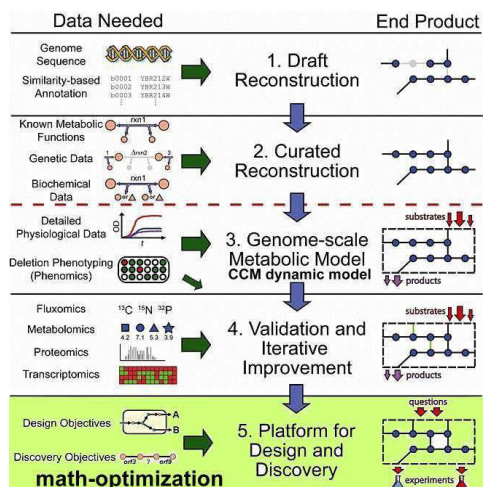
**Figure 7-6:** "Example of local solution - flux distribution (absolute values in mmol/gDW/hr) of two *E. coli* mutants for succinate and biomass production maximization (MINLP solution with the basic CONSTR of eq.(5),  $DF = 95$ ), that is: (a) gene #1 knockout ( $= -2.90 \cdot 10^{-12}$ ,  $= -16.384$ ,  $L = 2.41 \cdot 10^{-5}$ ); (b) gene #10 knockout ( $= -2.90 \cdot 10^{-12}$ ,  $= -16.384$ ,  $L = 2.21 \cdot 10^{-5}$ ). The reversible succinate-to-fumarate transformation (fluxes #44 and #89) is the main reaction responsible for succinate production maximization. Units are in mmol/gDW/hr, and in g-biomass/gDW/hr for biomass formation. Adapted from [42] with the courtesy of CABEQ II."



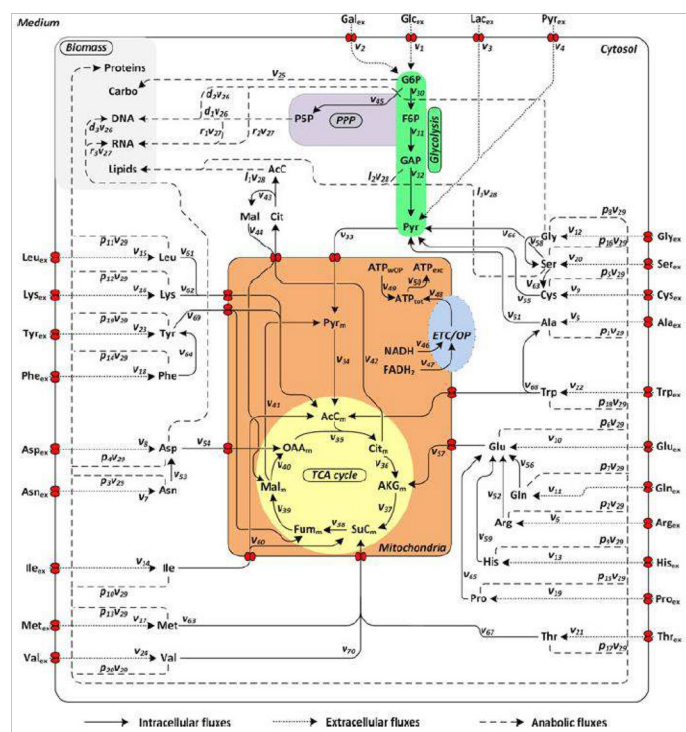
**Figure 7-7:** Carbohydrate metabolism in *E. coli* wild-strain [96]. The reduced model of Edwards and Palson [4] includes 72 metabolites, and 95 reactions. Adapted after [23].



**Figure 7-8:** [left] Carbohydrate metabolism (CCM) reaction pathway in the *E. coli* wild-strain, used by Edwards and Palson [4] to construct their kinetic model. [right] The Pareto-front of simultaneously maximum biomass and SUCC production given by GMO with knockout genes obtained by Maria et al. [42]. The numbers indicate the removed genes from the CCM (Table 7-2). The removed genes correspond to the cut (removed) reactions in the CCM schemes of the left side.



**Figure 7-9: The phases and data utilized in generating a metabolic reconstruction (that is a GMO).** The genome-scale metabolic reconstruction process can be summarized in four major phases, each of the latter phases building off the previous one. The fifth phase is use of the complete reconstruction for practical purposes (see 'Uses of Metabolic Models' of Orth et al. [339]). Characteristic of the reconstruction process is the iterative refinement of reconstruction content that is driven by experimental data and occurs in phases 2to4. For each phase, specific data types are necessary and these range from highthroughput data types, e.g., metabolomics, to detailed studies characterizing individual components, e.g., biochemical data for a particular reaction. For example, the genome annotation can provide a parts list of a cell, while genetic data can provide information about the contribution of each gene product toward a phenotype when removed or mutated. The **GMO** generated from each reconstruction phase can be utilized and applied to examine a growing number of questions with the final product having the broadest applications. Adapted after [339]. Source: <https://journals.asm.org/doi/10.1128/ecosalplus.10.2.1>



**Figure 8-1: The simplified CCM schemes used by Niklas et al. [227] to evaluate the cell main metabolic fluxes of interest, under dynamic conditions.** Notations: PPP = pentose phosphate pathway; TCA = tricarboxylic acid; ETC= electron transport chain, OP =oxidative phosphorylation, Carbo carbohydrates: Glc = glucose, Gal = galactose, Lac =lactate, Pyr= pyruvate, G6P = glucose 6-phosphate, P5P = pentose 5-phosphate, F6P =fructose 6-phosphate, GAP = glyceraldehyde 3-phosphate, AcC = acetyl coenzyme A, Cit = citrate, AKG = alpha-ketoglutarate, SuC = succinyl coenzyme A, Fum = fumarate, Mal =malate, OAA = oxaloacetate, ATP = adenosine triphosphate, ATP(OP) ATP from oxidative phosphorylation, ATP(wOP) = ATP without oxidative phosphorylation, ATPtot = total ATP,ATPexc = ATP in excess, NADH = nicotinamide adenine dinucleotide, FADH2 = flavin adenine dinucleotide. Standard abbreviations are used for the amino acids. Indices: m =mitochondrial, ex = extracellular.5 8.2

## Chap. 5. Case study no. 1: The use of a hybrid WCVV-GRC structured kinetic model to optimize a SCR-TPFB bioreactor used for mercury uptake from wastewaters by immobilized *E. coli* cells cloned with mer-plasmids.

### 5.1. Symbols used in the chap. 4

$a_G$  - G-L specific interfacial area

$a_L$  - L-G specific interfacial area (identical to  $a_G$ )

$a_S$  - L-S specific interfacial area

$A_j$  - atomic (molecular) mass of species  $j$

$a, b$  - rate constants in the Hill-type kinetic expression

$c_j$  - species  $j$  concentration

$D_j$  - diffusivity of species  $j$  in a certain phase

$D$  - cell content dilution rate (i.e. cell-volume

logarithmic growing rate)

$d_b$  - bubble average diameter

$d_p$  - particle diameter

$d_r$  - reactor diameter

$F$  - feed flow rate

FL - Liquid feed flow-rate

$g$  - gravitational acceleration

$[\text{Hg}_L^{2+}]$  - Concentration of the mercury ions in the liquid

(bulk) phase of the bioreactor

$K_m$  - Michaelis-Menten constants

$k_g$  - G-L mass transfer coefficient (on gas side)

$k_h$  - Henry constant

$k_l$  - L-G mass transfer coefficient (on liquid side)

$k_s$  - L-S mass transfer coefficient (on liquid side)

$k$  - rate constants

$n_H$  - Hill-coefficient

$n_{pD}, n_{pR}$  - partial orders of reaction

$n_j$  - number of moles of species  $j$

$n_s$  - number of species in the cell

$N_A$  - Avogadro number

$p$  - overall pressure

$$Re_L = (\Sigma_L d_p^4 \rho_L^3) / \mu_L^3 \text{ - Reynolds number (liquid)}$$

$Rg$  - universal gas constant

$r_j$  - species  $j$  reaction rate

$$Sc_L = \mu_L / (\rho_L D_{S,L}) \text{ - Schmidt number (liquid)}$$

$Sh = (k_s d_p) / D_{S,L}$  - Sherwood number

$T$  - temperature

$t$  - time

$t_c$  - cell-cycle time

$u_g$  - gas superficial velocity

$u_L$  - liquid superficial velocity

$V$  - volume

$v_M$  - maximum reaction rate

$X$  - Biomass in the bioreactor

$Y_j$  - molar ratio of species  $j$  to the rest of species in the mixture

### Greeks

$\alpha, \omega$  - stoichiometric coefficients

$\beta, \psi$  - constants used in evaluation of particle effectiveness in Table 5-8

$\epsilon_g$  - volume fraction of the gas in the bed

$\epsilon_L$  - volume fraction of the liquid in the bed

$\epsilon_p$  - particle porosity

$\epsilon_s$  - volume fraction of particles in the bed

$\Phi$  - optimisation objective function

$\phi$  - Thiele modulus

$\phi_c$  - Carman shape factor (Trambouze et al., 1988)

$\eta_j$  - effectiveness factor of reaction  $j$

$\mu_L$  - dynamic viscosity of the liquid

$\rho$  - density

$\pi$  - osmotic pressure

$\sigma$  - interfacial tension

$\Sigma_L$  - power dissipated per unit mass of liquid

$\tau_p$  - particle tortuosity

### Superscript

\* - saturation

Index

app - apparent

cell - referring to the *E. coli* cell

cyt - cytoplasm

ef - effective

env - environment

G - referring to gas, or at G-L interface

in - inlet

L - referring to liquid, or at the L-G interface

max - maximum

o - initial

p - particle

ref - reference value

s - referring to particle, or at liquid (L) – solid (S)

interface, or referring to the steady-state

trans - referring to the transport

Abbreviations

G-L - gas-liquid

G• - gene

Gmer - *mer* plasmids generating *mer* operons into the cell

GmerX (or GX) - *mer* genes (X = R,T,A,D)

GRC - genetic regulatory circuit

HSMDM - hybrid structured modular dynamic (kinetic)  
models

L-S - liquid-solid

MetG, MetP - metabolites

M-M Michaelis Menten

NADPH - nicotinamide adenine dinucleotide phosphate

NutG, NutP - nutrients

P• - protein

PmerX - *mer* proteins

QSS - quasi-steady-state

RSH - compounds including thiol redox groups

S - substrate

SCR - Semi-continuous reactor

TF - transcription factor

TPFB - three-phase fluidized bioreactor

WCVV - variable volume whole-cell

X-biomass

[.] - Concentration

## The chapter purpose

This chapter exemplifies the use of a complex HSMDM to solve an engineering problem at an industrial pilot scale, that is the use of a complex cell WCVV structured kinetic model of the *mer*-operon GRC expression (Figure 5-2) in a HSMDM model to optimize a pilot-scale SCR-TPFB bioreactor (Figure 5-3)

used for mercury uptake from wastewaters by immobilized *E. coli* cells cloned with *mer*-plasmids. The developed HSMDM dynamic model is linking the cell-scale model part (including the dynamics of the nano-scale state variables/species) to the biological reactor macro-scale state variables for improving the both model prediction quality and its validity range. Eventually, the HSMDM model was used to *in-silico* design a GMO (i.e. an *E. coli* cloned with *mer*-plasmids in a degree to be determined) for improving its capacity for mercury uptake from wastewaters (Figure 5-4).

The cell dynamic model of the *E. coli* cloned bacterium is able to simulate the self-control of the **GRC** responsible for the *mer*-operon expression, and to predict:

- a) the influence of the TPFB bioreactor control variables [such as, the feed flow-rate (FL), the mercury ions concentration  $[Hg_L^{2+}]_{in}$  in the feeding liquid], and of the biomass concentration in the bioreactor  $[X]$ ;
- b) the influence of various bioreactor running parameters [such as, the size of the solid porous particles ( $dp$ ) of pumice on which the biomass is immobilized; the concentration  $[Gmer]$  of the *mer*-plasmides used in the cloned *E. coli* cells) on the bioreactor performance in uptake the mercury ions from wastewaters and in eliminating them as mercury vapours by the continuously sparged air into the reactor [1,2,40,63,241-243].

This HSMDM dynamic model is a worthy example of applying WCVV models, and the GERM-s properties (P.I.-s) described in the chap. 4 to adequately represent a complex modular GRC-s for the *mer*-operon expression in gramnegative bacteria (such as *E. coli* cells). The structured GRC-WCVV model was proposed by Maria [1,2] to reproduce 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. [244-246] experimental data, and the Barkay et al. [247] information on the *mer*-operon expression characteristics. Later, this cellular GRC model was included in a HSMDM dynamic model of the SCR-TPFB bioreactor by Maria et al. [40,63] in order to simulate its dynamics over a wide range of operating conditions, that is: FL = [0.01-0.04] L/min.;  $[Hg_L^{(2+)n}]_{in}$  = [10-40] mg/L;  $[X]$  = [250-1000] mg/L;  $dp$  = [1-4] mm;  $[Gmer]$  = [3-140] nM.

As evidenced by this application, the current trend in bioengineering is to use multi-layer (hybrid) models to extend the detailing degree of the developed bioreactor dynamic models, by also including the dynamics of the concerned cell key-species metabolism. Exemplification is made by coupling an unstructured dynamic model of a TPFB, used for mercury uptake from wastewaters by immobilized *E. coli* cells, with a cell simulator of the GRC controlling the mercuric ion reduction in the bacteria cytosol. The obtained results reported a significant improvement in the model prediction quality (ca. 3-12% in state variables, and up to 40% in reduction rate vs. experimental information) and in the detailing degree [i.e. simulation of 26+3 (cell+bulk) vs. only 3 (bulk) variables dynamics]. The major advantages of the hybrid model come from the possibility to predict the bacteria metabolism adaptation to environmental changes over several cell generations, and also the effect of cloning cells with certain plasmids to modify its behaviour under stationary or perturbed conditions.

Basically, this chapter no. 5 exemplifies the possibility to couple an unstructured TPFB dynamic model with macro-scale state variables [248] used for mercury uptake by immobilized *E. coli* cells on pumice millimetric size support, with a structured *E. coli* cell model of Maria [1,2]. The advantage of using such a hybrid (bi-level) modelling approach is related to the improvement of the prediction accuracy of the reactor performance / state variable dynamics, and of the prediction of the bacteria metabolism adaptation to environmental 'step'-like changes in the environmental mercury content  $[Hg_L^{2+}]_{env}$  through the modelled cell GRC related to the *mer*-operon expression, and by mimicking the whole-cell growth under balanced conditions. If promising, such an approach can support the idea of (i) improving the quality of process monitoring (control), and (ii) *in-silico* design cloned *E. coli* with an increased content of *mer*-plasmids. And all these by using complex HSMDM dynamic model of increased predictive power. The investigation is also supported by the tremendous improvement in the computing power over the last decades, and by the continuous expansion of the available information from cellular bio-omics databanks, and by steady efforts necessary to elaborate detailed cellular numerical simulators. Exemplifications of such modular GRC models used for the



*in-silico* design of GMO of industrial use includes several published case studies by Maria [23] (Figure 5-1). Due to the cell metabolism complexity, and existence of both cell-level control parameters, together with the bioreactor macro-level control variables, *in-silico* optimization of an industrial bioprocess by using GMO-s often translated in a multi-objective optimization problem [3,12,40,42,59,107], difficult to be solved by using common numerical algorithms. A couple of case studies exemplify the own positive experience with of using HSMDM including cell structured models of CCM, and of various GRC-s for optimizing industrial bioreactors, or for design of some GMO-s for improving certain bioprocesses of practical interest are presented by Maria [3,12,23,40,42,58,62].

### Mercury ion reduction in bacteria cells – the apparent kinetics

“Bacteria resistance to mercury is one of the most studied metallic-ion uptake and release process (see the review of Barkay et al. [247]) due to its immediate largescale application for mercury removal from industrial wastewaters [248-250]. The bacteria response to the presence of toxic mercuric ions in the environment  $Hg_{env}^{2+}$  is apparently surprising. Instead of building carbon- and energy-intensive disposal ‘devices’ into the cell (like chelate-compounds) to ‘neutralize’ the cytosolic mercury  $Hg_{cyr}^{2+}$  and thus maintaining a tolerable level, a simpler and more efficient defending system is used. The metallic ions are  $Hg_{cyr}^{2+}$  catalytically reduced to the volatile metal,  $Hg_{cyr}^0$ , less toxic and easily removable outside the cell (as  $Hg_L^0$  in the liquid environment of the TPFB) by simple cell membrane diffusion. 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  $Hg_{cyr}^{2+}$  in cytosol as  $Hg(SR)_2$ , and of NAD(P)H reductants able to convert it into neutral metal  $Hg_{cyr}^0$ . (see the overall reactions of (Table 5-1). A genetic regulatory circuit (GRC) responsible for the involved mer operon expression controls the whole process, by including seven genes of individual expression levels of 7 encoded proteins, of which expression is induced and adjusted according to the level of mercury  $Hg(SR)_2$ , and other metabolites into cytosol (Figure 5-2). 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, key-metabolites, and key- proteins), thus compromising the whole cell metabolism.

While the role of each *mer*-gene and *mer*-protein in the mercury ion reduction process is generally known, not all the regulatory loops of the *mer*-operon expression are perfectly understood, and the way by which the cell adapts itself to variations of mercuric ion concentrations  $Hg_{env}^{2+}$  in the environmen. Philippidis et al. [244-246] proposed a reduced apparent (un-structured) kinetic model, of Michaelis-Menten (M-M) type, to quickly simulate the main steps of mercury uptake by *E. coli*, that is: the membranar transport of environmental  $Hg_{env}^{2+}$  into the cell (of *rt* rate) and its reduction (*rP* rate, see Table 5-1). To highlight the slowest process step, separate experiments have been conducted with cultures of intact cells or ‘permeabilized’ cells (with a more permeable cell membrane to metallic ionic species). The results clearly showed that membranar permeation is the rate controlling step, being of one order of magnitude slower than the cytosolic mercuric ion reduction. Identification of rate constants of the two main reactions for cloned *E. coli* cells with an increasingly copynumbers of *mer*-plasmids, in the range of  $[Gmer] = 3-140$  nM, comparatively to  $[Gmer] = 1-2$  nM (for wild-types of *E. coli*) reveals the following aspects (Table 5-1):

- a) The rate constants are strongly dependent on the *mer*-plasmid (genes) level into the cloned cells of *E. coli*, the reaction mechanism being more complex than those suggested by the two apparent reaction rates (*rt andrP*) of Philippidis et al. [244-246]
- b) Such reduced kinetic models can approximately represent the overall mercury uptake in cells and the steady-state process efficiency, being useful for the bioprocess scale-up engineering calculations [248].
- c) The unstructured models can not represent the *mer*-GRC response to various inducers, the cell response to stationary or dynamic perturbations in the mercury level ( $Hg_L^{2+} = Hg_{env}^{2+}$ ) in the bioreactor liquid-phase, and in its constant feeding. Also, the apparent M-M kinetic model cannot explain and simulate/predict the self-regulation of the whole transport-reduction process, and how the *mer*-gene expression is connected to the cell volume growth and the cell content replication;
- d) The cellular uptake process may be improved by increasing the *merpermease* content into the cell, up to a limit of ca.  $[Gmer] = 80$  nM *mer* plasmids [244-246], higher than the permeation rate remains unchanged, thus preventing exhaustion of the cell metabolic resources. From the same reason, the cell regulatory system maintains an upper limit for the membranar transport of environmental  $Hg_{env}^{2+}$  into the cell (*rt*) irrespectively to the  $Hg_{env}^{2+}$  concentration in the environment, the cytosolic  $Hg_{cyr}^{2+}$  concentration never exceeding a 3500-5000 nM level.

To model these unsolved aspects, an extended structured cell model has been developed by Maria [1,2], of a WCVV type (see chap. 4), by including a GRC of 7 GERM-s linked by following the rules described in the chap. 4, and by accounting the few experimental information of Philippidis et al. [244-246], and of Barkay et al. [247]. The derived GRC to simulate the *mer*-operon expression, and the process self-control at a molecular level under isothermal and isotonic conditions.”

### The TPFB bioreactor model and its nominal operating conditions

„To exemplify the construction of the bi-level HSMDM structured model for the approached mercury removal process, the TPFB bioreactor of Deckwer et al. [248] was approached. The main characteristics and the nominal operating conditions of this TPFB bioreactor under a semi-continuous (SCR) operation are

presented in the (Table 5-2). As a first step in the engineering analysis of this TPFB bioreactor, it is important to emphasize, by using the model-based numerical simulations, its mercury removal performances under variate operating conditions. Thus the most influential factors on the mercury uptake efficiency are pointed-out and ranked for further operating decisions. The used lab-scale TPFB bioreactor of (Figure 5-3, and Figure 5-4, upright) includes a resistant *E. coli* cell culture. The bioreactor is completely

automated being able to maintain its control parameters of (Table 5-2) at their optimal set-point, by ensuring a constant pH, temperature, a constant inlet feed flow-rate, and inlet mercury concentration  $Hg_{in}$ , a constant sparkling air inlet feed flow-rate, and a constant concentration of nutrients used as C/N/P source for the biomass optimal growth. Initially, to study this bioprocess, Deckwer et al. [248] used *E. coli* bacteria immobilized on alginate beads, but further tests have been extended by using porous pumice granules of 0.9 mm to 4 mm diameter. The pumice carrier checked in the present paper is particularly attractive, the carrier exhibiting a high BET area and porosity, a large pore size (even higher than  $10 \mu m$ ), thus allowing a good diffusion of the substrate (mercuric ions) to the cells from inside the support. The operating conditions are tightly controlled, that is the liquid flow rate, the aeration rate (pO<sub>2</sub>), pH, and the temperature required by an equilibrated bacteria growth (Table 5-2). The sufficient supplied oxygen guarantees a good cell metabolism, and a high content of cytosolic NADPH necessary for mercury reduction. Beside, the continuously bubbling air plays also the role of volatile metallic mercury carrier, by removing it from the liquid system. Eventually, the mercury vapours from the air leaving the system are condensed

and recovered [248]. A background pollution of ca. 100 nM is considered in the input water (that is ca.  $0.02 \text{ mg L}^{-1}$ , which is smaller than the metabolic regulation threshold of  $0.05 \text{ mg L}^{-1}$ ), thus maintaining active the *mer*-operon into the *E. coli* cell. The biomass content of the support is variable (ca. 0.6-3 gX/L, according to [248,251], but a quasi-constant level of ca. 1 gX/L can be maintained by employing a purge/renewal system for the solid particles. At an industrial-scale, when treating polluted waters, the outlet gas (air) from the bioreactor, containing the volatile metallic mercury, is passed through an adsorption device, or through a desublimator system allowing the recover of metallic mercury [252]. To simulate the (semi-)continuous TPFB bioreactor performance, a dynamic ideal model was considered [253,254], by assuming homogeneous (perfectly mixed) liquid and gas phases, and a uniform distribution of the solid particles (of uniform characteristics) in the G-L fluidized bed. By accounting for only the apparent mercury uptake rate ( $r_{app}$ ) by the immobilized bacteria (of concentration  $c_x$  in Table 5-1) in the spherical solid carrier (of  $\epsilon_s$ ), the mercury mass balance in the liquid and gas phases are presented in the (Table 5-3). The TPFB reactor dynamic model includes terms referring to the mass balance in the bulk (liquid, L) phase, the gas phase (G), the interphase L-G transport ( $r_{trans}$ ) of the volatile mercury, and the bioprocess inside the solid particles ( $r_{app}$ ). The former one, also includes the diffusional resistance of the substrate (mercury ions) / product (dissolved metallic mercury) transport through the pumice support pores. More specifically, the mercury differential mass balance in the TPFB given in (Table 5-3) includes the following terms:

- a) The apparent mercury reduction rate, evaluated at the solid interface;
- b) The substrate (that is  $S = Hg_L^{2+} = Hg_{env}^{2+}$ ) diffusional transport in the particle, expressed by the effectiveness factor ( $\eta$ ) evaluated using the Thiele modulus for a Michaelis-Menten type reaction [33], the effective diffusivity ( $D_{s,ef}$ ) accounting for the molecular diffusion ( $D_{SL}$ ), the particle porosity ( $\epsilon_p$ ) and the tortuosity ( $\tau$ ) (other resistances being neglected [255];
- c) The apparent rate  $r_{j,app}$  was evaluated from solving (on every time-increment) the quasi-steady-state equality of mass fluxes at the solid-liquid (S-L) external interface (also including the external diffusion coefficient  $k_s a_s$  [256]

d) The liquid-to-gas transport rate  $r_{trans}$  of the metallic mercury ( $Hg_L^0$  to  $Hg_G^0$ ), evaluated from the quasi-steady-state equality of mass fluxes at bubbles interface, by accounting for the mass transfer coefficients  $k_L a_L$  (on the liquid film side; experimental value adopted), and  $k_G A_G$  (on the gas film side; value from application of the Sharma's relationship given by Trambouze et al. [254]. The resulted hybrid dynamic model of (Table 5-3), with the apparent M-M kinetics of (Table 5-1) allows simulating the transient operating conditions of the TPFB bioreactor by using an apparent Michaelis-Menten kinetics of Philippidis et al. [244-246].

This apparent model of (Figure 5-11) can only give a rough idea on the bioreactor dynamics, but it is unable to describe the biomass adaptation to environmental changes, that is variations in the both inlet feed flow-rate and inlet mercury load in the influent. To offer a prediction to such engineering requirements, a structured kinetic model of the mercury uptake in the *E. coli* bacteria at a cellular level is necessary. The next chapter describes the WCVV structured **HSMDM** cell model proposed by Maria [1,2] to simulate the dynamics of the *mer*-operon expression self-regulation in the wild, or modified *E. coli* under various environmental (bioreactor operating) conditions. By using the reactor model and the apparent Michaelis-Menten kinetics of Table 5-1, the TPFB dynamics has been simulated under nominal operating

conditions of Table 5-2, but every time varying one operating parameter, that is:

- a) The inlet  $[Hg_L^{2+}]$  concentration from the background pollution level ( $0.02 \text{ mg L}^{-1}$ ) to successively 1, 5, or  $10 \text{ mg L}^{-1}$  levels (Figure 5-5);
- b) The inlet liquid flow rate FL of 0.01, 0.02, and  $0.04 \text{ L/min}$  (Figure 5-6)
- c) The biomass load  $c_x$  on the solid support taken as 0.1, 0.25,  $1.0 \text{ gX/L}$  (referred to the liquid volume) for a constant fraction of the solid in the reactor (Figure 5-7)
- d) The use of particle average diameter  $d_p$  of 1 mm, or 4 mm respectively (Figure 5-8).

The results of these TPFB simulations over ca. 50 min. running time reveal that the liquid residence time in the reactor (related to the inlet liquid flow rate, FL) and the biomass content ( $c_x$ ) are the most influential operating parameters, being directly responsible for the realised mercury uptake conversion. For instance, by doubling the feed flow rate (from  $0.02$  to  $0.04 \text{ L min}^{-1}$ ), the uptake conversion can be reduced with ca. 5%. The particle size is also important, an increase in the average  $d_p$  leading to a higher resistance to the diffusional transport in pores and to a particle effectiveness diminishment (from 0.51 for  $d_p = 1 \text{ mm}$  to 0.15 for  $d_p = 4 \text{ mm}$  under nominal conditions). On the other hand, the biomass average load in the reactor can be adjusted by employing a continuous purge-renewal system of the solid particles. The mercury content in the input flow,  $[Hg_L^{2+}]$  in the range of  $1\text{-}40 \text{ mg/L}$ , has a significant influence on the mercury content of the output gas (Figure 5-5). As further proved, this last parameter is also of tremendous

importance for adaptation of bacteria metabolism when using wild or cloned cells with an increased copynumbers of *mer*-plasmids."

## **E.coli cell structured dynamic model for mercury uptake, and its use to optimize the TPFB bioreactor**

### **Generalities on the proposed extended HSMDM model**

A first approach of this bioprocess/bioreactor modelling is those of Maria [1,2]. "The resulted hybrid dynamic model of (Table 5-3), with the apparent M-M kinetics of (Table 5-1) allows simulating the transient operating conditions of the TPFB bioreactor, but of a prediction quality very approximate about the

bioreactor dynamics. The main drawback of this TPFB apparent model is coming from its inability to describe the biomass adaptation to environmental changes, that is variations in the both inlet feed flow-rate and inlet mercury load in the influent. And, self-understood this rough model can give no information about the dynamics of the inner cell species related to the *mer*-operon expression. To offer a prediction to such engineering requirements, a structured kinetic model of the mercury uptake in the *E. coli* bacteria at a cellular level is necessary. The next chapter describes the WCVV structured cell model proposed by Maria [1,2,40] to simulate the dynamics of the *mer*-operon expression self-regulation under various environmental conditions. This model was constructed and validated by using the experimental data of Philippidis et al. [244-246], and the experimental information of

Barkay et al.[247] on the *mer*-operon characteristics. The WCVV cell model was built-up by accounting the GERM-s library, their P.I.-s, and the linking rules presented in the chap. 4. The proposed *E. coli* cell model by Maria [1,2,40] includes the GRC

responsible for the control of the *mer*-operon expression and for the whole process of mercury ions removal. The proposed GRC includes 4 lumped genes (denoted by GR, GT, GA, GD in (Figure 5-2, Figure 5-3, Figure 5-4, Figure 5-9) 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, other metabolites, and proteins), thus compromising the whole cell metabolism.

The GRC model includes four GERM-s of simple but effective [G(PP)1] type (see the discussion of chap. 4) as follows (see Figure 5-10):

a) a GERM to regulate the  $Hg^{2+}$  transport across the cellular membrane, mediated by three proteins (PmerP, PmerT, and PmerC) from the periplasmic space, considered as a lumped permease PT in the model. Phillippidis et al. [244-246] 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  $Hg(SR)_2$ . Instantly, the GT lumped gene expression is induced by the regulatory protein PR, and easily 'smoothed' by the large 'ballast' effect of the proteome lump P. (see rule 6 of chap. 4.5).

b) a 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 if they are present in traces, that is low nM concentrations). This GERM acts as an amplifier of the *mer*-expression leading to a quick (over ca. 30 s) cell response by starting the *mer*-enzymes production. The encoding GR gene expression to obtain PR is controlled by the protein PD, present in small amounts into the cell.

c) a GERM to control the expression of PA enzyme responsible for the  $Hg(SR)_2$  (that is  $Hg_{cyr}^{2+}$ ) reduction to metallic mercury ( $Hg_{cyr}^0$ ) into cytosol (relatively non-toxic for the cell, easily removable through membranar diffusion into the bioreactor bulk liquid-phase. From here, metallic mercury ( $Hg_L^0$ ) is continuously removed by the sparged air as mercury vapours ( $Hg_g^0$ ), to be later recovered. The encoding GA gene expression is induced and controlled by the PT protein.

d) a GERM controlling the protein PD synthesis. This protein has a complex role, by maintaining a certain level of GR expression even when the mercury is absent in the cytosol [247].

e) a GERM controlling the replication of the lumped cell proteome (P) and genome (G) (of concentrations  $10^7$  nM, and 4500 nM, respectively) in the immobilized *E. coli* cells. These data are based on the Ecocyc [95] databank (Figure 5-12), thus mimicking the cell 'ballast' effect on the cell genes expression, and the all considered reactions. The need to include the cell content lump (the so-called 'cell ballast') in the WCVV model is legitimate by the possibility offered by such a structured cell model to reproduce the smoothing effect of perturbations leading to more realistic transient times (comparing 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 eq. (3,5,6) of chap. 4, and the discussion of Maria [21,23,175-176] }.

In total, the GRC dynamic model describing the *mer*-operon expression includes only 26 individual or lumped cellular species involved in 33 reactions (Figure 5-2, and Figure 5-10). The structured cell WCVV model is presented in the (Table 5-4). The cell model is coupled with the SCR -TPFB dynamic model

(Table 5-3) through the  $r_j$ ,  $r_{j,app}$  linked to the  $r_j(c_{j,s})$  occurring inside cell. The WCVV cell model includes not only the reactions and the dynamics of the *mer*- GRC, but also the enzymatic reactions directly responsible for the environmental mercury  $Hg_{env}^{2+}$  (also denoted by  $Hg_L^{2+}$ ) import into the cell as  $r_{j,app}$ , and for its reduction to cytosolic metallic mercury in cloned *E. coli* cells. All reactions in the cell model of (Table 5-4) are considered elementary, excepting some of them for which extended experimental information exists, that

is [1,2,40]

- a) a Michaelis-Menten rate expression for the mercuric ion permeation through the membrane 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.
- d) Dimmerization reactions of TF-s are considered to be much rapid than the enzyme synthesis, while equal concentrations of active gene (G) and inactive (GPP) forms of the generic gene G are considered at homeostasis to maximize the GERM efficiency (see chap. 4.3-4.5) (Figure 5-2, and Figure 5-10).
- e) The lumped proteome P, present in a large amount, is included in all gene expression rates, thus leading to more realistic evaluation of the GERM regulatory efficiency indices P.I.-s (chap. 4.3, 4.4) [21,23,174,189].
- f) The model rate constants are estimated from solving the cell stationary mass balances for nominal concentrations of observable species (eqn.(8) of chap. 4), but also from optimizing the GERM regulatory indices (eqn.(10) of chap. 4.3), for instance by adjusting 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 (see chap. 4.3-4.4 for an extended discussion).

The *E. coli* cell characteristics and the considered cell key-species homeostatic concentrations used in the rate constants estimation are given in (Table 5-5). Exceptions are the Michaelis-Menten rate constants for the mercury transport and its reduction in cytosol adopted from the Phillipidis et al.[244-246] kinetic data. Thus, the M-M rate constants depends on the amount of [Gmer] plasmids in the cloned *E. coli* cells (Figure 5-10, and Figure 5-11). Simple correlations are used to include this essential aspect in the model. As extensively discussed in the chap. 4, such a cell WCVV structured model presents a large number of advantages, being able to: (1).- 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 the (Figure 5-13) as a step response after a ,step'-like perturbation in the mercury level from  $[Hg_{env}^{2+}]_s = 0.1 \mu M$  to  $10 \mu M$  (ca. 2 mg/L), in a cloned *E. coli* cell with *mer*-plasmids of [Gmer]= 140nM, compared to a cell cloned with only [Gmer]= 3 nM. 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. Clearly, an *E. coli* cell with a higher content of *mer*-plasmids reacts much strongly to the perturbation, by quickly starting to produce the enzymes responsables for the mercury removal. (2).- Because the  $Hg^{2+}$  reduction rate constants are dependent on the *mer*-plasmid level into the cell, the WCVV GRC model can predict the maximum level of *mer*-plasmids that can be added to the cell genome for improving its mercury uptaking rate (and capacity) without exhausting the internal cell resources, thus putting in danger the cell survival. Consequently, it follows that this cell model allows the *in-silico* design of modified *E. coli* cloned with a suitable amount of *mer*-plasmids to improve its possibility of cleaning wastewaters by improving the mercury uptaking capacity. As an example, in the (Figure 5-14) are presented the cell key-component stationary levels, and concentration of mercury in the bioreactor bulk-phase for two GMOs: one cloned with 67 nM, and another one cloned with 140 nM. Simulations of Maria and Luta [40] revealed that as the *mer*-plasmids level is increasing, as the mercury uptake capacity in increasing. However, an upper limit exists (around 140 nM) over which the cell resources will be exhausted, putting its metabolism in danger. (3).- By coupling the structured cell model of (Table 5-4) with the three-phase continuous bioreactor model of Table 5-3 (with immobilized *E. coli* cells on pumice beads; see also the bioreactor model in the Figure 5-11), Maria et al. [40,63] have been able to determine the optimal operating policies of the bioreactor in relationship to the culture of cloned cells. Similar studies are reported by [63,241,242].

### Extended HSMDM model structure

“To model the mercury uptake by *E. coli* bacteria at a cellular level, Maria [1,2] proposed a WCVV (chap. 4.2, 4.2.1) structured model that reproduces the dynamics of the *mer*-operon expression under various environmental conditions. The model was constructed and validated by using the Philippidis et al. [244-246] experimental data, and the Barkay et al.[247] information on the *mer*-operon characteristics. The proposed GRC, of modular construction, includes seven GERM-s placed in an *E. coli* cell of known characteristics (size, cell cycle, copynumbers of all individual or groups of components; EcoCyc databank [95]). The whole-cell approach allows predicting the dynamic response of the involved *mer*-operon to perturbations by mimicking the cell growth and its response to external stimuli. The *mer*-GRC acts by adjusting the expression level of individual *mer*-genes (self-regulation), but also by balancing the expression of related genes through various regulatory loops activated by certain exo-/endogeneous inducers. The basic unit of the construction is the semi-autonomous GERM (chap. 4.3) that control one gene expression [24,138,174]. To not complicate the GRC dynamic model, the GERM module

must include the minimum number of individual/lumped species/reactions necessary to adequately represent the experimental information. For instance, the gene expression general schema of (Figure 4-9-left) can be translated into a modular structure of reactions, more and more reduced depending on the GERM representation importance. Maria [24], and Yang et al.[228] proposed a classification of the GERM units that includes negative feedback regulatory loops and a cascade control (see chap. 4.3), thus allowing application of the control theory principles to quantitatively evaluate the individual GERM and the whole GRC regulatory performance indices (P.I.-s, see chap. 4.4), such as: *stationary efficiency* to treat 'step'-like perturbations in key species concentrations; *dynamic efficiency* to treat 'impulse'-like perturbations (translated by the species recovering times of the steady-state / homeostasis); *responsiveness* to stimuli (expressed by the transient times from one steady-state to a new one); *species connectivity* (expressed by a synchronised response to perturbations); *system stability region and strength* (see [21,146,174] for details). The GRC for mercury uptake in *E. coli* proposed by Maria [1,2] is schematically illustrated in (Figure 5-1), contains seven GERM-s, as follows: (i) a GERM to regulate the  $Hg^{2+}$  transport across the cellular membrane, mediated by a lumped permease PT, which is synthesized by the GT gene expression induced

by the regulatory protein PR. Once the mercuric ions arrive in the cytosol, thiol redox buffers (such as glutathione of millimolar concentrations) form a dithiol derivative  $Hg(SR)_2$ ; (ii) a GERM to control the synthesis 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, by starting production of *mer*-enzymes. The gene GR expression to get PR is controlled by the protein PD, present in small amounts into the cell; (iii) a GERM to control the synthesis of PA reductase responsible for the  $Hg(SR)_2$  reduction to metallic mercury (relatively non-toxic for the cell, easily removable through membranar diffusion). The encoding gene GA expression is induced and controlled by the PT protein; (iv) a GERM controlling the protein PD synthesis, having the role of maintaining a certain GR expression level in the absence of mercury [247]; (v) a GERM controlling the replication of the lumped cell proteome (P) and genome (G) (of large concentrations), thus mimicking the cell 'ballast' and 'smoothing' effect of perturbations on the other gene expressions and reactions of the cell; (vi) two reaction modules, characterizing the  $Hg^{2+}$  import into the cell (activated / saturated by the environmental  $Hg^{2+}$ , mediated by PT and RSH), and its cytosolic reduction mediated by PR in the presence of the renewable co-enzyme NADPH (present in excess in the cell).

The WCVV model equations for the mercury uptake in *E. coli*, together with the general hypotheses of the WCVV model are presented in the (Table 5-4). Basically, the cell is considered an open system, of uniform content and negligible inner gradients. To not increase the number of parameters, the structured model includes GERM of simplest form (chap. 4.3; [24]) for all *mer*-genes, by using dimeric TF-s (of Protein:Protein type) to increase the GERM regulatory efficiency, as experimentally proved by [21,80,189]. The resulted HSMDM model includes 26 individual or lumped cellular species and 33 reactions. All reactions are considered elementary, except some of them for which extended experimental information exists, that is Michaelis-Menten rate expressions for mercury permeation into the cell, and its reduction in cytosol. A Hill type induction of the GR expression is adopted to rapidly amplify the *mer*-operon expression when mercuric ions are present in significant amounts inside cell. Dimerization reactions of TF-s are considered to be much rapid than the enzyme synthesis, while equal concentrations of active (G(i)) and inactive (G(i):TFTF) forms of the generic gene G(i) are considered at homeostasis to maximize the GERM efficiency [21-24]. The homeostatic characteristics of *E. coli* cells (belonging to a uniform culture) from the reactor, and the adopted species concentrations are presented in (Table 5-5).

Phillippidis et al. [244-246] have proved, on an experimentally basis, that the membranar permeation step is the rate-determining step for the mercury uptake by the bacteria, and they used cloned *E. coli* cells with a higher amount of *mer* plasmids (G<sub>mer</sub> from 3 to 140 nM) to increase the content of *mer*-genes producing *mer*-enzymes responsible for the mercury uptake process. However, the increased mercury reduction efficiency is obtained up to a certain limit of *mer*-plasmids (with a mercury ions import controlled by the permease PT) to not exhaust the internal cell resources thus putting in danger the cell survival. When simulating the TPFB reactor performance (of nominal conditions given in the Table 5-2) by employing a classical *unstructured* reactor model (see chap. 5.5.5), the biomass adaptation to variable input loads is not accounted for, that is the maximum rates  $v_{m,t}$ ,  $v_{m,p}$ , and the Michaelis inhibition constants  $K_{m,t}$ ,  $K_{m,p}$ , and  $K_{i,p}$  of (Table 5-1) are kept constant (eventually depending only on the G<sub>mer</sub> plasmid level). In fact,  $v_{m,t}$  and  $v_{m,p}$  rate constants depend on the cell enzyme levels of PA (reductase) and PT (permease) (Figure 5-2), which vary during the bacteria adaptation in the bioreactor under stationary or transient conditions.

To solve this problem by also offering a more detailed and robust prediction on the system dynamics / cell metabolism evolution, the TPFB bioreactor and the *mer*-GRC of *E. coli* cell models are necessary to be coupled in a hybrid structured HSMDM model. The two linked differential models are solved together, by

applying a mutual exchange of input/output parameters  $\{[Hg_{env}^{2+}], [Hg_{env}^0], PT, PA\}$  on every small time increment throughout the solution (integration) of the HSMDM model, as graphically represented in (Figure 5-2).

The solving rule involves successive integration steps with an adopted time-interval equal to the cell cycle (ca. 30 min), enough to obtain the steady-state of the TPFB reactor on every time-interval, as following: (i) The rule starts with solving the extended hybrid HSMDM *E. coli* cell model by using the known initial condition (the cell variables' initial state, or those from the end of the previous integration cycle), and by considering the current concentration of  $[Hg_{env}^{2+}]$  in the bioreactor liquid phase (the nutrients are considered in excess and of constant levels). Thus, the cell species dynamics over one cell cycle is obtained from solving the WCVV cellular model. (ii) Then, the TPFB reactor model is solved over the current time-interval by using the known initial conditions (i.e. the reactor state variables eventually from the end of the previous time-interval), and by considering the enzymes PT and PA concentrations resulted from the *E. coli* model solution. [PT] and [PA] are necessary for setting the maximum reaction rates  $v_{m,t} = k_7 C_{PT}$  and  $v_{m,P} = k_8 C_{PA}$  in the reactor model. The biomass level on the support is taken constant in the simulated case study, but an additional mass balance can be easily added to the reactor model if necessary.

The procedure is repeated of a large number of times, over hundreds of cell cycles. For instance, (Figure 5-13) displays the *E. coli* cell adaptation after a 'step'-like perturbation in the environmental  $[Hg_{env}^{2+}]_s$  (that is in the bioreactor bulk-phase) from the background level of  $0.1 \mu M$  to  $10 \mu M$  (ca. 2 mg/L), for the

case of cell cloned with  $[Gmer] = 3 \text{ nM}$  (full line), or with  $[Gmer] = 140 \text{ nM}$  (dash line) *mer*-plasmids. The transient state toward the cell new homeostasis usually lies over 15-20 cell cycles as long as the environmental stationary perturbation is maintained. The simulated species dynamic trajectories in this (Figure 5-13) reveal a vigorous response of the cell *mer*-species (especially of PT, and PA) to the perturbation in the environmental  $[Hg_{env}^{2+}]_s$  (i.e. liquid bulk-phase)."

### Fitting the extended HSMDM model parameters from experimental data

To overcome the high computational effort necessary to fit the large number of rate constants of this extended HSMDM model, and to increase their confidence and physical meaning, a *three-step procedure* was employed based on the available experimental data and additional information from literature, as

(a). **Mass transport parameters** of the TPFB reactor, that is interfacial partial transfer coefficients ( $k_s a_s, k_L a_L, k_G a_G$ ), effective diffusivity ( $D_{ef}$ ), and particle effectiveness factor ( $\eta$ ) have been estimated by using the experimental data of Deckwer et al. [248], and based on common correlations from the chemical engineering literature (Table 5-3) evaluated for the specified reactor operating conditions of (Table 5-2). (b). **Cell model parameters are estimated** by using the Philippidis et al. [244-246] kinetic data obtained from separate batch experiments with not-cloned *E. coli* cells. By using the defined cell nominal characteristics of (Table 5-5) (some key notations are:  $t_c$  = cell cycle time;  $V_{cyt_0}$  = born cell volume;  $c_x$  = biomass concentration in the bioreactor; P = cell lumped proteome; G = cell lumped genome; NutG, NutP = lumped nutrients for the G and P synthesis, respectively;  $[Gmer]$  = *mer*-plasmid levels. The stationary levels of the essential cell *mer*-proteins (PA, PR, PD), and for the involved TF-s (PRPR, PTPT, PAPA, PDPD) are taken from the literature data, and by maximizing the GERM-s regulatory efficiency P.I.-s (chap. 4.3 ('rate constants estimation'), and chap. 4.4) [21,175,176]. The resulted rate constants of (Table 5-4) have been estimated for the most severe experimental conditions of  $[Hg_{env}^{2+}]_s = 120 \mu M$ , and  $[Gmer] = 140 \text{ nM}$ . The used first guess of Hill-induction rate constants ( $n_{PD} = 1, n_{PR} = -0.5, n_H = 2, a = 3$ ) have been adopted at values recommended in the literature, by similarity with the Hill-induction of gene expression in genetic switches (see the remarks included in Table 5-4), while the Michaelis-Menten rate constants of mercury membranar transport ( $r_{max,t}, K_{mt}$ ), and its reduction rates' constants ( $r_{max,p}, K_{mp}, K_{ip}$ ) have been kept at the fitted values of Philippidis et al. [244-246] (Table 5-1). The reference concentration  $C_{Hg2cy,ref}$  was adopted at the average cytosolic level of mercuric ions detected by Philippidis et al. [246]. During model estimation, the GERM regulatory indices have been kept at their optimized levels, which correspond to: (i) Equal concentrations of catalytically active/inactive forms  $[G(j)]_s = [G(j)TFn]_s$  adopted at steady-state to ensure GERM maximum regulatory efficiency vs. perturbations (i.e. smallest sensitivities of the homeostatic levels vs. external perturbations [21,24,185], and (cap.4.4); (ii) adjustable optimum  $[TF]_s$  level (ca. 4 nM

here; Table 5-5) involved in the gene expressions to get the minimum recovering times after a dynamic perturbation in the key species [21,175,176,189] (chap. 4.3 ('rate constants estimation'), and chap. 4.4). Other adjustable parameters, such as the cell concentration in biomass ( $c_{cell}$ ), are tuned to fit the experimental cellular mercury reduction rate of Philippidis et al.[244].

(c). **The model validity is extended** over a wider experimental range, by covering the input  $[Hg_L^{2+}] = 0-100$  mg/L, and plasmid levels of  $[Gmer] = 3-140$  nM. All these have been realised by adjusting some key-parameters of the extended HSMDM model, and by applying two concomitant fitting criteria, that is: (1) the cytosolic mercury reduction rate should fit the experimental values of Philippidis et al. [244-246], obtained for 'permeabilized' cells (i.e.  $\text{Min} \|r_{Hg,exp} - r_{S,model}\|_{model}$ , see r8 of Table 5-4); (2) the TPFB reactor output  $c_{Hg,out}$  predicted by the unstructured model (Table 5-1, and Table 5-3, and 118 chap. 5.5.5) with using the Philippidis et al.[246] Michaelis-Menten parameters and the mass transfer terms, in order to fit with those of the structured (cell + reactor) extended HSMDM model of (Table 5-3, and Table 5-4). The Hill parameter  $b=2a^4$  was adjusted by using the following approximate linear dependence on the inlet mercury load:

$$a = a_{min} + (a_{max} - a_{min})(cHg_{in} - cHg_{in}^{min}) / (cHg_{in}^{max} - cHg_{in}^{min}),$$

(see the fitted values  $a_{min}$  and  $a_{max}$  in Table 5-4 – footnote d). The above simplified correlation tries to account for the influence of the environmental mercury on the induction characteristics of the *mer*-operon expression, that is a slow induction in the presence of low levels of  $Hg_L^{2+}$  inducer, and a sharp (sigmoidal) *mer*-operon expression response to high levels of  $[Hg_{cyt}^{2+}]$  inducer.

When a certain saturation level is reached, the limited cell resources will impose 'flattening' the metabolic *mer*-rates and the *mer*-protein levels, irrespectively to the increased amount of mercury in the environment [244-246].

The extended HSMDM model predictions are in a satisfactory agreement with the experimental data. Thus, the predictions of  $r_{Hg}$  by the extended HSMDM model of (Figure 5-15, A-C) (curves '2') roughly fall within the confidence band [',Eup', ',Elow'] of the experimental curves (',E') of Philippidis et al.[244] for three different cloned cell cases (confidence curves being plotted by taking constant the reported maximum relative error of ca. 19%). The *unstructured* model (cap.5.5.5, Table 5-1 plus Table 5-3) predictions (that is curves '1') reported apparent  $r_{Hg}$  of low adequacy, that is lower values due to the rough bioprocess representation by the M-M model, and due to the inclusion of the mass transport resistance between the three-phases in contact. The extended HSMDM model adequacy in terms of standard error 'Std' / average observed value 'Obs' ratio is evaluated for every cloned cell culture, leading to Std/Obs of 22.3%, 16.7%, 20.4%, 19.9%, 18.8% for  $[Gmer]$  levels of 3, 67, 78, 124, and 140 nM respectively, that is acceptable values comparable to the maximum experimental error. The structured vs. unstructured model outputs, in terms of outlet concentration of mercury ( $c_{Hg,out}$ ) are also in a fair agreement (curves not display here), that is Std/Obs values of 3.0%, 4.7%, 3.7%, and 12% for  $[Gmer]$  levels of 67, 78, 124, and 140 nM respectively. The model poor adequacy for the  $[Gmer] = 3$  nM data set and low  $[Hg_L^{2+}]$  in the environment (input) might be explained by the use of a less adapted *E. coli* cell than probably those studied by Phillipidis et al.[244], reflected by a smaller [PA] initial level of 600 nM (see the PA-curve in Figure 5-15-D). Once a higher level of Gmer-plasmids are introduced into the cell, and higher  $Hg_L^{2+}$  stimulus levels are present then, the cytosolic PA level is tripling.

The extended HSMDM structured model cellular rate constants (see the above point (b) , and Table 5-4) are in a good agreement with the reported data from literature, as remarked in the same (Table 5-4). The fitted rate constants multiplied by the reactant lead to reaction rates of the same order of magnitude with those reported in the literature for similar genetic processes, such as the **TF** (repressor monomer) dimerization, the **TF** binding to gene operator, or the mRNA (genes) synthesis reactions (Table 5-4) [1,2]. This observation sustains the physical meaning of some model parameters, thus increasing the **HSMDM** model robustness.

By extending the detailing degree of the bioreactor dynamic model at a cellular level, the resulted structured **HSMDM** model preserves the adequacy of the unstructured model, but adding the possibility to predict the cell species/fluxes dynamics. By offering details on the cell metabolism adaptation, on the '**intrinsic**' reduction rate, and possibilities to *in-silico* predict the modified cell response to various stimuli, such **HSMDM** models are clearly superior to the unstructured (apparent) bioreactor dynamic models. Eventually, extended **HSMDM** models worth the supplementary experimental and computational effort to derive them."



## Prediction of the bioreactor dynamics by using the extended HSMDM model compared to the unstructured model

"Simulation of the **TPFB** bioreactor with the *unstructured* model (including only 3 bulk variables) of (**Table 5-1** plus **Table 5-3**) can predict the dynamics of the mercury uptake (3 bulk variables) by the biomass and the resulted metallic mercury transfer in the liquid and gas phase of the **TPFB**, as displayed in (**Figure 5-16**) after a 'step'-like perturbation in the  $Hg^{2+}$  level of the input flow (**Figure 5-16**, up-left) from traces of mercury (100 nM) up to  $20 \text{ mg L}^{-1}$  (curves '1'). These predicted trajectories are close to those predicted by the structured **HSMDM** cell and reactor model (including 26 cellular + 3 bulk variables) (curves '2') for two types of cloned cell cases (with  $[Gmer] = 67 \text{ nM}$  and  $[Gmer] = 140 \text{ nM}$ ).

The difference between the two process models (structured vs. unstructured) consist in the possibility to 'visualise' the dynamics of the involved metabolic processes over several cell cycles as response to the external stimuli. As plotted in (**Figure 5-17**), the structured bi-level model can predict the sharp increase in the *mer*-genes (**GR, GT, GA, GD**) levels and, the encoded *mer*-proteins (**PR, PT, PA, PD**), and cytosolic ('intrinsic') reduction rate, as well as the cytosolic content of mercury for both cloned cell cultures following a step-perturbation in the influent  $Hg^{2+}$  load (at an arbitrary time  $t = 9000 \text{ min}$ ). As the plots indicate, the transient period is of ca. 450 min (over the interval marked [9000, 9450] min on the time-axis), that is over ca. 15 cell cycles. The cell levels of the involved *mer*-genes and of their encoded proteins are continuously changing, while the genetic information is transmitted from one cell generation to the next one. In accordance to its structure, the cell model can also predict the effect of increasing the content of some genes (or of removing them by gene knock-out technique, not approached here). Such modelling aspects are impossible to be represented by an overall / unstructured model. Moreover, the structured **HSMDM** model can predict the cell characteristics and operating conditions required to get an imposed efficiency of the bioreactor (not presented here; see an example for another process given by Maria and Renea [12]).

One disadvantage of the structured **HSMDM** approach is related to the large amount of required experimental information on the cell metabolism and on the involved GRC necessary to adequately fit the model parameters. Another disadvantage is related to the required larger computational time (up to hours on a common PC using a 'stiff' integrator) to get a reasonable prediction of the cell bioprocess evolution. However, this last constraint is not very strong for the process monitoring purposes, as long as most of industrial biological processes present a time constant much larger than the required simulation time. Beside, for quick control purposes, a reduced hybrid bi-level unstructured model can be used instead, being adjusted by using the experimental data. The next paragraph presents such a quick approach of the process control."

## Optimization of the bioreactor dynamics by using a reduced unstructured hybrid model

Even if much simpler, the *unstructured* (apparent) model of the bioprocess (**Table 5-1**), in more simple forms, coupled with those of the **TPFB** bioreactor (**Table 5-3**), is leading to a hybrid dynamic model, able to adequately predict the dynamics of the bioreactor main state-variables (that is 3 bulk variables, see **Figure 5-17**). This reduced hybrid unstructured model (**HUDM**) can adequately predict the dynamics of the mercury uptake by the biomass and the resulted metallic mercury transfer in the liquid and gas phase, as displayed in (**Figure 5-16**) after a 'step'-like perturbation in the  $Hg^{2+}$  level of the input flow (figure upleft) from traces of mercury (100 nM) up to  $20 \text{ mg L}^{-1}$  (curves '1'). Of course, such a reduced model can not 'visualise' the dynamics of the involved metabolic processes, and the cell adaptation as response to the external stimuli over several cell cycles.

As proved by Maria et al.[40,63], such a reduced **HUDM** model is enough adequate for a quick engineering evaluation of the bioreactor, that is: **(1)**. A parametric sensitivity analysis, to highlight and rank the most influential operating parameters of the bioreactor and, **(2)**. A rough optimization of the bioreactor operation. This paragraph will briefly exemplify these capabilities of the **HUDM** model for the studied bioprocess and **TPFB** bioreactor of this chap. 5. **(3)**. A quick control of the **TPFB** bioreactor by means of its main control variables: the biomass load  $[X]$  that is the amount of solid support with immobilized living cells in the bioreactor; the mercury concentration  $[Hg^{2+}_{L,in}]$  in the inlet feed flow-rate, the inlet feed flow-rate (FL) (equal to the output flow-rate); the millimeter solid support size ( $dp$ ); the solid fraction in the **TPFB** bioreactor ( $\mathcal{E}$  s).

The engineering analysis of this paragraph exemplify this concept of large flexibility, and easy handling of these reduced *unstructured* **TPFB** models. Besides, such reduced dynamic hybrid models are easy to be up-dated from experimental data, because they include a small number of rate constants in the M-M (or Monod) based kinetic models. Exemplification is made for a lab-scale **TPFB-2** bioreactor using *P. putida* sp. immobilized on porous alginate granules. Even if, conceptually and functionally,

this **TPFB-2** bioreactor does not differ much from the **TPFB** bioreactor (**Table 5-2**) analysed in the chap. 5.1-5.5, there are a couple of differences. The most important refer to the biomass support (which is alginate beads, and not pumice), and to the range of tested solid fraction in the bioreactor ( up to 10%, as recommended in the literature [257] ).

The constructive features, and the nominal operating conditions of the **TPFB-2** bioreactor are given in (**Table 5-6**).

The approached **TPFB-2** bioreactor, continuously operated, is that used by Deckwer et al.[248] to remove the mercury ions from wastewater at a bench-scale using *P. putida* sp. immobilized on porous alginate granules of 0.9 mm average diameter. A good diffusion of the substrate (mercuric ions) to bacteria is assumed to take place in the large support pores, while an equilibrated biomass metabolism is ensured by the added nutrients in the fed wastewater, and by a vigorous liquid aeration by continuously sparkling air. The liquid flow rate (or equivalently the liquid residence time), the aeration rate (in a high excess vs. metabolic requirements), pH, and temperature are tightly monitored to ensure a sustainable bioprocess. The biomass content of the support varies (up to 0.5-0.6 gX/g-support; [248]), but one assumes that a quasi-constant level can be maintained by employing a purge/renewal system of the '**biocatalyst**' particles (i.e. immobilized biomass). The outlet gas containing the volatile metallic mercury is passed through an adsorption device for recovering the mercury (a liquid metal).

To simulate the **TPFB-2** bioreactor dynamics and performances, the dynamic ideal three-phases (G-L-S) model of (**Table 5-7**) was adopted [253,254]. The solid particles with the immobilized biomass (of uniform characteristics) are considered uniformly distributed in the homogeneous liquid phase, due to the well-mixing conditions. The metallic mercury, resulting from its reduction in the bacteria cytosol, diffuses through the cell membrane in the bioreactor liquid phase, and then is transported as micro-drops (of ca. 5  $\mu$  m diameter [250]), or dissolved (solubility of 26  $\mu$  g L<sup>-1</sup> at 26°C) to the gas-liquid interface from where it passes as vapours into the homogeneous gas phase. Low amounts of mercury bioaccumulate into bacteria, being neglected in the model. Mercury mass balance in the liquid and gas phases includes the following terms (**Table 5-7**):

a) The diffusion equations to get species concentration profile inside the particle are replaced by the apparent mercury uptake rate ( $r_{app}$ ) by the immobilized bacteria in porous particles. Its value is evaluated by solving the quasi-steadystate equality of mass fluxes for an assumed kinetic model, by using the substrate concentration  $c_{Hg2s}$  at the solid-liquid external interface of the carrier on every (small) time interval (**Table 5-8**).

b) In the flux equality equation, the solid-liquid mass transfer coefficient  $ksas$  was experimentally evaluated by Deckwer et al.[248], even if common criteria relationships can also be employed (**Table 5-7**).

c) The substrate transport diffusional resistance in the particles' pores is expressed by an effectiveness factor ( $\eta$ ) evaluated by means of the Thiele modulus, particularized for a Michaelis-Menten rate expression (**Table 5-8**, [33]). The effective diffusivity ( $Def$ ), experimentally evaluated by Deckwer et al.[248], was used instead of those evaluated by means of common formulae from literature.

d) The transport rate  $r_{trans}$  of the metallic mercury to the gas phase is evaluated from the quasi-steady-state equality of mass fluxes at bubbles interface (on every small-time interval) for known gas-liquid mass transfer coefficients ( $kLaL$  on liquid film side, and  $k_g a_g$  on gas film side). While  $K_L a_L$  was experimentally evaluated by Deckwer et al.[248],  $k_g a_g$  results from application of criteria relationships, by averaging the gas volume fraction in the bed over three different formulae displayed in (**Table 5-7**; [254]).

The adopted **TPFB-2** reactor model allows simulating transient operating conditions of the bioreactor for a specified mercury reduction bioprocess kinetics. Three types of *unstructured* apparent kinetic models of M-M type from literature have been checked in the present analysis (**Table 5-8**), as followings:

a) a pseudo-first-order kinetics (**PFOM**), valid for small substrate (mercuric ions) concentrations in the bioreactor liquid-phase, that is lower than 1 mg/L. This situation corresponds to the reactor operation over a wide parametric range of [ $\phi_{ref}/10, 2 \phi_{ref}$ ], where  $\phi_{ref}$  = Thiele modulus in (**Table 5-8**). This extremely simplified kinetic model advantage is coming from its simplicity, being easily to be used for the bioreactor design and control, by including *only one* adjustable rate constant ( $k$ ).

b) a simple Michaelis-Menten (M-M) kinetic model (**MMM**) of a first order inhibition with substrate (**MM1= MMM**), including two adjustable constants according to the biomass characteristics. The **TPFB-2** reactor model solution can be easily obtained, even if additional calculations are necessary to derive the particle effectiveness and to solve the mass flux equality at

solid interface during the transient regime (**Table 5-8**). This model was proposed by Philippidis et al.[244-246] for mercury uptake in *E. coli* cells, the two rate constants being adjusted according to the *mer* plasmid content in the cloned cells (in the range of 3-140 nM *mer*-plasmids, **Table 5-1**; [1,2]).

c) an extended Michaelis-Menten kinetic model, of a better adequacy, including a second order inhibition with substrate (**MM2= PHM**), with three adjustable rate constants.”

Here it is to notice that, even if simple M-M type kinetic models are used, solving the **TPFB-2** reactor model requires a large computational effort to ensure a satisfactory precision when deriving the particle effectiveness, or when solving the mass flux equality at solid-liquid interface during the transient regimes (a nonlinear equation). To overcome these problems, the second order inhibition term was included as an adjustable parameter  $\phi = 1 + (c_{Hg^{2+}}/K_i)^2$  (that is the **PHM** kinetic model). According to the experimental information of Deckwer et al. [248] an average value of  $\langle \phi \rangle = 2$  can be adopted for inlet substrate loads in the tested range of [1, 20] mg/L. The supplied oxygen in a large excess ensures an equilibrated cell metabolism and a high content of cytosolic NADPH necessary for mercury reduction. Traces of mercury in the fed wastewater maintain the *mer*-operon active into bacteria cell, as for the „wild” bacteria (*E. coli*, *Pseudomonas sp.*).

### Parametric sensitivity analysis of the TPFB-2.

”The parametric sensitivity analysis of the **TPFB-2** reactor employed for testing the mercury removal efficiency by using immobilized bacteria (*E. coli*, or *Pseudomonas sp.*) on a suitable porous support is an essential engineering analysis to be used in further process scale-up, optimization and control stages. The use of the reduced hybrid **HUDM** model with including an apparent M-M model is very suitable for such a quick engineering analysis of the bioreactor.

In the simulated case, the very effective bacteria metabolism allows an efficient reduction of the cytosolic mercury (92-99% conversion), excretion and transport of the resulted volatile metal to the gas phase for a wide range of pollutant loads in the wastewater. By keeping a reasonable small size of particles (1-2 mm), the most important control parameters appear to be the solid fraction and the feed flow rate of the reactor.

To determine the optimal operating conditions of the **TPFB-2** reactor and its control possibilities, an essential analysis step is to determine the most important factors that influence the process in order to adjust the reactor efficiency. This can be easily done by the so-called sensitivity analysis (review of Varma et al.[231]) by evaluating the time-dependent sensitivity of a state variable 'y' with respect to a parameter  $\phi$ , i.e.  $s(y; \phi) = \partial y / \partial \phi$  (in absolute terms), or  $S(y; \phi) = \phi_{ref} / y_{ref} s(y; \phi)$  (in relative terms; 'ref' = the reference/nominal operating conditions in the parametric space).

Five control parameters ( $\phi$ ) of the **TPFB-2** reactor are here considered, that is: **(1)** the biomass load [X] that is the amount of solid support with immobilized living cells in the bioreactor; **(2)** the mercury concentration  $[Hg^{2+}]_{L,in}$  in the inlet feed flow-rate; **(3)** the millimeter solid support size (dp); **(4)** the inlet feed flow-rate (FL) (equal to the output flow-rate); **(5)** the solid fraction in the **TPFB** bioreactor ( $\epsilon_s$ ). For such a purpose, instead of using the extended *structured* **HSMMDM** model, very computative, the reduced **HUDM** model with including *unstructured* bioprocess kinetics can be used for a quick approximate analysis of the **TPFB-2** bioreactor. Of course, the reduced unstructured **HUDM** model can not give any information about the cell metabolism and *mer*-species adaptations over tents of cell cycles following the bioreactor dynamics.

In the present application, the global sensitivities of mercury uptake efficiency have been derived by using the finite difference method with repeated simulations of the **TPFB-2** reactor for various values of the operating control variables, as listed in the below (**Outline 5-1**). It is self-understood that, bioreactor dynamic simulations are made by successively varying each of the control parameter, one after one, by keeping constants the other parameters at their nominal values listed in the same table.

The control parameters ( $\phi$ ) have been varied around the set-point (the reference value 'ref') in the aprox. range of  $[\phi_{ref}/10, 2\phi_{ref}]$ . The sensitivity analysis of the **TPFB-2** reactor was repeated by checking the all 3 unstructured kinetic models [**PFOM**, **MM1**, **PHM=MM2**]of (**Table 5-8**). The comparison of the simulation results allows to highlight the influence of the following control parameters on the bioreactor efficiency in removing the mercury from the wastewater, that is: **(1)** the biomass load  $c_x$  in the reactor (**Figure 5-18**); **(2)** the mercury ion load in the fed wastewater  $Hg_{in}^{2+}$  (**Figure 5-19**); **(3)** the average diameter  $d_p$  of the solid particles playing the role of biomass support (**Figure 5- 20**); **(4)** the inlet flow rate  $F_L$  (**Figure 5-21**); **(5)** the solid fraction  $\epsilon_s$  in the reactor (**Figure 5-22**). Being in high excess, variation of aeration rate was not included in the analysis. The graphs

display the dynamic evolution of the mercury concentrations in the liquid and gas phases, for every analysed reactor operation case. Simulations reproduced the transient operating period of the **TPFB-2** from its initial condition of mercury absence in all phases, until the reactor steady-state (**QSS**). The resulted global sensitivities of the mercury removal efficiency in the **TPFB-2** are presented by Maria et al.[63]. This sensitivity analysis of the **TPFB-2** bioreactor leads to the following conclusions:

a) The most adequate **PHM (MM2)** apparent kinetic model is supposed to give accurate predictions of the process evolution and reactor performances. The differences between **MM1** and **MM2** predictions are negligible in all cases. By contrast, as expected, the **PFOM** very global kinetic model offers very weak, and very biased predictions, even for small substrate concentrations in the reactor (much smaller than 1 g/L).

b) The same prediction discrepancy between the unstructured **MM1** and **MM2** models (from one side), and the **PFOM** global kinetic model (from the other side, see (**Table 5-8**) results from plotting the reactor steady-state concentrations in the gas-phase and liquid-phase for various operating parameters (**Figure 5-23**). As depicted, the **PFOM** predictions usually underestimate the **TPFB-2** reactor removal efficiency.

The weakness of **PFOM** unstructured global model is even more evident when predicting the bioreactor efficiency in (**Figure 5-23-A-E**) for a wide range of the control variables, their predictions being very different from those predicted by the **MM1-MM2(PHM)** models. Consequently, special precautions should be taken when a reduced unstructured kinetic model is used for a quick on-line control of the continuous bioreactor.”

c) According to the **MM2 (PHM)** bioprocess apparent model, the most influential parameter for the mercury uptake efficiency in the **TPFB-2** is the solid fraction ( $\varepsilon_s$ ) in the reactor, its relative sensitivity being 115%/100% (all parametric sensitivities  $s(y;\phi) = \partial y / \partial \phi$  mentioned above are given by Maria et al.[63]). For instance, by doubling the solid fraction, the mercury uptake conversion can be increased with more than 5%. Of next importance are the inlet flow rate  $F_L$  (related to the liquid residence time), and the particle diameter  $d_p$  controlling the diffusional resistance in pores and particle effectiveness). The use of solid particles larger than 2 mm is not recommended due to such reasons. The biomass load  $C_x$  in the reactor, and the inlet mercury load  $Hg_{in}^{Hg^{2+}}$  in the wastewater play a smaller role in the tested parameter ranges, because the sufficient amount of biomass can sustain the quasi-uniform process efficiency. In fact, in the tested parametric space, the mercury concentration within the reactor bulk-liquid [ $Hg_L^{2+}$ ] never exceeds the critical load of 10-15 mg/L, value above which, mercury import into the cell is lowered by the genetic *mer*-regulatory circuit as proved experimentally by Philippidis et al.[244-246] for the *E. coli* cell case. Simulations of Maria [1,2] proved a quick adaptation of the cell metabolism for wild or cloned bacteria aiming at maintaining a quasi-uniform mercury removal efficiency within large loads of substrate. Of course, the unstructured, reduced **HUDM** model can not give any information about the cell metabolism and *mer*-species adaptations over tents of cell cycles following the bioreactor dynamics. Only the extended structured **HSMDM** model of the **TPFB** is able to provide this kind of information.

d) In the bioreactor dynamic model (**Table 5-7**), the uptake rate ( $r_{app}$ ) is closer to the transport rate ( $r_{trans}$ ) at high conversions, thus indicating an efficient mercury removal from the liquid phase with an approximate rate of ca. 0.18 mg/L/min which, is much higher than those realized by the sorption in biomaterials removal method (that is 0.03 mg/L/min; [258]).

As a general conclusion, the use of hybrid *unstructured* Michaelis-Menten kinetic models coupled with the bioreactor model to obtain the reduced **HUDM** models, can offer interpretable results for a quick but very rough engineering evaluation of the **TPFB** bioreactor. The bias in conversion predicted by the very approximate linear **PFOM** model is up to 7-8% in the investigated parametric range, but can be higher for a wider range of operating conditions. Even if convenient for designing an adaptive controller of the process, such **HUDM HUDM** kinetic models are not recommended for the accurate bioprocess analysis, reactor design and optimization. The extended *structured* **HSMDM** model, even if requiring more intensive experimental and computational steps to adjust its parameters and to simulate the multi-phase system dynamics, is preferred in all the engineering analysis steps. That is because such a *structured* model can predict not only the bioreactor state-variables dynamics, but also the biomass adaptation to environmental changes, that is variations in the both inlet feed flowrate and inlet mercury load in the influent. Besides, such an extended **WCVV** structured cell model is able to simulate the dynamics of the *mer*-operon expression, and its self-regulation in the wild, or modified bacteria under various environmental (bioreactor operating) conditions. Such a structured **HSMDM** model can guide the engineers and biologists to obtain GMO bacteria of higher performances for the studied bioprocess.

Mass Balance and State Equations	Remarks
$\frac{dC_j}{dt} = \frac{1}{V} \frac{dn_j}{dt} - D C_j = g_j(C, k)$	continuous variable dynamic model representing the cell growing phase (ca. 80% of the cell cycle)
$\frac{1}{V} \frac{dn_j}{dt} = r_j(C, k) ; j = 1, \dots, n_s$	
$V(t) = \frac{RT}{\pi} \sum_{j=1}^{n_s} n_j(t)$	Pfeffer's law in diluted solutions
$D = \frac{1}{V} \frac{dV}{dt} = \left( \frac{RT}{\pi} \right) \sum_j \left( \frac{1}{V} \frac{dn_j}{dt} \right)$	$D$ = cell content dilution rate = cell volume logarithmic growing rate
$\frac{RT}{\pi} = \frac{V}{\sum_{j=1}^{n_s} n_j} = \frac{1}{\sum_{j=1}^{n_s} C_j} = \frac{1}{\sum_{j=1}^{n_s} C_{j0}}$ constant.	constant osmotic pressure ( $\pi$ ) constraint
$\left( \sum_j C_j \right)_{cyt} = \left( \sum_j C_j \right)_{env}$	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 ( $\pi_{cyt} = \pi_{env} = \text{constant}$ ).	
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 ; $V = V_0 e^{D_s t}$ ; $t_c$ = duration of the cell cycle.	
g. Homeostatic stationary growth of $(dC_j / dt)_s = g_j(C_s, k) = 0$ .	
h. Perturbations in cell volume are induced by variations in species copynumbers under the isotonic osmolarity constraint: $V_{perturb} / V = (\sum n_j)_{perturb} / (\sum n_j)$ .	

Notations: T = absolute temperature; R = universal gas constant; V= cell (cytosol) volume;  $\pi$  = 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.

**Table 4-1:** The concepts and the basic hypotheses of WCVV dynamic modelling framework in living cells of variable volume (adapted from [24]).

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

**Table 4.2 :** Some advantages when using the holistic WCVV framework when modelling [21-24,175,176].

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

**Table 4-3 :** The regulatory efficiency performance indices **P.I.**-s proposed to evaluate the perturbation treatment efficiency by a generic **GERM** of (Figure4-9, or Figure 4-7) type, following the definitions of Maria [24]. Abbreviations: Min = to be minimized; Max = to be maximized. Note: k(syn) and k(decline) refers to the  $\rightarrow P \rightarrow$  overall reaction. Notations: 'n' = nominal value; 's' = stationary value; (\*) 40 see eq. (11) and [185] for the monodromy matrix **A** calculation;  $\lambda(i)$  = i-th eigenvalues of the Jacobian matrix  $J_c = (\partial h(C, k) / \partial C)_s$ , defined before eq.(9); **A** = monodromy matrix, defined by eq. (11);  $\tau(j)$  = species 'j' recovering time of its **QSS**-level, with an accepted tolerance (usually 1-5%); Nut= nutrient;  $Re(\lambda)$  = real part of '  $\lambda$  '; **AVG** = average; **STD** = standard deviation (st. dev.);  $C_j$  = species 'j' concentration; **RD** = dynamic regulatory (recovering) index (equivalent with the recovering rate after a dynamic perturbation); **QSS** = quasi-steady-state; **P** denotes the key-protein expressed in the analysed generic **GERM**; = the sensitivity of NutP(j) vs- concentration C(i) of species 'i'. Adapted after [23].

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

**Table 4-4:** The nominal (homeostatic **QSS**) *E. coli* cell conditions, and the recovering rates of a **[G(P)1]** gene expression regulatory module after a -10% impulse perturbation in the **[P]s** of 1000 nM at an arbitrary  $t=0$ . Cell initial volume of the considered *E. coli* cell, is of  $V_{cyt,0} = 1.66 \times 10^{-15}$  L. Adapted from [24] by the courtesy of CABEQ JI.

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

**Table 4-5:** Comparison of the species recovering rates for a **GERM** modelled with a **[G(P)1]** comparatively to a **[G(P)0]** model. The nominal conditions are those of **Table 4-4** for the **high ballast** cell case, but with **[G]s** = 1 nM. NG = negligible. Notations: **NutP** and **NutG** are substrates used in the synthesis of metabolites **MetP** and **MetG**, respectively. These metabolites are used for **P** and **G** synthesis, respectively. **G**= a generic gene (DNA); **P** = the protein encoded by **G**; **M** = RNA; **GP** = the inactive complex of **G** with **P**;  $C_j$  = species 'j' concentration; **cyt** = cytoplasm; 'o' = initial; 's' index refers to the **QSS**; NG = negligible. Adapted from

[175,176].

Reference value under nominal conditions of <u>Table 5-6</u>	Variation interval of the operating parameters
$c_{X,ref} = 524$ mg/(L liq.)	varied parameter: $c_X = [X] = [0.1; 0.2; 0.3; 0.5; 0.75; 1; 2] \times c_{X,ref}$
$[Hg_{in}^{2+}]_{ref} = 10$ mg/L	varied parameter: $[Hg_{in}^{2+}] = [0.1; 0.2; 0.3; 0.5; 0.75; 1; 2] \times [Hg_{in}^{2+}]_{ref}$
$d_{p,ref} = 0.09$ cm	varied parameter: $d_p = [0.1; 0.2; 0.3; 0.5; 0.75; 1; 2] \times d_{p,ref}$
$F_{L,ref} = 0.02$ L/min	varied parameter: $F_L = [0.1; 0.2; 0.3; 0.5; 0.75; 1; 2] \times F_{L,ref}$
$\varepsilon_{s,ref} = 0.79$ g/L	varied parameter: $\varepsilon_s = [0.1; 0.2; 0.3; 0.5; 0.75; 1; 2] \times \varepsilon_{s,ref}$

**Outline 5-1:** The values of the operating control variables used to simulate the performances of the **TPFB-2** bioreactor, to perform its parametric sensitivity analysis. Adapted after [63]. Remarks: uptake conversion under nominal (reference) conditions is  $x_{Hg,ref} = 0.925$ .

The dilution rates are:  $= D_L = F_L / V_L$  (for the bulk liquid), and  $D_G = F_G / V_G$  (for the sparger air).

<b>Membranar transport of <math>Hg_{env}^{2+}</math>:</b>			
$Hg_{env}^{2+} \xrightarrow[-2H^+]{+2RSH, (PT)} Hg_{cyt}^{2+}$			
<b>Rate expression:</b>	<b>Parameters:</b>		
$r_t = \frac{v_{m,t} c_{Hg_{env}^{2+}}}{K_{mt} + c_{Hg_{env}^{2+}}}$ , (membranar transport)	[Gmer], nM	$r_{max,t}$ , nM min <sup>-1</sup> mgX <sup>-1</sup>	$K_{mt}$ , nM
$v_{m,t} = r_{max,t} A_{Hg} c_X$ ;	3	8.2	3800
$r_{max,t} = k_7 c_{PT}$ (see <a href="#">Table 5-4</a> )	67	13.4	4700
	78	17.5	3600
	124	20.6	5100
	140	19.8	6000
<b>Mercury ion reduction:</b>			
$Hg(SR)_2 \xrightarrow[-2NADP^+]{+2NADPH, (PA)} Hg_{cyt}^0 + 2RSH$			
<b>Rate expression:</b>	<b>Parameters:</b>		
$r_P = \frac{v_{m,P} c_{Hg_{cyt}^{2+}}}{K_{mP} + c_{Hg_{cyt}^{2+}} + c_{Hg_{cyt}^{2+}}^2 / K_{iP}}$ ,	[Gmer], nM	$r_{max,P}$ , nM min <sup>-1</sup> mgX <sup>-1</sup>	$K_{mP}$ , nM
(reduction in cytosol)	3	45	12600
$v_{m,P} = r_{max,P} A_{Hg} c_X$ ;	67	152	15900
$r_{max,P} = k_8 c_{PA}$ (see <a href="#">Table 5-4</a> )	78	168	22800
	124	221	15500
	140	305	19500
			96700
			91500
			87400
			75200
			93100

**Table 5-1:** Un-structured (apparent, reduced) kinetic model of Michaelis-Menten type for mercury ions reduction by *E. coli* (after [244-246]). Notations: substrate  $S = Hg_{env}^{2+} = Hg_{env}^{2+}$ ; **PT** = lumped permease for the membranar transport of  $Hg_{env}^{2+}$ ; into cytosol  $Hg_{cyt}^{2+}$ ; **PA** = lumped reductase of cytosolic mercury ions  $Hg_{cyt}^{2+}$  to metallic mercury  $Hg_{cyt}^0$ ; RSH = low molecular-mass thiol redox cytosolic buffers; NADPH = nicotinamide adenine dinucleotide phosphate; subscripts: 'env' = environmental; 'cyt' = cytosol. Other notations are given in the symbols list of chap. 5.1. After [40].



<b><u>Reactor characteristics</u></b> [Deckwer et al., 2004][248]	
Liquid volume	$V_L = 1 \text{ L}$
Height / diameter ( $d_r$ )	0.2 m / 0.08 m
Volumetric fraction of particles in the fluidized-bed	$\varepsilon_s = 2 \text{ g L}^{-1}$ (0.8 [Deckwer et al., 2004]; up to $3\text{-}4 \text{ g L}^{-1}$ [Berne and Cordonnier, 1995]) [251]
<b><u>Nominal operating conditions</u></b> [Deckwer et al., 2004; Maria, 2010][1,2,248]	
Temperature; pressure; pH	37 °C ; 1 atm; pH = 7.2
Liquid inlet/outlet flow rate	$F_L = 0.02 \text{ L min}^{-1}$
Liquid inlet concentrations [ $\text{Hg}^{2+}$ , $\text{Hg}^0$ ] (polluted water)	$[c_{\text{Hg}_{env}^{2+}}, c_{\text{Hg}_{env}^0}]_{\text{in}} = [20, 0] \text{ mg L}^{-1}$
Liquid background pollution [ $\text{Hg}^{2+}$ , $\text{Hg}^0$ ]	$[c_{\text{Hg}_{env}^{2+}}, c_{\text{Hg}_{env}^0}] = [0.02, 0] \text{ mg L}^{-1}$
Gas (air) inlet/outlet flow rate	$F_G = 0.2748 \text{ moles min}^{-1}$
Gas inlet concentration of $\text{Hg}^0$	0 $\text{mg L}^{-1}$
Gas (air) superficial velocity	$u_G = 0.0233 \text{ m s}^{-1}$
<b><u>Pumice particle diffusion characteristics</u></b> [Wagner-Döbler et al., 2000; Whithan and Sparks, 1986]	
Particle specific density (dry) [249,261]	$\rho_s = 2400 \text{ kg m}^{-3}$
Particle average diameter	$d_p = 1 \text{ mm}$
Catalyst porosity / BET area	$\varepsilon_p = 0.8 / S_{\text{BET}} = 0.5 \text{ m}^2 \text{ g}^{-1}$
Apparent G-L mass transfer coefficient (on gas side)(Footnote b)	$k_G a_G = 0.617 \text{ s}^{-1}$
Apparent G-L mass transfer coefficient (on liquid side)	$k_L a_L = 0.022 \text{ s}^{-1}$ [Deckwer et al., 2004][248]
Apparent L-S mass transfer coefficient (on liquid side)	$k_s a_s = 0.0114 \text{ s}^{-1}$ [Deckwer et al., 2004]
Catalyst tortuosity (Footnote c)	$\tau_p = 1.42$
$\text{Hg}^{2+}$ effective diffusivity in support (experimental)	$D_{ef} = 4.5 \cdot 10^{-10} \text{ m}^2 \text{ s}^{-1}$ [Deckwer et al., 2004]
$\text{Hg}^{2+}$ diffusivity in air	$D_G = 0.1259 \cdot 10^{-4} \text{ m}^2 \text{ s}^{-1}$ [Lambert, 1996][262]
Henry constant for $\text{Hg}^{2+}$	$He = \exp(8.54 - 4150/T)/101.3 \text{ atm (mg/L)}^{-1}$ [Deckwer et al., 2004][248]

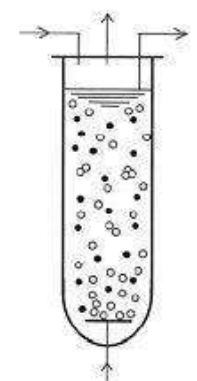
(a) fluid physical properties correspond to those of water and air at the operating conditions.

(b) see calculation rule of [Table 5-3](#).

(c) particle tortuosity was evaluated by using [263] [Shen and Chen, 2007] formula  $\tau_p^2 = \varepsilon_p / (1 - \sqrt[3]{1 - \varepsilon_p})$

**Table 5-2:** Nominal operating conditions and characteristics of the semicontinuous (SCR) three-phase fluidized bed reactor (TPFB) used for mercuric ions uptake by the immobilized *E. coli* cells on pumice porous support  
Footnote

a). Notations are given in the chap. 5.1.

Mercury mass balance equations	Observations
<p>- liquid phase: <math>\frac{dc_j}{dt} = \frac{F_L}{V_L}(c_{j,in} - c_j) + \alpha r_{app} + \omega r_{trans}</math>; species <math>j = \text{Hg}_L^{2+}, \text{Hg}_L^0</math>; <math>\alpha = \begin{cases} -1 &amp; \text{for } j = \text{Hg}_L^{2+} \\ 1 &amp; \text{for } j = \text{Hg}_L^0 \end{cases}</math>; <math>\omega = \begin{cases} 0 &amp; \text{for } j = \text{Hg}_L^{2+} \\ -1 &amp; \text{for } j = \text{Hg}_L^0 \end{cases}</math></p> <p>- gas phase: <math>\frac{dc_{j,G}}{dt} = \frac{F_G}{V_G}A_j(Y_{j,in} - Y_j) + \frac{V_L}{V_G}r_{j,trans}</math>; species <math>j = \text{Hg}_G^0</math></p> <p>- apparent <math>r_{j,app}</math> reaction rate is evaluated from equality of L-S mass transfer and surface reaction rate (at quasi-steady-state):  <math>r_{j,app} = k_s a_s (c_j - c_{j,s}) = \eta_j r_j(c_{j,s})</math></p> <p>For Michaelis-Menten rate expressions <math>r = \frac{v_m c_S}{K_m + c_S}</math>, the solution <math>r_{app}</math> results from solving the equation:  <math>r_{app}^2 - r_{app}(\eta v_m + c_{S,s} k_s a_s + K_m k_s a_s) + c_{S,s} k_s a_s \eta v_m = 0</math>,                      with the constraint: <math>0 \leq r_{app} \leq r_{max} = \frac{v_m c_{S,s}}{K_m + c_{S,s}}</math>; (usually <math>c_{S,s} \approx c_S</math>);</p> <p>In the present case, <math>r = r_T</math> (Table 5-1; substrate <math>S = \text{Hg}_L^{2+} = \text{Hg}_{env}^{2+}</math>):</p> <p>- apparent <math>r_{trans}</math> transport rate of <math>\text{Hg}_L^0</math> (indexed by 'Hg0') is evaluated from equality of G-L transfer rate at gas interface (at quasi-steady-state):  <math>r_{trans} = k_G a_G (c_{Hg0,L}^* k_H - p_{Hg0,G}) = k_L a_L (c_{Hg0,L}^* - c_{Hg0,L})</math>, that is:  <math>r_{trans} = (c_{Hg0,L} - p_{Hg0,G} / k_H) / [1 / (k_G a_G k_H) + 1 / k_L a_L]</math>.</p>	 <p>- at <math>t = 0</math>, <math>c_j = c_{j,0}</math>;                      - hypotheses: dynamic operation, isothermal, iso-pH, perfectly mixed liquid and gas phases, of constant volume; immobilized cell on porous pumice; uniform concentration of solid and biomass in the reactor (by purging).</p>
<p><b>Model parameters</b></p> <p>- biocatalyst effectiveness factors are calculated for every reaction <math>j</math> with the relationships for spherical particles [Doran, 1995][33]</p> $\eta_j = \frac{\eta_{j0} + \beta \eta_{j1}}{1 + \beta}, \quad \beta = \frac{K_m}{c_{S,s}}$ $\eta_{j0} = \begin{cases} 1 & \text{for } 0 \leq \phi_{j0} \leq 0.577 \\ I - \left[ \frac{1}{2} + \cos\left(\frac{\Psi + 4\pi}{3}\right) \right]^3 & \text{for } \phi_{j0} > 0.577 \end{cases}$ $\eta_{j1} = \frac{1}{3\phi_{j1}^2} (3\phi_{j1} \coth(3\phi_{j1}) - 1),$ <p>- <math>k_s a_s</math> is evaluated by using the relationships [Trambouze et al., 1988][254]</p> $k_s = Sh \times D_{S,L} / d_p; \quad a_s = (6 \varepsilon_s) / d_p$	<p><b>Observations</b></p> $\Psi = \arccos\left(\frac{2}{3\phi_{j0}^2} - 1\right)$ <p>Thiele moduli are:</p> $\phi_{j0} = \frac{d_p}{6\sqrt{2}} \sqrt{\frac{v_{m,j}}{D_{S,ef} c_{S,s}}}$ $\phi_{j1} = \frac{d_p}{\delta} \sqrt{\frac{v_{m,j}}{D_{S,ef} K_m}}$ $D_{S,ef} = \frac{\varepsilon_p D_{S,L}}{\tau_p}$ $\Sigma_L = g(u_L + u_G) / \varepsilon_L$ $\varepsilon_L = 1 - \varepsilon_s$ $u_L = F_L / (\pi d_T^2 / 4)$

**Table 5-3:** Three-phase fluidized bed (TPFB) reactor model [1,2,5,33,40,253,254]. Index „s” = concentration at the particle surface (considered at the particle surface (considered identical to those in the bulk-phase)).



**Table 5-4:** The structured variable cell-volume whole-cell (WCVV) model proposed for the reduction of cytosolic mercury ions  $Hg_{cyt}^{2+}$  to metallic mercury  $Hg_{cyt}^0$  in cloned *E. coli* cells. Adapted from [1,2,40]. Notations are given in the chap.5.1. **TF** = transcription factor.

Reaction	Rate expression (Footnote a)	Estimated model parameters (units in min and nM) (Footnote b)	Remarks
<b>Genome and proteome replication</b>			
$NutG \xrightarrow{P} MetG$	$k_1^c NutG^c P$	$1.5489 \times 10^{-9}$	<b>(Footnote g)</b>
$NutP \xrightarrow{P} MetP$	$k_2^c NutP^c P$	$2.3875 \times 10^{-9}$	
$MetG \xrightarrow{P} G$	$k_3^c MetG^c P$	$5.1710 \times 10^{-13}$	
$MetP \xrightarrow{G} P$	$k_4^c MetP^c G$	$3.4237 \times 10^{-7}$	
$G + P \rightarrow GP$	$k_5^c G^c P$	$\sim 10^{-2}$	
$GP \rightarrow G + P$	$k_6^c GP$	$10^5$	
<b>Mercury ion import into cytosol</b>			
$Hg_{env}^{2+} \xrightarrow[+2RSH, (PT)]{-2H^+} Hg_{cyt}^{2+}$	$\frac{k_7^c P T^c Hg_{env}^{2+}}{K_{mt} + c Hg_{env}^{2+}}$	$k_7 = 2.9052 \times 10^3$ [PT] <sub>s,o</sub> = 2823 nM $r_{max,t} = k_7^c P T$	$r_{max,t}, K_{mt}$ of <b>Table 5-1</b> [Philippidis et al., 1991a] [244]
<b>Mercury ion reduction</b>			
$Hg(SR)_2 \xrightarrow[+2NADPH, (PA)]{-2NADP^+} Hg_{cyt}^0 + 2RSH$	$\frac{k_8^c P A^c Hg_{cyt}^{2+}}{K_{mP} + c Hg_{cyt}^{2+} + c^2 Hg_{cyt}^{2+} / K_{iP}}$	$k_8 = 1.2016 \times 10^5$ [PA] <sub>s,o</sub> = 1800 nM $r_{max,P} = k_8^c P A$	$r_{max,P}, K_{mP}, K_{iP}$ of <b>Table 5-1</b> [Philippidis et al., 1991a] [244]
<b>Volatile mercury <math>Hg^0</math> removal</b>			
$Hg_{cyt}^0 \rightarrow Hg_{env}^0$	$k_9^c Hg_{cyt}^0$	$1.0105 \times 10^4$	
<b>Gene GR expression module</b>			
$MetG \xrightarrow{Hg_{cyt}^{2+}, P, PD, PR} GR$	$K_{GR} = c_{Hg}^{n_H} c_{Hg}^{2cy.ref}$	$k_{10} = 4.5821 \times 10^{-11}$ $b = 2a^4$ $a = \text{variable}$ <b>(Footnote d)</b>	$c_{Hg}^{2cy.ref} = \text{avg. experimental value}$ (Philippidis et al., 1991c) [246] $n_{PR}$ of [Voit, 2005]; [169]
	$k_{10}^c P^c c_{MetG}^c c_{PD}^{n_{PD}} c_{PR}^{n_{PR}} \times$		

**Table 5-4:** (continued).

$\frac{\left(1 + b \left(c_{Hg_{cyt}^{2+}} / c_{Hg_{2cy.ref}}\right)^{n_H}\right)}{\left(K_{GR} + c_{Hg_{cyt}^{2+}}\right)}$		$c_{Hg_{2cy.ref}} = 1866$	$n_H = 2-4$ similar to genetic switch induction of [Laurent et al., 2005; Gonze, 2010][265, 266]
$MetP \xrightarrow{GR} PR$	$k_{11} c_{MetP} c_{GR}$	$n_{PD} = 1; n_{PR} = 0.5; n_H = 2$	
		$3.3227 \times 10^{-10}$	
$GR + PRPR \rightleftharpoons GRPRPR$	$k_{12} c_{GR} c_{PRPR}$		(Footnote f)
	$k_{13} c_{GRPRPR}$	$2.5000 \times 10^4$	(Footnote c)
		$10^5$	
$2PR \rightleftharpoons PRPR$	$k_{14} c_{PR}^2$	$4.0000 \times 10^1$	(Footnote e)
	$k_{15} c_{PRPR}$	$10^5$	(Footnote c)
<b>Gene GT expression module</b>			
$MetG \xrightarrow{P, PR} GT$	$k_{16} c_{MetG} c_P c_{PR}$	$1.0725 \times 10^{-16}$	(Footnote g)
$MetP \xrightarrow{GT} PT$	$k_{17} c_{MetP} c_{GT}$	$4.8253 \times 10^{-9}$	
$GT + PTPT \rightleftharpoons GIPTPT$	$k_{18} c_{GT} c_{PTPT}$	$2.5000 \times 10^4$	(Footnote f)
	$k_{19} c_{GIPTPT}$	$10^5$	(Footnote c)
$2PT \rightleftharpoons PTPT$	$k_{20} c_{PT}^2$	$5.0210 \times 10^{-2}$	(Footnote e)
	$k_{21} c_{PTPT}$	$10^5$	(Footnote c)
<b>Gene GA expression module</b>			
$MetG \xrightarrow{P, PT} GA$	$k_{22} c_{MetG} c_P c_{PT}$	$3.7998 \times 10^{-18}$	(Footnote g)
$MetP \xrightarrow{GA} PA$	$k_{23} c_{MetP} c_{GA}$	$3.1379 \times 10^{-9}$	
$GA + PAPA \rightleftharpoons GAPAPA$	$k_{24} c_{GA} c_{PAPA}$	$2.5000 \times 10^4$	(Footnote f)
	$k_{25} c_{GAPAPA}$	$10^5$	(Footnote c)
$2PA \rightleftharpoons PAPA$	$k_{26} c_{PA}^2$	$1.2345 \times 10^{-1}$	(Footnote e)
	$k_{27} c_{PAPA}$	$10^5$	(Footnote c)
<b>Gene GD expression module</b>			

Table 5-4: (continued).

$MetG \xrightarrow{P, PA} GA$	$k_{28}c_{MetG}c_{P}c_{PA}$	$5.9584 \times 10^{-18}$	(Footnote g)
$MetP \xrightarrow{GD} PD$	$k_{29}c_{MetP}c_{GD}$	$3.3227 \times 10^{-10}$	
$GD + PDPD \rightleftharpoons GDPDPD$	$k_{30}c_{GD}c_{PDPD}$	$2.5000 \times 10^4$	(Footnote f)
	$k_{31}c_{GDPDPD}$	$10^5$	(Footnote c)
$2PD \rightleftharpoons PDPD$	$k_{32}c_{PD}^2$	40	(Footnote e)
	$k_{33}c_{PDPD}$	$10^5$	(Footnote c)

**Footnotes:**

(a) variable volume formulation of reaction rates  $(dn_j/dt)/V$ .

(b) identified for [Table 5-2](#) conditions, and  $[Gmer] = 140$  nM.

(c) adopted value of ca.  $(10^6-10^7)D_s$  [Klipp et al., 2009; Morgan et al., 2004; Maria, 2005b, 2009b][1,2,24,84,113]

(d) the following linear dependence was fitted from data:

$a = a_{\min} + (a_{\max} - a_{\min})(c_{Hg,in} - c_{Hg,in}^{\min}) / (c_{Hg,in}^{\max} - c_{Hg,in}^{\min})$ , with  $a_{\min} = 2.6$ ,  $a_{\max} = 4.1$ , and  $a = 3.0$  for  $[Gmer] > 124$  nM. For the outside of the correlation range of  $[c_{Hg,in}^{\min}, c_{Hg,in}^{\max}] = [10, 100]$  mg L<sup>-1</sup>, the value  $a = 1.5$  should be adopted.

(e) the reaction rate is of ca.  $10^4$  1/nM/min, being similar to the **TF** (monomer repressor) dimerization with a rate constant of ca.  $10^2$  1/nM/min [Salis and Kaznessis, 2005][80]

(f) the reaction rate is of ca.  $10^4$  1/nM/min, being similar to the **TF** binding to the gene operator, with a rate constant of ca.  $10^2$  1/nM/min [Salis and Kaznessis, 2005][80]

(g) the reaction rate is of ca.  $10^1-10^2$  1/nM<sup>2</sup>/min, being similar to the mRNA (genes) synthesis reactions, of rate constants of ca.  $10^4$  1/nM/min [Wang et al., 2010][144]

Table 5-4: (continued).

<b>Monoculture cell:</b>			
Cell cycle time [Trun and Gottesman, 1990] [259]		$t_c = 30$ min (wild); up to 4-6 h (mutants)	
Born cell volume [Kubitschek, 1990] [191]		$V_{cyt,o} = 0.6 \times 10^{-15}$ L	
Cell content average dilution rate		$D_s = \ln(2) / t_c = 1.3863$ 1/h	
Cell concentration in biomass		$c_{cell} = 2 \cdot 10^{12}$ cells/gX (adjustable; adopted $5 \cdot 10^{12}$ cells/gX by [Philippidis et al., 1991a]	
Biomass concentration in the bioreactor		$c_X = 1$ g/L (ca. 0.6-3 g/L: [Deckwer et al., 2004; Berne and Cordonnier, 1995][248,251]	
Cell density in the culture medium [Mackay, 2001][260]		$\rho_{cell} = 1$ kg/(L cell)	
<b>Species concentrations in nM</b> (referring to the cell volume; see the model of <a href="#">Table 5-4</a> )( <a href="#">Footnote a</a> )	<b>Reference value</b> (units in nM)	<b>Simulated homeostasis for inlet [Hg<sup>2+</sup>] = 20 mg/L</b> (units in nM)	<b>Observations</b>
Mercury in cytoplasm, [Hg <sub>cyt</sub> <sup>2+</sup> ]	avg. 1866 0-3430	212.7	[Philippidis et al., 1991c][246]
[NutG] = Lumped nutrients used for the lumped genome <b>G</b> synthesis, (ref. to the environmental volume)	$3 \cdot 10^7$	$3 \cdot 10^7$	[Morgan et al., 2004][113]
[NutP] = Lumped nutrients used for the lumped proteome <b>P</b> synthesis, (ref. to the environmental volume)	$3 \cdot 10^8$	$3 \cdot 10^8$	[Morgan et al., 2004][113]
[MetP] = Lumped metabolites used for the lumped proteome <b>P</b> synthesis	$3 \cdot 10^8$	$3 \cdot 10^8$	[Morgan et al., 2004][113]
[MetG] = Lumped metabolites used for the lumped genome <b>G</b> synthesis	$2.01 \cdot 10^7$	$2.01 \cdot 10^7$	<a href="#">Footnote (b)</a>
Lumped genome, [G] (active) = [GP](inactive)	4500/2	4500/2	<a href="#">Footnotes (c,d)</a>
Lumped proteome, [P]	$10^7$	$10^7$	<a href="#">Footnote (e)</a>
<b>Mer</b> -operon plasmid concentration [Gmer]	1-2 (wild) 3-140 (cloned)	3 (cloned)	[Philippidis et al., 1991b-c] [244-246]
Genes expressing the <b>mer</b> -proteins, [GR], [GT], [GA], [GD]	[Gmer]/3	[4.8,16.4,25.6,31.0]	[Maria, 2010] [1,2]
Catalytic inactive forms of <b>mer</b> genes, [GRPRPR], [GTPTPT], [GAPAPA], [GPDPPD]	[Gmer]/3	[0.2, 2.1, 8.2, 22.0]	[Maria, 2010]; <a href="#">Footnotes (d,f)</a>
Proteins initiating and controlling the <b>mer</b> operon expression, [PR], [PD]	100	[19.9,84.2]	Fitted from data [Maria, 2010] [1,2]
Lumped permease PT= PmerT + PmerP + PmerC	500-7000 (avg. 2800)	1021.8	<a href="#">Footnote (g)</a>

**Table 5-5:** *E. coli* cell characteristics and the considered cell key-species concentrations. After [1,2,40].

for the membranar transport of $Hg_{env}^{2+}$			
[PA] = Reductase of the cytosolic mercury ions $Hg_{cyt}^{2+}$	500-4000 (avg. 1800)	1022.0	Fitted from data; <a href="#">Footnote (h)</a>
[PXPX] with X= R, T, A, D	4		
Proteic dimers of <i>mer</i> -proteins with the role of <b>TF</b> -s in the <i>mer</i> -gene expression.	(1-100)	4	<a href="#">Footnote (i)</a>

**Footnotes:**

- (a) Inner cell concentrations are evaluated with the formula of [24] [Maria, 2005b]:  $c_j = (\text{copynumbers of species "j" per cell}) / (N_A V_{cyt})$ , where  $N_A = 6.022 \times 10^{23}$  is the Avogadro number,  $V_{cyt}$  = average volume of cell (ca.  $1.5 V_{cyt,o}$ ). Born cell volume  $V_{cyt,o} = 0.6 \times 10^{-15}$  L.
- (b) calculated from the state-law constraint for an isotonic and isothermal cell system (subscript 'env' refers to the bulk liquid phase environment):
- $$\sum_j^{all} c_{MetGj,cyt} + \sum_j^{all} c_{MetPj,cyt} + \sum_j^{all} c_{Gj,cyt} + \sum_j^{all} c_{Pj,cyt} = \sum_j^{all} c_{j,env}$$
- (c) The considered K-12 strain of *E. coli* genome includes ca. 4500 genes [EcoCyc, 2005][95]
- (d) The maximum regulatory effectiveness of expression takes place for equal active and inactive **G**-forms at steady-state [Maria, 2005b, 2006, 2007], i.e.  $[G_j]_s = [G_j TF_j]_s$ , where **TF** denotes the transcription factor adjusting the gene **G<sub>j</sub>** activity. Index "s" denotes the quasi-steady-state (QSS)[24,78,174]
- (e) *E. coli* genome includes ca. 1000 ribosomal proteins of 1000-10000 copies, ca. 3500 non-ribosomal proteins of avg. 100 copies, and ca. 4500 polypeptides of avg. 100 copies [EcoCyc, 2005][95]
- (f) As reported by Philippidis et al. [1991a], the level of *mer*-genes represents only a fraction of the inserted *mer*-plasmids of **[G<sub>mer</sub>]** concentrations. As an average, this fraction is ca. 1/3 (estimated by Maria [2010])[1,2]
- (g) lumped permease **PI** concentration is evaluated based on the experimental observations [Barkay et al., 2003], with the approximate formula (in nM):  $[PI] \approx 1.5 \{ [PA] + 15 \} + [PR]$ ; (see the details of Maria [2009b, 2010])[1,2]
- (h) At cell homeostasis, the **PA** reductase level (of average 1800 nM) is comparable with those of the permease **PI** (the Michaelis-Menten constant being 3700 nM for  $Hg_{cyt}^{2+}$  binding reaction [Barkay et al., 2003][247])
- (i) Evaluated by optimizing the *mer*-operon **GRC** holistic properties [Maria, 2009b]. See also chap. 4, and [1,2]

Table 5-5: (continued).



<b>Reactor characteristics:</b>	
Liquid volume	$V_L = 1 \text{ L}$
Height / diameter ( $d_p$ )	0.2 / 0.08 m / m
Volumetric fraction of particles in the fluidized-bed	$\epsilon_s = 0.79 \text{ g/L}$ (0.1-1, [Trambouze et al., 1988])[254]
<b>Nominal operating conditions:</b>	
Temperature; pressure; pH	30 °C; 1 atm; pH = 7
Inlet/outlet liquid flow rate	$F_L = 0.02 \text{ L/min}$
Inlet liquid concentrations [ $\text{Hg}^{2+}$ , $\text{Hg}^0$ ] (polluted water)	$[c_{\text{Hg}_{env}^{2+}}, c_{\text{Hg}_{env}^0}]_{in} = [10, 0] \text{ mg/L}$
Inlet/outlet gas (air) flow rate	$F_G = 0.2811 \text{ moles/min}$
Inlet gas concentration of $\text{Hg}^0$	0
Superficial gas (air) velocity	$u_G = 0.0233 \text{ m/s}$
<b>Alginate particle and diffusion characteristics:</b>	
Specific particle density (wet)	$\rho_s = 1035 \text{ kg/m}$ [Rassis et al., 1998][267]
Average particle diameter	$d_p = 0.9 \text{ mm}$
Biomass concentration in the reactor	$c_X = 524 \text{ mg/L}$ (250, [Philippidis et al., 1991a])[244]
Effective diffusivity of $\text{Hg}^{2+}$ in immobilizate	$D_{ef} = 4.5 \cdot 10^{-10} \text{ m}^2/\text{s}$ (experimental)
$\text{Hg}^{2+}$ diffusivity in air	$D_G = 0.1259 \cdot 10^{-4} \text{ m}^2/\text{s}$ [Lambert, 1996][262]

**Table 5-6:** The characteristics of the **TPFB** bioreactor analyzed in the chap.5.5.5. for the mercury uptake by an *Ps. putida* cell culture immobilized on porous alginate beds [63]. The nominal operating conditions correspond to those of the **TPFB** bioreactor used for mercuric ions reduction on immobilized cells by Deckwer et al.[248].

<p><b>Liquid phase:</b></p> $\frac{dc_{Hg2L}}{dt} = \frac{F_L}{V_L} (c_{Hg2L,in} - c_{Hg2L}) - r_{app}$ $\frac{dc_{Hg0L}}{dt} = \frac{F_L}{V_L} (c_{Hg0L,in} - c_{Hg0L}) + r_{app} - r_{trans}$ <p><b>Gas phase:</b></p> $\frac{dc_{HgG}}{dt} = \frac{F_{G,air}}{V_G} A_{Hg} (Y_{Hg0G,in} - Y_{Hg0G}) + \frac{V_L}{V_G} r_{trans}$	<ul style="list-style-type: none"> <li>- dynamic, isothermal, iso-pH operation;</li> <li>- perfectly mixed liquid and gas phases, of constant volume;</li> <li>- <i>Ps. putida</i> immobilized on porous carrier;</li> <li>- quasi-constant concentration of solid and biomass in the reactor;</li> <li>- quasi-constant activity of the biocatalyst</li> </ul> <p>(<math>c_{Hg2L}</math>, <math>c_{Hg0L} = [Hg^{2+}]</math> and <math>[Hg^0]</math> in the bulk liquid, respectively)</p>
<p><b>Transport rate of metallic mercury from liquid to the gas phase:</b></p> <p>- evaluated from equality of G-L transfer rate at gas interface (at quasi-steady-state):</p> $r_{trans} = k_G a_G (c_{Hg0L}^* k_H - p_{Hg0G}) = k_L a_L (c_{Hg0L} - c_{Hg0L}^*) =$ $(c_{Hg0L} - p_{Hg0G} k_H^{-1}) / [(k_G a_G k_H)^{-1} + (k_L a_L)^{-1}]$ <p>Parameter estimation:</p> <ul style="list-style-type: none"> <li>- <math>k_G = 6.6 D_G / d_b</math> (for <math>Hg^0</math>); <math>a_G = 6 \varepsilon_G / d_b</math> (spherical bubbles); <math>k_G a_G = 0.568</math> 1/s (nominal)</li> <li>- <math>d_b = 26 d_r (g d_r^2 \rho_L / \sigma)^{-0.5} (g d_r^3 \rho_L^2 / \mu_L^2)^{-0.12} (\mu_G / \sqrt{g d_r})^{-0.12}</math></li> <li>- gas holdup <math>\varepsilon_G = V_G / (V_L + V_G)</math> is averaged over three correlations [Trambouze et al., 1988][254]</li> </ul> $\frac{\varepsilon_G}{(1 + \varepsilon_G)^4} = 0.14 \left( \frac{u_G \mu_L}{\sigma} \right) \left( \frac{\rho_L \sigma^3}{g \mu_L^4} \right)^{7/24} \left( \frac{\rho_L}{\rho_L - \rho_G} \right)^{15/24} \left( \frac{\rho_L}{\rho_G} \right)^{5/72};$ $\frac{\varepsilon_G}{(1 + \varepsilon_G)^4} = 0.25 \left( \frac{u_G \mu_L}{\sigma} \right) \left( \frac{\rho_L \sigma^3}{g \mu_L^4} \right)^{7/24}; \quad \varepsilon_G = \frac{u_G}{30 + 2u_G}, \quad (u_G \text{ in cm/s})$ <ul style="list-style-type: none"> <li>- <math>k_H = p_{Hg0}^* / c_{Hg0L}^* = \exp(8.54 - 4150/T) / 101.3</math>, atm L/mg (for <math>Hg^0</math>) [Deckwer et al., 2004][248]</li> </ul>	
<p><b>G-L mass transfer on the liquid side:</b> <math>k_L a_L = 0.022</math> 1/s (experimental) [Deckwer et al., 2004][248]</p>	
<p><b>L-S mass transfer on the liquid side:</b> <math>k_s = Sh \times D_L / d_p</math>, (for <math>Hg_L^{2+}</math>) [Trambouze et al., 1988][254]</p> <p><math>a_s = (6 \varepsilon_s) / d_p</math> (spherical particles); <math>k_s a_s = 0.0114</math> 1/s (experimental) [Deckwer et al., 2004][248]</p>	

**Table 5-7:** The reduced mathematical model of the **TPFB** bioreactor of Deckwer et al. [248], analyzed in the chap.5.5.5.

<p><b>First-order kinetic model (PFOM):</b></p> <ul style="list-style-type: none"> <li>- <math>r = k \cdot c_{Hg2e} \cdot c_X</math></li> <li>- <math>r_{app}</math> is evaluated from equality of L-S mass transfer and reaction rate at quasi-steady-state:</li> </ul> $r_{app} = k_s a_s (c_{Hg2e} - c_{Hg2s}) = \eta \cdot k \cdot c_X \cdot c_{Hg2s}$ $r_{app} = \frac{c_{Hg2e}}{(k_s a_s)^{-1} + (\eta \cdot k \cdot c_X)^{-1}}$	<p><b>Parameters:</b></p> $k = v_m / K_s$ $v_m = 0.03864 \text{ mg/min/mgX}$ $K_s = 4.52 \text{ mg/L}$ $\eta = \sqrt{\frac{D_{ef}}{d_p^2 \cdot k_v}} ; k_v = k \cdot c_X$
<p><b>Michaelis-Menten (M-M) kinetic model (MMM, and PHM):</b></p> <ul style="list-style-type: none"> <li>- <math>r = \frac{v_m \cdot c_{Hg2e} \cdot c_X}{K_s + \varphi \cdot c_{Hg2e}}</math></li> <li>- <math>r_{app}</math> is evaluated from equality of L-S mass transfer and reaction rate at quasi-steady-state:</li> </ul> $r_{app} = k_s a_s (c_{Hg2e} - c_{Hg2s}) = \eta \frac{v_m \cdot c_{Hg2s}}{K_s + \varphi \cdot c_{Hg2s}}$ <p>resulting: <math>r_{app} = (-b \pm \sqrt{b^2 - 4c}) / 2</math>, where:</p> $b = -(\eta \cdot v_m + c_{Hg2e} \cdot k_s a_s + K_s \cdot k_s a_s)$ $c = \eta \cdot v_m \cdot c_{Hg2e} \cdot k_s a_s$ <p>with:</p> $0 \leq r_{app} \leq \frac{v_m \cdot c_{Hg2s}}{K_s + \varphi \cdot c_{Hg2s}}$	<p><math>\varphi = 1</math> for <b>MM1</b> kinetics of 1-st order substrate inhibition; <math>\varphi = 2</math>, for <b>MM2</b> kinetics with enhanced substrate inhibition; <math>\varphi \approx 2</math> is an approximation of the 2-nd order inhibition.</p> <p><math>\varphi = 1 + \left(\frac{c_{Hg2e}}{K_i}\right)^2</math>, for Haldane kinetics</p> <p><b>(PHM):</b>  <math>K_i = 0.62 \text{ mg/L}</math> and average <math>c_{Hg2e}</math> from simulations.</p> <ul style="list-style-type: none"> <li>- effectiveness factor (spherical particles) <b>(note a):</b></li> </ul> $\eta = \frac{\eta_0 + \beta \eta_1}{1 + \beta}, \quad \beta = \frac{K_s}{c_{Hg2s}} \approx \frac{K_s}{c_{Hg2e}}$ $\eta_0 = \begin{cases} 1 & \text{for } \theta \leq \phi_0 \leq 0.577 \\ 1 - \left[ \frac{1}{2} + \cos\left(\frac{\Psi + 4\pi}{3}\right) \right]^3 & \text{for } \phi_0 > 0.577 \end{cases}$ $\eta_1 = \frac{1}{3\phi_1^2} (3\phi_1 \coth(3\phi_1) - 1)$
<p><b>(note a)</b> <math>\Psi = \arccos\left(\frac{2}{3\phi_0^2} - 1\right)</math> ; Thiele moduli are: <math>\phi_0 = \frac{d_p}{6\sqrt{2}} \sqrt{\frac{v_m}{D_{ef} c_{Hg2s}}}</math></p> $\phi_1 = \frac{d_p}{6} \sqrt{\frac{v_m}{D_{ef} K_s}}$	

**Table 5-8:** Tested apparent kinetic models ( $r_{app}$ ) for mercury ion reduction by using immobilized *Ps. putida* cells on alginate, after [33,248]. Notation:  $c_{Hg2e} = c_{Hg2L}$  = environmental  $Hg^{2+}$  concentration;  $2 c_{Hg2s} = Hg^{2+}$  concentration at the solid surface. These  $r_{app}$  reaction rates are part of the **TPFB** bioreactor model of (Table 5-7), analyzed in the chap.5.5.5.

The FBR initial conditions		
Parameter	Nominal (initial) value	Obs.
Biomass initial concentration ( $c_{x,0}$ )(gDW/L)	0.16 Experimental data of [Chen, 2020] (Figure. 6-8-c)	with the courtesy of [Chen, 2020][268]
Batch time ( $t_f$ )	3780 min (63 h)	That is 63 h
Cell content dilution rate ( $\mu$ ), (1/min.)	$1.25 \cdot 10^{-5}$ - 0.015	Estimated 0.0017 [Chassagnole et al., 2002][43]
Feed flow-rate ( $F_L$ )	0.015 L h <sup>-1</sup>	maintained quasi-constant
Bioreactor liquid initial volume ( $V_{L,0}$ )	0.5 L (initial)	Variable, due to the continuous feeding of the <b>FBR</b> .
Glucose feeding solution concentration $c_{glc}^{feed}$	3330.5 mM	maintained constant by [Chen, 2020][268]
Initial glucose concentration in the bioreactor $c_{glc}^{ext}$ at ( $t = 0$ ).	194.53 mM Experimental data of [Chen, 2020](Fig. 6-7)	[Chen, 2020][268]
Temperature / pH	37°C / 6.8	[Chen, 2020][268]
Bioreactor capacity [ $\max(V_L)$ ], and facilities	3 L, automatic control of pH, DO, temperature	[Chen, 2020][268]
Biomass density ( $\rho_x$ )	565.5 gDW (L cytosol) <sup>-1</sup>	[Chassagnole et al., 2002][43]
Initial concentrations for the glycolytic cell species (in mM)	$c_{F6P}(t=0) = 0.6003$ $c_{FDP}(t=0) = 0.2729$ $c_{PEP}(t=0) = 2.6729$ $c_{PYR}(t=0) = 2.6706$ $c_{ATP}(t=0) = 4.27$ [AMDTP] <sub>total</sub> = 5.82	measured by [Chassagnole et al., 2002][43]
Initial concentrations for the TRP synthesis operon species (in $\mu$ M)	$c_{OR}(t=0) = 0.01$ $c_{OT}(t=0) = 3.32$ (nM) $c_{MRNA}(t=0) = 0.01$ $c_E(t=0) = 928$ (nM) $c_{TRP}(t=0) = 0.164$	measured by [Bhartiya et al., 2006][289] this paper, data of [Chen, 2020]

**Table 6-1:** The nominal initial operating conditions of the **FBR** used by Chen [268] to collect the kinetic data of the **TRP** synthesis by using a suspended culture of genetically modified *E. coli* cells (T5 strain).

Species mass balance	Auxiliary relationships, and estimated rate constants
<p>Glucose,</p> $\frac{dc_{glc}^{ext}}{dt} = \frac{FL_j}{V_L(t)} (c_{glc,j}^{feed} - c_{glc}^{ext}) - \frac{c_x(t)}{\rho_x} V_1$ <p><math>c_{glc,j}^{feed}</math> = control variables to be optimized;  <math>j = 1, \dots, N_{div}</math> (equal time-arcs);  <math>c_{glc}^{ext}(t=0)</math> is given in (Table 6-1) for the nominal FBR of [Chen, 2020][268]                      For the optimal FBR with adopted <math>N_{div} = 5</math>, the feeding policy is (Footnote a):</p> $c_{glc,j}^{feed} = \begin{cases} c_{glc,0}^{feed} & \text{if } 0 \leq t < T_1 \\ c_{glc,1}^{feed} & \text{if } T_1 \leq t < T_2 \\ c_{glc,2}^{feed} & \text{if } T_2 \leq t < T_3 \\ c_{glc,3}^{feed} & \text{if } T_3 \leq t < T_4 \\ c_{glc,4}^{feed} & \text{if } T_4 \leq t < t_f \end{cases}$ <p>Species inside the cell</p> $\frac{dc_{f6p}}{dt} = V_1 - V_2 - \mu c_{f6p}$ $\frac{dc_{fdp}}{dt} = V_2 - V_3 - \mu c_{fdp}$ $\frac{dc_{pep}}{dt} = 2V_3 - V_4 - y_{trp}(2V_3) - \mu c_{pep}$ $\frac{dc_{pyr}}{dt} = V_4 - V_5 - \mu c_{pyr}$ $\frac{dc_{atp}}{dt} = -V_1 - V_2 + 2V_3 + V_4 - V_6 - \mu c_{atp}$	<p>i) <math>c_{amp} + c_{adp} + c_{atp} = c_{amdt} = \text{constant [3,307,308]}</math>                      ii) <math>c_{adp}</math> results from solving the thermodynamic equilibrium relationship <math>c_{atp}c_{amp} = Kc_{adp}^2</math>, that is:  <math display="block">c_{adp}^2 \frac{K}{c_{atp}} + c_{adp} - c_{amdt} + c_{atp} = 0</math>                      iii) <math>\mu</math> = cell dilution rate (Table 6-1)                      iv) The initial values of cell species concentrations are given in Table 6-1 (see also Footnote (b))                      v) The lump <math>c_{tca}</math> of Figure 6-3 includes species belonging to the TCA cycle. There are no measurements on this lump, so it was excluded from data fitting.                      vi) The adopted value for <math>y_{trp}</math> by [Maria, 2021] is: [3]  <math>y_{trp} = r_{syn,trp} / r_{syn,pep} = 1/43.63</math> (at QSS)[295] Stephanopoulos and Simpson, 1997]; <math>y_{trp}</math> was re-estimated from simulating from experimental data by [Maria, 2021], resulting: <math>y_{trp} = 0.467</math>                      vii) See (Table 6-3) for the <math>V_1</math>-<math>V_6</math> flux expressions.</p>

**Table 6-2:** Mass balance of the cell glycolytic key-species, and of the FBR control variables (GLC, FL) for the optimally operated (time step-wise feeding policy) FBR. Adapted from [3,25,39,58,62].

<p>Liquid volume dynamics</p> $\frac{dV_L}{dt} = F_{L,j}; V_L(t=0) = V_{L,0} \text{ in (Table 6-1); } j = 1, \dots, N_{div} \text{ (equal time-arcs)}$	<p>viii) “For the adopted <math>N_{div} = 5</math>, the feeding policy is (see the <a href="#">Footnote a</a>):</p> $F_{L,j} = \begin{cases} F_{L,0} & \text{if } 0 \leq t < T1 \\ F_{L,1} & \text{if } T1 \leq t < T2 \\ F_{L,2} & \text{if } T2 \leq t < T3 \\ F_{L,3} & \text{if } T3 \leq t < T4 \\ F_{L,4} & \text{if } T4 \leq t < t_f \end{cases}$
<p>Biomass dynamics</p> $\frac{dc_x}{dt} = \frac{\mu_x c_{glc} c_x}{(a_x \exp(b_x t))^{N_x}};$ $c_x(t=0) = c_{x,0} \text{ in (Table 6-1);}$	<p>ix) The biomass growth inhibition corresponds to a modified Contois model [Carlsson and Zambrano, 2014]. The estimated rate constants by [Maria, 2021] are:  <math>\mu_x = 1.05 \cdot 10^{-4}</math> (1/min.mM);  <math>a_x = 10.19</math>;  <math>b_x = 1.8036 \cdot 10^{-2}</math> (1/min);  <math>N_x = 7.334 \cdot 10^{-2}</math></p>
<p><b>Footnote a:</b> For the adopted <math>N_{div} = 5</math>, the <math>j=1, \dots, N_{div}</math> time-arcs approx. switching points are: T1= 12.5 h.; T2= 25 h.; T3= 37.5 h.; T4= 50 h.; <math>t_f = 63</math> hrs”. The <math>F_{L,0} - F_{L,4}</math> time step-wise feed flow-rates are to be determined together with the other control variables (that is <math>c_{glc,j}^{feed}</math>) to ensure the <b>FBR</b> optimal operation.</p> <p><b>Footnote b:</b> The initial concentrations of cell species (F6P, FDP, PEP, PYR, ATP), and of the biomass are given in (Table 6-1).</p>	

Table 6-2: (continued).

“Reactions	Rate expressions	Estimated rate constants (units in mM, min)
<p><b>GLC import system</b></p> <p>Modification for the T5 strain</p> $\text{glc} + \text{pep} \rightarrow \text{f6p} + \text{pyr}$ $\text{pyr} + \text{atp} \rightarrow \text{pep} + \text{adp} + \text{h}$ $\text{glc} + \text{atp} \rightarrow \text{f6p} + \text{adp} + \text{h}$	$V_1 = r_{uptake} = \rho_x / c_x \cdot \frac{r_{uptake}^{max} c_{glc}^{ext}}{K_{PTS,a1} + c_{glc}^{ext}}$	$r_{uptake}^{max} = 1.1191$ (1/min) $K_{PTS,a1} = 3487.5$ (mM) $K_{PTS,a2} = 0$ $K_{PTS,a3} = 0$
$\text{f6p} + \text{atp} \rightarrow \text{fdp} + \text{adp} + \text{h}$	$V_2 = r_{PFK} = \frac{(V_1/V_{2m}) c_{f6p}^\delta}{\left( K_{2m}^\delta + K_{2m}^\delta \left[ \frac{K_R^{amp}}{K_T^{atp}} \right]^n \left( \frac{c_{atp}}{c_{amp}} \right)^n + c_{f6p}^\delta \right)}$	$\delta = 1.0437$ $n = 2$ $V_{2m} = 0.062028$ (mM/min) $K_{2m} = 6.16423$ (mM) $K_R^{amp} = 25$ $\mu\text{M}$ $K_T^{atp} = 60$ $\mu\text{M}$

Table 6-3: Reaction rate expressions V1-V6 of the hybrid model of (Table 6-2), describing the dynamics of the cellular glycolytic species according to the kinetic model of Maria [3,25], and of Chassagnole et al.[43]. In the present study, this glycolysis kinetic model was modified by replacing the PTS system (V1 flux) for the GLC uptake with those of the mutant T5 *E. Coli* strain tested by Maria and Renea [12]. The model rate constants were estimated by Maria [3] to fit the experimental data of Chen [268] presented in (Table 6-1 and Figures 6-4, 6-6, 6-7). Species abbreviations are given in the abbreviation list.

Reactions	Rate expressions	Estimated rate constants (units in mM, min)
$\text{fdp} + 2 \text{ adp} (+ 2 \text{ nad} + 2 \text{ p}) \leftrightarrow 2 \text{ pep} + 2 \text{ atp} (+ 2 \text{ nadh} + 2 \text{ h} + 2 \text{ h}_2\text{o})$	$V_3 = k_3 c_{fdp}^\alpha - k_{3p} c_{pep}^\beta$	$k_3 = 4602.3 \text{ (1/min)}$ $k_{3p} = 31.917 \text{ (1/min)}$ $\alpha = 0.05$ $\beta = 3$
$\text{pep} + \text{ adp} + \text{ h} \rightarrow \text{ pyr} + \text{ atp}$	$V_4 = v_{PK} = \frac{(V_1/V_{4m})c_{pep}^\gamma}{\left( K_{4m}^\gamma + K_{4m}^\gamma \left[ \frac{K_R^{fdp}}{K_{T,PK}^{atp}} \right]^m \left( \frac{c_{atp}}{c_{fdp}} \right)^m + c_{pep}^\gamma \right)}$	$\gamma = 1.331879$ $m = 4$ $V_{4m} = 0.1333655 \text{ (mM/min)}$ $K_{4m} = 1.146443 \text{ (mM)}$ $K_R^{fdp} = 0.2 \text{ (mM)}$ $K_{T,PK}^{atp} = 9.3 \text{ (mM)}''$

Table 6-3: (continued)

Reactions	Rate expressions	Estimated rate constants (units in mM, min)
$\rightarrow \text{ products (accoa, succ, lac, etoh, ac, ...)}$	$V_5 = \frac{k_5 c_{pyr}^{n_{consum, pyr}}}{K_{consum, pyr} + c_{pyr}}$	$k_5 = 693.3544 \text{ (1/min)}$ $K_{consum, pyr} = 395.525 \text{ (mM)}$ $n_{consum, pyr} = 2.6814$
$\rightarrow \text{ adp} + \text{ h}$	$V_6 = k_6 c_{atp}$	$k_6 = 552.38 \text{ (1/min)}$
$\text{p} \leftrightarrow \text{ atp} + \text{ amp}$	$c_{atp} c_{amp} = K c_{adp}^2$ Obs. I) [Termonia and Ross, 1981a, 1981b] indicated experimental evidence of a very fast reversible reaction catalyse <i>AKase</i> , the equilibrium being quickly reached. ii) $k_6$ constant takes values according to the micro-organism phenotype (related to the gene encoding the enzyme <i>AT</i> that catalyse this reaction). iii) $c_{amp} + c_{adp} + c_{atp} = c_{amdtp} = \text{constant}$ [Termonia and Ross, 1981a, 1981b][307,308] iv) $c_{adp}$ results from solving the thermodynamic equilibrium relationship: $c_{atp} c_{amp} = K c_{adp}^2$ , that is: $c_{adp}^2 \frac{K}{c_{atp}} + c_{adp} - c_{amdtp} + c_{atp} = 0$ . “	$K = 1$

Table 6-3: (continued)

Rate expression:	Kinetic model parameters (units in mM, μM, min)
$\frac{dc_{OR}}{dt} = k_1 c_{OT} C_1(c_{TRP}) - k_{d1} c_{OR} - \mu c_{OR}$	$k_1 = 59.062, 1/\text{min.mM};$ $k_{d1} = 0.5443, 1/\text{min};$
$\frac{dc_{MRNA}}{dt} = k_2 c_{OR} C_2(c_{TRP}) - k_{d2} c_{MRNA} - \mu c_{MRNA}$	$k_2 = 17.796, 1/\text{min};$ $k_{d2} = 14.094, 1/\text{min}$
$\frac{dc_E}{dt} = k_3 c_{MRNA} - \mu c_E$	$k_3 = 1.157, 1/\text{min}$
$C_1(c_{TRP}) = \frac{K_{i,1}^{n_H}}{K_{i,1}^{n_H} + c_{TRP}^{n_H}}; C_2(c_{TRP}) = \frac{K_{i,2}^{1.72}}{K_{i,2}^{1.72} + c_{TRP}^{1.72}}$	$K_{i,1} = 3.53, \mu\text{M}$ $n_H = 1.92$ $K_{i,2} = 0.04, \mu\text{M}$ (see <a href="#">Footnote c</a> )
$\frac{dc_{TRP}}{dt} = (c_{PEP} c_E)^g \frac{\mu_T c_{TRP}}{(a_T \exp(b_T t))^{N_T}} - \mu c_{TRP}$	$g = -0.32$ $\mu_T = 0.36365, 1/\text{min};$ $a_T = 3.9923;$ $b_T = 0.017153, 1/\text{min};$ $N_T = 0.071515$
(see <a href="#">Footnotes a-b</a> )	
<b>Footnote a:</b> The adopted modification for the TRP-synthesis inhibition replaces the C3 variable of [Bhartiya et al., 2006] model (not displayed here) with a modified Contois model, with including a power-law inhibition with TRP-growth at the denominator. [289]	
<b>Footnote b:</b> The nitrogen source in the TRP synthesis is considered in excess and included in the $\mu_T$ constant. To be connected to the glycolysis kinetic model, the PEP species dynamics, generated by the glycolysis model, was explicitly included in the TRP synthesis rate as a substrate. [Maria, 2021] [3]	
<b>Footnote c:</b> The initial concentrations of the TRP-operon species (OR, MRNA, E, TRP) are given in (Table 6-1).	

**Table 6-4:** Species mass balance in the TRP- operon expression kinetic model of Bhartiya et al.[289] were modified by Maria et al.[62] to better fit the experimental data, as followings: i) PEP (from glycolysis) is the substrate of TRP-synthesis, and the node coupling this synthesis with the glycolysis module; ii) a novel model for the TRP-synthesis inhibition was proposed and identified from experiments. The model rate constants are estimated by Maria [3] to fit the experimental data of Chen [268] (Figures 6-4, 6-6, 6-7, 6-8) collected from a FBR with using the modified *E. Coli* T5 strain, and the 'nominal' operating conditions of (Table 6-1). Species notations (TRP, OR, OT, MRNA, E) are given in the abbreviation list. QSS = quasi-steady-state."

<i>E.coli</i> strain	V1 flux (at the initial FBR conditions) (mM/min)	Total GLC consumption over the batch time(g)	TRP-production of FBR (mM/min)
Maria et al. [36] ("wild" strain)	$1.2485 \cdot 10^2$	360	0.001- 0.04 (not optimized FBR)
Maria [46] (T5 strain) (Table 6-1)	$1.2526 \cdot 10^4$	567	0.048 (nominal, not optimized FBR)
This paper (T5 strain)	$1.2526 \cdot 10^4$	532	0.06 and higher (*) (optimized FBR)

**Table 6-5:** Efficiency of the modified *E.coli* T5 strain for GLC-uptake, and for the TRP production in the tested FBR of (Table 6-1). Adapted from Maria and Renea [12].

(\*) By following the same optimal feeding policy.



abbrevi- ation	Metabolite	Formula	Charge	CAS Number
<b>13dpg</b>	3-Phospho-D-glyceroyl phosphate	$C_3H_4O_{10}P_2$	-4	38168-82-0
<b>2pg</b>	D-Glycerate 2-phosphate	$C_3H_4O_7P$	-3	None
<b>3pg</b>	3-Phospho-D-glycerate	$C_3H_4O_7P$	-3	None
<b>6pgc</b>	6-Phospho-D-gluconate	$C_6H_{10}O_{10}P$	-3	None
<b>6pgl</b>	6-phospho-D-glucono-1,5-lactone	$C_6H_9O_9P$	-2	None
<b>ac</b>	Acetate	$C_2H_3O_2$	-1	71-50-1
<b>ac[e]</b>	Acetate (extracellular)	$C_2H_3O_2$	-1	71-50-1
<b>acald</b>	Acetaldehyde	$C_2H_4O$	0	75-07-0
<b>acald[e]</b>	Acetaldehyde (extracellular)	$C_2H_4O$	0	75-07-0
<b>accoa</b>	Acetyl-CoA	$C_{23}H_{34}N_7O_{17}P_3S$	-4	72-89-9
<b>acon-C</b>	cis-Aconitate	$C_6H_3O_6$	-3	585-84-2
<b>actp</b>	Acetyl phosphate	$C_2H_3O_5P$	-2	19926-71-7
<b>adp</b>	ADP	$C_{10}H_{12}N_5O_{10}P_2$	-3	58-64-0
<b>akg</b>	2-Oxoglutarate	$C_5H_4O_5$	-2	328-50-7
<b>akg[e]</b>	2-Oxoglutarate (extracellular)	$C_5H_4O_5$	-2	328-50-7
<b>amp</b>	AMP	$C_{10}H_{12}N_5O_7P$	-2	61-19-8
<b>atp</b>	ATP	$C_{10}H_{12}N_5O_{13}P_3$	-4	56-65-5
<b>cit</b>	Citrate	$C_6H_5O_7$	-3	77-92-9
<b>co2</b>	CO2	$CO_2$	0	124-38-9
<b>co2[e]</b>	CO2 (extracellular)	$CO_2$	0	124-38-9
<b>coa</b>	Coenzyme A	$C_{21}H_{32}N_7O_{16}P_3S$	-4	85-61-0
<b>dhap</b>	Dihydroxyacetone phosphate	$C_3H_5O_6P$	-2	57-04-5
<b>e4p</b>	D-Erythrose 4-phosphate	$C_4H_7O_7P$	-2	585-18-2
<b>etoh</b>	Ethanol	$C_2H_6O$	0	64-17-5
<b>etoh[e]</b>	Ethanol (extracellular)	$C_2H_6O$	0	64-17-5
<b>f6p</b>	D-Fructose 6-phosphate	$C_6H_{11}O_9P$	-2	643-13-0
<b>fdp</b>	D-Fructose 1,6-bisphosphate	$C_6H_{10}O_{12}P_2$	-4	488-69-7
<b>for</b>	Formate	$CHO_2$	-1	64-18-6
<b>for[e]</b>	Formate (extracellular)	$CHO_2$	-1	64-18-6
<b>fru[e]</b>	D-Fructose (extracellular)	$C_6H_{12}O_6$	0	57-48-7
<b>fum</b>	Fumarate	$C_4H_2O_4$	-2	110-17-8
<b>fum[e]</b>	Fumarate (extracellular)	$C_4H_2O_4$	-2	110-17-8

**Table 7-1:** The considered metabolites in the Edwards and Palsson [4] model of the **CCM** in the wild *Escherichia coli*.

<b>g3p</b>	Glyceraldehyde 3-phosphate	$C_3H_5O_6P$	-2	142-10-9
<b>g6p</b>	D-Glucose 6-phosphate	$C_6H_{12}O_9P$	-2	56-73-5
<b>glc-D[e]</b>	D-Glucose (extracellular)	$C_6H_{12}O_6$	0	50-99-7
<b>gln-L</b>	L-Glutamine	$C_5H_{10}N_2O_3$	0	56-85-9
<b>gln-L[e]</b>	L-Glutamine (extracellular)	$C_5H_{10}N_2O_3$	0	56-85-9
<b>glu-L</b>	L-Glutamate	$C_5H_9NO_4$	-1	56-86-0
<b>glu-L[e]</b>	L-Glutamate (extracellular)	$C_5H_9NO_4$	-1	56-86-0
<b>glx</b>	Glyoxylate	$C_2HO_3$	-1	298-12-4
<b>h2o</b>	H2O	$H_2O$	0	7732-18-5
<b>h2o[e]</b>	H2O (extracellular)	$H_2O$	0	7732-18-5
<b>h</b>	H+	H	1	12408-02-5
<b>h[e]</b>	H+ (extracellular)	H	1	12408-02-5
<b>icit</b>	Isocitrate	$C_6H_7O_7$	-3	30810-51-6
<b>lac-D</b>	D-Lactate	$C_3H_5O_3$	-1	10326-41-7
<b>lac-D[e]</b>	D-Lactate (extracellular)	$C_3H_5O_3$	-1	10326-41-7
<b>mal-L</b>	L-Malate	$C_4H_5O_5$	-2	97-67-6
<b>mal-L[e]</b>	L-Malate (extracellular)	$C_4H_5O_5$	-2	97-67-6
<b>nad</b>	Nicotinamide adenine dinucleotide	$C_{21}H_{26}N_7O_{14}P_2$	-1	53-84-9
<b>nadh</b>	Nicotinamide adenine dinucleotide - reduced	$C_{21}H_{27}N_7O_{14}P_2$	-2	58-68-4
<b>nadp</b>	Nicotinamide adenine dinucleotide phosphate	$C_{21}H_{25}N_7O_{17}P_3$	-3	53-59-8
<b>nadph</b>	Nicotinamide adenine dinucleotide phosphate - reduced	$C_{21}H_{26}N_7O_{17}P_3$	-4	2646-71-1
<b>nh4</b>	Ammonium	$H_4N$	1	14798-03-9
<b>nh4[e]</b>	Ammonium (extracellular)	$H_4N$	1	14798-03-9
<b>o2</b>	O2	$O_2$	0	7782-44-7
<b>o2[e]</b>	O2 (extracellular)	$O_2$	0	7782-44-7
<b>oa</b>	Oxaloacetate	$C_4H_3O_5$	-2	328-42-7
<b>pep</b>	Phosphoenolpyruvate	$C_3H_3O_6P$	-3	138-08-9
<b>pi</b>	Phosphate	$HO_4P$	-2	14265-44-2
<b>pi[e]</b>	Phosphate (extracellular)	$HO_4P$	-2	14265-44-2
<b>pyr</b>	Pyruvate	$C_3H_3O_3$	-1	127-17-3
<b>pyr[e]</b>	Pyruvate (extracellular)	$C_3H_3O_3$	-1	127-17-3
<b>q8</b>	Ubiquinone-8	$C_{10}H_{17}O_4$	0	1339-63-5
<b>q8h2</b>	Ubiquinol-8	$C_{10}H_{19}O_4$	0	56275-39-9
<b>r5p</b>	alpha-D-Ribose 5-phosphate	$C_5H_9O_8P$	-2	4300-28-1
<b>ru5p-D</b>	D-Ribulose 5-phosphate	$C_5H_9O_8P$	-2	4151-19-3

Table 7-1: (continued)

<b>Table 7-1. (continued)</b>				
<b>s7p</b>	Sedoheptulose 7-phosphate	$C_7H_{13}O_{10}P$	-2	None
<b>succ</b>	Succinate	$C_4H_4O_4$	-2	110-15-6
<b>succ[e]</b>	Succinate (extracellular)	$C_4H_4O_4$	-2	110-15-6
<b>succoa</b>	Succinyl-CoA	$C_{25}H_{35}N_7O_{19}P_3S$	-5	604-98-8
<b>xu5p-D</b>	D-Xylulose 5-phosphate	$C_5H_9O_8P$	-2	None

Table 7-1: (continued)

No.	Reaction name	Equation	Net flux (*)	Net flux (**)	Lower limit	Upper limit
1	acetaldehyde dehydrogenase (acetylating)	[c] : acald + coa + nad <=> accoa + h + nadh	0	-1.3756·10 <sup>-11</sup>	-1000	1000
2	acetaldehyde reversible transport	acald[e] <=> acald[c]	0	-7.8444·10 <sup>-12</sup>	-1000	1000
3	acetate kinase	[c] : ac + atp <=> actp + adp	0	-6.1391·10 <sup>-12</sup>	-1000	1000
4	aconitase (half-reaction A, Citrate hydro-lyase)	[c] : cit <=> acon-C + h2o	6.0072	6.0072	-1000	1000
5	aconitase (half-reaction B, Isocitrate hydro-lyase)	[c] : acon-C + h2o <=> icit	6.0072	6.0072	-1000	1000
6	acetate reversible transport via proton symport	ac[e] + h[e] <=> ac[c] + h[c]	0	-6.1391·10 <sup>-12</sup>	-1000	1000
7	adenylate kinase	[c] : amp + atp <=> (2) adp	0	5.2637·10 <sup>-11</sup>	-1000	1000
8	2-Oxoglutarate dehydrogenase	[c] : akgl + coa + nad -> co2 + nadh + succoa	5.0644	5.0644	0	1000
9	2-oxoglutarate reversible transport via symport	akgl[e] + h[e] <=> akgl[c] + h[c]	0	-2.8422·10 <sup>-12</sup>	-1000	1000
10	alcohol dehydrogenase (ethanol)	[c] : etoh + nad <=> acald + h + nadh	0	-5.9117·10 <sup>-12</sup>	-1000	1000
11	ATP maintenance requirement	[c] : atp + h2o -> adp + h + pi	8.39	8.39	8.39	1000
12	ATP synthase (four protons for one ATP)	adp[c] + (4) h[e] + pi[e] <=> atp[c] + (3) h[c] + h2o[c]	45.514	45.514	-1000	1000
13	Biomass Objective Function (with GAMS)	[c] : (1.496) 3pg + (3.7478) accoa + (59.8100) atp + (0.3610) asp + (0.0709) f6p + (0.1290) g3p + (0.2050) g6p + (0.2557) gln-L + (4.9414) glu-L + (59.8100) h2o + (3.5470) nad + (13.0279) nadph + (1.7867) oaa + (0.5191) pep + (2.8328) pyr + (0.8977) r5p -> (59.8100) adp + (4.1182) akgl + (3.7478) coa + (59.8100) h + (3.5470) nadh + (13.0279) nadp + (59.8100) pi	0.8739	0.8739	0	1000

Table 7-2 (1 of 5): The considered reactions in the Edwards and Palsson [4] model of the CCM in the wild *Escherichia coli*. Notation: [e] = environment; [c] = cytosol (units correspond to mmol/gDW/hr; biomass formation is expressed as g-biomass/gDW/hr, or in 1/hr).

14	CO2 transporter via diffusion	co2[e] <=> co2[c]	-22.8098	-22.8098	-1000	1000
15	citrate synthase	[c] : accoa + h2o + oaa -> cit + coa + h	6.0072	6.0072	0	1000
16	cytochrome oxidase bd (ubiquinol-8: 2 protons)	(2) h[c] + (0.5) o2[c] + q8a2[c] -> (2) h[e] + h2o[c] + q8[c]	43.5989	43.5989	0	1000
17	D-lactate transport via proton symport	h[e] + lac-D[e] <=> h[c] + lac-D[c]	0	-5.0022·10 <sup>-12</sup>	-1000	1000
18	enolase	[c] : 2pg <=> h2o + pep	14.7161	14.7161	-1000	1000
19	ethanol reversible transport via proton symport	etoh[e] + h[e] <=> etoh[c] + h[c]	0	-5.9117·10 <sup>-12</sup>	-1000	1000
20	Acetate exchange	[c] : ac <=> ac[e]	0	6.1662·10 <sup>-12</sup>	0	1000
21	Acetaldehyde exchange	[e] : acald <=> acald[e]	0	7.8028·10 <sup>-12</sup>	0	1000
22	2-Oxoglutarate exchange	[c] : akgl <=> akgl[e]	0	2.9574·10 <sup>-12</sup>	0	1000
23	CO2 exchange	[e] : co2 <=> co2[e]	22.8098	22.8098	-1000	1000
24	Ethanol exchange	[e] : etoh <=> etoh[e]	0	5.9292·10 <sup>-12</sup>	0	1000
25	Formate exchange	[e] : fcr <=> fcr[e]	0	6.0134·10 <sup>-11</sup>	0	1000
26	D-Fructose exchange	[e] : fru <=> fru[e]	0	3.0363·10 <sup>-10</sup>	0	1000
27	Fumarate exchange	[e] : fun <=> fun[e]	0	3.1576·10 <sup>-10</sup>	0	1000
28	D-Glucose exchange	[e] : glc-D <=> glc-D[e]	-10	-10	-10	1000
29	L-Glutamine exchange	[e] : gln-L <=> gln-L[e]	0	3.0640·10 <sup>-10</sup>	0	1000
30	L-Glutamate exchange	[e] : glu-L <=> glu-L[e]	0	2.8755·10 <sup>-10</sup>	0	1000
31	H+ exchange	[e] : h <=> h[e]	17.5309	17.5309	-1000	1000
32	H2O exchange	[e] : h2o <=> h2o[e]	29.1758	29.1758	-1000	1000
33	D-Lactate exchange	[e] : lac-D <=> lac-D[e]	0	5.0309·10 <sup>-12</sup>	0	1000
34	L-Malate exchange	[e] : mal-L <=> mal-L[e]	0	3.1553·10 <sup>-10</sup>	0	1000
35	Ammonium exchange	[e] : nh4 <=> nh4[e]	-4.7653	-4.7653	-1000	1000

Table 7-2 (2 of 5): (continued)

36	O <sub>2</sub> exchange	[e] : o <sub>2</sub> <=> o <sub>2</sub> [e]	-21.7995	-21.7995	-1000	1000
37	Phosphate exchange	[e] : pi <=> pi[e]	-3.2149	-3.2149	-1000	1000
38	Pyruvate exchange	[e] : pyr <=> pyr[e]	0	5.0478·10 <sup>-12</sup>	0	1000
39	Succinate exchange	[e] : succ <=> succ[e]	0	4.6482·10 <sup>-13</sup>	0	1000
40	fructose-bisphosphate aldolase	[c] : fbp <=> dhap + g3p	7.4774	7.4774	-1000	1000
41	fructose-bisphosphatase	[c] : fbp + h <sub>2</sub> o -> f6p + pi	0	5.6732·10 <sup>-11</sup>	0	1000
42	formate transport via proton symport (uptake only)	for[e] + h[e] -> for[c] + h[c]	0	3.7922·10 <sup>-10</sup>	0	1000
43	formate transport via diffusion	for[c] -> for[e]	0	4.3935·10 <sup>-10</sup>	0	1000
44	fumarate reductase	[c] : fum + q8h2 -> q8 + succ	0	497.47	0	1000
45	Fructose transport via PEP.Pyr PTS (f6p generating)	frm[e] + pep[c] -> f6p[c] + pyr[c]	0	3.1681·10 <sup>-10</sup>	0	1000
46	fumarase	[c] : fum + h <sub>2</sub> o <=> mal-L	5.0644	-5.0644	-1000	1000
47	Fumarate transport via proton symport (2 H)	fum[e] + (2) h[e] -> fum[c] + (2) h[c]	0	3.0476·10 <sup>-16</sup>	0	1000
48	glucose 6-phosphate dehydrogenase	[c] : g6p + nadp <=> 6pgl + h + nadph	4.96	4.96	-1000	1000
49	glyceraldehyde 3-phosphate dehydrogenase	[c] : g3p + nad + pi <=> 13dpg + h + nadh	16.0235	16.0235	-1000	1000
50	D-glucose transport via PEP.Pyr PTS	glc-D[e] + pep[c] -> g6p[c] + pyr[c]	10	10	0	1000
51	glutamine synthetase	[c] : atp + glu-L + nh4 -> adp + gln-L + h + pi	0.2235	0.2235	0	1000
52	L-glutamine transport via ABC system	atp[c] + gln-L[e] + h <sub>2</sub> o[c] -> adp[c] + gln-L[c] + h[c] + pi[c]	0	3.1609·10 <sup>-16</sup>	0	1000
53	glutamate dehydrogenase (NADP)	[c] : glu-L + h <sub>2</sub> o + nadp <=> akgl + h + nadph + nh4	-4.5419	-4.5419	-1000	1000
54	glutaminase	[c] : gln-L + h <sub>2</sub> o -> glu-L + nh4	0	4.9750·10 <sup>-12</sup>	0	1000
55	glutamate synthase (NADPH)	[c] : akgl + gln-L + h + nadph -> (2) glu-L + nadp	0	5.5288·10 <sup>-12</sup>	0	1000
56	L-glutamate transport via proton symport, reversible (periplasm)	glu-L[e] + h[e] <=> glu-L[c] + h[c]	0	-2.9559·10 <sup>-12</sup>	-1000	1000

Table 7-2 (3 of 5): (continued)

57	phosphogluconate dehydrogenase	[c] : 6pgc + nadp -> co2 + nadph + ru5p-D	4.96	4.96	0	1000
58	H <sub>2</sub> O transport via diffusion	h <sub>2</sub> o[e] <=> h <sub>2</sub> o[c]	-29.1758	-29.1758	-1000	1000
59	isocitrate dehydrogenase (NADP)	[c] : icit + nadp <=> akgl + co2 + nadph	6.0072	6.0072	-1000	1000
60	Isocitrate lyase	[c] : icit -> glx + succ	0	1.8071·10 <sup>-10</sup>	0	1000
61	D-lactate dehydrogenase	[c] : lac-D + nad <=> h + nadh + pyr	0	-5.1159·10 <sup>-12</sup>	-1000	1000
62	malate synthase	[c] : accoa + glx + h <sub>2</sub> o -> coa + h + mal-L	0	1.8071·10 <sup>-10</sup>	0	1000
63	Malate transport via proton symport (2 H)	(2) h[e] + mal-L[e] -> (2) h[c] + mal-L[c]	0	3.0499·10 <sup>-16</sup>	0	1000
64	malate dehydrogenase	[c] : mal-L + nad <=> h + nadh + coa	5.0644	5.0644	-1000	1000
65	malic enzyme (NAD)	[c] : mal-L + nad -> co2 + nadh + pyr	0	4.0902·10 <sup>-11</sup>	0	1000
66	malic enzyme (NADP)	[c] : mal-L + nadp -> co2 + nadph + pyr	0	5.8181·10 <sup>-11</sup>	0	1000
67	NADH dehydrogenase (ubiquinone-8 & 3 protons)	(4) h[c] + nadh[c] + q8[c] -> (3) h[c] + nad[c] + q8h2[c]	38.5346	38.5346	0	1000
68	NAD transhydrogenase	[c] : nad + nadph -> nadh + nadp	0	6.2334·10 <sup>-11</sup>	0	1000
69	ammonia reversible transport	nh4[e] <=> nh4[c]	4.7653	4.7653	-1000	1000
70	o <sub>2</sub> transport via diffusion	o <sub>2</sub> [e] <=> o <sub>2</sub> [c]	21.7995	21.7995	-1000	1000
71	pyruvate dehydrogenase	[c] : coa + nad + pyr -> accoa + co2 + nadh	9.2825	9.2825	0	1000
72	phosphofruktokinase	[c] : atp + f6p -> adp + f6p + h	7.4774	7.4774	0	1000
73	pyruvate formate lyase	[c] : coa + pyr -> accoa + for	0	6.0134·10 <sup>-11</sup>	0	1000
74	glucose-6-phosphate isomerase	[c] : g6p <=> f6p	4.8609	4.8609	-1000	1000
75	phosphoglycerate kinase	[c] : 3pg + atp <=> 13dpg + adp	-16.0235	-16.0235	-1000	1000
76	6-phosphogluconolactonase	[c] : 6pgl + h <sub>2</sub> o -> 6pgc + h	4.96	4.96	0	1000
77	phosphoglycerate mutase	[c] : 2pg <=> 3pg	-14.7161	-14.7161	-1000	1000
78	phosphate reversible transport via proton symport	h[e] + pi[e] <=> h[c] + pi[c]	3.2149	3.2149	-1000	1000

Table 7-2 (4 of 5): (continued)

79	phosphoenolpyruvate carboxylase	[c] : $co_2 + h_2o + pep \rightarrow h + oaa + pi$	2.5043	2.5043	0	1000
80	phosphoenolpyruvate carboxykinase	[c] : $atp + oaa \rightarrow adp + co_2 + pep$	0	$2.8585 \cdot 10^{-11}$	0	1000
81	phosphoenolpyruvate synthase	[c] : $atp + h_2o + pyr \rightarrow amp + (2) h + pep + pi$	0	$5.2692 \cdot 10^{-11}$	0	1000
82	phosphotransacetylase	[c] : $accoa + pi \rightleftharpoons actp + coa$	0	$6.2328 \cdot 10^{-12}$	-1000	1000
83	pyruvate kinase	[c] : $adp + h + pep \rightarrow atp + pyr$	1.7582	1.7582	0	1000
84	pyruvate reversible transport via proton symport	$h[e] + pyr[e] \rightleftharpoons h[c] + pyr[c]$	0	$-5.0022 \cdot 10^{-12}$	-1000	1000
85	ribulose 5-phosphate 3-epimerase	[c] : $ru5p-D \rightleftharpoons xu5p-D$	2.6785	2.6785	-1000	1000
86	ribose-5-phosphate isomerase	[c] : $r5p \rightleftharpoons ru5p-D$	-2.2815	-2.2815	-1000	1000
87	succinate transport via proton symport (2 H)	$(2) h[e] + succ[e] \rightarrow (2) h[c] + succ[c]$	0	$1.1830 \cdot 10^{-10}$	0	1000
88	succinate transport out via proton antiport	$h[e] + succ[c] \rightarrow h[c] + succ[e]$	0	$1.2295 \cdot 10^{-10}$	0	1000
89	succinate dehydrogenase (irreversible)	[c] : $q8 + succ \rightarrow fum + q8h_2$	5.0644	5.0644	0	1000
90	succinyl-CoA synthetase (ADP-forming)	[c] : $atp + coa + succ \rightleftharpoons adp + pi + succoa$	-5.0644	-5.0644	-1000	1000
91	transaldolase	[c] : $g3p + s7p \rightleftharpoons e4p + f6p$	1.497	1.497	-1000	1000
92	NAD(P) transhydrogenase	$(2) h[e] + nadh[c] + nadp[c] \rightarrow (2) h[c] + nad[c] + nadph[c]$	0	$5.7674 \cdot 10^{-10}$	0	1000
93	transketolase	[c] : $r5p + xu5p-D \rightleftharpoons g3p + s7p$	1.497	1.497	-1000	1000
94	transketolase	[c] : $e4p + xu5p-D \rightleftharpoons f6p + g3p$	1.1815	1.1815	-1000	1000
95	triose-phosphate isomerase	[c] : $dhap \rightleftharpoons g3p$	7.4774	7.4774	-1000	1000

Table 7-2 (5 of 5): (continued)

**Footnotes:** (\*) Fluxes correspond to equilibrated stationary growth with a glucose uptake rate of  $-10$  mmol/gDW/hr, and anoxygen uptake rate of  $-1000$  mmol/gDW/hr. (\*\*) Flux values are obtained by means of LP procedure from solving the singlelevel optimization problem of biomass production maximization in *Escherichia coli* cells under the basic CONSTR of eq. (5) (not the global optimum).

Knockout genes	$v_{13}$ (biomass rate, hr <sup>-1</sup> )	$v_{39}$ (succinate rate, mmol gDW <sup>-1</sup> hr <sup>-1</sup> )	$v_{39}/v_{50}$ (Note b)	Lagrange function (L)	$\overline{CV}_{eq}$ index	$v_{39} >$ $v_{39}^{min}$
<i>two genes removed under basic constraints CONSTR</i>						
61, 73	2.8477·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	4.88·10 <sup>-7</sup>	9.71·10 <sup>-6</sup>	
51, 82	2.8566·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	1.71·10 <sup>-7</sup>	5.75·10 <sup>-6</sup>	
45, 51	2.6734·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	2.39·10 <sup>-7</sup>	6.79·10 <sup>-6</sup>	
45, 55	2.7374·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	5.80·10 <sup>-7</sup>	1.05·10 <sup>-5</sup>	
10, 94	2.7726·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	3.92·10 <sup>-7</sup>	8.70·10 <sup>-6</sup>	
9, 61	3.0328·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	6.01·10 <sup>-7</sup>	1.07·10 <sup>-5</sup>	
1, 84	2.7496·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	2.94·10 <sup>-7</sup>	7.53·10 <sup>-6</sup>	
51, 82	2.9069·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	1.19·10 <sup>-5</sup>	4.79·10 <sup>-5</sup>	
51, 54	2.9085·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	1.55·10 <sup>-7</sup>	5.47·10 <sup>-6</sup>	
1, 78	2.9555·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	3.16·10 <sup>-6</sup>	2.47·10 <sup>-5</sup>	
1, 3	3.2408·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	1.68·10 <sup>-7</sup>	5.70·10 <sup>-6</sup>	
(82, 83), Etc.	2.7578·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	2.60·10 <sup>-7</sup>	7.08·10 <sup>-6</sup>	
<i>four genes removed under basic constraints CONSTR</i>						
1, 10, 61, 73	2.9187·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	5.69·10 <sup>-7</sup>	1.04·10 <sup>-5</sup>	
61, 73, 51, 54	2.8809·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	7.35·10 <sup>-7</sup>	1.19·10 <sup>-5</sup>	
25, 45, 55, 78	2.9001·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	2.10·10 <sup>-5</sup>	6.37·10 <sup>-5</sup>	
1, 30, 51, 94	2.8650·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	1.46·10 <sup>-5</sup>	5.31·10 <sup>-5</sup>	
43, 1, 78, 41	2.9872·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	2.12·10 <sup>-7</sup>	6.40·10 <sup>-6</sup>	
85, 73, 41, 1	3.1386·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	3.51·10 <sup>-7</sup>	8.23·10 <sup>-6</sup>	
(78, 42, 59, 76), Etc.	3.0323·10 <sup>-12</sup>	1.6384·10 <sup>1</sup>	1.6384	5.03·10 <sup>-7</sup>	9.85·10 <sup>-6</sup>	
<i>two genes removed by three more additional constraints <sup>(a)</sup></i>						
61, 73	2.3229·10 <sup>-12</sup>	1.6514·10 <sup>1</sup>	1.6514	4.84·10 <sup>2</sup>	3.05·10 <sup>-1</sup>	Yes
51, 82	2.4282·10 <sup>-12</sup>	1.6498·10 <sup>1</sup>	1.6498	6.28·10 <sup>2</sup>	3.48·10 <sup>-1</sup>	Yes
1, 10	2.4588·10 <sup>-12</sup>	1.6497·10 <sup>1</sup>	1.6497	4.77·10 <sup>2</sup>	3.03·10 <sup>-1</sup>	Yes
60, 45	4.7780·10 <sup>-12</sup>	1.4549·10 <sup>1</sup>	1.4549	8.40·10 <sup>2</sup>	4.02·10 <sup>-1</sup>	Yes
91, 22	2.9067·10 <sup>-12</sup>	1.2613·10 <sup>1</sup>	1.2613	4.16·10 <sup>2</sup>	2.83·10 <sup>-1</sup>	Yes
61, 73	8.7392·10 <sup>-1</sup>	4.6207·10 <sup>-12</sup>	~ 0	1.85·10 <sup>-6</sup>	1.89·10 <sup>-5</sup>	No

**Table 7-3:** Various local solutions of the MINLP problem for simultaneous succinate and biomass production maximization in mutant *E. coli* cells by removing two or four genes from the CCM of the cell wild strain

of Edwards and Palsson [4] model. The  $\overline{CV}_{ineq}^2$  inequality constraint violation index eq.(7) is zero for all below solutions (that is, the

all inequality constraints are fulfilled). The  $\overline{CV}_{eq}$  index is given by eq.(7).

















## Case study no. 2: the use of a hybrid modular CCM cell-scale structured kinetic model coupled with a FBR classical dynamic model (including macro-scale state variables) to maximize the TRP production

### 1.1. Symbols used in the chap. 6

$c_i, [i]$  - Species 'i' concentration

$c_x^{feed}$  - Biomass concentration

$c_{GLC}^{feed}$  - Glucose feeding solution concentration

$c_{glc,0}^{ext} = c_{GLC}^{ext}(t=0)$  - Initial glucose concentration in the bioreactor

$c_{glc,j}^{feed}$  - Glucose feeding solution concentration over the time-arc 'j'

$c_{GLC}^{ext}$  - Glucose concentration in the bulk-phase

FL liquid feed flow rate in the bioreactor

$K, k_j, K_j, K, n, V_{2M}, V_{4m},$

$r_j^{max}, a_x, b_x, N_x,$

$r_{uptake}^{max},$

$K_{PTS,a1}, K_{PTS,a2}, K_{PTS,a3},$  - Reaction rates, and/or equilibrium constants of the kinetic model

$V_{2m}, g, K_R^{amp}, K_T^{atp}, \mu T,$

$a_T, b_T, N_T$  etc

$r_i$  - Species (i) reaction rate

$t, tf$  - Time, batch time

$V_1 - V_6$  - metabolic fluxes in the glycolysis (Tables 6-2, 6-3,

and Figure 6-3)

$V_L$  -liquid volume in the bioreactor

$y_{ica}, y_{trp}$  -stoichiometric coefficients

$\alpha, \beta, \gamma, \delta$  -Reaction rate constants

$\mu$  -Cell content dilution rate, that is  $\ln(2) / tc$ , where  $tc$

denotes the cell cycle

$\Omega$  - FBR optimization objective function

$\rho_x$  - biomass density

### **Subscripts**

0,o - Initial

cell - Referring to the (inside) cell

ext - External to cell (i.e. in the bulk phase)

f - final

inlet - In the feed

x - Biomass

### **Abbreviations**

13dpg, pgp - 1,3-diphosphoglycerate

3pg - 3-phosphoglycerate

2pg - 2-phosphoglycerate

AA - amino-acids

Accoa, acetyl-CoA - acetyl-coenzyme A

AC - acetate

ADP, adp - adenosin-diphosphate

*AK-ase* - adenylate kinase

ALE - adaptive laboratory evolution

AMP, amp - adenosin-monophosphate

ATP, atp - adenosin-triphosphate

*ATP-ase* - ATP monophosphatase

BR - Batch reactor

CCM - central carbon metabolism

CIT - citrate

CSTR - continuously stirred tank reactor

DO - dissolved oxygen

DW - dry mass

**E** -

enzyme anthranilate synthase in the TRP synthesis model.

**ETOH** - ethanol

**ext** - External to the cell (i.e. in the bulk phase)

**FBR** - Fed-batch bioreactor

FDP, fdp - fructose-1,6-biphosphate

F6P, f6p - fructose-6-phosphate

GalP/Glk - galactose permease/glucokinase

G3P, g3p, GAP,

gap, 3PG, 3pg

- glyceraldehyde-3-phosphate

2PG, 2pg - 2-phosphoglycerate

G6P, g6p - glucose-6-phosphate

GLC, glc - glucose

Glc(ex), [GLC]ext - Glucose in the environment (bulk phase)

**GMO** - genetic modified micro-organisms

**GRC** - genetic regulatory circuits

*HK-ase* - hexokinase

HSMDM - hybrid structured modular dynamic (kinetic) models

JWS - Silicon Cell project of Olivier and Snoep [49]

LAC, lac - lactate

Max (x) - Maxim of (x)

MGM -The reduced kinetic model of glycolysis developd by Maria [25]

MMA - the adaptive random optimization algorithm of Maria [71,72]

MRNA, mRNA - tryptophan messenger ribonucleic acid during its encoding gene dynamic transcription, and translation;

NAD(P)H - nicotinamide adenine dinucleotide (phosphate) reduced

NLP - Nonlinear programming

ODE - ordinary differential equations set

OR - the complex between **O** and **R** (aporepressor of the **TRP** gene)

OT - the total **TRP** operon

P, Pi - Phosphoric acid

PEP, pep - phosphoenolpyruvate

13DPG=PGP - 1,3-diphosphoglycerate

*PFK-ase* - phosphofructokinase

PK-ase - pyruvate kinase

PTS - phosphotransferase, or the phosphoenolpyruvate glucose phosphotransferase system

PYR, pyr - Pyruvate

QSS - quasi-steady-state

**R5P** - ribose 5-phosphate

mRNA - messenger ribonucleic acid



SUCC, suc - succinate

TCA, tca - tricarboxylic acid cycle

TF - gene expression transcription factors

TRP, Trp, trp - tryptophan

X - biomass

Wt. - Weight

[.] - Concentration

## Bioprocess generalities

"An exceptional example of multiple applications of extended structured **HSMDM** dynamic models is offered by Maria [3,12]. As exemplified in the chap. 4 of this work, hybrid kinetic models, linking structured cell metabolic processes to the dynamics of macroscopic variables of the bioreactor, are more and more used in the engineering evaluations to derive more precise predictions of the process dynamics under variable operating conditions. Depending on the cell model complexity, such a math tool can be used to evaluate the metabolic fluxes in relationships to the bioreactor operating conditions, thus suggesting ways to genetically modify the micro-organism for certain purposes. Even if development of such an extended dynamic model requires more experimental and computational efforts, its use is advantageous.

The approached probative example of this chapter refers to a **HSMDM** model able to simulate the dynamics of nano-scale variables from several pathways of the central carbon metabolism (**CCM**) of *E. coli* cells, linked to the macroscopic state variables of a fed-batch bioreactor (**FBR**) used for the tryptophan (**TRP**) production. The used *E. coli* strain was modified to replace the **PTS**-system for glucose (**GLC**) uptake with a more efficient one. The study presents multiple elements of novelty: **(i)** the experimentally validated modular model itself, and **(ii)** its efficiency to *in-silico* derive an optimal operation policy of the **FBR**, with a higher accuracy compared to the classical empirical (heuristically) optimization rules using apparent unstructured kinetic models of the bioprocess [11,268,269].

Over the last decades, there is a continuous trend to develop more and more effective bioreactors [253,269] to industrialize important biosyntheses for producing fine-chemicals used in the food, pharmaceutical, or detergent industry, by using free-suspended or immobilized cell cultures in suitable bioreactors, as reviewed by Maria [11]. The batch (**BR**), semi-batch (fed-batch, **FBR**), a serial sequence of **BR**-s [8], or the continuously operated fixed-bed, or three-phase fluidized-bed bioreactors (with immobilized biocatalyst), etc., are successfully used to conduct biosyntheses aiming to replace complex chemical processes,

energetically intensive and generating toxic wastes [3,5,11,58,270].

Among applications, it is to mention fermentative processes for production of organic acids, alcohols, vinegar, amino-acids (TRP, cysteine, lysine, etc.), proteins, yeast, hydrogen, food products, and food additives, recombinant proteins (monoclonal antibodies), etc., by using bioreactors with microbial (cell) cultures, or enzymatic reactors [269,270], by integrating genetic and engineering methods

[240,271].

Bioreactors with microbial / animal cell cultures (suspended or immobilized) have been developed in simple or complex constructive / operating alternatives as reviewed by [11,272,273], to mention only few of such review works. In spite of their larger volumes, the continuously mixing aerated tank reactors (**CSTR**), operated in **BR**, **FBR**, or continuous modes, are preferred for bioprocesses requiring a high oxygen transfer, and a rigorous temperature/pH control. This is why, an effective **FBR** was used in the approached case study of **TRP** production, as being more flexible like operating regime alternatives.

From the engineering point of view, in addition to the production capacity optimization, there are several important problems to be addressed, that is: **(1)** The key-point in screening among bioreactor alternatives and operating modes. The answer to this problem is related to the maintenance of the bioprocess optimal

conditions that ensure a high biomass activity (free or immobilized on a suitable porous support), by supporting its growth

to compensate its natural biodegradation, and the risk to disintegrate the flocks or the support through mechanical shearing induced by the mixing, thus leading to the biomass leakage and washout. **(2)** Development of optimal operating policies of the adopted/given bioreactor, based on an available process dynamic (kinetic) model (extended or reduced) derived from on-/off-line measurements. The model-based optimal operation of the bioreactor can be applied in two ways: **(2a)** 'off-line', in which an optimal operating policy is determined by using an adequate kinetic model (usually a deterministic one, based on the process mechanism), previously identified from separate experiments; in this alternative, extended/complex dynamic models of the bioprocess/bioreactor can be used, not being restricted by the 'real-time' application, and **(2b)** 'on-line', with using an extremely simplified dynamic model (an apparent-empirical one, often of a simple polynomial form) of the bioprocess/bioreactor, and a classic state-parameter estimator, based on the online recorded data. Of course, the (2b) alternative, even if simple and with a 'realtime' application is very approximate, being often inadequate, and requires frequent (during bioreactor operation) empirical model updating [13,14,18,274- 276].

The current (default) approach to solve the model-based design, optimization and control problems of the industrial biological reactors is the use of *unstructured* models of Monod type (for cell culture reactors) or of Michaelis- Menten type (if only enzymatic reactions are retained) that ignores detailed representations of cell processes [23,31,32,34,241,269]. See an example in chap. 5.5.5. The applied engineering rules are similar to those used for chemical processes and inspired from the nonlinear system control theory [13,15,31-38,277,278]. However, by accounting for only key process variables (biomass, substrate and product concentrations), these models do not properly reflect the metabolic changes, being unsuitable to accurately predict the cell response to environmental perturbations by means of (self-) regulated cell metabolism.

The alternative is to use *structured* kinetic models, by accounting for cell metabolic reactions and component dynamics. Such deterministic models lead to a considerable improvement in the predictive power, with the expense of incorporating a larger number of species mass balances including parameters (rate constants) difficult to be estimated from often incomplete data and, consequently, difficult to be used for industrial scale purposes [1,2,21-24,40,41,63].

An alternative compromise, also tested in this chapter, and in the previous chap. 5, is to use **hybrid dynamic models** that combine unstructured with structured process characteristics, linked to the macroscopic state variables of the bioreactor dynamic model, to generate more precise predictions [21-23,40,63,279,280]. The idea of hybrid kinetic models is to inter-connect groups of process variables belonging to at least two hierarchical levels of model details. The resulted *composite (hybrid)* model is able to simulate the bioreactor dynamics simultaneously at various levels of detail. Thus, the dynamics of the bioreactor *macroscopic* state variables (i.e. species present in the liquid bulk) is simulated concomitantly to the *nanoscopic* variables describing the cell metabolic processes of interest, because the macro-/nano-scale variables are closely linked, as long as some cell metabolites are imported/excreted from/in the bioreactor bulk. Even if such a complex / extended dynamic model, including some complex cell metabolic pathways requires more experimental and computational efforts to be built-up and identified from structured kinetic data, the resulted *hybrid* (bi-level, macroscopic and nanoscopic) dynamic model presents major and remarkable advantages, as listed (a-j) and discussed in the chap.4.1.

In fact, such a hybrid structured cell dynamic model must include only the essential parts of the central carbon metabolism (CCM) (**Figure 4-1, Figure 4-23, Figure 4-37, Figure 4-38, Figure 4-42, Figure 4-43, and Figure 4-46**) by incorporating the pathway responsible for the target metabolite synthesis, and the lumped modules of the cell core, that is: the **glycolysis**, the **GLC** uptake system (i.e. the phosphotransferase (PTS), or an equivalent system), the **ATP**-recovery system, the pentose-phosphate pathway (**PPP**), the tricarboxylic acid cycle (**TCA**), and other metabolic pathway modules (if necessary in simulations). See for instance [39,42,58,68].

A special interest was given to the accurate modelling of the **glycolysis** dynamics and its self-regulation [25,39,59,61] as long as most of the glycolysis intermediates are starting nodes for the internal production of lot of cell metabolites (e.g. amino-acids, SUCC, CIT; TRP) [3,11,42,58,62].

This need to have good quality structured cell models to simulate the dynamics (and regulation) of the bacteria CCM became a subject of very high interest over the last decades, allowing *in-silico* design of **GMO-s** with desirable characteristics of various applications [21-24].

As a result, an impressive large number of valuable *structured deterministic* models (based on a mechanistic description of

the metabolic enzymatic reactions taking place among individual or lumped species) have been proposed in the literature to simulate the cell **CCM** dynamics, with including tenths-to-hundreds of key species. Here, it is worth mentioning the *E. coli* model of Edwards and Palsson [4] used by [25,42,43,44,45,46,47] for various purposes, or the *S. cerevisiae* glycolysis model of Teusink et al. [48], or the **JWS** platform of Olivier and Snoep [49], and the **MPS** platform of Seressiotis and Bailey [50] to simulate

cell metabolism (dynamics and/or fluxes), to mention only few of them. Simulation platforms, such as E-cell [51,52,], or V-cell [53], accounting for thousands of species and reactions, display extended capabilities to predict the dynamics of the cell metabolism under various conditions, based on EcoCyc, KEGG, Prodigic, Brenda and other bio-omics databanks (review of Maria [22,23]). A worthwhile **CCM**-based dynamic or stationary models were reported by Maria et al. [3,23,25,42] and schematically represented in (**Figure 4-1**, **Figure 4-23**, **Figure 4-37**, **Figure 4-38**, **Figure 4-42**, **Figure 4-43**, and **Figure 4-46**). Deterministic kinetic models using continuous variables has been developed by Maria [25] for the glycolysis, and by [43,54-57] for the **CCM**. Such models can adequately reproduce the cell response to continuous perturbations, the cell model structure and size being adapted based on the available -omics information.

In spite of such tremendous modelling difficulties, the development of *structured reduced deterministic* (rather than stochastic) models [21-23] able to adequately reproduce the dynamics of some **CCM** complex metabolic syntheses [20,25,26], but also the dynamics of the genetic regulatory circuits (**GRC**-s) [21] tightly controlling the metabolic processes reported significant progresses over the last decades [27,28]. Even if they are rather based on sparse information from various sources, unconventional statistical identification, and lumping algorithms [21-23,29,67,64], such *structured reduced deterministic* kinetic models have been

proved to be extremely useful for *in-silico* analyse and characterize the cell **CCM**, (especially the stationary metabolic fluxes), and for designing novel **GRC**-s conferring new properties/functions to the mutant cells [21-23,30].

Even if such extended structured models are currently used only for research purposes, being difficult to be identified, it is a question of time until they will be adapted for industrial / engineering purposes in the form of adaptable *structured hybrid models* **HSMDM**. The case study presented and discussed here proves this engineering applicative aspect of **HSMDM**-s.

This chapter is aiming to prove the feasibility and advantage of using this novel concept to couple an extended cell structured deterministic (modular) model with the macroscopic dynamic model of the bioreactor. The resulted hybrid dynamic model **HSMDM** was successfully used for engineering evaluations. Exemplification is made for optimization of the **FBR** used for the **TRP** production.

L-Tryptophan (**TRP**) is a high-value aromatic amino acid with important applications in food and pharma industry. **TRP** is an aromatic non-polar  $\alpha$ -aminoacid essential in humans, that is used in the cell biosynthesis of proteins, being also a precursor to the neuro-transmitter serotonin, of the melatonin hormone, and of vitamin PP [281].

The case study presented in this chapter uses a hybrid dynamic **HSMDM** model built-up by Maria [3] by linking a **CCM**-based structured kinetic model with a **FBR** ideal dynamic model. The resulted hybrid (bi-level) **FBR** model was used to *in-silico* determine the optimal (time step-wise) feeding policy of the **FBR**

used by Chen et al. [282] to study the **TRP**-synthesis by using a genetically modified *E. coli* T5 strain culture. The thus obtained optimal operating policy of the **FBR** has proven to be very effective, by ensuring maximization of the **TRP** production with involving only two key control variables, that is:

- a) the variable feed flow-rate  $FL_j$ , ( $j = 1, \dots, N_{div}$ ), and
- b) the variable feeding GLC concentration  $c_{glc,j}^{feed}$  ( $j = 1, \dots, N_{div}$ ),

Where  $N_{div} = 5$  are the '*time arcs*', that is the equal time-intervals in which the batch-time was divided. During each 'time-arc' (of equal lengths), the control variables are kept constant at optimal values determined from solving the below described optimization problem (i.e. maximization of the **TRP** production in this case).

The obtained optimal operating policy of the approached **FBR**, by using the extended **HSMDM** model, reported better

performances compared to the not-optimally operated **FBR** of Maria et al.[3,58], or of Chen [268].

The structured modular kinetic model of Maria [3,25] used in this numerical analysis includes four inter-connected modules characterizing the dynamics of the concerned cell pathways involved in the **TRP**-synthesis, that is:

- a) Module **[A]** – glycolysis;
- b) Module **[B]** – **ATP** recovery system;
- c) Module **[C]** – **TRP** synthesis (that is the **TRP**-operon expression);
- d) Module **[X]** – the suspended biomass growth.

This cellular structured bioprocess model was experimentally identified, and checked over extensive experiments conducted by several authors, that is [25,26,39,43] for the glycolysis, and by Chen et al. [268,282], and by Maria [3] for the **TRP** synthesis. Experimental data of Chen [268] for the **TRP**-synthesis are also used to validate the predictions of the hybrid **HSMDM** model [12].

The present study presents multiple elements of novelty. *i)* Although production of **TRP** by engineered *E. coli* has been extensively studied, the need of multiple precursors for its synthesis and the complex regulations of the biosynthetic pathways make the achievement of a high yield still very challenging [3]. This engineering problem was solved here by using a model-based (*in-silico*) approach, completed with a biological improvement of the used *E. coli* cell culture; *ii)* The derived optimal operating policy of the **FBR** is given on time-intervals (the so-called ‘time-arcs’) of equal length, and of a reduced number to be easily implemented. The control variables present optimal but constant levels over each time-arc (different between time-arcs) during the **FBR** operation. *iii)* The used biomass culture refers to a modified *E.coli* T5 strain. The characteristics of this strain were reflected in the rate constants estimated by Maria [3]. This T5 strain was produced by Chen et al.[282,283] to increase the **TRP** production in their bench-scale **FBR**. They performed genetic modifications of the **TRP** producer ‘wild’ strain S028. Basically, they remove the **PTS** import-system of **GLC**, in the ‘wild’ strain by replacing it by a more effective one based on the galactose permease/ glucokinase (**GalP/Glk**) uptake system, by modulating the gene expression of GalP/Glk. The resulted T5 strain showed an increase of the specific **TRP** production rate in a non-optimal **FBR** by 52.93% (25.3 mg/gDW biomass /h) compared to the initial strain [282], and by ca. 70% if the used **FBR** is optimally operated (this case study). *iv)* The results reveal the close link between the cell key-metabolites dynamics and the **FBR** operating conditions. *v)* The used hybrid bi-level kinetic **HSMDM** model is enough complex to adequately represent the dynamics of the **FBR** state-variables, that is: the biomass **[X]** growth in the bulk-phase, the **GLC** depletion in the bioreactor liquid-phase, the excreted **TRP** dynamics in the bulk-phase, and the dynamics of the excreted **PYR** but also the dynamics of the cell key-species involved in the concerned reaction pathway modules, that is: **[A]** glycolysis, **[B]** **ATP**-recovery system, **[C]** **TRP**-operon expression.”

### The used *E. coli* GMO strain

Although “production of **TRP** by engineered *E. coli* has been extensively studied, the need of multiple precursors for its synthesis, and the complex regulations of the biosynthetic pathways make the achievement of a high product yield still very challenging. The metabolic flux analysis of [268,283,284,] suggests that replacement of the **PTS** glucose uptake system in the wild *E. coli* with the galactose permease/ glucokinase (**GalP/Glk**) uptake system can double the **TRP** yield from glucose. Finally, these authors obtained a promising *E. coli* T5 strain which, tested in a bench-scale pilot **FBR** proved an increased **GLC** import capacity of the **GMO** *E. coli* together with an increased **TRP** yield by ca. 20% compared to an initial mutant S028 strain (that is 0.164 vs. 0.137 g TRP/g GLC), while the specific production rate was increased by 53% [282]. The cell flux analysis of Chen [268,283] indicated the doubling of fluxes responsive for the

**TRP** synthesis. Finally, a highly productive strain T5AA resulted, with a **TRP** production rate of 28.83 mg/gDW/h” [268,282,284,285]. More details on *E. coli* mutants presenting alternative routes for **GLC** uptake are given by [268,282,284,285-288].

## Experimental FBR bioreactor and the recorded kinetic data

“To estimate the rate constants of the hybrid structured kinetic model for the studied TRP synthesis with using the modified *E.coli* T5 strain of [3,268,282], Maria [3] used the experimental kinetic data of Chen [268] obtained in a benchscale three-phase (G-L-S) **FBR** operated under the so-called '*nominal*' (non-optimal) conditions displayed in (**Table 6-1**). The completely automated **FBR** of 1.5 L capacity includes a large number of facilities described in detail by Chen [268]. The *nominal non-optimal* operation of this bioreactor means addition of a controlled constant feed flow-rate of substrate solution (GLC) of a constant concentration, together with nutrients, additives (for the pH control), anti-bodies, etc. in recommended amounts (**Table 6-1**) along the entire batch. A reduced **FBR** scheme can be found in the (**Figure 6-1**).

To obtain kinetic data, samples have been taken from the **FBR** bulk during the batch (63 h), with a certain frequency (2 to 5 h), thus determining the concentration dynamics of the key-species of interest, that is: **X** (biomass), **GLC**, **TRP**, **PYR**. These recorded data are presented in (**Figures 6-4, Figure 6-6, Figure 6-7, and Figure 6-8**, see the blue points). Concerning the analytical techniques used to derive such measurements, the reader is referred to the work of Chen [268] (see also the Acknowledgement of Maria and Renea [12]).”

## The structured dynamic HSMDM model for the TRP production in a FBR

The **HSMDM** model developed by Maria [3], and valorized for engineering purposes by Maria and Renea [12] is a *hybrid* (bi-level) model including two inter-connected parts, that is: (1).- one '*classical*' part is simulating “the dynamics of the bioreactor *macroscopic* state variables (i.e. species present in the liquid bulk) and, (2).- one structured part is simulating the dynamics of the *nanoscopic* variables describing the cell metabolic processes of interest (for the 'wild', or the **GMO E. coli**). All these *simultaneous* dynamic simulations at various levels of detail are based on the differential mass balance of the macroscopic and nanoscopic state variables.

Dynamic simulation of the two **HSMDM** model parts (1-2) is mandatory performed concomitantly, because the macro-/nano-scale variables are closely linked and inter-related, as long as some cell metabolites are imported/excreted from/in the bioreactor bulk. Even if such a complex / extended dynamic model, including some complex cell metabolic pathways requires more experimental and computational efforts to be built-up and identified from structured kinetic data, the resulted *hybrid* (bi-level, macroscopic and nanoscopic) dynamic model presents major and remarkable advantages, as listed (issues a-j) and discussed in the chap.4.1. The both parts (1-2) of this **HSMDM** dynamic model are below described

in detail, module after module.

Being a metabolite of high practical importance, intense efforts have been invested to decipher its synthesis regulation mechanism in various microorganisms, for deriving an adequate dynamic model of its **QSS** or oscillatory synthesis to be used for engineering purposes. Some results includes the deterministic kinetic models of Maria et al.[62], and of Bhartiya et al.[289], while other studies [54] are rather focus on determining correlations between flux distribution, flux control, and the optimized enzyme amount distribution, but employing a too reduced kinetic model, not able to simulate most of the **CCM** reaction pathways, and the cell metabolic process dynamics.

The **TRP** synthesis regulation being a very complex process, a significant number of simplified kinetic models with lumped terms (species and/or reactions) have been proposed in the literature (see the review of Maria et al. [58,62]). Kinetic modelling of this complex process is even more difficult because, as proved by the following researchers [39,59,62,290,291,292], under certain **FBR** operating conditions, the **TRP**-synthesis can become an oscillatory process. Oscillations in the **TRP** synthesis are produced due to the concomitant activation and high order repression of the **TRP**-operon expression, together with a nonlinear demand for end product, making its expression to be cyclic. The cell growth and dilution rates [related to the cell cycle, and the liquid residence-time in a (semi-)continuous bioreactor] strongly influence the **TRP** system stability, as *in-silico* proved by Maria [39,58].

The adopted hybrid kinetic model is those of Maria [3] built-up using the kinetic data of Chen [268] collected in a **FBR** operated under the nominal (not-optimal) conditions of (**Table 6-1**), with using the T5 strain of *E. coli*. This complex structured kinetic model presented in (**Table 6-2, Table 6-3, and Table 6-4**) is a deterministic one. The **CCM**-based model core is the glycolysis dynamic model of Maria [25], validated by using literature data.

To keep the bi-level **HSMDM** hybrid model of Maria [3,39] adapted here of a reasonable extension, but also to facilitate estimation of its rate constants, this dynamic model accounts for only the key-species included in *four linked cell reaction modules* responsible for the **TRP**-synthesis, as followings: three structured modules {[**A**], [**B**], [**C**]} concern some essential **CCM** cell processes ( **Figure 6-3** for {[**A**], [**B**], [**C**]}, and **Figure 6-2** for {[**A**]} ); the fourth kinetic module concerns the biomass [**X**] growth dynamics in the **FBR** bulk. These interconnected four modules are also integrated in the **FBR** dynamic model, as followings:

**Module [A]** - glycolysis with a modified **GLC**-uptake system (due to the used modified *E. coli* T5 strain) (**Figure 6-2**);

**Module [B]** - **ATP**-recovery system. The pink rectangle in (**Figure 6-3**);

**Module [C]** - **TRP**-operon expression. The gray rectangle in (**Figure 6-3**)

**Module [X]** - The biomass growth kinetic model (in the **FBR** bulk-phase).

**Macroscopic FBR** dynamic model – describes the dynamics of the bioreactor state variables (**GLC, X, TRP, PYR, V**).”

A brief description of these parts of the **HSMDM** hybrid model is presented below. For more details, the reader should consult the works of Maria [3,12].

## The biomass [**X**] growth

”The cell culture in the bioreactor is considered to be homogeneous, and introduced as a lump [**X**] in the **FBR** model of (**Table 6-2**). A modified Contois model, adjusted by considering a power-law inhibition with the 1-st order growing biomass at the denominator [293], was proved to be the most adequate vs. the experimental data of (**Figure 6-7c**). To overcome the absence, at this modelling stage, of the predicted values of [**X**] and [**GLC**] (coming from the **FBR** model coupled with the glycolysis dynamic model), simulations of the biomass dynamics over the batch have been performed by using the experimentally recorded [**X**, and **GLC**] species trajectories (of Chen [268]), interpolated with the cubic splines functions (INTERP1 facility of Matlab™package). The estimated kinetic model of the biomass is given in (**Table 6-2**).

The [**X**] module is connected to the other cell processes, by influencing the **GLC** dynamics in the bulk phase through the **X**-growth rate (**Table 6-2**) that, in turn, influences the **GLC** import flux **V1** into the cell (**Table 6-3, Figure 6-2, and Figure 6-3**).”

## Module [A] - Glycolysis

In short, ”glycolysis module is a determined sequence of ten enzyme catalyzed reactions (see the reduced pathway of (**Figure 6-2**), and of (**Figure 6-3**) with only 6 lumped reactions) that converts glucose (**GLC**) into pyruvate (**PYR**). The free energy released by the subsequent **TCA** originating from **PYR** is used to form the high-energy molecules **ATP**, and **NADH** that support the glycolysis and most of enzymatic syntheses into the cell [294]. Adequate modelling of the glycolysis dynamics is important because the glycolytic intermediates provide entry/exit points to/from glycolysis. Thus, most of the monosaccharides, such as

fructose or galactose, can be converted to one of these intermediates, further used in subsequent pathways. For example, **PEP** is the starting point for the synthesis of essential aminoacids (**AA**) such as tryptophan, cysteine, arginine, serine, etc. [43,62,295,296].

Due to the tremendous importance of the glycolysis in simulating the cell **CCM**, intense efforts have been made both in its experimental study, and in modeling the dynamics of this process specifically in bacteria ( see the short reviews [25,39,297] ). The large number of glycolysis reduced or extended kinetic models proposed in the literature (see the review [25]) present a complexity ranging from 18-30 species, included in 48-52 reactions, with a total of 24-300 or more rate constants. Most of these models are however too complex to be easily identified from (often) few available kinetic data, and too complex to be further used for engineering calculations. Beside, with few exceptions, most of them can not satisfactorily reproduce the glycolytic oscillations occurrence on a mechanistic basis.” [39,59].

The adopted glycolysis kinetic model of Maria [25,39] even if of a reduced form, “by accounting only for **9** key-species in lumped reactions with including **17** easily identifiable rate constants belonging to **V1-V6** metabolic fluxes (**Figure 6- 2, Figure**

**6-3**, **Table 6-2**, and **Table 6-3**) has been proved to adequately reproduce the cell glycolysis under steady state, oscillatory, or transient conditions depending on: **(i)** the defined glucose concentration level/dynamics in the bioreactor bulk (liquid) phase, **(ii)** the total **A(MDT)P** cell energy resources, and **(iii)** the cell phenotype characteristics related to the activity of enzymes involved in the **ATP** utilization and recovery system {here denoted as **Module [B]**}. A detailed discussion about the operating conditions leading to glycolytic oscillations are extensively presented by Maria et al.[39,59,62]. This is why, the **FBR** and the glycolysis dynamic models have to be considered together (**Table 6-2**, and **Table 6-3**) when simulating the dynamics of the **[GLC]** in the **FBR** bulk-phase, and of the cell metabolites of interest {**F6P**, **FDP**, **PEP**, **PYR**, **ATP**, **TRP**-operon expression} into the cell. The adopted rate expressions for the glycolysis main fluxes **V1-V6** presented in (**Table 6-2**, and **Table 6-3**) are those of the basic model, excepting those of the **GLC** import system (**V1**), modified to match the T5 *E. coli* strain kinetic data [3]. It is worth mentioning that, even if not the case here, under certain conditions (that is external/environmental, and internal/genomic factors), glycolysis and **TRP**-synthesis can become oscillatory processes” [25,58,59,61,62]. According to the experimental data, the produced **TRP** (that is the denoted **Module [C]** here) is excreted (**Figure 6-3**) through a process described by Chen [268]. The **PYR** key-metabolite concentration in the cell is regulated through complex mechanism [298,299], the excess being excreted, as experimentally proved by Chen [268].

”Starting from the extended reaction pathway of (**Figure 4-1**), and from the **CCM** models of Chassagnole et al.[43], of Edwards and Palsson [4], and of Maria et al. [42], and by applying chemical engineering lumping techniques, Maria [25] proposed a valuable reduced dynamic model of glycolysis displayed in (**Figure 6- 2**), and denoted by **MGM**, by accounting only for **9** key-species in **7** lumped reactions, with including **17** easily identifiable rate constants belonging to **V1-V6** metabolic fluxes (**Figure 6-2**, and **Figure 6-3**, left-side). The **MGM** rate constants have been identified by Maria [25] with using the experimental kinetic data of [26,43] obtained from a **FBR** including a ‘wild’ *E.coli* culture, operate with ‘pulse-like’ addition of the substrate (**GLC**). When using the modified *E.coli* in the **FBR**, Maria [3] adjusted the **MGM** rate constants by using the {**GLC**, **TRP**, **PYR**, **X**} experimental kinetic curves recorded over the **FBR** batch (**TRP**, and **PYR**, being two excreted metabolites by the cells in the growing medium).

The **MGM** model has been proved to adequately reproduce the cell glycolysis under steady state, oscillatory, or transient conditions according to: **(i)** the defined glucose concentration dynamics in the bioreactor, **(ii)** the total **A(MDT)P** cell energy resources; **(iii)** the cell phenotype characteristics (related to the activity of enzymes involved in the **ATP** utilization and recovery system) [39,59,62]. Here **A(MDT)P** denotes the lump of the following species: **ATP** = adenosin-triphosphate; **ADP** = adenosin-diphosphate; **AMP** = adenosinmonophosphate. This is why, the **FBR** and the **MGM** glycolysis dynamic models have to be considered together [3] when simulating the dynamics of the **[GLC]** in the **FBR** bulk-phase, and of the cell metabolites of interest {**F6P**(fructose-6- phosphate), **FDP**(fructose-1,6-biphosphate), **PEP**(phosphoenolpyruvate), **PYR**(Pyruvate), **ATP**} into the cell. The adopted rate expressions for the glycolysis main metabolic fluxes **V1-V6** in the here discussed **HSMDM** hybrid model are those of the basic **MGM** model.”

### Module [B] - the ATP recovery system

As revealed by the reactions figured in the pink rectangle of (**Figure 6-3**), “the efficiency and the dynamics of the **ATP** recovery system is essential for the reaction rates of the whole **CCM**, as long as **ATP** plays a catalytic-chemical energy provider role. As underlined by Maria et al. [39,59,62], among the involved parameters, an essential factor is the **k6** reaction rate (determined by the *ATP-ase* characteristics in **Figure 6-3**) and included in the glycolysis model of (**Table 6-2**, and **Table 6-3**). The involved enzymes characteristics are directly related to the cell phenotype (that is cell genomic) controlling the **[AMDTP]** total energy resources level. To not complicate the simulations, the **[AMDTP]** level was kept unchanged in the present analysis at an average value given in (**Table 6-1**), as suggested by Chassagnole et al.[43]. The adopted kinetic model for the glycolysis (that is the **V1-V6** reaction rates expressions of (**Figure 6-3**, **Table 6-2**, and **Table 6-3**), and the equilibrium relationships for the **ATP-ADP-AMP** system (**V6**, and equilibrium relationships) given in (**Table 6-2**, and **Table 6-3**) were imported from the literature [3,25,39]. This kinetic model was validated by Maria [3], based on experimental checks to fairly represent the dynamics and the thermodynamics of the internal modules **[A]**, and **[B]** also in the modified *E. coli* T5 strain. Maria [25,39] proved that this **ATP** recovery model fairly represent the dynamics and the thermodynamics of such an important internal module. Rate constants were identified concomitantly with those of module **[A]**, in the same way.”

As an observation, the two modules **[A]**, and **[B]** are inter-connected by sharing the **ATP** species, while the module **[A]** and

**[X]** are inter-connected by sharing **{X, and GLC}** species. Thus, the dynamics of species belonging to the three inter-connected modules **{[A], [B], and [X]}** can be simulated concomitantly, according to the reduced reaction pathway of **(Figure 6-3)**.

### Module [C] - TRP operon expression system

“The adopted *in-silico* evaluation of the **TRP** synthesis of Maria [3] is based on a simplified pathway of the **TRP**-operon expression, as displayed in **(Figure 6-3)**, derived from various studies reviewed by Maria et al.[62]. Modelling the **TRP** synthesis on a deterministic (mechanism-based) approach is difficult because this cellular process is known as being, under certain conditions, a **QSS**, or an oscillatory one [39,62,289,290,291]. However, to avoid extended models, difficult to be estimated and used, most of the reduced dynamic models from literature do not distinguish the process components from the regulatory components, and lumped reactions/species are considered instead, the regulatory performance being included via adjustable model parameters and terms. Kinetic models trying to reproduce the **TRP**-operon expression self-regulation [290,291] are too extended to be of use for engineering evaluations purposes. Due to the process complexity, some modelling approaches [54] are rather focus on determining correlations between flux distribution, flux control, and the optimized enzyme activity distribution, by employing a too reduced kinetic model, not able to simulate most of the **CCM** key-modules and, consequently, its predictive power is very limited. Due to such reasons, in the present analysis, simulations of the **TRP** synthesis were performed by using the reduced **CCM**-based kinetic model of Maria [39,62].

The adopted dynamic model of Maria [3] for the **TRP** synthesis (**TRP**-operon expression) is given in **(Table 6-4)**. This kinetic model is derived from those of Bhartiya et al. [289]. The operon expression regulation terms (C1,C2) were kept unchanged. Only the **TRP** mass balance was changed according to the below **(a-d)** reasons. The rate constants of the considered **OR, mRNA, TRP, E** key-species mass balances were re-estimated by using the experimental kinetic data of Chen [268] given in **(Figure 6-4, Figure 6-6, Figure 6-7, and Figure 6-8)**. The **TRP** mass balance of the Bhartiya et al.[289] model was modified and re-estimated stepby- step as followings:

**a)** An explicitly connection of the **TRP**-module to the glycolysis **module [A]** pathway was introduced through the **PEP** precursor sharing node (in **Figure 6-3**). Consequently, **PEP** is included as a substrate in the **TRP** mass balance ( $C_{TRP}/dt$ ) in **(Table 6-4)**, while the **PEP** consumption term is also considered in the **PEP** balance of the glycolysis model according to the recommended fluxes ratios of Stephanopoulos and Simpson [295], as a first guess **(Table 6-2)**. Analysis of this model suggests that intensifying **TRP** synthesis clearly depends on the glycolysis intensity (that is the magnitude of the average concentrations of the glycolytic species), and on its dynamics (**QSS**, or oscillatory) [25,59]. In fact, as remarked by [284,286], the **PEP** precursor is the limiting factor for the **TRP**-synthesis. This is why, intense efforts have been made to increase its production by glycolysis intensification. This can be realized by optimizing the **FBR** operating policy (as in the present numerical analysis), and/or by using (also in this analysis) a genetically modified *E. coli* T5 strain culture of [268,282].

**b)** The **TRP**-synthesis model of Bhartiya et al.[289] **(Table 6-4)** includes two terms for the **TRP**-product inhibition, that is the C3-term (of allosteric-type), plus a Michaelis-Menten term. Our tests proved that these terms do not adequately fit the **TRP** experimental kinetic data of **(Figure 6-8)**. This is why, the product inhibition term in the **TRP** balance of **(Table 6-4)** has been replaced by the most adequate Contois-type model, with considering a power-law inhibition of the 1-st order growing **TRP** at the denominator. Eventually, the rate constants of the **TRP** kinetic **module [C]**, the **PEP** consumption stoichiometry, and the rate-constants of the all **modules [A]**, and **[B]** were re-estimated (refined) with using the whole (complete) hybrid **FBR** model by using the all available experimental kinetic trajectories of the key-species offered by Chen [268](see the acknowledgement of Maria [3]), and given in **(Figure 6-4, Figure 6-6, Figure 6-7, and Figure 6-8)**.

**c)** The initial guess of the rate constants of the **TRP** **module [C]** were adopted from the literature. Finally, this rough estimate was refined with using the experimentally recorded **TRP** species dynamic trajectory The required **PEP**, and **GLC** dynamic trajectories were transferred from the simulated **{FBR-dynamic model, and module [A], and module [B]}**, all being available at this point.

**d)** By contrast to the literature, in the **TRP** balance of **(Table 6-4)**, an activation inhibition term was considered by bringing together the substrate (**PEP**), and the first key-enzyme (anthranilate synthase, **E**) who trigger the **TRP** synthesis [3].

Such an approach was proved to better fit the experimental data  $c_{trp}(t_u)$ ,  $u=1,\dots,n$  (no. of data, i.e. 17) of **(Figure 6-8)**, and



to confer more flexibility/adjustability to the kinetic model. The estimated ' $g$ ' constant, of a small negative value, reflects the slightly inhibition of the **TRP**-synthesis with the substrate **PEP**, as suggested in the literature." [3]

### The FBR dynamic model

"All the above described four kinetic model modules are integrated in the **FBR** dynamic model. To not complicate the numerical simulations, the **FBR** model adopted by Maria [3] is a classical one, that is the **FBR** ideal model of Moser [253], which fairly describe the key-species dynamics during the batch at a macroscopic level (in the liquid bulk phase). The bioreactor initial conditions and the time step-wise dynamics of the two control variables [that is (1) added **GLC** substrate solution, and (2) the feed flow-rate] will be further explored to derive a desired optimum operation policy of the studied **FBR**. of (Table 6-1).

The bioreactor ideal model main hypotheses are the followings [253]: i) isothermal, iso-pH, iso-DO operation; ii) it is self-understood that nutrients, additives, antibiotics, and pH-control compounds are added initially and during **FBR** operation to ensure the optimal grow of the biomass, as indicated by Chen [268]; iii) aeration in excess over the batch to ensure an optimal biomass maintenance, and to contribute to the liquid homogeneity; iv) perfectly mixed liquid phase (with no concentration gradients), of a volume increasing according to the liquid feed flow-rate time-varying policy; v) the limits of the liquid feed flow-rate ( $F_{L,j}$  in Table 6-2) are adjusted to not to exceed the bioreactor capacity  $\text{Max}(VL)$  in (Table 6-1); vi) negligible mass resistance to the transport of oxygen and compounds into the liquid and biomass flocks (if any); vii) **GLC** substrate is initially added in the bioreactor and during the batch according to an optimal feeding policy to be determined; viii) the feed flow-rate during the batch,  $F_{L,j}$  is varied according to an optimal feeding policy to be determined for every *time-arc* index ' $j$ ' in the eq. (5).

The dynamic model is *hybrid* (bi-level) because it connects the macro-state variable of the **FBR** (biomass **X**, **GLC**, **TRP**, **PYR**) with the cell nano-scale key-variables (**GLC**, **F6P**, **FDP**, **PEP**, **PYR**, **ATP**, Table 6-2, and Table 6--3) of the glycolysis, and of the **ATP** recovery system, and those (**TRP**, **OR**, **OT**, **MRNA**) of the **TRP** operon expression (Table 6-4). The all four kinetic modules are linked to the macroscopic **FBR** dynamic model through the formulated mass balances in (Table 6-2, Table 6-3, and Table 6-4).

From a mathematical point of view, in a general form, the **FBR** dynamic hybrid model of (Table 6-2, Table 6-3, and Table 6-4) translates to a set of **12** differential mass balances (**ODE** set) written for the key-species of the **FBR** in the following form: Species in the bulk-phase: ' $i$ ' denotes species present in the **FBR** bulk; ' $j$ ' denotes the **FBR** feeding time-arcs;  $j=1, \dots, N_{div}$

$$\frac{dc_i}{dt} = \frac{F_{L,j}}{V_L} (c_{inlet,i,j} - c_i) \pm r_i(c(t), c_0, k); c_{i,0} = c_i(t=0), \quad (1)$$

Key-species inside cells:

$$\frac{dc_i}{dt} = \pm r_i(c(t), c_0, k) - \mu c_i; c_{i,0} = c_i(t=0), \quad (2)$$

' $i$ ' denotes species inside cells, that is (**GLC**, **F6P**, **FDP**, **PEP**, **PYR**, **ATP**) for the glycolysis, and (**OR**, **MRNA**, **E**, **TRP**) for the **TRP** operon expression.

Biomass in the bulk phase.

$$\frac{dX}{dt} = r_x(c(t), c_{x,0}, k); c_{x,0} = c_x(t=0). \quad (3)$$

Liquid volume dynamics:

$$\frac{dV_L}{dt} = F_{L,j}; V_{L,0} = V_L(t=0). \quad (4)$$

In the eqn. (1),  $c_{inlet,i,j}$  refers to the concentration of the species ' $i$ ' in the feeding solution, constant over the time-interval ' $j$ ' ( $j = 1, \dots, N_{div}$ ). In the present case only **GLC** is fed in the **FBR** during the batch. The reaction rate **ri** expressions together with

the associated rate constants and other details are given in (Table 6- 3, and Table 6-4). In eqn. (1-3),  $\mathbf{c}$  = vector of species concentrations;  $\mathbf{c0}$  = initial value of  $\mathbf{c}$  (at time  $t=0$ ) given in (Table 6-1);  $\mathbf{k}$  = vector of the model rate constants. The reactor content dilution (determined by the increasing  $\mathbf{VL}$  in eqn. (4) is due to the continuously added  $\mathbf{FL}$  term.

In the eqn. (1),  $\mathbf{GLC}$  and  $\mathbf{FL}$  are the control variables. The optimal,  $F_{L,j}$  to be determined are given on time step-wise values over  $j=1, \dots, N_{div}$  time-arcs. For instance, for the adopted  $N_{div} = 5$ , the  $j=1, \dots, N_{div}$  time-arcs switching points given in the eq. (5) are:  $T1 = t_f / N_{div}$  (12.5 h.);  $T2 = 2 t_f / N_{div}$  (25 h.);  $T3 = 3 t_f / N_{div}$  (37.5 h.);  $T4 = 4 t_f / N_{div}$  (50 h.);  $t_f = 63$  h. More specifically:

Feed flow-rate policy:

$$F_{L,j} = \begin{cases} F_{L,0} & \text{if } 0 \leq t < T1 \\ F_{L,1} & \text{if } T1 \leq t < T2 \\ F_{L,2} & \text{if } T2 \leq t < T3 \\ F_{L,3} & \text{if } T3 \leq t < T4 \\ F_{L,4} & \text{if } T4 \leq t < t_f \end{cases} \quad (5)$$

Similarly, for the adopted  $N_{div} = 5$  equal time-arcs:

Feeding policy for the  $\mathbf{GLC}$  solution concentration:

$$c_{glc,j}^{feed} = \begin{cases} c_{glc,0}^{feed} & \text{if } 0 \leq t < T1 \\ c_{glc,1}^{feed} & \text{if } T1 \leq t < T2 \\ c_{glc,2}^{feed} & \text{if } T2 \leq t < T3 \\ c_{glc,3}^{feed} & \text{if } T3 \leq t < T4 \\ c_{glc,4}^{feed} & \text{if } T4 \leq t < t_f \end{cases} \quad (6)$$

To not complicate the engineering calculus, the main assumption in the eqns. (5-6) is that: On each time step-wise 'arc', index  $j=1, \dots, N_{div}$ , the control variables,  $F_{L,j}$  and  $c_{glc,j}^{feed}$  are kept constant. Of course, the values on each time-arc do not have to be necessarily equal to each other.

**The 'nominal' FBR not-optimal operating conditions.** Under these conditions of Chen [268], the control variables,  $F_{L,j}$  and  $c_{glc,j}^{feed}$  are kept constant on each time-arc at the non-optimal values given in (Table 6-1). Moreover, they are also equal between, that is:

$$F_{L,0} = F_{L,1} = F_{L,2} = F_{L,3} = F_{L,4}, \text{ and}$$

$$c_{glc,0}^{feed} = c_{glc,1}^{feed} = c_{glc,2}^{feed} = c_{glc,3}^{feed} = c_{glc,4}^{feed}$$

**FBR optimal operating conditions.** By contrast, under the optimal conditions studied in this paper, the suitable time step-wise values  $F_{L,0} - F_{L,4}$ , and those of  $c_{glc,0}^{feed} - c_{glc,4}^{feed}$  are to be determined together (simultaneously) to reach the optimum of an objective function (maximum of **TRP** production here)." Multiobjective **FBR** optimization is also possible (see [6,241]), but is beyond the scope of this research.

### Rate constants estimation for the HSMDM hybrid model

In short, "the methodology used by Maria [3] to estimate the adopted bilevel modular dynamic model consists in a sequence of a *trial and error* steps, by adjusting the literature information (reaction rate expressions and constants characterizing the dynamics of cell metabolic species of interest) to fit the available experimental kinetic data recorded from the above described **FBR**. The sequence of computational steps is summarized over the next sub-sections of this chapter. In total, the developed hybrid

structured kinetic model includes **49** rate constants to be estimated from the experimental kinetic curves of **4** observed species (**GLC, TRP, PYR, X**), each species time-trajectory including **17** uniformly distributed recorded points (**Figures 6-4, Figures 6-6, Figures 6-7, Figures 6-8**). This estimation problem is equivalent to a nonlinear programming one (**NLP**) of high difficulty [29] due to its high dimension, and high nonlinearity of the dynamic model and their constraints.

To avoid unfeasible local estimates of the **NLP** problem, Maria [3] used a sequential approach. A rough estimate of the kinetic rate constants for the **modules [A+B+C+X]** given in (**Tables 6-2, Tables 6-3, Tables 6-4**) was generated by using a step-by-step (module-after-module) approach, with also accounting for the shared species {**PEP** by modules **[A+C]; X**, and **GLC** for modules **[A+B+X]**}. If missing during simulations, the experimental **TRP, GLC**, or **X** time-trajectories were taken instead (interpolated with the cubic splines **INTERP1** facility of Matlab™ package[3]).

Finally, the thus obtained rate-constants were refined by means of a standard weighted least square criterion [29] with considering the whole **FBR** hybrid **HSMDM** model, with including the all 4 inter-connected modules **[A+B+C+X]**. To reduce the problem size, only **27** independent model rate constants were accounted during estimation (from the total of **49** rate constants). A number of (**49-27**) rate constants have been adopted from the literature [58,62]. Eventually, the all rate constants have been refined by Maria [3], and presented in (**Tables 6-2, Tables 6-3, Tables 6-4**). The thus identified **FBR** hybrid structured dynamic model fit very well the experimental data as indicated by the comparative plots of (**Figures 6-4, Figures 6-6, Figures 6-7, Figures 6-8**).

As a parenthesis, the multi-modal **NLP** estimation problem solved by Maria [3] is a difficult one, being highly nonlinear, with including nonlinear constraints defining a non-convex domain. For such large-size non-convex estimation problems, the commercial optimization routines usually encounter difficulties in

reaching the feasible global solution with an acceptable reliability. This is why, a very effective **NLP** solver has been used instead, that is the adaptive random search **MMA** of Maria [71,72] implemented on the Matlab™ numerical calculus platform by Maria [72]. The **NLP** solution was checked by using several (randomly generated) *initial guesses* for the rate-constants. A stiff integrator (ODE15S routine of Matlab™ package) has been used to solve the **ODE** dynamic model with a high accuracy.

A comparison of the model estimated rate constants for the modified T5 *E. coli* strain from using the **FBR** experimental data of Chen [268], with those of the same model but estimated for experiments using the 'wild' *E.coli* strain was presented by Maria [3]. As expected, most of the estimated rate constants present

similar values for some general reaction steps. However, due to the mentioned modifications of the used *E.coli* T5 strain in the adopted kinetic model, important differences between the two strains of this bacteria are reported for: **(i)** The rate expressions and parameters of the **GLC** import system (that is flux **V1** in **Table 6-2, Table 6-3**, and chap.6.5.2, and chap.6.5.3); **(ii)** The biomass growing dynamics (**Table 6-2**), and **(iii)** The **TRP**-synthesis **module [C]**, in both parameters and rate expressions (**Table 6-4**). As another observation, for the nominal (not-optimal) **FBR** experimental conditions (**Table 6-1**) used by Chen [268], the species dynamics belonging to inside the cell, and to the external liquid-phase tend to reach a quasi-steady-state (**QSS**) that corresponds to a balanced cell growth(homeostasis) in the bioreactor."

### 6.5.7. Ways to intensify the TRP production in the FBR

"As revealed by the concerned literature [3,58,59,62,268], intensifying the **TRP** synthesis strongly depends on a couple of internal/external factors, as followings: **(a)** the glycolysis intensity (mainly, the **GLC** uptake flux, and the average levels of glycolytic species), transmitted via **TRP** to the **module [C]** via the shared **PEP** intermediate; **(b)** the glycolysis dynamics (**QSS**, or oscillatory behaviour). On the other hand, as pointed out by Maria [39,59], in turn, the glycolysis intensity is controlled by several cell internal and external factors, as followings:

a) The **GLC** import system efficiency (flux **V1** in (**Figures 6-2, 6-3**)) is regulated and triggered by the external concentration of glucose, and by the subsequent **PEP** and **PYR** synthesis (see the kinetic model of (**Tables 6-2&6-3**). The regular **GLC**-uptake system, that is the **PTS** translocation system in the 'wild' strain (of a complex reaction rate expression discussed by [3,25,43]) was replaced in the present studied *E. coli* T5 strain, as mentioned in the chap. 6.3, with a more efficient one [3,12] able to speed-up at least 2x the **GLC**-uptake flux into the cell [268]. Such a modified **GLC**-import was modelled by a simple Michaelis-Menten kinetics in the model of (**Table 6-3**), by accounting for the well-known **GLC** substrate inhibition.

b) The **GLC** quick import and conversion to the precursor **PEP** requires important amounts of regenerable **ATP** and, a rapid enough **ATP** to **ADP** conversion rate, as well as its quick regeneration. The re-estimated rate

c) constants of the kinetic **module [B]** (pink rectangle in (**Figure 6-3**), and chap. 6.5.3), concomitantly with those of the kinetic **module [A]** from the experimental data coming from the **FBR** operated with modified *E. coli* cells implicitly will ensure the requirement that **A(MDT)P** energy system to be able to support the cell glycolysis (see **V2**, **V4**, and **V6** fluxes expressions in (**Table 6-3**), and the **ATP** mass balance in (**Table 6-2**)). On the other hand, limited **A(MDT)P** energy resources which exist into the cell will slow-down the **GLC** import if the **ATP** use/regeneration is not working fast enough [300]. Such an **A(MDT)P** resource is linked to the micro-organism phenotype. Here, the total **A(MDT)P** was adopted (**Table 6-1&6-3**) at the average level recommended by Chassagnole et al.[43].

d) Additionally, due to the enzymes **ATP-ase** and **AK-ase** characteristics related to the bacteria genome and cell phenotype (**Figure 6-3**), a limited **ATP** conversion rate can sustain the glycolytic reactions, while the **ATP** recovery rate is limited by the enzymes participating to the **A(MDT)P** inter-conversion reactions (that is the **K** and **k6** rate constants in the kinetic model of (**Table 6-3**)). This is why, **k6** rate constant has been re-estimated here to fit the experimental data, as suggested by Maria et al.[59,61].

e) At the same time, glycolysis being a systemic process, with a complex regulatory structure, its dynamics (oscillatory, transient, or **QSS**) is also related to the rate constants of their all-involved reactions. Consequently, all these rate constants have been considered in the final estimation step of the whole **FBR** hybrid **HSMDM** kinetic model. Similarly, Silva and Yunes [301] found that glycolysis [**QSS**, or oscillatory] are only possible if the external concentration of **GLC**, and the maximum reaction rates controlled by the enzymes *PFKase* and *GKase* (that control the **V1** and **V2** reactions in (**Figure 6-3**)) are within specific intervals. From the same reason, the rate constants related to the **GLC**-uptake system in the modified *E.coli* cell (modified **V1** flux in (**Table 6-3**)) were reestimated to match the experimental kinetic data.

f) As a corollary of the issue (**d**), Maria et al. [39,58,59,62] determined the operating conditions leading to glycolytic oscillations, or **QSS** by varying the external factor [**GLC**]<sub>ext</sub>; and some internal factors such as the total [**AMDTP**] level, and the **k6** rate constant of (**Table 6-3**). Such an investigation was not necessary here, because no oscillatory process was identified in the present operating case.

g) Not least, simulations of Maria [3,39] revealed that **TRP**-synthesis efficiency is also strongly influenced by external factors, related to the **FBR** operating regime, namely: **(i)**The cell dilution (taken into account, as ' $\mu$ ' in the approached hybrid kinetic model (**Table 6-2**); **(ii)**The **GLC** concentration in the external (bulk) phase ( $c_{glc}^{ext}$  in (**Table 6-2**)); **(iii)** The optimal operating policy for the control variables. In this paper, such an operating policy will correspond to the time stepwise variation of the feed flowrate [ $F L_j$  in eqn.(5)], and of the **GLC** feeding concentration [ $c_{glc,j}^{feed}$  in eqn.(6)]."

## The fed-batch bioreactor (FBR) optimization problem

### Problem preliminary examination

"To support further engineering calculations, a reasonable extended hybrid modular structured **HSMDM** model was developed by Maria [3], by expressing the macroscopic state-variable species dynamics (i.e., biomass **X**, substrate **GLC**, and the product **TRP**) governing the studied **FBR** performances (**Table 6-1**), as a function of the key-intra-cellular species dynamics related to the cell **CCM** metabolic fluxes responsible for the **TRP** synthesis. This link is realized by means of (**GLC**, **X**, **PEP**, **ATP**) model key-species (chap. 6.5). The main modification made by Maria and Renea [12], when adopting this hybrid dynamic **HSMDM** model, refers to the introduction of a variable **FBR** feeding both in the feed flowrate [eqn. (5), and (**Table 6-2**)], and in the **GLC** feeding solution concentration (eqn. (6), and (**Table 6-2**)).

The reasonable compromise between the hybrid model details (number of accounted intra-cellular species and reaction pathways) and its predictive power has been realised by using only the cell key **modules [A-B-C]** of interest (**Figure 6-3**, in a lumped form, chap. 6.5.2, chap. 6.5.3, and chap. 6.5.4) linked to the bulkphase macroscopic species (that is suspended **X**, and **GLC**) (chap. 6.5.2, chap. 6.5.3, and chap. 6.5.5). The fair adequacy of the resulted dynamic model (**Table 6-2 - 6-4**) vs. the experimental data was proved by Maria [3]. Consequently, this hybrid **HSMDM** model becomes suitable for further engineering evaluations of the reactor and bioprocess efficiency, as the case here.

The optimal **FBR** operation derived in this **case study no. (2)** is more complex than the simple non-optimal (*'nominal'*) operation of Chen [268] from (**Table 6-1**). Mainly, the feed flowrate and **GLC** concentration in the feeding solution are no longer kept constant. By contrast

a) the batch-time is divided in **Ndiv** (equal time-'arcs') of equal lengths

the control variables are kept constant only over every *'time-arc'* at optimal values for each time-arc determined from solving an optimization problem (i.e., maximization of the **TRP** production in this case). The time-intervals of equal lengths  $\Delta t = t_f / N_{div}$  are obtained by dividing the batch time  $t_f$  into  $N_{div}$  parts  $t_{j-1} \leq t \leq t_j$ , where  $t_j = j \Delta t$  are switching points (where the reactor input is continuous and differentiable). Time-intervals for the present case study with an adopted  $N_{div} = 5$  are shown in the 'Liquid volume dynamics' row of (**Table 6-2**), and in its Footnote (a)."

## Formulation of the optimization problem

### Selection of the FBR control variables

By analysing the **FBR** hybrid model of (**Table 6-2**), "completed by the reaction rates expressions and parameters given in the (**Table 6-3&6-4**), the natural option is to choose as control variables those with the higher influence on the biological process, and easily to handle. In the present case, according to the discussion of chap. 6.5.5, and chap. 6.5.7, two control variables were chosen, namely those related to the reactor feeding, that is:

a) The substrate  $c_{glc,j}^{feed}$  ( $j=1, \dots, N_{div}$ ) whose concentration play the major role in the cell glycolysis efficiency and **TRP** production

b) The liquid feed flow-rate,  ${}^F L_j$  ( $j=1, \dots, N_{div}$ ) with **GLC** solution who is directly linked to the **GLC** feeding, and responsible for the reactor content dilution (the dilution rate being defined as  $D = F_L / V_L$ ).

In the present optimization strategy, each control variable is kept constant over each time-arc (index 'j') of the batch. Of course, they are not necessarily equal between different time-arcs. For  $N_{div} = 5$ , in total there are  $5 \times 2 = 10$  unknown eqn. (**7**) to be determined by optimization, under certain constraints, that is (**Table 6-2**):

$$c_{glc,j}^{feed} \quad (j=1, \dots, N_{div}) \quad (7)$$

The **FBR** initial state is given in (**Table 6-1**) for both inside cell, and bulkphase species. Those of the **FBR** control, and the bulk-phase variables, that is, the initial liquid flow rate, and the substrate initial concentration [as shown in **Table 6-2**, and eqn. (**5-6**)] are included as unknown variables in the **FBR** optimization, that is:"

$$F_{L,0} = F_L(t=0) = {}^F L, 0 \text{ in eqn. (5)} \quad (8)$$

$$[GLC]_0 = c_{glc}^{ext}(t=0) = c_{glc,j}^{feed} \text{ in eqn. (6)} \quad (9)$$

#### 6.6.2.2. FBR optimization - objective function ( $\Omega$ ) choice

"By considering the control variables indicated in eqn. (**7**), the **FBR** optimization consists of determining its optimal initial load, simultaneously with its feeding policy for every time-interval during the batch eventually leading to maximization of [**TRP**] production during the batch, that is: Find the control variables values of eqn. (**7-9**), to reach

$$Max\Omega, \text{ where: } \Omega = Max[TRP(t)], \text{ with } (t) \in [0, t_f] \quad (10)$$

The  $TRP(t)$  dynamics in eqn. (**10**) is model-based evaluated, by solving the **ODE** dynamic model of the **FBR** [eqn. (**1-6**)] over the whole batch time  $(t) \in [0, t_f]$ ."

#### 6.6.2.3. Optimization problem constraints

The optimization problem (**10**) is "subjected to the following multiple constraints:

a. The **FBR** model eqn.(**1-6**) including the bioprocess kinetic model (**Tables 6-2- 6-4**);

b. The **FBR** initial condition from (**Table 6-1**), excepting for  ${}^F L, 0$  and which are determined from solving the optimization problem (the initial guess is taken from the same (**Table 6-1**));

c. To limit the excessive consumption of **GLC** substrate, and to prevent the hydrodynamic stress due to the limited reactor volume, feasible searching ranges are imposed to the control/decision variables, that is:

$$(GLC)_{inlet, \min} = 1000 \text{ (mM)} \leq (GLC)_{inlet, j} \leq (GLC)_{inlet, \max} = 4500 \text{ (mM); (10)}$$

$$FL_{, \min} = 0.01 \text{ (L/h)} \leq FL_{, j}; FL_{, \max} = 0.04 \text{ (L/h) (11)}$$

d. physical meaning of searching variables:

$$^F L_j, > 0; c_{glc} j^{feed} \geq 0, (j = 1, \dots, N_{div}) \quad (12)$$

e. physical meaning of state variables:

$$c_i(t) \geq 0 \text{ (} i = 1, \dots, \text{no. of species in the model) (13)}$$

f. limit the maximum cell resources in **AMDTP**

$$[ATP](t) < Total[AMDTP] \text{ of (Table 6-1),}$$

with  $[ATP](t)$  obtained from solving the **FBR** model eqn. (1-6) (14)

As an observation, the imposed ranges for the control variables are related to not only the implementation facilities, but also to economic reasons, meaning minimum substrate consumption, reduced dilution of the reactor content, and an effective bioreactor control.

## Selecting the number of time-arcs (Ndiv), and of the operating alternative

The adopted **FBR** operating policy alternative of chap. 6.6.2.1 is the simplest, and easiest to implement operating mode for the two control variables. "It implies a time stepwise variable feeding of the bioreactor, over an adopted ( $N_{div} = 5$  here) equal time-arcs that covers the whole batch time. Each time-arc 'j' ( $j = 1, \dots, N_{div}$ ) is characterized by optimal levels of the feed flowrate  $^F L_j$ , and  $c_{glc} j^{feed}$  of the **GLC** concentration

in the feeding solution [see eqn.(7-9)].

This type of **FBR** operation, despite its simplicity and easy to be implemented, it still includes enough degrees of freedom to offer a wide range of **FBR** operating facilities that, in principle, might be investigated, for instance (see also the discussion of Maria [11]):

- by choosing unequal time-arcs, of lengths to be determined by the optimization rule.
- by considering the whole batch time as an optimization variable.
- by increasing the number of equal time-arcs ( $N_{div}$ ) to obtain a more *refined* and versatile **FBR** operating policy but keeping the same non-uniform feeding policy (that is of the two control variables), as adopted here.
- by considering the search min/max limits of the control variables as unknown (to be determined).
- by feeding the bioreactor with a variable feed flowrate, but with a **GLC** solution of an uniform concentration over a small/large number ( $N_{div}$ ) of time-arcs.

Most of these alternatives are not feasible, by presenting a large number of disadvantages, as extensively discussed by Maria and Renea [12].

The optimization alternative used in this work is the best, because:

- is simple, by accounting onlu two control variables (chap. 6.6.2.1),
- it accounts a relatively small number of time arcs, that is  $N_{div} = 5$ . with equal time-arc-lengths of  $t_f / (N_{div}) = 63/5$  h.

The alternatives (**a-e**) are not approached here from the following reasons:

- Alternatives (**a-c**) are not good options, because as ( $N_{div}$ ) increases, the necessary computational effort grows significantly (due to considerable increase in the number of searching variables), thus hindering the quick (real-time) implementation of the derived **FBR** operating policy. Additionally, multiple optimal operating policies can exist for the resulted

over-parameterized constrained optimization problem of a high nonlinearity, thus increasing the difficulty to quickly locate a feasible globally optimal solution of the **FBR** optimization problem.

A brief survey of the **FBR** optimization literature [302,303] reveals that a small number ( $N_{div} < 10$ ) is commonly used for such **FBR** due to the abovementioned reasons.

Additionally, as the ( $N_{div}$ ) increases, the operating policy is more difficult to implement, since the optimal feeding policy requires a larger number of stocks with feeding substrate solutions of different concentrations, separately prepared to be fed for every time-arc of the **FBR** operation (a too expensive alternative). Also, the NLP optimization problem is more difficult to solve because of the multimodal objective function, leading to multiple solutions difficult to discriminate and evaluate. This is the case, for instance of an obtained infeasible optimal policy requiring a very high  $[X]$ , difficult to be ensured due to limitations in keeping necessary levels of the related running parameters of the bioreactor (that is dissolved oxygen, nutrients, pH-control substances, anti-bodies, etc.). Besides, **FBR** operation with using a larger number of small time-arcs ( $N_{div}$ ) can raise special operating problems when including **PAT** (Process Analytical Technology) tools [304].

b) The alternative **(d)** is unlikely because it might indicate unrealistic results, as explained at point **(c)** of chap.6.6.2.3. In our numerical analysis, carefully documented upper bounds of control variables were tested to ensure the practical implementation of the optimal operating policy.

c) The alternative **(e)** is also not feasible, even if a larger ( $N_{div}$ ) will be used. That is because, it is well-known that the variability of the **FBR** feeding over the batch time-arcs is the main degree-of-freedom used to obtain **FBR** optimal operating policies of superior quality” [3,6,11,12,99]. By giving up with the variable feed flowrate and substrate concentration, sub-optimally **FBR** operating policies will be obtained, of low performances.

### The used numerical solvers

The time-evolution of the accounted species in the **HSMDM** model (those from inside cell, and those from the bulk-phase) governed by the mass balance of eqn.(1-2) ”is obtained by solving the **FBR** dynamic model eqn.(1-6) with the initial condition of  $C_{j,0} = C_j(t = 0)$  of (Table 6-1) for the inside cell species, except bulk  $[GLC]_0$  to be determined from the **FBR** optimization, as indicated by eqn. (7,9). The imposed batch time  $t_f$ , and the optimal medium conditions are those of (Table 6-1). The dynamic model solution was obtained with a high precision, by using the high-order stiff integrator ('ode15s') of the MATLAB™, with suitable quadrature parameters to keep the integration error very low.

Because the math form of the **FBR** hybrid model eqn.(1-6), the optimization objective eqn.(10), and the problem constraints eqn.(11-14) (chap.6.6.2.3) are all highly nonlinear, the formulated optimization problem eqn.(7-10) translates into a nonlinear optimization problem (**NLP**) with a multimodal objective function and a non-convex searching domain. To obtain the global feasible solution with enough precision, the multi-modal optimization solver **MMA** of Maria [29,71,72] has been used, as being proved in previous works to be more effective compared to the common (commercial) algorithms. The computational time was reasonably short (minutes-hours) using a common PC, thus offering a reasonable quick implementation of the obtained **FBR** optimal operating policy.

### The optimization problem solution particularities

The obtained optimal operating policy of the **FBR**, for the optimization problem formulated in the chap. 6.6.2.2, with the control variables of chap. 6.6.2.1, and the constraints of chap. 6.6.2.3, and adopted **Ndiv** in chap. 6.6.2.4 is given in (Figure 6-6) for the feeding policy of the **GLC** concentration  $C_{glc,j}^{feed}$  ( $j=1,...,5$ ), and in (Figure 6-7-a) for the feed flow-rate,  $F L_j$  ( $j = 1,...,5$ ). It is to observe that, due to the above formulated engineering problem, the **FBR** optimal operating policy will be given for every time-intervals (of equal lengths) uniformly distributed throughout the batch-time. This optimization problem solution will be analyzed in more detail in the next chapter 6.7.

Such an optimal time stepwise variable feeding of the bioreactor presents advantages and inherent disadvantages. The advantages are related to the higher flexibility of the **FBR** operation, leading to a higher productivity in TRP as proved in the next chapter. Beside, the imposed limits of the control variables prevent excessive substrate consumption with any benefit, or an excessive reactor content dilution.

As a disadvantage, the **FBR-s** with such a time-variable control are more difficult to operate than the simple **BR-s**, as long as the time stepwise optimal feeding policy requires different stocks of feeding substrate solutions of different concentrations to be

used over the batch. This is the price paid for achieving **FBR** best performances. This need to previously prepare different substrate stocks to be fed for every 'time-arc' (that is a batch-time division in which the feeding is constant) is offset by the net higher productivity of **FBR** compared to those of a simple **BR** as below discussed, and pointed-out in the literature [5,6,241,303,305]. In fact, the best operating alternative (**FBR** vs. **BR**) is related to many others economic factors (operating policy implementation costs, product cost compared to its production costs, product price fluctuation, etc.), not discussed here."

## Optimization results and discussion

The obtained optimization problem solution „(of the type discussed in the chap. 6.6.2.6) is given in (**Figure 6-6-top**, curve 2) for the **GLC** feeding concentrations, and in (**Figure 6-7-a**, curve 2) for the feed flowrate. Thus optimally operated **FBR** displays the bulk [**TRP**] dynamics of (**Figure 6-8**, curve2). The corresponding dynamics of cell glycolytic species during the batch is displayed in (**Figure 4**), while those belonging to the TRP-operon expression in (**Figure 5**). The dynamics of species present in the reactor liquid phase are presented in (**Figure 6-6**) for **GLC**, and in (**Figure 6-7-c**) for the biomass (**X**). In these figures, the species dynamics plotted for the optimal **FBR** operation (black curves 2, i.e., the model prediction) are compared to those corresponding to the nominal, non-optimal **FBR** operation (blue curve 1 of Maria [3]), and with the experimental blue points of Chen [268]. The both operating policies (optimal 1, and the not optimal 2) are obtained with using the same modified *E.coli* T5 strain of Chen et al. [268,282]."

By analysing the resulted **FBR** optimal operating policy (plots no. 2 in (**Figures 6-4-6-8**)) compared to those of the sub-optimal (nominal) operation of Chen [282], several observations can be derived, as followings:

a) By using the same **FBR** but operated under the nominal (non-optimal) conditions of (**Table 6-1**), the modified *E.coli* T5 strain reported a higher **GLC**-uptake rate, and a **TRP**-production much higher (with ca. 50%, [282]) compared to the "wild" strain, as revealed by the comparative analysis given in (**Table 6-5**).

b) The efficiency of the optimally operated **FBR** (this paper) in the **TRP** production is significantly higher (with ca. 20%) compared to the same **FBR** but sub-optimally (nominally) operated (**Table 6-5**), even if the same modified *E.coli* T5 strain was employed in both cases. The same conclusion also results by comparing the **TRP** final concentrations in the **FBR** bulk given in (**Figure 6-6**, **Figure 6-7**, **Figure 6-8**) for the two operating policies (optimal-variable feeding vs.-not optimal/uniform fed). If one add to this 20% production increase due to the optimally operated **FBR**, to the 50% due to the use of the GMO bacteria, it results a total of 70% increase in the TRP production compared to the "wild" strain use.

c) The optimal **FBR** operation reported a similar dilution of the reactor content, as revealed by (**Figure-6-7 b**) for the two operating alternatives of the **FBR** (optimal, and not-optimal). By contrast, the substrate **GLC** is better used, as proved by (**Table 6-5**). The **GLC** consumption in (**Table 6-5**) was computed with the following relationship

$$m_{GLC} = \sum_{j=1}^{N_{div}} c_{glc,j}^{feed} F_{L,j} \Delta t_j ; \Delta t_j = t_f / N_{div} \quad (15)$$

d) As expected, a higher TRP-productivity requires a higher GLC consumption as the case when using a modified *E. coli* T5 strain instead the "wild" type. As revealed by (Table 6-5), the GLC consumption is influenced by the **FBR** operating mode, even if the same cell strain is used. As indicated by our present analysis given in (Table 6-5), the GLC overall consumption for the optimal (variable feeding) **FBR** operation is roughly similar to that of a non-optimally (uniform feeding) **FBR** operation. Not surprisingly, the optimal operating mode requires a slightly lower GLC consumption (with ca. 6%). That is because it's better used during the batch.

e) The comparative analysis of the glycolytic species dynamics in (Figure 6-4) reveals close trajectories (even quasi-identical for F6P, FDP species), with any accumulation tendency, for both nominal (not-optimal, curves 1), or optimal (curves 2) **FBR** operation. By contrast, the intermediate PEP species is formed in high amounts but then is quickly consumed in the subsequent TRP synthesis, thus tending to reach a QSS. The more intensive GLC import for the optimal **FBR** operation (curve 2) and its successive transformation over the glycolysis pathway, and TRP operon expression is reflected by a higher ATP consumption compared to the non-optimal **FBR** operation. The PYR metabolite is consumed in the TCA cycle and excreted in the bulk-phase (fairly predicted by our kinetic model in (Figure 6-4), thus matching the experimental data).



f) The comparative analysis of the TRP-operon expression species dynamics in (Figure 6-5) reveals very close trajectories between the two alternative FBR operation. The exception is the excreted TRP, which displays a different dynamics for the nominal (not-optimal, curves 1), or optimal (curves 2) FBR operation. Such a result can be explained by the operon expression mechanism, involving a tight control via its inhibition terms presented in (Table 6-4).

g) The comparative plots of GLC concentration dynamics in the FBR bulk-phase are presented in (Figure 6-6). They indicate similar decreasing trajectories for the both investigated FBR operating alternatives: i) nominal (notoptimal, curves 1), or optimal (curves 2). Such a result can be explained by the same GLC-uptake mechanism of the modified *E. coli* T5 strain. In the optimal case (curves 2) the GLC consumption is higher, due to a higher TRP productivity. The curve 2 unevenness is linked to the variable feeding with GLC of the optimally operated FBR [see the feeding plots in the top part of (Figure 6-6)].

h) The comparative plots of the biomass dynamics in the FBR bulk-phase are presented in (Figure 6-7 c). They reveal similar increasing trajectories, for the both investigated FBR operating alternatives: i) nominal (not optimal, curves 1), or optimal (curves 2). In the optimal operation case, the biomass growth is more intense (reaching a 10% in the bioreactor, close to the admissible limit [254,257]), due to a significant higher GLC-uptake, and a better GLC use during the batch, thus offering more favourable biomass growth conditions.

i) The TRP concentration dynamics in the bulk-phase is plotted in (Figure 6-8) for the both investigated FBR operating alternatives: i) nominal not-optimal operation of (Table 6-1), that is curves 1, compared to the experimental data (•, blue) of Chen [268], or ii) optimal FBR operation (curve 2). The TRP higher final concentration leads to a higher productivity for the optimally operated FBR (see above observation no. a). Such a result proves that the optimal time stepwise FBR feeding [that is, the GLC feeding curve 2 in (Figure 6-6-top), and the feed flow-rate policy of (Figure 6-7-a)] is superior to the non-optimal uniform feeding of the bioreactor, leading to a better .GLC use, even if, the overall GLC consumption [see the above observation no. d] is similar for both nominal, and optimal FBR operation. The better GLC use for the optimal FBR operation is also proved by the less produced secondary metabolite PYR in ((Figure 6-4), curve 2), and by a smaller QSS concentration for the PEP intermediate ((Figure 6-4), curve 2), quickly transformed in the final product TRP.

## Chap. 7. Case study no. 3: the use of a hybrid CCM cell-scale structured kinetic model coupled with a BR classical dynamic model (including macro-scale state variables) to maximize both biomass and succinate production by the insilico design of GMO *E. coli* cells.

### Symbols used in the chap. 7

$A_j$  - reaction affinity [73]

$C_a^b$  -  $a! / b!(a-b)!$  = number of ways of choosing 'b' objects from a collection of 'a' objects without regard to order

$CV_{eq}$  - equality constraint violation index

$CV_{ineq}$  - inequality constraint violation index

$c_i$  - species 'i' concentration

$f_j$  - individual objective functions

KG - number of removed genes

L - Lagrange function of the optimization problem

$\lambda_{eq}, \lambda_{ineq}$  - Lagrange multipliers of the equality and inequality constraints

M - number of fluxes in the considered metabolic pathway

N - number of metabolites in the considered metabolic pathway

$S$  - stoichiometric matrix,  $S_{ij}$  i.e. the stoichiometric coefficient of the metabolite 'i' in the reaction 'j'.

$S_j$  - synthetic accessibility of an output 'j', i.e. the minimal number of metabolic reactions needed to produce the component 'j' from the network inputs

$S_i$  - total synthetic accessibility index,  $S_i = \sum_j S_{ij}$

T - temperature

t - time

$v, v_j$  - vector of stationary metabolic fluxes, or reaction rate

$y_j$  - Boolean variables

$w_j$  - weights of individual objective functions

### Greeks

$\beta_j$  - exponent of  $v_j$  flux in the **NLP** objective function

$\Phi$  - objective function

$\lambda_i^{stoich}$  - dual variable associated with the stoichiometric constraint involving reaction 'i'

$\mu_j$  - dual variable associated with any restriction of the flux  $v_j$

### Index

low - lower limit

max - maximum

min - minimum

up - upper limit

**Observation:** species abbreviations together with their chemical formula are given

in the (Table 7-1).

## General remarks about CCM kinetic models, and ways used to solve the problem to in-silico design GMO-s of industrial interest by using a gene knock-out strategy.

" The case study presented in this chapter brilliantly exemplifies how a complex HSMDM, including a very extended **CCM** from literature, can be used for both bioinformatics and engineering purposes. More specifically, this kinetic model was used to *in-silico* design of an *E.coli* bacterium by using a pareto-optimal front methodology of obtain maximum production of biomass and succinate in batch bioreactor (BR)

In general, is well known that bioprocess optimization by genetically modifying the microorganism characteristics is an intensively investigated subject due to the immediate economic interest of the large-scale industrial biosynthesis. A large variety of alternatives using elaborated experimental procedures, assisted by *in-silico* cell design based on topological, or dynamic models, have emerged [21-23]. By using the extended CCM model of Edwards and Palsson [4] for the wild strain of *E. coli*, the present study *in-silico* investigates the possibility of using a mixed-integer nonlinear programming (MINLP) approach to determine the optimal metabolic fluxes in a design GMO in respect to industrial (engineering) multi objective criteria associated to gene knockout strategies. The advantage of the proposed math power-law type criterion is coming from the possibility to account, in a simple way, for the flux nonlinear interactions and for the complex constraints of the highly nonlinear optimization problem. The combinatorial rule is included in the iterative MINLP solver, while a large number of constraints can increase the chance to obtain a reduced set of viable gene-knockout solutions for a given metabolic network. Multiple gene deletion alternatives are thus identified, allowing a cell high growth rate with maximizing externally imposed chemical production targets. In the present case study, exemplification is made for the case of designing a GMO *E. coli* cell culture able to realize maximization of both biomass and succinate production in a BR, by using a reduced CCM kinetic/topological model from the literature. Comparatively to the linear optimization procedures (LP) that solves a combinatorial problem in a bi-level optimization approach, of dimensionality

sharply increasing with the number of removed genes, the MINLP alternative combined with the Pareto-front technique leads to superior results by considering an adjustable nonlinear influence of fluxes to the optimization goals, its performance being less dependent on the number of knockout genes.

One important industrial application of GMO-s is those of maximizing the production of a target metabolite in industrial bioreactors. In the present case study, the problem is to use a structured CCM model of a known wild strain of *E. coli* to *in-silico* design (by using a theoretical gene knockout strategy) an *E. coli* GMO able to maximize the production of biomass concomitantly with maximization of the succinate (SUCC) production. Succinate is an important biosynthesis product with multiple uses in the pharmaceutical and food industry [Wikipedia, 2022].

The thus formulated optimization problem is a multi-objective one (Figure 4-31, Figure 4-32), in the presence of the stoichiometric and flux limitation constraints. To solve it, Maria et al. [42] used a structured CCM kinetic model of a moderate size for the *E. coli* cell (wild-strain) proposed by Edwards and Palsson [4] accounting for 72 metabolites (Table 7-1) involved in 95 reactions (Table 7-2) (see the reduced pathway schemes in (Figure 7-7), and (Figure 7-8)) to *in-silico* determine what genes (and the corresponding reactions) should be removed (the so-called 'Gene knockout' procedure) from the bacteria genome to realize maximization of both biomass and succinate production by the GMO *E. coli* cell. These two optimization objectives are opposite. Thus, a trade-off solution should be chosen among the geometrical locus of all problem solutions realizing the best tradeoff between the two contrary objectives (the so-called Pareto-optimal front, [309]).

Beside simulation of the cell key-species dynamics, such a structured cell simulator of their CCM allows identification of genome modifications leading to the improvement of the formulated optimization objectives. Simultaneous removal of several genes and, self-understood, of their corresponding reactions (catalyzed by the encoded enzymes of the removed genes) from the metabolic pathway of (Figure 7-8) is also possible.

The below formulated MINLP optimization problem is difficult to be solved not only because multiple solution may exist, but the searching domain is not convex, thus reducing the possibility to obtain the global optimum of the problem [310].

To solve this optimization problem, Maria et al. [42] applied an elegant Pareto-optimal front procedure, and a genetic algorithm to obtain the Pareto-optimal front (see the curve of Figure 7-8-right) including the geometrical locus of all problem solutions realizing the best tradeoff between the two contrary objectives.

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 (indicated in parenthesis in (Figure 7-8-right)). Due to the very high complexity of the problem (dozens of hours of computing time), only a few number of alternative solutions have been checked. Results indicated that a better succinate production is obtained when synthesis of formate, acetate, lactate, ethanol, and glutamine are suppressed (see the removed reactions in (Figure 7-8-left)).

Over the last decades, biotechnology and bioengineering is developing new research directions for improving the metabolic performances of the microorganisms used in the process industry. The new approach, known in the literature as the concept 'From gene to product' [271] is based on the application of fundamental science knowledge (biology, biochemistry) and engineering science approaches (in particular the principles, concepts, rules, and algorithms of the (bio)chemical engineering [21-23,74], (Figure 4-2, Figure 4-3, Figure 4-31, Figure 4-32), and the text before eq. (9) of chap. 4.2.1.3, and (Figure 4-20) to understand the cell metabolism, species interactions, and the genetic regulatory circuits (GRC) responsible for regulation of the cell metabolic biochemical reactions. The result is the possibility to *in-silico* re-design the cell metabolism to derive novel microorganisms, genetically modified (GMO), by conferring new properties and functions to the mutant cells (i.e., desired 'motifs'), with applications in various fields, such as improving industrial bioprocesses (biosynthesis, pollutant biotreatment, drug industry), designing of novel measure devices (biosensors, bioindicators), or in medicine (gene therapy).

Numerical simulation of metabolic cell processes, at a topological or dynamic (kinetic) level, in a holistic, modular, compartmented, lumped or extended approach is necessary to *in-silico* (based on mathematical models) design of GMO-s, by combining knowledge from various modern fields, such as synthetic biology,

systems biology, genetic circuit engineering, molecular bioengineering, but also classical (bio)chemical engineering, and nonlinear systems theory [1, 2, 21-24, 174, 278, 311-313]. Living cells are evolutionary, autocatalytic, self-adjustable structures able to convert nutrients from environment into additional copies of themselves during the cell cycle. In spite of tremendous progresses made in cell process analysis and in development of bio-omics databanks [23], various approaches exist in analysing

and modelling the genome replication, cell metabolism, and the multiple regulatory functions of the cell syntheses [21-23]. Different analyses are justified by the very high complexity of metabolic processes, implying thousands of species and ten of thousands of (self) catalytic reactions, enzymes, co-enzymes, activators, inhibitors, transcriptional factors, proteic-oligomers, intermediates, regulatory and signalling chains, motility, membranar and internal transport, gene transcription, morphogenesis, and cellular differentiation, all in an inter-connection difficult to decipher. As a consequence, the topological cell models have been preferred, on a first step, by applying the so-called 'Metabolic Control Analysis' (MCA) [73,314] to investigate the sensitivity of the stationary cell system vs. external perturbations.

The *in-silico* (math-based, engineering approach) to analyse and design novel GMO-s is based on developing dynamic simulators of the essential cellular processes (CCM), by applying physico-chemical laws and principles, (bio)chemical engineering principles, concepts, and rules, and the algorithms of the nonlinear systems regulation theory, and including both stationary and kinetic information on metabolic processes [21-23,74,278]. However, the availability of enzymatic reaction kinetic information was fragmented, and consequently such dynamic models rarely include more than hundreds-to-thousand modelled biochemical reactions. In contrast, due to very large quantity of qualitative, less structured bio-omic information, the attention turned to developing methods to analyse the relative importance of various metabolic events / reactions for the whole cell metabolism. In this context, the developed metabolic 'Flux Balance Analysis' (FBA), or the 'Metabolic Flux Analysis' (MFA) [188], and the 'Elementary Mode Analysis' (EMA) [179] allowed evaluating the cell metabolism efficiency and how resources are used, and also the minimum set of enzymes required by the cell growth with preserving the physiological functions and system invariants (the so-called 'modes' derived from the null space of the stoichiometric matrix [173]). Similarly to EMA, the 'Extreme Pathways Analysis' (ExPA) [315] determines the '*solution space*' within which fall all possible steady-state flux distributions of the metabolic network, by means of a constraint-based approach (by using the species differential mass balance re-written at QSS, and indices based on the maximum reaction rates). The algebraic calculus is based on Kirchhoff's first law (i.e., the production and utilization rate of a metabolite must balance at steady-state), and the second law (i.e. the free-energy change around a biochemical loop must be zero).

Rigorous statistical methods can be applied to identify the relations between metabolites in a reaction network [316,317], by performing a '*modal matrix analysis*' to assess which metabolites could be grouped ('*pooled*'), and by developing reduced topological and dynamic models by using these pooled (lumped) metabolites. In the end, they recommend the type of aggregate variables to be used for kinetic model development when '*sufficient*' experimental information is unavailable. Kauffman et al. [318] used this '*modal matrix analysis*' to pool metabolites and to extract the dynamic characteristics of a biological network (see their application on the human red blood cell metabolism). They also show how '*dynamic phase planes, statistical time-lagged correlation analysis, and temporal decomposition*' can be used to relate the cell biochemical mechanistic details to the overall metabolic functions. As such methods are based on analysing the quasisteady-state (QSS) metabolism of a cell (i.e., under-balanced growth, at homeostasis), hybrid stationary-dynamic models have also been developed. For instance, Mahadevan et al. [319] introduced the so-called 'Dynamic Flux Balance Analysis' (DFBA), which incorporates rate of change of flux constraints from analysing the evolution of the metabolic flux distribution over time. Dynamic interpretations of the flux control (sensitivity) coefficients of the MCA have also been studied. e.g., by Tusek and Kurtanjek [47]. The metabolic flux distribution is obtained either from the cell kinetic model written at QSS conditions, or by the classical metabolic engineering methods of Stephanopoulos et al. [188].

The cell metabolic fluxes are the (enzymatic) reaction rates evaluated at QSS, that is at the cell homeostasis (equilibrated growth).

FBA is a classical but still very powerful method to determine the stationary (QSS) distribution of metabolic fluxes (for a known metabolic pathway), and also to relate any change in the environmental conditions, or in the cell structure (including genome modification) to the ways by which the environmental nutrients are used inside the cell [188,320,321]. The FBA is based on the stoichiometric mass balance constraints under steady-state conditions, of type  $s_v = 0$  (where 'S' is the stoichiometric matrix including the stoichiometric coefficients of the metabolites in the reaction pathway and,  $Sv = 0$  is the vector of stationary metabolic fluxes, including internal, transport, and the growth fluxes). In the FBA, the exchange fluxes are assigned to those metabolites that enter or leave the particular network only, with constraints ranging from negative to positive values, accordingly. Those metabolites that are consumed within the network are not assigned any exchange flux value. As the number

of fluxes is much higher than the number of measured fluxes and mass balance constraints, the feasible set of solutions is defined by the intersection of the null space (i.e., the vector space for  $v$ ) and max-min type of constraints imposed to the fluxes, i.e.  $v_{j,min} \leq v_j \leq v_{j,max}$ ,  $j=1,\dots,M$  (where M is the number of fluxes in the considered metabolic pathway) [4]. In principle, if a sufficient number of constraints (including the kinetic details) are available, a 'single point' solution in the flux-space can be obtained. In practice, measurements of some net stationary fluxes allow obtaining the least squares solution of the problem with linear constraints, by means of a simple matrix calculation [320,321]. The FBA stoichiometric constraints can also be used to correct (reconcile) the measured fluxes affected by gross errors [322]. More elaborated and precise approaches use both stationary and dynamic information on the cellular utilization of a C-labelled substrate with the carbon isotopes (either  $^{13}\text{C}$  or  $^{14}\text{C}$ ) to extract supplementary information on metabolic fluxes in terms of isotopomer distribution, easier to measure experimentally [323-327]. In fact, FBA and MCA are closely connected because a modification of the cell stationary fluxes due to a certain internal/external perturbation factor (or due to a cell genome modification, e.g., by gene knockout, or by cloning procedure) can be transcribed in terms of stationary sensitivities [328].

**FBA** can highlight the most effective and efficient pathway through the network in order to achieve a particular objective function. In fact, multiple stationary flux solutions (homeostatic **QSS**) may exist according to the environmental conditions and the cell adaptation characteristics (dependent on its genome, and GRC-s). The multiple cellular objectives are associated to the system action to perform regulatory, metabolic, homeostatic, and phenotypic functions that realize the best tradeoff between proliferation and differentiation from one side, and cellular functions and robust growth from other side, with the optimal use of the available resources, with shortest transient times, and highest P.I.-s of the GRC-s (see chap. 4.2, 4.4, 4.5, outline 4-1). For instance, Selvarasu et al. [329] used a weighted multi-objective optimization rule to identify synergistically switching pathways for multi-product strain improvement [330].

To solve this complex optimization problem of getting the optimal geneknock-out strategy to obtain the GMO of desired characteristics, various multiobjective optimization strategies have been proposed, because it is still unclear how to combine and/or prioritize mutually competing objectives to achieve a true optimal solution, or how to select from the very large number of *Pareto optimal solutions* those that realizes the 'best' tradeoff for the cell designers' preferences. For a multiobjective optimization problem, a Pareto-optimal solution is one where any improvement in one objective can only take place at the cost of another objective. Consequently, for continuous variables, an infinity of Pareto-optimal solutions might exist [309]. On the other hand, a significant re-routing of flux directions and cycle fluxes are reported when switching from one objective to another within system constraints.

Following the review of Nagrath et al. [331] to find an optimal set of stationary net fluxes for a defined number of genes encoding the enzymes that participate in the considered metabolic reactions, several methods can be followed. One alternative is to use the linear programming (LP) to find the maximum of a weighted linear combination of fluxes of type  $\max(Wv)$  [ where vector 'W' includes the chosen weights  $v$ ; is the vector of stationary metabolic fluxes (reaction rates), including internal, transport, and the growth fluxes] [332,333]. However, in contrast to EMA and ExPA, only a single solution results in the end, even if additional linear constraints (other than the stoichiometric balance  $Sv = 0$ , where 'S' is the stoichiometric matrix including the stoichiometric coefficients of the metabolites in the reaction pathway) limiting the fluxes are added to the LP formulation. By varying the weights, or by applying an iterative weighting procedure, a still reduced number of Pareto-optimal solutions are usually obtained (especially when Pareto-frontier is non-convex). Besides, the weight selection (usually between 0 and 1 for scaled objectives) in association to physical meanings is difficult.

A similar route, the so-called '*Goal Programming*', uses sets of upper and lower weights to optimize the composite objective. Even if a larger number of Pareto-optimal solutions are thus obtained, the method suffers from the same disadvantages. Alternatively, the '*Linear Physical Programming*' method [332] replaces the a-priori prioritization of cellular objectives (fluxes) by successively relaxing (smoothing) the explicit flux constraints, from very strong ('*highly desirable*') to very soft ('*unacceptable*'), and by minimizing the weighted distances from the solution to the boundaries. The number of classes defining the preference degree for each objective is still a subjective decision, somehow equivalent to inspecting various weights in the multi-objective optimization but realized in a more comprehensive way. The procedure leads to a larger number of Pareto-optimal solutions due to the possibility to gradually relax the associated LP problem with an increased physical significance of the imposed constraints.

A similar transformation of the multi-objective optimization in a LP problem in the presence of stoichiometric, enzyme

maximum amounts, and solvent capacity constraints, was presented by Vera et al. [334]. By performing a nonlinear (logarithmic) transformation of the power-law reaction rate expressions (the so-called '*S-systems*'), the steady-state metabolic fluxes can be optimized vs. species concentrations and enzyme activities, by applying an evolutionary algorithm [335]. Even if attractive, the procedure requires the knowledge of kinetic expressions of the involved reactions, while the S-type representation of cellular processes, even if computationally convenient, suffer from a number of limitations [169,174].

Often several Pareto-optimal cell flux solutions are available for the same cellular system, representing design alternatives, from which one can subjectively selected to offer the '*best trade-off*' among multiple objectives. An important application of such a Pareto-optimal technique is the design of mutant cells (GMOs) by testing the effect of gene knockouts on the stationary metabolic fluxes of the cell. The computation rule is the following: the enzymatic flux that correlates to the gene that needs to be removed is given a constraint value of zero. Then, the reaction that the particular enzyme catalyzes is completely removed from the analysis. In such a way, a large number of strategies of gene deletions can be tested, by inspecting the feasibility of each problem solution (obtained GMO) in respect to certain constraints or objectives.

The simplest approach of solving this GMO design problem, by using an optimal gene knock-out (deletion) strategy is the FBA, able to inspect the successive deletion of only one gene [4,336], by retaining the optimal solution vs. a certain linear objective function. The simultaneous deletion of more than one gene leads to a very extended combinatorial problem, a total number of  $C_M^{KG}$  trials being necessary for removing KG genes from the total of M genes. For such a multiple gene removal case, a more sophisticated optimization procedure, described in this chapter, should be applied.

It is also to mention that while the gene knockout procedure is trying to drain the cell resources to the over-production of the desired metabolites by cutting alternative metabolic pathways, other proposed cell optimization techniques try to re-design complex regulatory circuits (GRC-s) to compensate the removed cell functions.

To solve such more difficult gene-deletion (knockout) *in-silico* problems, a completion to FBA is the so-called '*minimization of metabolic adjustment*' method (MOMA) [332]. MOMA employs quadratic programming to identify the closest point in the flux space to the FBA wild-type point, compatible with the gene deletion constraints. MOMA displays a significantly higher correlation than FBA, being of use for predicting the behavior of the perturbed metabolic networks, whose growth performance is in general sub-optimal. However, as observed by Wunderlich and Mirny [337], the FBA, EMA, and MOMA are unable to separate the role of topology and other parameters in the network objective function, while EMA is computationally very expensive and provide '*little insight into why certain mutations are lethal, whereas others are tolerated*'. They proposed to use the so-called '*total synthetic accessibility index*',  $s_i = \sum_j S_j$  evaluated from summing the synthetic accessibility of the outputs, defined as the minimal number of metabolic reactions needed to produce a component 'j' from the reaction network inputs. If an enzyme knockout does not change the index  $s_i$ , that is the biomass can be produced without extra metabolic cost, then this mutant is viable. If  $s_i$  becomes infinite, at least one essential component of the biomass cannot be produced from network inputs, and therefore the gene knockout leads to a lethal phenotype.

Finally, it is to observe that the topology plays a central role in determining network function and malfunction [337], and the viability of the sets of mutants. However, the gene knockout *in-silico* procedure through FBA is intrinsically incomplete as long as it is difficult to separate the contribution of topology from the contributions of kinetic and equilibrium characteristics of the cell CCM-system. Also, inferences among genes are not accounted for, while validation of FBA conclusions by means of simulating GRCs, and dynamic response to perturbations is necessary [20,174,280,312].

Because the main bioengineering objective through gene knockouts (usually maximization of production of a certain metabolite, i.e. the so-called '*dual problem*') is associated at a cellular level with a biological objective (usually maximization of the biomass yield, i.e. the so-called '*primal problem*'), one worthy alternative is to formulate the problem as a bi-level programming problem [310]. Such an approach is justified by the observation that the yields for some metabolites are far below their theoretical maximum given certain nutrient flux entering into the cell. Linear constraints of the bi-level optimization problem impose fixed substrate uptake, fulfillment of the network stoichiometric balance, upper/lower limits of fluxes, and other balance relationships derived from the physical meaning of the model variables. Following the mathematical rules, the '*primal*' problem, aiming at maximizing the bioengineering objective subjected to maximizing the cellular objective in the presence of linear constraints, is equivalent to solving the associated '*dual*' problem, of LP type, aiming at maximizing only one composed objective function in the presence of the original and additional constraints. When a gene knockout strategy is investigated, Boolean variables are

added to each flux, leading to a mixed-integer LP (that is the MILP) problem. By limiting the number of knockouts, the solution consists in a set of retained genes and the associated optimum values of stationary fluxes. Exemplification of this procedure ('*opt Knock*' of Burgard et al. [310]) is made for optimizing production of succinate and lactate in *E. coli* GMO cells by using the CCM extended model of Edwards and Palsson [4], accounting for 436 species involved in of 720 reactions (Figure 7-7). Even if effective, application of *opt Knock* requires solving a very large combinatorial problem, and multiple solutions might exist for the same multiple optimization objectives. Besides, the right choice of max/min boundary values of the auxiliary variables increases difficulties in applying this procedure.

The aim of this chapter is to present the original approach of Maria et al. [42] in using a **MINLP** numerical optimization procedure in solving the above formulated multi-objective cell metabolism optimization problem, in a nonlinear power-law formulation  $Max(\prod v_j^{\beta_j})$ , by including the adjustable influence of fluxes for reaching the composite goal [here  $v$  is the vector of stationary metabolic fluxes (reaction rates), including internal, transport, and the growth fluxes]. Even if being similar to the weighted multi-objective and goal optimization (if a logarithmic transformation is applied) [309], this procedure includes the possibility to account for nonlinear interactions among fluxes and nonlinear constraints without losing any property by math transformations. The **MINLP** procedure was then used to identify multiple gene deletion combinations that allow a maximum cell growth rate with maximizing externally imposed chemical (product) production targets, as an alternative to the unsatisfactory, and too computative combinatorial MILP procedure. The modified MINLP algorithm below presented is simple to be applied, without requiring specification of auxiliary variables, and easily extendable to solve a large variety of nonlinear multi-objective optimization problems of the same type, in a relatively simple way [73,334,338].

Exemplification is made here for the case of finding sets of knockout genes that will ensure simultaneous maximization of succinate (SUCC, target metabolite) and biomass production in *E. coli* GMO cells, by using the CCM model of Edwards and Palsson [4] (the reduced variant of (Figure 7-1, Figure 7-7, and Figure 7-8) accounting for 72 species (Table 7-1) involved in 95 reactions (Table 7-2). As multiple feasible solutions exist for this optimization problem, a step-by-step increase of the number of constraints might lead to a reduction in the gene knockout alternatives when using the MINLP formulation. Even if only linear constraints have been included in the tested case study, the nonlinear multi-objective formulation can be easily extended by accounting for flux interdependencies, the use of energy charge, carbon and nitrogen recoveries at steady state, or cell regulatory / thermodynamic properties" (chap. 4.2, 4.4, 4.5), thus allowing to reduce the number of viable GMO-s found as feasible solutions of the math problem. A step-by-step experimental-computational (LP) procedure to design and check GMO of interest was proposed by Orth et al. [339] (Figure 7-9).

### The reduced CCM model of Edwards and Palsson [4] for the wild-strain of *Escherichia coli*

In fact, "for a common micro-organism, the glycolysis together with the phosphotransferase (PTS)-system, or an equivalent one used for the GLC-uptake into the cell, and with the pentose-phosphate pathway (PPP), and with the tricarboxylic acid cycle (TCA), all these are part of the so-called central carbon metabolism (CCM), (Figure 4-23, Figure 7-1, Figure 7-8) [42].

The stoichiometric reduced model of the CCM of *E. coli* K-12 is those proposed by Edwards and Palsson [4], and Orth et al. [339]. The reduced variant includes 72 metabolites (Table 7-1) participating in 95 reactions (Table 7-2, and Figure 7-1), the stationary net fluxes being limited by specified minimum / maximum values (that is -1000 / +1000 mmol/gDW/hr.). The fluxes correspond to an equilibrated cell growth, with a glucose uptake rate of -10 mmol/gDW/hr, and an oxygen uptake rate of -1000 mmol/gDW/hr. The model was obtained by lumping the extended CCM dynamic model of Edwards and Palsson [4] that includes 720 reactions and 436 metabolites. In the extended variant, unconstrained uptake routes for inorganic phosphate, carbon dioxide, oxygen, sulphate, potassium, sodium, and ammonia are provided, and the capacity constraints were used to define the reaction reversibility. Lower limits for the internal fluxes were set to zero for all irreversible fluxes, and all reversible fluxes were upper bounded at a large value. Transport fluxes for metabolites not available in the media were always restricted to zero, while forward and backward reactions result in positive and negative fluxes respectively.

Due to the applied lumping procedure, the reduced CCM model approached here contains a lot of overall reactions that sum 'elementary' metabolic steps. For instance, the stationary rate of biomass production,  $v_{biomass} = v_{13}$ , results as a sum of many contributory steps leading to the overall stoichiometry given in (Table 7-2) marked with green.

In fact, for the present CCM optimization analysis, maximization of only two fluxes are of interest, that is: (obj1) maximum of biomass production,  $v_{biomass} = v_{13}$ , in (Table 7-2) marked with green.

(obj2) maximum of SUCC production,  $v_{succinate} = v_{39}$ , in (Table 7-2) marked with yellow.

Burgard et al. [310] solved this dual-optimization problem of succinate production maximization ( $v_{succinate}$ ), subjected to biomass production maximization ( $v_{biomass}$ ) in the presence of linear constraints of the pathway stoichiometry, based on the stoichiometric mass balance constraints under steady-state conditions, of type  $Sv = 0$  (where 'S' is the stoichiometric matrix including the stoichiometric coefficients of the metabolites in the reaction pathway, and  $v$  is the vector of stationary metabolic fluxes, including internal, transport, and the growth fluxes).

In such a way, Burgard et al. [310] obtained a large number of gene knockout solutions, such as: (a) removed genes no. {61,73}, or (b) removed genes no. {1,10,61,73} (Table 7-2) for the reaction/flux numbers leading to a GMO with a reduced CCM, but with  $v_{succinate} = 11$  mmol/gDW/hr, and  $v_{biomass} = 0.3$  1/hr.; or (c) removed genes {3,50,82,83} leading to a GMO with  $v_{succinate} = 15$  mmol/gDW/hr, and  $v_{biomass} = 0.16$  1/hr., etc. see the representation of some solutions in (Figure 7-2). On the other hand, the size of the used MILP combinatorial problem increases extremely much with the number of removed genes. Simulations and experiments also revealed existence of a non-linear relationship between  $v_{succinate}$  and  $v_{biomass}$ , that is large  $v_{succinate}$  (of maximum 16.4 mmol/gDW/hr. in the studied cell growth conditions) corresponds to a too small  $v_{biomass}$ , and vice-versa. It clearly appears that the two optimization objective (obj1), and (obj2) above formulated are opposite. Consequently, it was concluded that several sub-optimal solutions can exist when designing a mutant cell, according to the considered sets of removed genes."

This is why it is of high interest to solve this multi-objective optimization problem in a more systematic way. The below chapters are dealing with presenting the original and effective approach of Maria et al. [42].

### Formulate the optimization problem to maximize the SUCC and biomass production in a BR

When designing an optimal phenotype (GMO) for a cell of known characteristics (of a CCM with known relationships with the cell genome and proteome), "a bi-/multi-objective cell flux optimization problem should be formulated, by accounting for several goals [340]

- a) maximize ATP production to determine conditions of optimal metabolic energy efficiency.
- b) minimize nutrient uptake by determining the conditions under which the cell will perform its metabolic functions while consuming the minimum amount of available nutrients.
- c) minimize redox production by finding conditions where the cells operate to generate the minimum amount of redox potential, and minimum adenylate energy charge (ATP, ADP, AMP) necessary to draw the inner cell syntheses.
- d) minimize the Euclidean norm of fluxes, i.e. the sum of the fluxes allowing to channel the metabolites as efficiently as possible through the metabolic pathways.
- e) maximize target metabolite production, by optimizing the cell capabilities to produce a certain compound of practical (industrial) interest.
- f) maximize the biomass production ensuring the cellular network to evolve and proliferate.

It is to observe that the last two goals, i.e. *maximize the biomass and metabolite production*, are competing (opposite) objectives in a cell due to the requirement of using the resources for a maximum responsiveness to the environmental changes rather than for the overproduction of a specific compound. In a design GMO strain, a certain tradeoff should be realized between cell growth and forced metabolite production to preserve the cell growth and proliferation objectives.

Referring to the CCM model of (Table 7-1, and Table 7-2), the single level optimization problem can be formulated in terms of biomass production maximization ('cellular objective'), in the form:

$$[v_1, \dots, v_M] = \arg \text{Max } \Phi = v_{biomass} (= v_{13}) \quad (1)$$

$$\text{s.t. } \sum_{j=1}^M S_{ij} v_j = 0, \quad i = 1, \dots, N$$

$$v_{j,min} \leq v_j \leq v_{j,max}, \quad j = 1, \dots, M$$



In the eq.(1),  $M = 95$  fluxes and  $N = 72$  metabolites corresponds to the considered CCM metabolic system (the flux  $v_j$  numbering correspond to those of (Table 7-2, and Figure 7-1). The way to formulate the stoichiometric matrix ('S') and the species mass balance for the considered metabolic reactions under stationary conditions, in the form of a linear set  $S \mathbf{v} = \mathbf{0}$  is exemplified in the Chap. 7.7.

Other objectives can be considered when solving the CCM optimization problem, with a similar formulation, for instance a linear combination of target metabolic fluxes [4,332]. Successive solutions derived under various environmental conditions allow for instance to determine the correct sequence of byproduct secretion under increasingly anaerobic conditions [341]. When the same objective is associated with gene knockout alternatives, estimated fluxes can offer valuable information on the essential genes in the cell (lethality of gene knockout) [4,332]. Additional linear constraints to the optimization problem usually account for environmental requirements (e.g.nutrient limitation) [333].

When not only the 'cellular objective' (biomass) is optimized, but also the production of a certain metabolite (the so-called 'bioengineering / chemical objective'), a bi-level optimization problem results (e.g. by using succinate as target metabolite), of the following math form:

$$\begin{aligned}
 & \text{Primal: } [v_1, \dots, v_M] = \arg \text{Max } \Phi = v_{\text{succinate}} \\
 & \text{s.t. Max } v_{\text{biomass}} \\
 & \text{s.t. } \sum_{j=1}^M S_{ij} v_j = 0, \quad i = 1, \dots, N \\
 & v_{j,\min} \leq v_j \leq v_{j,\max}, \quad j = 1, \dots, M; \quad v_{\text{biomass}} \geq v_{\text{biomass\_target}}
 \end{aligned} \tag{2}$$

Where  $v_{\text{biomass\_target}}$  is the minimum level of biomass production which has to be realized by the 'optimized' cell. The notation 's.t.' (that is 'subjected to') denotes the problem constraints. The problem (2) unknowns are the CCM fluxes  $[v_1, \dots, v_M]$ . One alternative to solve this 'primal' problem is to transform it in an equivalent LP problem (called 'dual' problem), of the following form:

$$\begin{aligned}
 & \text{Dual: } [v_1, \dots, v_M] = \arg \text{Max } \Phi = v_{\text{succinate}} \\
 & \text{s.t. } v_{\text{biomass}} = \mu_{\text{biomass}} v_{\text{biomass\_target}} \\
 & \sum_{j=1}^M S_{ij} v_j = 0, \quad i = 1, \dots, N; \quad \sum_{i=1}^N \lambda_i^{\text{stoich}} S_{i,\text{biomass}} + \mu_{\text{biomass}} = 1 \\
 & \sum_{i=1}^N \lambda_i^{\text{stoich}} S_{i,j} + \mu_j = 0, \quad j \neq \text{biomass}; \quad \lambda_i^{\text{stoich}} \in \mathbb{R} \\
 & v_{j,\min} \leq v_j \leq v_{j,\max}, \quad j = 1, \dots, M; \quad v_{\text{biomass}} \geq v_{\text{biomass\_target}} \\
 & \mu_{j,\min} \leq \mu_j \leq \mu_{j,\max}, \quad (\text{for reversible reactions, but not secr\_only}) \\
 & \mu_j \geq \mu_{j,\min}, \quad (\text{for reversible reactions, and secr\_only}) \\
 & \mu_j \leq \mu_{j,\max}, \quad (\text{for irreversible reactions, and not secr\_only}) \\
 & \mu_j \in \mathbb{R}, \quad (\text{for irreversible reactions, and secr\_only}) \quad \text{Actif}
 \end{aligned} \tag{3}$$

In the eq.(3), the following notations have been used

'secr\_only' denotes transport fluxes for metabolites that can only be secreted from the network;

$\mu_j$  = dual variable associated with any restriction of the corresponding flux

$v_j$  in the primal problem;

$\lambda_i^{\text{stoich}}$  = dual variable associated to the stoichiometric constraints.

Such a transformation of the problem eq. (2) into the problem eq. (3) is possible due to observation that if the optimal solutions of primal and dual problems are bounded, their objective functions must be equal at optimality [342-345].

However, the previous LP problem eq.(3) raises several complications when additional constraints are formulated and/or extended cellular CCM kinetic/stationar model is approached, making the problem preparation very laborious, and computational very intensive to solved it. For instance, in the Burgard et al. [310] formulation, two additional constraints have to be added to the problem eq.(2) while finding the optimal gene knockout strategy. In this former case, the problem eq.(2) requires also searching for the optimal additional Boolean variables  $v_j \in \{0,1\}$  that multiply the fluxes (that is  $v_j v_j$  instead of  $v_j$ ) which, in turn, will complicate the formulation eq.(3). Besides, other LP algorithm disadvantages have to be mentioned, as follows: (i) the LP transformation of primal problem is valid only for the bi-level optimization and not for several objective functions (as usually is the case); (ii) exclusion of nonlinear constraints; (iii) effective solution of the dual LP problem requires the correct setting of upper/lower bounds of the dual variables, i.e.  $\mu_{j,\min}, \mu_{j,\max}$ ; (iv) when more than 2-3 genes are simultaneously removed from the CCM network, the resulted LP combinatorial problem becomes extremely computational intensive, and practically ineffective and impossible to be solved for cells including a large number of genes.

Consequently, to solve the multi-objective LP or NLP problems to obtain a GMO with a reduced number of genes but with optimized characteristics, various alternatives can be approached, such as: (a) min-max procedure (that is the maximum production of the target metabolite, with minimum violation of the stoichiometric constraints), or (b) single-composite function that includes all the objective functions multiplied by given weights [309], or (c) the Pareto-optimal front method (this chapter) [6,309,346-351]. There is no general approach for such a choice because the decision is case-dependent [334]. For instance, if the individual objective functions  $f_j$  are scaled in the same range, a composite objective function can be defined as a linear combination,  $\Phi_{LP} = \sum_j w_j f_j$  where the adopted weights usually satisfy the conditions  $\sum_j w_j = 1$  [4,309]. However, nonlinear combinations of individual  $f_j$  are also possible, depending on their physical significance [334].

For instance, one possibility retained for comparison is to transform the two objectives of the problem eq.(2) in a single level LP optimization by using the following composite function:

$$\begin{aligned}
 [v_1, \dots, v_M] &= \arg \text{Max } \Phi_{LP} = w_{succinate} v_{succinate} + w_{biomass} v_{biomass} \quad (= w_{39} v_{39} + w_{13} v_{13}) & (4) \\
 \text{s.t. } \sum_{j=1}^M S_{ij} v_j &= 0, \quad i = 1, \dots, N \\
 v_{j,\min} &\leq v_j \leq v_{j,\max}, \quad j = 1, \dots, M.
 \end{aligned}$$

The optimization problem eq.(4) considers the two main objectives, that is

**(obj1)** maximum of biomass production  $v_{biomass} = v_{13}$ , in (Table 7-2) marked with green. This macro-scale variable is related to the performances of a BR, or another bioreactor.

**(obj2)** maximum of SUCC production,  $v_{succinate} = v_{39}$ , in (Table 7-2) marked with yellow.

The weights  $\{w_{39}$ , and  $w_{13}\}$  in the composite function  $\Phi_{LP}$  of eq.(4) are adopted according to the relative importance given to the two oposite objective functions (that is (obj1), and (obj2)).

For the present case study, Maria et al. [42] investigated another route to achieve the multi-objective optimization of cell fluxes, by formulating a nonlinear programming (NLP) problem, with using a power-law type composite objective function, of the following form:

$$\begin{aligned}
 (5) \quad [v_1, \dots, v_M] &= \arg \text{Max } \Phi_{NLP}; \quad \Phi_{NLP} = \prod_{j=1}^M v_j^{\beta_j}, \quad \beta_j \in R \\
 \text{s.t. CONSTR:} & \\
 \sum_{j=1}^M S_{ij} v_j &= 0, \quad i = 1, \dots, N \quad \text{(or } \mathbf{Sv} = \mathbf{0}) \\
 v_{j,\min} &\leq v_j \leq v_{j,\max}, \quad j = 1, \dots, M
 \end{aligned}$$

The individual fluxes can be included ( $\beta_j \neq 0$ ) or not ( $\beta_j = 0$ ) in the optimization rule, in an (un)scaled form, with the exponent  $\beta_j$  sign and magnitude depending on the maximization / minimization goal, and on its relative importance in the CCM metabolism. As another observation, when some fluxes (index 'exp') are measured, a supplementary equality constraint should be added to eq.(5), of the form,  $\mathbf{S}_{unk} \mathbf{v}_{unk} = -\mathbf{S}_{exp} \mathbf{v}_{exp}$  where 'unk' index denotes the unknown vector of fluxes.

In fact, if a logarithmic transformation is applied to the goal function  $\Phi_{NLP}$  in eq.(5), an equivalent weighted multi-objective LP optimization problem of type 'MAX(  $wv$  )' is obtained. However, the NLP formulation is not fully equivalent with the LP one, as long as the linear transformation distorts the flux contribution to the main goal, and it can not represent the nonlinear interdependencies among fluxes. For instance, various nonlinear objectives of power-law type can be formulated according to the desired modification of phenotype, e.g.: (i) maximum of  $v_{succinate}$  with minimum of  $v_{biomass}$ , that is 'Max  $\Phi_{NLP} = v_{succinate}^{\beta_5} / v_{biomass}$ ' (ii) maximum of a series of scaled fluxes  $v_k / v_l v_m$ , in the composite form of

$\text{Max } \Phi_{NLP} = (v_k / \|v\|_2) (v_l / \|v\|_2) (v_m / \|v\|_2) \dots$ , where  $\|v\|_2$  denotes the Euclidean norm of the vector 'v' including all the CCM fluxes; (iii) maximum of a certain metabolic flux  $v_k$  related to the corresponding overall production of entropy into the cell, that is  $\text{Min } \Phi_{NLP} = (-v_k) / \sum_j (v_j A_j / T)$  (where:  $A_j$  = reaction affinity, T = absolute temperature). For other nonlinear objectives and constraints of the above mentioned CCM optimization problem, the reader is referred to the work of Heinrich and Schuster [73]. Besides, the advantage of the NLP formulation consists in the possibility to also include nonlinear constraints derived from the imposed properties of the metabolic pathway, by accounting for the gene inferences [313], for regulatory network properties [21-23,174] (chap. 4.4, 4.5), or other thermodynamic properties [330].

In the present case study, Maria et al. [42] used a particular form of the NLP problem eq.(5) for maximizing both the succinate and biomass positive fluxes, of the form:

$$[v_1, \dots, v_M] = \arg \text{Max } \Phi_{NLP} = \arg \text{Min } (-\Phi_{NLP}) ; \Phi_{NLP} = v_{succinate}^{\beta_5} v_{biomass} \cdot \beta_5 > 0 \quad (6)$$

The degree of freedom (DF) of the NLP problem depends on the number  $NEQ < N$  of equality constraints of the stoichiometric  $Sv = 0$  balance set (here N = number of metabolites in the considered CCM metabolic pathway). The constraints are accounted during the solution search in a simple way, for instance by evaluating the constraint violation degree by means of two original indices:  $CV_{eq}$  and  $CV_{ineq}^2$ , defined as followings:

$$CV_{eq} = \left( \sum_{i=1}^N \sum_{j=1}^M S_{ij} v_j \right) ; \overline{CV}_{eq} = CV_{eq} / N ; CV_{eq}^2 = (CV_{eq})^2 \quad (\text{equality constraints}),$$

$$CV_{ineq}^2 = \sum_{j=1}^M C_{j,up}^2 + \sum_{j=1}^M C_{j,low}^2 ; \overline{CV}_{ineq}^2 = (CV_{ineq}^2) / M \quad (\text{inequality constraints}), \quad (7)$$

where:

$$C_{j,up}^2 = \begin{cases} (v_j - v_{j,max})^2, & \text{if } v_j > v_{j,max} \\ 0, & \text{else} \end{cases}$$

$$C_{j,low}^2 = \begin{cases} (v_{j,min} - v_j)^2, & \text{if } v_j < v_{j,min} \\ 0, & \text{else} \end{cases}$$

When solving the NLP problem eq.(6), the constraint violations 'penalize' the objective function by means of an extended Lagrange function, of the form:

$$[v_1, \dots, v_M] = \arg \text{Min } L = -\Phi_{NLP} + \lambda_{eq} CV_{eq}^2 + \lambda_{ineq} CV_{ineq}^2 \quad (8)$$

where the Lagrange multipliers  $\lambda_{eq}$ ,  $\lambda_{ineq}$  are chosen to be zero if the constraints are not violated, and receive positive values (constant, or increasing / decreasing numbers according to the search failure / success) [346,352]. The constraints are scaled according to the objective function range. In the present study, as  $\Phi_{NLP}$  in eq. (6) is the product of two cellular fluxes, squared indices of constraint violation are included, with a uniform weight of  $\lambda_{eq} = \lambda_{ineq} = 1$

To make our results comparable to similar case studies from literature, the considered constraints in the optimization problem of biomass and target metabolite SUCC production maximization are those indicated by Burgard et al. [310] for the extended CCM model, but adapted to the reduced CCM model structure of (Table 7-1, Table 7-2, and Figure 7-1), that includes the glucose uptake balance, maintenance requirements, and the minimum level of biomass production. In math terms, such requirements are translated as followings (with the notations of (Table 7-2):

$$\begin{bmatrix} v_{pts} + v_{glc} = v_{glc\_uptake} \\ v_{biomass} \geq v_{biomass\_target} \\ v_{atp} \geq v_{atp\_maintenance} \end{bmatrix} \Rightarrow \begin{bmatrix} v_{pts} = v_{glc\_uptake} \\ v_{biomass} \geq v_{biomass\_target} \\ v_{atp} \geq v_{atp\_maintenance} \end{bmatrix} \Rightarrow \begin{bmatrix} |v_{28}| = |v_{50}| \\ v_{13} \geq v_{13}^{min} \\ v_{12} \geq v_{11} \end{bmatrix} \quad (9)$$

where: M = the number of considered fluxes (stationary reactions) in the CCM metabolic network (that is 720 in the extended model, and 95 in the reduced model); N = number of metabolites (that is 436 in the extended model, and 72 in the reduced

model);  $v_j$  = generic metabolic flux;  $S_{ij}$  = stoichiometric coefficient of the metabolite 'i' in the reaction 'j'. In the previous formulations,  $v_{glc\_uptake}$  is the basic glucose uptake scenario,  $v_{atp\_maintenance}$  is the non-growth consumption flux associated to the ATP maintenance requirement, while  $v_{biomass\_target}$  is the minimum level of biomass production imposed by the GMO designer (theoretically being higher than zero, or even zero) [310]. Adaptation of these metabolic constraints reported some modified relationships due to the used lumped species and reactions. Thus, the equality constraint [ $v_{pts} + v_{glk} = v_{glc\_uptake}$ ] refers to phosphotransferase ( $glc-D[e] + pep[c] \rightarrow g6p[c] + pyr[c]$ ) and glucokinase ( $[c] : atp + glc-D \rightarrow adp + g6p + h$ ) fluxes of the extended CCM model. In the reduced CCM model  $v_{pts}$ , refers to GLC import via PTS-reaction (system) in (that is the  $v_{30}$  flux in (Table 7-2). As long as the *E. coli* core CCM reduced  $m^{j-1}$  does not include the  $glk$ /glucokinase flux, this constraint reduces to [ $v_{pts} + v_{glk} = v_{glc\_uptake}$ ] constraint as long as  $v_{glk}$  is not accounted for. The constraint [ $v_{biomass} \geq v_{biomass\_target}$ ] was implicitly considered during the MINLP optimization problem eq. (7-9) of maximum biomass production so,  $v_{biomass\_target}$  in the constraints eq. (9) is usually set to zero."

### Solving the FBA optimization problem to determine various gene knockout strategies to design optimized GMO-s

"To determine possible optimal GMO phenotypes of the analysed *E. coli* microorganism, Maria et al. [42] applied a FBA coupled with the multi-objective optimization of type eq.(7-9), by estimating the optimal stationary fluxes [ $v_1 \dots v_M$ ] associated with a proposed *sub-set* of genes encoding the enzymes participating to the obtained reduced CCM metabolic pathway of the design GMO (due to the applied gene knock-out procedure). To point-out the importance of the number and structure of the problem constraints, a step-by-step strategy to identify the optimal *E. coli* mutant for biomass and succinate production was developed. By setting the number KG of genes which have to be removed from the cell, the basic MINLP rule consists in simultaneously finding the removed genes and the optimal fluxes, in a problem formulation similar to eq.(6-8), that is:

$$[v_1, \dots, v_M, y_1, \dots, y_M] = \arg \min \Phi_{NLP} = -\Phi_{NLP} + \lambda_{eq} CV_{eq}^2 + \lambda_{ineq} CV_{ineq}^2$$

s.t. CONSTR: (10)

$$\sum_{j=1}^M S_{ij} v_j = 0, \quad i = 1, \dots, N;$$

$$v_{j,min} y_j \leq v_j \leq v_{j,max} y_j, \quad j = 1, \dots, M; \quad y_j = \{0,1\};$$

$$\sum_{j=1}^M (1 - y_j) = KG; \quad y_{13} = y_{biomass} = 1; \quad y_{39} = y_{succinate} = 1$$

When one gene (and its encoded enzyme, index 'j') is removed from the cell CCM, the associated flux ( $v_j$ ) is also omitted from the reaction pathway, by simply setting to zero their associated Boolean variable  $y_j = 0$ . The obtained GMO *E. coli* mutant presents optimized fluxes from the point of view of the engineering requirements [the above (obj1), and (obj2)], but not necessarily will preserve the main cell functions that ensure the cell maintenance and survival. Consequently, every thus derived math solution has to be metabolically viable, being

interpreted as physical meaning before its validation."

### Solving gene knockout LP, or MINLP problem by only accounting for the stoichiometric constraints:

One starts to solve the optimal GMO design problem eq. (6-9) by finding optimal fluxes in the *E. coli* cell in the presence of only basic stoichiometric balance set constraints  $S v = 0$  of the CCM metabolic pathway given in (Table 7-2). "The stoichiometric matrix 'S' of size [N = 72 species × M = 95 reactions], in the Matlab™ code is presented in (Table 7-4).

The problem is solved by using both the LP formulation eq. (4) (with the adopted weights  $w_{succinate} = w_{biomass} = 1$ ), and the NLP formulation eq.(5) with the CONSTR set of constraints and  $\beta_s = 1$ . The results, presented in (Figure 7-3), indicate practically the same (unique) solution, irrespective of the used method. As expected, the resulted very large values for the reversible succinate-to-fumarate transformation (fluxes #44 and #89) is the main reaction responsible for succinate production maximization. But this theoretical solution does not necessarily ensure the cell viability, and other dynamic / thermodynamic constraints should be further considered.

When genes are *in-silico* knockout from the cell CCM, the optimal GMO. design problem solution ceases to be unique, and several alternatives might exist. Theoretically, a number of  $C_M^{KG}$  mutant cells results from removing **KG** genes from the total of  $M$ . For  $KG = 1$ , the number of knockout trials equals the number of genes ( $M$ ) controlling the CCM. This number increases sharply with  $KG$ , following an approximately power law given by the Stirling formula  $N! = (N/e)^N \sqrt{2\pi N}$  [353].

If one successively removes one gene (from gene #1 to gene #95,  $KG = 1$ ), every time evaluating the optimal fluxes with the criterion eq. (4), the obtained optimal LP solutions are presented in the (Figure 7-4-a), by using the LP solver of Matlab™ math computational platform [347]. It is to observe that a large number of

alternatives exhibit the same performance index, that is  $\Phi_{LP} = v_{succinate} + v_{biomass} = 16.384$  mmol/gDW/h, even if two additional constraints ( $|v_{28}| = |v_{50}|$ , and  $v_{12} \geq v_{11}$ ) and have been added (Figure 7-4-b). This result is similar to those of Edwards and Pals son [4] suggesting that 'a large number of the central metabolic genes can be removed without eliminating the capability of the metabolic network to support growth under the conditions considered', due to the interconnectivity of the metabolic reactions.

The same single-gene knockout rule was repeated by using the MINLP.criterion eq. (10) with the basic stoichiometric constraints included in the Lagrange function 'L'. The results, presented in (Figure 7-5), indicate the same conclusion as those obtained from using the LP criterion, that is a large number of genes can be removed by keeping the succinate production at the highest level of  $v_{succinate} = 16.384$  mmol/glhf (Figure 7-5-b). The inequality constraints are all the time fulfilled, while the violation index  $\overline{CV}_{eq}$  of equality constraints is roughly negligible (more precise solutions are possible but with the expense of a significant supplementary computational effort). Slightly violation of equality constraints ( $\overline{CV}_{eq}$  index) is sharply penalized by large Lagrange functions  $L$  in the (Figure 7-5-c). It is to remark that many removed genes lead practically to the same optimal fluxes into the cell, as for instance the removed genes #1 and #10 in (Figure 7-6) (corresponding to the reactions #1 and #10 in (Figure 7-1). These removed reactions block the EtOH production (an unnecessary metabolite), being also removed in one of the mutant *E. coli* cells obtained, and experimentally validated by Burgard et al. [310].

To investigate multiple gene knockout alternatives, two or four genes have been concomitantly removed from the CCM pathway when solving the MINLP problem eq. (6-9). It is to observe that the number of possible solutions of the same quality increases very much (by displaying almost the same objective function 'L') 229 Some of the *in-silico* mutant *E. coli* cells obtained by means of MINLP criterion eq. (10) are presented in (Table 7-3). As marked in the (Figure 7-1) some of the solutions are expected, for instance the use of cell resources for succinate production maximization by blocking formation of ethanol (removed fluxes #1 and #10), lactate. (removed flux #61), formate (removed flux #73), glutamine (removed fluxes #51 and #54) etc. (see some of these solutions in the Figure 7-2). Also, from the mathematical point of view it appears that F6P production by two alternative routes (#50 and #45) is redundant, and one of them should be removed. Such multiple solutions require a physical meaning evaluation to check viability of each *in-silico* obtained GMO (mutant) *E. coli* cell. For instance, the removed gene set # (82,83) is not viable as long as the flux #82 is responsible for the PYR production, and its elimination will not ensure the essential energetic TCA pathway (Krebs cycle) of the cell. As the visual inspection of a larger number of solutions is difficult for complex cell system cases, an automatic rule is preferable, by using, for instance, the Wunderlich and Mirny [337] synthetic accessibility concept to identify unfeasible cases of non-viable cells CCM pathways when products cannot be synthesized from the network inputs. As another observation, Burrard et al. [310] have found several Pareto-optimal GMO *in-silico* solutions that maximize the succinate production by eliminating the oxygen uptake reactions (#36, #70). This alternative was not identified by our procedure after a significant large number of trials, probably due to the reduced form of the CCM model requiring the use of oxygen in the Q8 production (#16), which is essential for the here considered CCM metabolism.

From the numerical point of view, solving the associated MILP, (eq. (4) plus CONSTR of eq. (10)) or LP (eq. (4)) multi-objective problem leads to an extended combinatorial calculus when optimizing the sum of fluxes with also removing certain genes from the CCM. The advantage of using the MINLP formulation eq. (6-9) is coming from the concomitant random search for optimal fluxes and gene knockout alternatives during the same iterative rule, with the risk of missing gene knockout alternatives of similar quality in terms of the objective function. This risk can be reduced when additional constraints are added to the optimization problem, or when a suitable flux prioritization (by means of the exponents  $\beta_j$ ) is formulated.

Despite the large number of advantages, the proposed MINLP formulation inherently presents some limitations. Similarly, to the weighted multi-objective and goal optimization cases [29,309], the proposed criterion leads to a reduced number of Pareto-

optimal GMO solutions (see the extended discussion of Nagrath et al. [331]). Two such Pareto-optimal solutions, corresponding to the removed gene set (1,10,61,73) are displayed in the (Figure 7-2). An increased number of solutions can be obtained if a repeated application of the MINLP procedure is performed by using different sets of weights / exponents (varied within certain limits), with the expense of a considerable supplementary computational effort."

### Solving Gene Knockout LP or MINLP Problem by Accounting for the Stoichiometric Constraints, and Some Additional Constraints

To check the effect of introducing new constraints to the MINLP optimization problem in order to reduce the large number of gene knockout set of solutions, three additional constraints eq. (9) suggested by Burgard et al. [310] are added to the multi-objective optimization problem eq. (10). "Some of the MINLP solutions, for the case of 2, or 4 simultaneously removed genes from the wild-type CCM of the *E. coli* cell, are presented in (Table 7-3). The results reveal several conclusions, as follows:

a) Multiple GMO-s are obtained for the *E. coli* as possible gene knockout alternatives from its CCM, in order to simultaneously maximize the biomass and succinate production, valid for removing one, two, four or more possible genes from the CCM metabolic network. When deriving the Pareto-optimal front solutions (two of them are represented in (Figure 7-2) for the all removed genes #1,#10,#61,#72), the slow convergence of the used MINLP algorithm may lead to approximate (sub-optimal) solutions also, as indicated by the approximate fulfilment of the equality constraints, with an acceptable average error of [ $10^{-6}$  up to 1.5] flux units compared to the feasible range of fluxes [-1000, +1000]. However, the GMO sub-optimal solutions can be easily identified from the constraint fulfilment analysis and eventually removed. The solver used by the MINLP algorithm was the adaptive random search of Maria [29,71,72], and a modification of the Nelder-Mead algorithm (routine 'FMINSEARCH' of the Matlab computational platform [347]. The use of other solvers (e.g., Evolutionary algorithms [71,72]) might improve the solution quality, but with the expense of a much larger computational effort.

b) By inspecting the most relevant GMO solutions given in (Figure 7-2, and Table 7-3) it is to observe that high biomass production rates correspond to low production in succinate, and vice-versa. As previously mentioned, these two optimization objectives are opposites. These results obtained by Maria et al. [42] are in a perfect agreement with the experimental data and findings of Burgard et al. [310] from (Figure 7-2). The obtained Pareto-optimal front of (Figure 7-2) displays a nonlinear dependence between the two opposite stationary fluxes, that is:

(obj1). maximum of biomass production  $v_{biomass} = v_{13}$ , in (Table 7-2 marked with green).

(obj2). maximum of SUCC production,  $v_{succinate} = v_{39}$  in (Table 7-2, marked with yellow).

For instance, small values of  $v_{biomass}$  corresponds to ca.  $v_{succinate} = 17$  mmol/gDW/hr.

c) The analysis of the physical meaning of CCM fluxes in the design GMO *E. coli* can indicate the viability of the optimization problem solution (design GMO). As previously mentioned, and marked in (Figure 7-1), succinate. production maximization can be achieved by removing some metabolic unessential steps, e.g., synthesis of ethanol, lactate, formate, glutamine, etc. which is in a fully agreements with the in-silico results of Burgard et al. [310] with using an extended cell CCM model (i.e., removed genes no. 1, 3, 10, 50, 61, 73, 82, 83)." However, such a check of cell viability for each of the *in-silico* obtained GMO is extremely laborious, and an automatic rule might be adopted (for instance, the 'synthetic accessibility rule' of Wunderlich and Mirny [337]). Other imposed constraints to the CCM-system can also be used in this respect, as previously discussed.

### Solving Gene Knockout MINLP Problem with Imposing a Minimum Limit for the Succinate High Production

The conflicting succinate and biomass production in *E. coli* cells indicates that several mutant cell solutions can be obtained, some with large  $v_{succinate}$  and low  $v_{biomass}$  levels, and vice-versa. To design cells with a high succinate production rate rather than a high biomass production rate, one alternative is to 'artificially' introduce a constraint to the MINLP problem, by defining a lower limit for the succinate production rate, e.g.,  $v_{succinate} = v_{39} > v_{39}^{min} = 10$  mmol/gDW/hr. The results indicate a still large number of possible GMO-s with reduced CCM. networks, of close efficiency. Some of the obtained *E. coli* cell design GMO-s with

these supplementary constraints are presented in (Table 7-3), when removing 2 or 4 genes from the CCM. For instance, the solution with removed genes # {91and22} [ $v_{39}, v_{13}, L$ ] = [12.6,  $2.90 \cdot 10^{-12}$ , 416] is close to the solution with removing genes # {61,73} [ $v_{39}, v_{13}, L$ ] = [16.5,  $2.32 \cdot 10^{-12}$ , 484]. However, the most plausible GMO solution is that corresponding to the removed genes # {61,73}, thus blocking the production of LAC and FOR without disturbing the main metabolic pathway (as is the case when removing the flux #91 of (Figure 7-1).

An alternative 'artificial' thresholds for some fluxes of the CCM to 'force' a certain GMO solution, is to increase  $\beta_s = 2^{\beta_j \neq 0}$  of some fluxes in the composite MINLP objective function eq. (5-6). For instance, by adopting in eq. (6), other *E. coli* GMO cell strains can be obtained (two of them are displayed in (Table 7-3). Not only derivation of mutant cells of high succinate productivity is thus favored, but also the precision of the Pareto-optimal solution is roughly doubled (i.e., smaller  $\overline{CV}_{eq}$  index) with using the same computational effort. The benefit of such an adjustable flux weight in achieving the composite goal is obvious, by the expense of requiring an extended computational effort when investigating the adjustable relative prioritizations of fluxes in the MINLP formulation. Comparatively to the LP approach, the advantage of MINLP with including nonlinear correlations among fluxes and constraints, recommends the MINLP approach as a worthy algorithm for designing GMO cells with desired optimized characteristics reporting higher performances of a biological BR.

### Conclusions About the Use of HSMDM for the Optimal Design of GMO-S with Desired Characteristics

Application of a HSMDM model, including an extended structured CCM model of the analysed microorganism, and of a MINLP numerical procedure to *in-silico* find theoretical gene knockout alternatives (that is GMO-s) that optimize several formulated objectives (for instance, to maximize the production of a target metabolite, like SUCC here, and of the biomass in a BR) has been proved to be a very promising and effective numerical tool for *in-silico* design mutant cells with desired characteristics. The MINLP computational strategy described in this chapter can overcome the time-consuming complex combinatorial MILP problem solver (that corresponds to a multi-layer LP formulation) and can save considerable computing time by superposing the knockout rule to the basic NLP optimization approach. However, the identified multiple solutions of the MINLP problem, explained by the cell metabolism complexity, must be further 'filtered' by adding supplementary (non)linear constraints, other than the regular stoichiometric ones, leading to a considerably reduction in the number of gene-knockout alternatives.

The use of the LP formulation  $Max(wv)$ , with subjective weights 'w' allocated to the target fluxes 'v', or transformation of primal problem in a dual LP problem (for only bi-level optimization) is laborious, requiring formulating and prepare the derived dual problem. Moreover, the optimization problem solution is dependent on the adopted upper/lower bounds of the dual variables, while the resulting LP combinatorial problem when removing several genes from a large possible number becomes very competitive. Recent improvements of the LP numerical algorithm led to a better description of the optimal solution set, but do not overcome the very competitive combinatorial problem.

The advantage of the proposed power-law type MINLP multi-objective function eq. (5-6,10), and of a structured HSMDM model (with including an extended CCM stationary/dynamic model) is coming from the possibility to account, in a simple way, for the flux nonlinear interactions and complex constraints as also mentioned in the literature [333,334,338]. The combinatorial rule is included. in the iterative MINLP solver, while the larger number of considered (nonlinear) constraints can increase the chance to obtain a reduced set of feasible gene knockout / GMO solutions for a given metabolic CCM network. The preferred MMA random search of Maria [71,72] can offer a higher reliability in finding a global solution (if any) of the optimal-flux-gene-knockout problem, with also providing the opportunity for the integer variables to span their range of possible values during the search of optimal metabolic fluxes. In such a manner, a continuous evaluation of the effects of removing various genes during the MINLP solver iterations is realized. Because the random searches are usually slowly convergent near the problem solution, approximate solutions are usually retained, with an acceptable precision of fulfilling the problem constraints. However, derivation of a larger set of optimal. GMO solutions by using an adjustable relative prioritization of fluxes in the MINLP formulation will lead to an extended computational effort.

In any variant, the resulting multiple gene knockout GMO solutions have to be validated from several points of view, both theoretically (physical meaning) and experimentally. To reduce the number of MINLP solutions, careful formulation of the optimization problem constraints is crucial. Gene inference [312], feasible reaction pathways [337], or any information on protein-gene interactions, and on the genetic regulatory circuits (GRC, chap. 4.2-4.5) can be used for such purposes. On the other

hand, several criteria to check the design optimal GMO cells for viability can be used, for instance the Wunderlich and Mirny [337] synthetic accessibility index, or additional information on the gene inferences.

A more systematic rule for designing mutant cells should be based on using hybrid stationary dynamic HSMDM models to incorporate stationary and kinetic information on the metabolic fluxes distribution over time.

As will be discussed in the chap. 8 and proved by the examples (case studies) of this book, the use of HSMDM models with incorporating structured dynamic/stationary models (deterministic, with continuous variables) of the cell CCM, offers multiple advantages in solving bioengineering and bioinformatic problems, such as:

a) *In-silico* (math model based) simulation of the bioreactor dynamics with an increased accuracy, simultaneously at a macroscopic- scale (bioreactor state variables), but also at a cell-scale (i.e. the main metabolites of the CCM). Such an advantage of the HSMDM models can immediately be used for various engineering analyses, such as bioreactor design, optimization, control, and can direct the *in-silico* research to design a GMO of optimized characteristics.

b) Simulation of the mercury operon (*mer*-GRC) expression and relate the dynamics of *mer*-GRC cell-variables to those of the semi-continuous (SCR), three-phase fluidized (TPFB) bioreactor dynamics, in order to optimize the mercury uptake from wastewaters, and to predict the self-adaptation of the cell metabolism over dozens of cell cycles to the variable conditions of the bioreactor (case study no.1, chap. 5). *In-silico* design of *E. coli* GMO (cloned with *mer*-plasmids) to optimize the SCR bioreactor performances.

c) Simulation of the TRP-operon (tryptophan) expression and relate the dynamics of *trp*-GRC cell-variables to the dynamics of cell-metabolites involved in the CCM, and to the macro-scale dynamics of the fed-batch (FBR) bioreactor variables, in order to optimize the FBR operating policy, and to predict the increase of the TRP-productivity due to the use of a design GMO (case study no.2, chap. 6).

d) The use of a hybrid CCM cell-scale structured kinetic model coupled with a BR dynamic model to maximize both biomass and succinate production by the *in-silico* design of GMO *E. coli* cells, with using an optimal gene knock-out numerical strategy (case study no.3, chap. 7).

## Annex - Metabolic Stoichiometric Balance

Chemical and biochemical kinetics are based on the postulate that a reaction rate  $v_j$  can be expressed as a unique (usually nonlinear) function of the involved species concentrations  $c_i$ , of all participating chemical species at a certain time 't'. When (bio)chemical reactions are the only cause of concentration changes, that is the transport processes are negligible, the concentration dynamics, in a reacting system of constant volume, is given by the following mass balance ODE set [173].

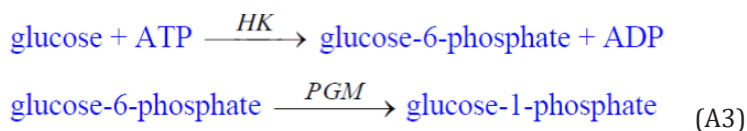
$$\frac{dc_i}{dt} = \sum_{j=1}^M S_{ij}v_j, \quad i = 1, \dots, N \quad (A1)$$

When the (bio)chemical system subsists in a steady state, the balance equation (A1) in a matrix formulation becomes:

$$S v(c^*) = 0 \quad (A2)$$

Where '\*' superscript denominates the steady-state values of the species concentrations. To exemplify the way to relate the stoichiometric matrix 'S' to the steady state reaction rates (denominates as fluxes), the following metabolic reactions are considered [73].





These two reactions are catalyzed by the enzyme's hexokinase (HK, EC 2.7.1.1) and phosphoglucomutase (PGM, EC 5.4.2.2), respectively. By attaching the stoichiometric matrix of the two reactions (indexed with 1 and 2), the steady-state mass balance (A2) for the system (A3) can be written as:

$$S v = \begin{pmatrix} HK & PGM \\ gluc & -1 & 0 \\ G6P & 1 & -1 \\ G1P & 0 & 1 \\ ATP & -1 & 0 \\ ADP & 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (\text{A4})$$

In the CCM mass balance, all the involved reacting species have been included. To be feasible and consistent, the mass balance has to be completed with the input/output fluxes into/from the system (see (Figure 4-40), and (Figure 4-41), and [21-22]). For larger biochemical systems, this procedure is neither necessary nor useful, allowing checking the molecular mass conservation relationships, to derive large models, and to study some properties of the system, related to the conservation relations and the reaction invariants, as long as the stationary flux vector is the null space of the matrix 'S' [67,73].

## Chap. 8. Conclusions and perspectives of using HSMDM for solving complex engineering problems

The HSMDM kinetic models presented in the case studies no. (1-3) of this book are structured and are **hybrid** because they are extended on *two levels* of the bioprocess representation: (1) one on a nano-scope, cellular scale, which uses as state variables the concentrations of the cell key-species; (2) the other is on a macro-scope scale of the bioreactor bulk-phase, which refers to the dynamics of the bioreactor state variables (i.e. the directly measurable concentrations of metabolites in the liquid phase, by using classical analytical methods). It should be noted that, in such a hybrid kinetic model, the cellular state variables are directly connected to those of the bioreactor through the cellular import/export (excretion) processes of substrates / nutrients / metabolites through the cell membrane. While, the dynamics of macro-molecular species, necessary for the estimation calculations of the hybrid kinetic model, can be easily measured by using current analytical techniques, by contrast, the dynamics of species (metabolites) at a cellular level are much difficult to measure, both at QSS or under dynamic conditions, although there have been remarkable recent advances in this field [227].

These advances allow the simultaneous measurement of a large number (600) of cellular metabolites and to monitor their evolution over time. These refined measurements allow evaluation of metabolic fluxes, under QSS or dynamic conditions, of an extended/reduced CCM scheme presented in (Figure 8-1), or (Figure 4-1 and Figure 4-23), by using the classical calculation method of Stephanopoulos et al. [188] schematically figured in (Figure 8-3). The three case studies presented and extensively discussed in this book proved that HSMDM structured **hybrid models** are valuable tools to be used in engineering/bioinformatic analyses for various purposes, such as:

1. Simulation of the bioreactor dynamics with an increased accuracy, simultaneously at a macroscopic (bioreactor state variables) scale, but also at a cell-scale (for the main metabolites of the CCM). Thus:

- 1.1. In the case study (no. 1)(chap. 5), the HSMDM structured hybrid model is able to predict the dynamics of [26(cell

species) + 3(bulk species)] vs. only [3 (bulk species)] by a classical macroscopic SCR bioreactor model, while covering a wider range of input substrate [Hg<sup>2+</sup>] loads, with using design cloned *E. coli* cells with various amounts of mercury-plasmids [Gmer] of (3-140 nM). Eventually this HSMDM model was used to *in-silico* design optimally cloned *E. coli* cells able to optimize the mercury removal from wastewaters. Eventually this HSMDM model was used by Maria and Luta [40] to derive the optimal operating policy of a FBR leading to maximize the mercury removal from wastewaters.

1.2. In the case study (no. 2) (chap. 6), the HSMDM structured hybrid model is able to predict the dynamics of [11(cell species) + 4(bulk species)] vs. only [3 (bulk species)] by a classical macroscopic FBR model, while covering a wider range of control variables (feed flowrates of the GLC solution), and various GMO *E. coli* cells strains. The extended bi-level hybrid kinetic model was proven to fairly represent the dynamics of an experimentally studied FBR under a nominal (uniform feeding) operating policy, for both macroscopic state variables and for the cell key species of the CCM reaction modules related to the TRP production and excretion in the FBR bulk, that is: [A] glycolysis, [B] ATP recovery system, [C] TRP operon expression, and biomass [X] growth. This extended HSMDM model was validated by using the recorded data from a lab-scale FBR over a long batch time (63 h). Eventually this HSMDM model was used by Maria and Renea [12] to derive the optimal operating policy of a FBR leading to TRP production maximization.

1.3. In the case study (no. 3) (chap. 7), the HSMDM structured hybrid model is able to predict the metabolic fluxes dynamics of [72(cell species), involved in 95 reactions + 2(bulk species, i.e. biomass, SUCC)] vs. only [2 (bulk species)] by a classical macroscopic BR model, while covering a wider range of control variables (GLC in the bioreactor), and various GMO *E. coli* cells strains. The extended bi-level hybrid kinetic model was proven to fairly represent the dynamics of a studied BR under a nominal operating policy, for both macroscopic state variables (GLC, SUCC, X) and for the cell key species of the CCM reaction modules related to the succinate (SUCC) production and excretion in the BR bulk. This extended HSMDM model was experimentally validated by Edwards and Palsson [4]. Eventually this HSMDM model was used by Maria et al. [42] to optimize a BR, by simultaneously maximizing the production of both biomass, and succinate by using this extended structured CCM model from literature.

2. Allows off-line evaluation, with a higher accuracy, of the optimal operating policy of the analysed bioreactor using a free, or an immobilized cell culture.

3. Allows various bioinformatics analyses, such as:

3.1. Highlight the close connection between the adaptation of cellular metabolism (cellular fluxes, according to the metabolic engineering principles of Stephanopoulos et al. [188]) and the dynamics of macroscopic state variables of the bioreactor.

3.2. An immediate application of (3.1) refers to the *in-silico* design of GMO-s suitable for the industrial applications objectives. This aspect is exemplified in the all three case studies.

As again proved by the three case studies discussed here, the use of the main concepts, principles, and rules of the (bio) chemical engineering (BCE) [3,21,24,60,74,175,176] were proved to be beneficial when developing extended structured cell kinetic models used to solve difficult biochemical engineering, or bioengineering problems. By re-emphasizing the same idea, the use of BCE principles, concepts, and computational techniques, combined with nonlinear systems regulation theory techniques, advanced mathematical and numerical calculation methods, has proven to be extremely beneficial when developing extended hybrid structured (kinetic) cellular models (HSMDM), used to solve difficult biochemical or bioengineering problems [22-23,178,278]. Here is to mention, from the literature, others successful applications of HSMDM models in solving (bio)chemical engineering problems. For instance:

(a) the HSMDM model used to optimize a FBR employed for the mono-clonal anti-bodies (mcAb) production using a hybridoma cell culture [11]

(b) The HSMDM model is used to optimize a FBR employed for mannitol. production from fructose based on a multi-enzymatic inter-connected system.

(c) the case studies of chap.11 of [23], and [9]

(d) The HSMDM model is used to optimize a BR employed for the yeast growth under oscillatory conditions [354-356].

The reduced CCM reaction schema of *S. cerevisiae* to produce biomass is presented in (Figure 8-4). Under certain operating conditions of the BR, the dynamics of all metabolic species become oscillatory (Figure 8-5). That is because one of the intermediate (G), becomes an “oscillatory node” [357], being produced and consumed with different rates, simultaneously adjusted by rapid (+) feed backwards loops, and by slow (-) feed backwards loops. When S is in a big concentration in the cell, the rates ( $r_{SE}$  and  $r_{GE}$ ) are large, thus producing large amounts of E (ethanol). As a consequence, E is stored, later being transformed into R and X following a slow reaction. The enzyme T is considered to be present in small amounts, not being included in the biomass X. This enzyme is not degraded by the rT2 reaction, but only inactivated. The reduced kinetic model of this structured CCM model, together with its rate constants are given in (Figure 8-6). Simulations of the BR by using this HSMDM model with the initial condition of (Figure 8-6) is given in (Figure 8-5). It is to observe that the oscillation node 'G' exhibits a 'flat teeth' and 'flip flop' behaviour, while the dynamics of species further away from the oscillation node reports a periodic sharp increase/decrease.

Once again, the three approached case studies of this book exemplify how the use of a hybrid kinetic model of HSMDM type, based on the cellular CCM, of moderate size, and including several inter-connected modules of reaction pathways and GRC-s, is a modern, accurate and valuable computing tool when developing structured simulators for various engineering applications such as:

- (a) Analysis of cellular metabolic fluxes under variable operating conditions of the bioreactor
- (b) Optimizing the synthesis of the desired metabolite by optimizing the operation of the bioreactor and / or by genetic modification of the cell culture
- (c) *In-silico* reprogramming (based on mathematical model) of cellular metabolism by designing GMO-s with desired characteristics [21-23,178]
- (d) A rapid analysis of the cellular metabolism (i.e., the main metabolic fluxes), leading to an assessment of the use of the substrate into the cell, and of the conditions for the occurrence of cellular oscillations, or those leading to a quasi-stationary operation (QSS) of the bioreactor.
- (e) Structural interpretations of metabolic changes in GMO-s in direct connection with the operation mode of the bioreactor
- (f) More precise optimization of the bioreactor operation (see chap. 6.6, as an example)
- (g) Obtaining simple reduced models by using the BCE algorithms and rules [29,64,67], 'locally' valid (in the operating parameters - time space), for an effective on- / off-line control of the analyzed bioreactor
- (h) More accurate extrapolations of the bioprocess dynamics in the studied bioreactor, over dozens of cell cycles
- (i) Determination of the bioreactor optimal operating conditions, in respect to multiple objectives, aiming to minimize the substrate consumption, but maintaining a high productivity for the desirable metabolite (e.g., TRP in the case study 2, or SUCC in the case study 3).

## References

1. Maria G (2009b) 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.
2. Maria G (2010) A dynamic model to simulate the genetic regulatory circuit controlling the mercury ion uptake by *E. coli* cells. *Revista de Chimie (Bucharest)* 61(2): 172-186.
3. Maria G (2021) A CCM-based modular and hybrid kinetic model to simulate the tryptophan synthesis in a fed-batch bioreactor using modified *E. coli* cells. *Computers & Chemical Engineering* 153: 107450-107466.
4. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci* 97(10): 5528-5533.
5. Maria G (2012) Enzymatic reactor selection and derivation of the optimal operation policy, by using a model-based modular simulation platform, *Comput Chem Eng* 36: 325-341.

6. Maria G, Crisan M (2017) Operation of a mechanically agitated semicontinuous multi-enzymatic reactor by using the Pareto-optimal multiple front method. *J Process Control* 53: 95-105.
7. Maria G (2020B) Model-based optimization of a batch reactor with a coupled bi-enzymatic process for mannitol production, *Computers & Chemical Engineering* 133: 106628-106635.
8. Maria G, Peptănară IM (2021) Model-based optimization of mannitol production by using a sequence of batch reactors for a coupled bi-enzymatic process - A dynamic approach. *Dynamics-Basel* 1: 134-154.
9. Maria G, Renea L, Maria C (2022) Multi-objective optimization of the fed-batch bi-enzymatic reactor for mannitol production. *Dynamics-MDPI* 2: 270-294.
10. Straathof AJJ, Adlercreutz P (2005) *Applied biocatalysis*, Harwood Academic Publ., Amsterdam, Netherlands.
11. Maria G (2020A) Model-based optimization of a fed-batch bioreactor for mAb production using a hybridoma cell culture, *Molecules-Basel - Organic Chemistry* 25(23): 5648-5674.
12. Maria G, Renea L (2021) Tryptophan production maximization in a fedbatch bioreactor with modified *E. coli* cells, by optimizing its operating policy based on an extended structured cell kinetic model, *MDPI Bioengineering-Basel* 8(12): 210-247.
13. DiBiasio D (1989) Introduction to the control of biological reactors. In: *Chemical engineering problems in biotechnology*. In: Shuler ML (ed.), American Institute of Chemical Engineers, New York, USA, pp. 351-391.
14. Bonvin D (1998) Optimal operation of batch reactors-a personal view. *J Process Control* 8(5-6): 355-368.
15. Smets IY, Claes JE, November EJ, Bastin GP, van Impe JF (2004) Optimal adaptive control of (bio)chemical reactors: past, present and future. *J Process Control* 14(7): 795-805.
16. Srinivasan B, Bonvin D, Visser E, Palanki S (2003) Dynamic optimization of batch processes: II. Role of measurements in handling uncertainty. *Comput Chem Eng* 27(1): 27-44.
17. Martinez E (2005) Batch-to-batch optimization of batch processes using the STATSIMPLEX search method, *Proc. 2<sup>nd</sup> Mercosur Congress on Chemical Engineering*. Rio de Janeiro, Costa Verde, Brasil, p. 14-18.
18. Rao M, Qiu H (1993) *Process control engineering: A textbook for chemical, mechanical and electrical engineers*, Gordon and Breach Science Publ., Amsterdam, Netherlands.
19. Hartig F, Keil FJ, Luus R (1995) Comparison of optimization methods for a fed-batch reactor. *Hung J Ind Chem* 23: 81-160.
20. Styczynski MP, Stephanopoulos G (2005) Overview of computational methods for the inference of gene regulatory networks. *Comp Chem Eng* 29(3): 519-534.
21. Maria G (2017A) A review of some novel concepts applied to modular modelling of genetic regulatory circuits, *Juniper publ.*, Irvine, CA, USA.
22. Maria G (2017B) Deterministic modelling approach of metabolic processes in living cells - a still powerful tool for representing the metabolic process dynamics, *Juniper publ.*, Irvine, CA, USA.
23. Maria G (2018) In-silico design of Genetic Modified Micro-organisms (GMO) of industrial use, by using Systems Biology and (Bio)Chemical Engineering tools, *Juniper publ.*, Irvine, USA.
24. Maria G (2005B) Modular-Based Modelling of Protein Synthesis Regulation, *Chemical and Biochemical Engineering Quarterly* 19(3): 213-233.
25. Maria G (2014a) In-silico derivation of a reduced kinetic model for stationary or oscillating glycolysis in *Escherichia coli* bacterium, *Chemical & Biochemical Engineering Quarterly* 28(4): 509-529.
26. 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.
27. Xiong J (2006) *Essential bioinformatics*, Cambridge University Press, Cambridge, UK.
28. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, et al. (2010) OptFlux: an opensource software platform for in silico metabolic engineering. *BMC Syst Biol* 4: 45.
29. Maria G (2004) A review of algorithms and trends in kinetic model identification for chemical and biochemical systems, *Chemical and Biochemical Engineering Quarterly* 18: 195-222.
30. Wu WH, Wang FS, Chang MS (2011) Multi-objective optimization of enzyme manipulations in metabolic networks considering resilience effects. *BMC Systems Biology* 5: 145.
31. Reuss M (1986) Computer control of bioreactors present limits and challenges for the future. In: Morari M, McAvoy TJ (Eds.), *Proc. 3<sup>rd</sup> Intl Conf on Chemical Process Control - CPCIII, Asilomar (USA)*, Elsevier, Amsterdam, Netherlands, pp. 12-17.
32. Banga JR, Alonso AA, Singh PR (1994) Stochastic optimal control of fed-batch bioreactors, *AIChE Annual Meeting, San Francisco, USA*, p. 13-18.

33. Doran PM (1995) Bioprocess engineering principles, Elsevier, Amsterdam, Netherlands.
34. Sarkar D, Modak JM (2005) Pareto-optimal solutions for multi-objective optimization of fed-batch bioreactors using nondominated sorting genetic algorithm. *Chem Eng Sci* 60(2): 481-492.
35. Henson MA (2010) Model-based control of biochemical reactors, In: Levine W (Ed.), *The control handbook*, Taylor and Francis (2<sup>nd</sup> edition), New York, USA.
36. Henson MA, Muller D, Reuss M (2004) Combined metabolic and cell population modelling for yeast bioreactor control. In: Allgöwer F (Ed.), *Proc. IFAC Symposium on Advanced Control of Chemical Processes*, Hong Kong.
37. Bodizs L, Titica M, Faria N, Srinivasan B, Dochain D, et al. (2007) Oxygen control for an industrial pilot-scale fed-batch filamentous fungal fermentation. *Journal of Process Control* 17(7): 595-606.
38. Ashoori A, Moshiri B, Khaki SA, Bakhtiari MR (2009) Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. *Journal of Process Control* 19: 1162-1173.
39. Maria G (2020C) In-silico determination of some conditions leading to glycolytic oscillations and their interference with some other processes in *E. coli* cells. *Frontiers in Chemistry* 8: 526679-526693.
40. 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(11): 98-115.
41. Roeva O, Pencheva T, Tzonkov S, Arndt M, Hitzmann B, et al. (2007) Multiple model approach to modelling of *Escherichia coli* fed-batch cultivation extracellular production of bacterial phytase, *Journal of Biotechnology* 10(4): 592-603.
42. Maria G, Xu Z, Sun J (2011) Multi-objective MINLP optimization used to identify theoretical gene knockout strategies for *E. coli* cell. *Chemical & Biochemical Engineering Quarterly* 25(4): 403-424.
43. Chassagnole C, Noisommit RN, Schmid JW, Mauch K, Reuss M (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering* 79(1): 53-73.
44. Usuda Y, Nishio Y, Iwatani S, Van Dien SJ, Imaizumi A, et al. (2010) Dynamic modeling of *Escherichia coli* metabolic and regulatory systems for aminoacid production. *Journal of Biotechnology* 147(1): 17-30.
45. Kadir TAA, Mannan AA, Kierzek AM, McFadden J, Shimizu K (2010) Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification. *Microbial Cell Factories* 9: 88.
46. Ceric S, Kurtanjek Z (2006) Model identification, parameter estimation, and dynamic flux analysis of *E. coli* central metabolism. *Chem Biochem Eng Q* 20(3): 243-253.
47. Tusek AJ, Kurtanjek Z (2009) Model and global sensitivity analysis of *E. coli* central metabolism, In: Troch I, Breitenacker F (Eds.), *Proc. 6<sup>th</sup> Vienna Conference on Mathematical Modelling MATHMOD*, Beč, Austria, pp.253.
48. Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden, et al. (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? *Testing biochemistry. Eur J Biochem* 267(17): 5313-5329.
49. Olivier BG, Snoep JL (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* 20(13): 2143-2144.
50. Seressiotis A, Bailey JE (1986) MPS: An algorithm and data base for metabolic pathways synthesis. *Biotechnol Lett* 8: 837-842.
51. 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.
52. Tomita M (2001) Whole-cell simulation: a grand challenge of the 21<sup>st</sup> century. *Trends in Biotechnology* 19(6): 205-210.
53. Slepchenko BM, Schaff JC, Macara I, Loew LM (2003) Quantitative cell biology with the Virtual Cell. *Trends in Cell Biology* 13(11): 570-576.
54. Schmid JW, Mauch K, Reuss M, Gilles ED, Kremling A (2004) Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in *Escherichia coli*, *Metabolic Engineering* 6(4): 364-377.
55. Costa RS, Machado D, Rocha I, Ferreira EC (2009) Large scale dynamic model reconstruction for the central carbon metabolism of *Escherichia coli*, In: Omatu S, et al. (Eds.): *Distributed Computing Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living, Proc. IWANN conf., Salamanca (Spain)*, LNCS 5518, Springer-Verlag, Berlin, Germany, pp. 1079-1083.
56. Costa RS, Machado D, Rocha I, Ferreira EC (2010) Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis-Menten and approximate kinetic equations. *BioSystems* 100(2): 150-157.
57. Machado D, Zhuang KH, Sonnenschein N, Herrgård MJ (2015) Current challenges in modeling cellular metabolism, *Frontiers in Bioengineering and Biotechnology* 3: 193.
58. Maria G, Mihalachi M, Gijiu CL (2018c) In silico optimization of a bioreactor with an *E. coli* culture for tryptophan production by using a structured model coupling the oscillating glycolysis and tryptophan synthesis. *Chemical Eng Res and Design* 135: 207-221.
59. Maria G, Mihalachi M, Gijiu CL (2018b) Model-based identification of some conditions leading to glycolytic oscillations in *E. coli* cells, *Chemical and Biochemical Engineering Quarterly* 32(4): 523-533.
60. Maria G, Gijiu CL, Maria C, Tociu C, Mihalachi M (2018d) Importance of considering the isotonic system hypothesis when modelling the self-control of gene expression regulatory modules in living cells. *Current Trends in Biomedical Engineering & Biosciences* 12(2).
61. Maria G, Mihalachi M, Gijiu CL (2018e) Chemical engineering tools applied to simulate some conditions producing glycolytic oscillations in *E. coli* cells. U P B

- Sci Bull Series B - Chemie 80: 27-38.
62. Maria G, Gijiu CL, Maria C, Tociu C (2018a) Interference of the oscillating glycolysis with the oscillating tryptophan synthesis in the E. coli cells. *Computers & Chemical Engineering* 108(4): 395-407.
63. Maria G, Luta I, Maria C (2013) Model-based sensitivity analysis of a fluidised-bed bioreactor for mercury uptake by immobilised *Pseudomonas putida* cells, *Chemical Papers* 67(11): 1364-1375.
64. Maria G (2019) Numerical methods to reduce the kinetic models of (bio)chemical processes, Printech Publ, Bucharest, Romania, pp. 815 pages.
65. Edwards K, Edgar TF, Manousiouthakis VI (1998) Kinetic model reduction using genetic algorithms. *Comp & Chem Eng* 22: 239-246.
66. Martinez EC, Beltramini LJ (1990) Lumping upon time-scales: Modeling upon topological factors. *Chem Eng Sci* 45(8): 2103-2108.
67. Maria G (2005) Relations between apparent and intrinsic kinetics of programmable drug release in human plasma. *Chemical Engineering Science* 60(6): 1709-1723.
68. Dorka P (2007) Modelling batch and fed-batch mammalian cell cultures for optimizing MAb productivity, MSc diss., University of Waterloo, Canada.
69. Miskovic L, Tokic M, Fengos G, Hatzimanikatis V (2015) Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes. *Current Opinion in Biotechnology* 36: 146-153.
70. Kurata H, Sugimoto Y (2018) Improved kinetic model of *Escherichia coli* central carbon metabolism in batch and continuous cultures, *Journal of Bioscience and Bioengineering* 125(2): 251-257.
71. Maria G (2003) ARS combination with an evolutionary algorithm for solving MINLP optimization problems. In: Hamza MH (Ed.), *Modelling, Identification and Control*; IASTED/ACTA Press: Anaheim, CA, USA, pp. 112-117.
72. Maria G (1998) Adaptive Random Search and Short-Cut Techniques for Process Model Identification and Monitoring, Proc. FOCAP098 Int. Conf. on Foundations of Computer Aided Process Operations, Snowbird (USA), pp. 351-359.
73. Heinrich R, Schuster S (1996) *The regulation of cellular systems*, New York, Chapman & Hall, USA.
74. Maria G (2017) Application of (bio) chemical engineering principles and lumping analysis in modelling the living systems. *Current Trends in Biomedical Engineering & Biosciences* 1(4).
75. Haraldsdottir HS, Thiele I, Fleming RMT (2012) Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophys J* 102(8): 1703-1711.
76. 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.
77. Blass LK, Weyler C, Heinzle E (2017) Network design and analysis for multi-enzyme biocatalysis. *BMC Bioinformatics* 18(1): 366.
78. 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(4): 353-373.
79. Heinemann M, Panke S (2006) Synthetic Biology - putting engineering into biology. *Bioinformatics* 22(22): 2790-2799.
80. Salis H, Kaznessis Y (2005) Numerical simulation of stochastic gene circuits, *Computers & Chemical Engineering* 29(3): 577-588.
81. Kaznessis YN (2006) Multi-scale models for gene network engineering. *Chemical Engineering Science* 61(3): 940-953.
82. 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.
83. Klipp E, Nordlander B, Krüger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. *Nature Biotechnology* 23(8): 975-982.
84. Klipp E (2009) Timing matters. *FEBS Letters* 583(24): 4013-4018.
85. Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, et al. (2009) *Systems biology. A textbook*, Wiley-VCH, Weinheim, Germany.
86. Chen MT, Weiss R (2005) Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signalling elements from *Arabidopsis thaliana*. *Nat Biotechnol* 23(12): 1551-1555.
87. Tian T, Burrage K (2006) Stochastic models for regulatory networks of the genetic toggle switch. *Proc Natl Acad Sci USA* 103(22): 8372-8377.
88. Sotiropoulos V, Kaznessis YN (2007) Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *BMC Syst Biol* 1-7.
89. Tomshine J, Kaznessis YN (2006) Optimization of a stochastically simulated gene network model via simulated annealing. *Biophys J* 91(9): 3196-3205.
90. 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.
91. Shimizu TS, Bray D (2002) *Computational cell biology - The stochastic approach*, MIT Press Math6X9, Cambridge, USA.
92. Hucka M, Bergmann FT, Dräger A, Hoops S, Keating SM, et al. (2015) *Systems Biology Markup Language (SBML) Level 2 Version 5: Structures and Facilities*

- for Model Definitions. *J Integrative bioinformatics* 12(2): 271.
93. Ballet P, Zemirline A, Marce L (2001) Cellular automata, reaction-diffusion and multiagents systems for artificial cell modeling, Departement d'Informatique, Universite de Bretagne Occidentale, France.
94. Kinoshita A, Nakayama Y, Tomita M (2001) In silico analysis of human erythrocyte using E-Cell system, ICSB 2001 - 2nd Int. Conf. on Systems Biology, California Institute of Technology, Pasadena (CA), USA.
95. EcoCyc (2005) Encyclopedia of *Escherichia coli* K-12 genes and metabolism, SRI Intl., The Institute for Genomic Research, Univ. of California at San Diego, USA.
96. KEGG (2006) A collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for a large number of cells and micro-organisms, Kanehisa Laboratories, Kyoto, Japan.
97. Kanehisa M, Furumichi M, Sato Y, Ishiguro WM, Tanabe M (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research* 49(D1): D545-D551.
98. NRCAM (2002) The National Resource for Cell Analysis and Modeling. Virtual cell modeling and simulation framework, The University of Connecticut Health Center, Richard D. Berlin Center for Cell Analysis & Modeling, Farmington, USA.
99. Moraru I, Schaff JC, Slepchenko BM, Loew LM (2002) The Virtual Cell: An integrated modeling environment for experimental and computational cell biology. *Annals of the New York Academy of Sciences* 971: 595-596.
100. Peters M, Eicher JJ, Van Niekerk DD, Waltemath D, Snoep JL (2017) The JWS online simulation database. *Bioinformatics* 33(10): 1589-1590.
101. Westerhoff HV (2006) Engineering life processes live: the Silicon cell. In: ESCAPE-16 Conference, Garmisch-Partenkirchen, Germany.
102. Bartol TM, Stiles JR (2002) MCell: A general Monte Carlo simulator of cellular micro physiology. In: Computational Neurobiology Lab, The Salk Institute, California, USA.
103. Ichikawa K (2000) A-Cell: a platform for the construction of biochemical reaction and electrical equivalent circuit models in a biological cell and a neuron. In: Fuji Xerox Co. Ltd, Kanagawa, Japan.
104. Ichikawa K (2001) A-Cell: graphical user interface for the construction of biochemical reaction models. *Bioinformatics* 17(5): 483-484.
105. CRGM (2002) Research Center in Molecular Genetics, Institute of Histology and General Embryology, University of Bologna, Italy.
106. Guantes R, Poyatos JF (2006) Dynamical principles of two-component genetic oscillators. *PLoS Computational Biology*, 2(3): 188-197.
107. Maria G (2014b) Extended repression mechanisms in modelling bistable genetic switches of adjustable characteristics within a variable cell volume modelling framework. *Chemical & Biochemical Engineering Quarterly* 28(1): 35-51.
108. Chen T, He HL, Church GM (1999) Modeling gene expression with differential equations. *Pacific Symposium of Biocomputing*, Hawaii, USA, pp. 29-40.
109. Drengstig T, Jolma IW, Ni XY, Thorsen K, Xu XM, et al. (2012) A Basic Set of Homeostatic Controller Motifs. *Biophys J* 103(9): 1-11.
110. Komasilovs V, Pentjuss A, Elsts A, Stalidzans E (2017) Total enzyme activity constraint and homeostatic constraint impact on the optimization potential of a kinetic model. *Biosystems* 162: 128-134.
111. Matsuura T, Tanimura N, Hosoda K, Yomo T, Shimizu Y (2017) Reaction dynamics analysis of a reconstituted *Escherichia coli* protein translation system by computational modeling. *PNAS* 114(8): E1336E1344.
112. Zhang R, Ehigie JO, Hou X, You X, Yuan C (2017) Steady-State Preserving simulation of genetic regulatory systems. *Computational and Mathematical Methods in Medicine*, pp. 1-16.
113. Morgan JJ, Surovtsev IV, Lindahl PA (2004) A framework for whole cell mathematical modelling. *J Theor Biology* 231: 581-596.
114. Surovtsev IV, Morgan J, Lindahl PA (2007) Whole-cell modeling framework in which biochemical dynamics impact aspects of cellular geometry. *J Theor Biol* 244(1): 154-166.
115. Maria G, Berger D, Nastase S, Luta I (2012a) Modelling alternatives of the irinotecan release from functionalized mesoporous-silica supports. *Microporous and Mesoporous Materials* 149(1): 25-35.
116. Maria G, Stoica AI, Luta I, Stirbet D, Radu GL (2012b) Cephalosporin release from functionalized MCM-41 supports interpreted by various models. *Microporous and Mesoporous Materials* 162(11): 80-90.
117. Sherlock JP (2016) Chemical engineering matters. Report of the I Chem E technical vice president, London.
118. Woinaroschy A, Ofiteru DI, Lavric V (2009) Exploratory investigation of bioprocesses sustainability improvement by multicriteria-multilevel optimization. *Environmental Engineering and Management Journal* 8(3): 521-526.
119. Woinaroschy A (2016) A paradigm-based evolution of chemical engineering. *Chinese Journal of Chemical Engineering* 24(5): 553-557.
120. Gani R (2004) Computer-aided methods and tools for chemical product design. *Chem Eng Res Des* 82(11): 1494-1504.
121. Gani R, Bařdyga J, Biscans B, Brunazzi E, Charpentier JC, et al. (2020) A multi-layered view of chemical and biochemical engineering. *Chemical Engineering Research and Design* 155: 133-145.
122. (2022) *Bioinformatics* Wikipedia, last accessing.

123. Silva MR (2006) Bioinformatics tools for the visualization and structural analysis of metabolic networks. PhD Thesis, TU Braunschweig, Germany.
124. Janowski SJ (2013) VANESA - A bioinformatics software application for the modeling, visualization, analysis, and simulation of biological networks in systems biology applications. PhD thesis, Faculty of Technology at the Bielefeld University, Germany.
125. Eden P (2003) Course on: Bioinformatics and Computational Biology. Faculty of Medicine, Lund University, Sweden.
126. Lesk AM (2002) Introduction to Bioinformatics. In: Oxford Univ Press, UK.
127. Baxevanis AD, Ouellette BF (2001) Bioinformatics - A Practical Guide to the Analysis of Genes and Proteins. Wiley.
128. Lengauer T (2002) Bioinformatics - From genomes to drugs, Wiley, USA.
129. Baldi P, Brunak S (2001) Bioinformatics. In: The Machine Learning Approach. (2<sup>nd</sup> edn.), MIT Press, Cambridge, Massachusetts, USA.
130. Isaev A (2004) Introduction to mathematical methods in Bioinformatics. In: Springer, Germany.
131. Eidhammer I, Jonassen I, Taylor WR (2004) Protein Bioinformatics - An algorithmic approach to sequence and structure analysis. In: Wiley, USA.
132. Xiong J (2006) Essential bioinformatics. Cambridge univ press, UK.
133. Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? An introduction and overview. In: Yearbook of Medical Informatics, Thieme Medical Publishers, USA, pp. 83-99.
134. Michal G (2014) Roche co. report, Basel CH.
135. Torres NV, Voit EO (2002) Pathway analysis and optimization in metabolic engineering. In: Cambridge University Press, Cambridge (MS), UK.
136. Cornish BA (2016) Biochemical evolution- The pursuit of perfection. Garland Science, New York, USA.
137. 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.
138. Savageau MA (2002) Alternatives designs for a genetic switch: Analysis of switching times using the piecewise power-law representation. Math Biosciences 180(1-2): 237-253.
139. Nandy SK (2017) BioXYZ, what is XYZ? Curr Trends Biomedical Eng & Biosci 3(2).
140. Sutherland W (2005) Optimization- this beguilingly simple idea allows biologists not only to understand current adaptations, but also to predict new designs that may yet evolve. Nature pp. 435.
141. Banga JR (2008) Optimization in computational systems biology. 6<sup>th</sup> Simon Stevin lecture on optimization in engineering, Leuven-Heverlee, NL. BMC Systems Biology 2: 47.
142. Klamt S, Stelling J (2002) Combinatorial complexity of pathway analysis in metabolic networks. Technical report, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse Germany, 29(1-2): 1-8.
143. Hecker, M, Lambeck S, Toepfer S, Van Someren E, Guthke R (2009) Gene regulatory network inference: Data integration in dynamic models-A review. BioSystems 96(1): 86-103.
144. Wang X, Li Y, Xua X, Wang Y (2010) Toward a system-level understanding of microRNA pathway via mathematical modeling. Biosystems 100(1): 31-38.
145. Zhdanov VP (2009) Bistability in gene transcription: Interplay of messenger RNA, protein, and nonprotein coding RNA. Biosystems 95(1): 75-81.
146. 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.
147. Hlavacek WS, Savageau MA (1997) Completely uncoupled and perfectly coupled gene expression in repressible systems. J Mol Biol 266: 538-558.
148. Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. J Mol Biol 323(5): 785-793.
149. De Jong H (2002) Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. Journal Computational Biology 9(1): 67-103.
150. De Jong H, Guze JL, Hernandez C, Page M, Sari T, et al. (2003) Hybrid modelling and simulation of genetic regulatory networks: a qualitative approach, In: Maler O, Pnueli A (eds.), Hybrid Systems: Computation and Control, HSCC, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. 2623, Germany.
151. Endy D, Brent R (2001) Modelling cellular behaviour. Nature 409(6818):391-395.
152. Hasty J, McMillen D, Isaacs F, Collins JJ (2001a) Computational studies of gene regulatory networks: in numero molecular biology. Nature Reviews Genetics 2(4): 268-279.
153. Hasty J, Isaacs F, Dolnik M, McMillen D, Collins JJ (2001b) Designer gene networks: Towards fundamental cellular control. Chaos 11(1): 207-220.
154. Hasty J, McMillen D, Collins JJ (2002) Engineered gene circuits. Nature 420(6912): 224-230.
155. Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. PNAS 100(13): 7714-7719.
156. Kaern M, Blake WJ, Collins JJ (2003) The engineering of gene regulatory networks. Annu Rev Biomed Eng 5: 179-206.
157. Mochizuki A (2005) An analytical study of the number of steady states in gene regulatory networks. Journal of Theoretical Biology 236(6): 291-310.
158. Samad HE, Khammash M, Petzold L, Gillespie D (2005) Stochastic modelling of gene regulatory networks. Int J Robust Nonlinear Control 15(15): 691-711.
159. Tyson JJ, Albert R, Goldbeter A, Ruoff P, Sible J (2008) Biological switches and clocks. J R Soc Interface 5: S1-S8.



160. Widder S, Schicho J, Schuster P (2007) Dynamic patterns of gene regulation I: Simple two gene systems. *J Theor Bio* 246(3): 395-419.
161. Liu XM, Xie HZ, Liu LG, Li ZB (2009) Effect of multiplicative and additive noise on genetic transcriptional regulatory mechanism. *Physica A* 388(4): 392-398.
162. Kaznessis YN (2011) SynBioSS-Aided Design of Synthetic Biological Constructs. *Methods in Enzymology* 498: 137-152.
163. Van Someren EP, Wessels LFA, Backer E, Reinders MJT (2003) Multi-criterion optimization for genetic network modelling. *Signal Processing* 83(4): 763-775.
164. Xu Z, Zheng P, Sun J, Ma Y (2013) ReacKnock: Identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. *Plos One* 8(12): e72150.
165. Xu Z, Sun X, Yu S (2009) Genome-scale analysis to the impact of gene deletion on the metabolism of *E. coli*: constraint-based simulation approach. *BMC Bioinformatics* 10(Suppl 1): S62.
166. Xu Z, Sun J, Wu Q, Zhu D (2017) Find\_tfSBP: find thermodynamics - feasible and smallest balanced pathways with high yield from large-scale metabolic networks. *Nature Scientific Reports* 7(1):17334.
167. Gilbert D, Heiner M, Jayaweera Y, Rohr C (2017) Towards dynamic genome-scale models. *Briefings in Bioinformatics* 20(4): 1167-1180.
168. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335-338.
169. Voit EO (2005) Smooth bistable S-systems. *IEE Proc Syst Biol* 152(4): 207-213.
170. Maria G, Morgan J, Lindahl P (2002) Kinetic simulation of hypothetical "Mechanical Cells" that maintain intracellular homeostasis while growing autocatalytically on environmental nutrients present in variable amounts. *J Theor Biology*.
171. CellML Group (2002) CellML: Related biological languages. Bioengineering research group, The University of Auckland, New Zealand.
172. Hedley WJ, Nelson MR, Bullivant DP, Nielson PF (2001) A short introduction to CellML. *Philosophical Transactions of the Royal Society of London A* 359(1783): 1073-1089.
173. Aris R (1969) Elementary chemical reactor analysis. In: Prentice-Hall, New Jersey, USA.
174. Maria G (2007) Modelling bistable genetic regulatory circuits under variable volume framework. *Chemical and Biochemical Engineering Quarterly* 21(4): 417-434.
175. Maria G, Scoban AG (2017) Setting some milestones when modelling gene expression regulatory circuits under variable-volume whole-cell modelling framework. 1. Generalities. *Revista de Chimie (Bucharest)* 68(12): 3027-3037.
176. Maria G, Scoban AG (2018) Setting some milestones when modelling gene expression regulatory circuits under variable-volume whole-cell modelling framework. 2. Case studies, *Revista de Chimie (Bucharest)* 69(1): 259-266.
177. Maria G, Maria C, Tociu C (2018) A comparison between two approaches used for deterministic modelling of metabolic processes and of genetic regulatory circuits in living cells. *U P B Sci Bull Series B Chimie* 80(1): 127-144.
178. Maria G, Gijiu CL, Crişan M, Maria C, Tociu C (2019) Model-based re-design of some genetic regulatory circuits to get Genetic Modified Microorganisms (GMO) by using engineering computational tools (a minireview). *Current Trends in Biomedical Engineering & Biosciences* 18(3).
179. 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.
180. Stelling J (2004) Mathematical models in microbial systems biology. *Current opinion in Microbiology* 7(5): 513-518.
181. Crampin EJ, Schnell S (2004) New approaches to modelling and analysis of biochemical reactions. pathways and networks, *Progress in Biophysics & Molecular Biology* 86(1): 1-4.
182. Kholodenko BN, 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.
183. Das S, Caragea D, Welch SM, Hsu WH (2010) Handbook of research and computational methodologies in gene regulatory networks. In: Medical information science reference. IGI Global, New York.
184. Myers CJ (2009) Engineering genetic circuits. In: Chapman & Hall, CRC Press, Boca Raton, USA.
185. 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.
186. Bailey JE (1991) Towards a science of metabolic engineering. *Science* 252(5013): 1668-1675.
187. Nielsen J (1998) Metabolic engineering, techniques for analysis of targets for genetic manipulations. *Biotechnol Bioeng* 58(2-3): 125-132.
188. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Metabolic Engineering. Principles and Methodologies. In: Academic Press, San Diego CA, USA.
189. Maria G (2009) Building-up lumped models for a bistable genetic regulatory circuit under whole-cell modelling framework. *Asia-Pacific Journal of Chemical Engineering* 4: 916-928.
190. Allen GC, Kornberg A (1991) The priB gene encoding the primosomal replication n-protein of *E. coli*. *J Biological Chemistry* 266(18): 11610-11613.
191. Kubitschek HE (1990) Cell volume increase in *Escherichia coli* after shifts to richer media. *Journal of Bacteriology* 172(1): 94-101.

192. Wallwork SC, Grant DJW (1977) Physical Chemistry. In: Longman, London, UK.
193. Kimball JW (2022) Cell metabolism. In: Tufts and Harvard Univ, USA.
194. Hudder B, Yang Q, Bolting B, Maria G, Morgan JJ, et al. (2002) Computational Modeling of Iron Metabolism in Mitochondria. NIST Conference on 'Systems Biology Approaches to Health Care: Mitochondrial Proteomics, Gaithersburg, USA, p. 1.
195. Hodgkin AL, Huxley AF (1952) A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J Physiol* 117(4): 500-544.
196. Noble D (1962) A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pace-maker potentials. *Journal of Physiology* 160(2): 317-352.
197. Wolkenhauer O, Mesarovic M (2005) Feedback dynamics and cell function: Why Systems Biology. *Mol Biosyst* 1: 14-16.
198. Kitano H (2002) Computational systems biology. *Nature* 420: 206-210.
199. Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: Systems Biology. *Annu Rev Genomics Hum Genet* 2: 343-372.
200. Bower JM, Bolouri H (2001) Computational Modeling of Genetic and Biochemical Networks. In: MIT Press, Cambridge, USA.
201. Atauri P, Curto R, Puigjaner J, Cornish-Bowden A, Cascante M (1999) Advantages and disadvantages of aggregating fluxes into synthetic and degradative fluxes when modelling metabolic pathways. *Eur J Biochem* 265(2): 671-679.
202. Szedlaczek ES, Aricescu AR, Havsteen BH (1996) Time-dependent control of metabolic systems by external effectors. *J theor Biol* 182(3): 341-350.
203. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA* 94(3): 814-819.
204. Gillespie DT, Mangel M (1981) Conditioned averages in chemical kinetics. *Jl Chemical Physics* 75: 704-709.
205. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, et al. (2002) *Molecular Biology of the Cell*. (4<sup>th</sup> edn.), Garland Science, New York.
206. Tyson JJ, Novak B, Odell GM, Chen K, Thron CD (1996) Chemical kinetic theory: Understanding cell cycle regulation. *Trends Biochem Sci* 21(3): 89-96.
207. Maria G (1989) An Adaptive Strategy for Solving Kinetic Model Concomitant Estimation-Reduction Problems. *Canadian Journal of Chemical Engineering* 67(5): 825-832.
208. Filippi C, Greffe JL, Bordet J, Villermaux J, Barnay JL, et al. (1986) Tendency modeling of semibatch reactors for optimization and control. *Chem Eng Sci* 41(4): 913-920.
209. McCroskey PS (1988) A knowledge-based approach for the development of complex kinetic mechanisms, Ph.D. Diss, Carnegie-Mellon Univ, Pittsburgh, Pennsylvania, USA.
210. Maria G, Rippin DW T (1996) Recursive robust kinetics estimation by using mechanistic shortcut technique and a pattern-recognition procedure. *Comp & Chem Eng* 20(1): S587-S592.
211. Valdes-Perez RE (1990) Machine discovery of chemical reaction pathways. Ph. D. Diss, Carnegie-Mellon Univ. Pittsburgh, Pennsylvania.
212. Valdes-Perez RE (1991) On the concept of stoichiometry of reaction mechanisms. *J Phys Chem* 95: 4918-4921.
213. Valdes-Perez RE (1992) A necessary condition for catalysis in reaction pathways. *J Phys Chem* 96: 2394-2396.
214. Knight JP (1995) Computer-Aided Design Tools to Link Chemistry and Design in Process Development. Ph. D. Diss, Massachusetts Institute of Technology, United States.
215. Mavrovouniotis ML, Bonvin D (1994) Towards design of reaction paths, FOCAPD94 Conference, Snowmass Village (CO), Colorado, United States.
216. Mavrovouniotis ML (1995) Product and process design with molecular-level knowledge, In: Stephanopoulos G (Ed.), *Artificial intelligence approaches in product and process engineering*, Academic Press, New York, USA.
217. Prickett SE, Mavrovouniotis ML (1997) Construction of complex reaction systems. I. Reaction description language. II. Molecule manipulation and reaction application algorithms. *Comp & Chem Eng* 21(11): 1219-1254.
218. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle J, et al. (2003) The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19(4): 524-531.
219. Pedersen F, Tyle H, Niemela JR, Guttmann B, Lander L, et al. (1994) Environmental hazard classification, Tema Nord Report 1994:589, NCM, c.
220. Xu S, Nirmalakhandan N (1998) Use of QSAR models in predicting joint effects in multicomponent mixtures of organic chemicals. *Water Res* 32(8): 2391-2399.
221. Buxton GV, Greenstock CL, Helman WP, Ross AB (1988) Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals in aqueous solution. *J Phys Chem Ref Data* 17: 513-759.
222. NIH Database (2004) National Institute of Health database.
223. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV (2002) Untangling the wires: A strategy to trace functional interactions in signalling and gene networks. *Proc Natl Acad Sci USA* 99(20): 12841-12846.

224. Qian Y, McBride C, Del Vecchio D (2017) Programming Cells to Work for Us, Annual Review of Control, Robotics, and Autonomous Systems.
225. Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, et al. (2004) Programmable cells: Interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA* 101(22): 8414-8419.
226. UG (2017) citing Leroy Hood, Inst. Systems Biology, Seattle, The University of Glasgow, Centre for Mathematics Applied to the Life Sciences, Scotland.
227. Niklas J, Schrader E, Sandig V, Noll T, Heinze E (2011) Quantitative characterization of metabolism and metabolic shifts during growth of the new human cell line AGE<sub>1</sub>HN using time resolved metabolic flux analysis. *Bioprocess Biosyst Eng* 34(5): 533-545.
228. 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.
229. 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.
230. Smigelschi O, Maria G (1986) A Modified Half-Interval Search Procedure for Solving Multivariate Nonlinear Process Models. *Hungarian Journal of Industrial Chemistry* 14(4): 453-462.
231. Varma A, Morbidelli M, Wu H (1999) Parametric sensitivity in chemical systems, Cambridge University Press, Cambridge (MS), United Kingdom.
232. Sauro HM, Kholodenko BN (2004) Quantitative analysis of signalling networks. *Prog Biophys Mol Biol* 86(1): 5-43.
233. Liao JC, Lightfoot Jr EN (1987) Extending the quasi-steady state concept to analysis of metabolic networks. *J Theor Biol* 126(3): 253-273.
234. Hofmeyr JHS (2001) Metabolic control analysis in a nutshell. *Proc. 2<sup>nd</sup> Intl Conf on Systems Biology* Madison, WI.
235. Vance W, Arkin A, Ross J (2002) Determination of causal connectivities of species in reaction networks. *Proceedings of the National Academy of Science USA* 99: 5816-5821.
236. Maria G (2022) Despre școala de inginerie și tehnologie (bio)chimică de la Universitatea Politehnică din București, (About the school of biochemical engineering and technology at University Politehnica of Bucharest), Printech Publ, Bucharest (800 pag.)
237. Schuetz R, Kuepfer L, Sauer W (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in E coli, *Molecular Systems Biology* 3: 119.
238. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cell: evaluating the consequences of constraints. *Nature Reviews Microbiology* 2(11): 886-897.
239. Maria G, Crisan M, Maria C (2016) Parameter estimation of the (Bio)chemical process kinetic models, Printech Publ, Bucharest, Romania, pp. 528.
240. Hempel DC (2006) Development of biotechnological processes by integrating genetic and engineering methods. *Eng Life Sci* 6(5): 443-447.
241. Scoban AG, Maria G (2016) Model-based optimization of the feeding policy of a fluidized bed bioreactor for mercury uptake by immobilized P. putida cells. *Asia-Pacific Journal of Chemical Engineering* 11(5): 721-734.
242. Maria G, Scoban AG (2017b) Optimal operating policy of a fluidized bed bioreactor used for mercury uptake from wastewaters by using immobilized P. putida cells, *Current Trends in Biomedical Engineering & Biosciences*, Juniper publ, Irvine, CA, USA, 2(4): 555594.
243. Maria G, Szedlaczek S, Laslo AC, Zaharia E (2005) Modular-based simulation of regulated protein synthesis, 2nd Croatian-German Conference on Enzyme Reaction Engineering, Dubrovnik (Croatia), poster P9, pp. 21-24.
244. Philippidis GP, Malmberg LH, Hu WS, Schottel JL (1991a) Effect of gene amplification on mercuric ion reduction activity of Escherichia coli, *Applied & Environmental Microbiology* 57(12): 3558-3564.
245. Philippidis GP, Schottel JL, Hu WS (1991b) Mathematical modelling and optimization of complex biocatalysis. A case study of mercuric reduction by Escherichia coli, *Expression systems and processes for DNA Products*, National Science Foundation Report ECE-8552670, University of Minnesota, USA.
246. Philippidis GP, Schottel JL, Hu WS (1991c) A model for mercuric ion reduction in recombinant Escherichia coli. *Biotech Bioeng* 37(1): 47-54.
247. Barkay T, Miller SM, Summers AO (2003) Bacterial mercury resistance from atoms to ecosystems. *FEMS Microbiology Reviews* 27(2-3): 355-384.
248. Deckwer WD, Becker FU, 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.
249. Wagner-Döbler I, von Canstein H, Li Y, Timmis K, Deckwer WD (2000) Removal of mercury from chemical wastewater by microorganisms in technical scale. *Env Sci & Technol* 34: 4628-4634.
250. Leonhäuser J, Röhrich M, Wagner-Döbler I, Deckwer WD (2006) Reaction engineering aspects of microbial mercury removal. *Eng Life Sci* 6(2): 139-148.
251. Berne F, Cordonnier J (1995) Industrial water treatment, Paris, Gulf Publishing Co, Houston, Texas, USA.
252. Ledakowicz S, Becker U, Deckwer WD (1996) Development of mercury biotransformation process in fluidised-bed reactor with immobilized microorganisms. In: Wijffels RH, Buitelaar RM, Bucke C, Tramper J (Eds.), *Immobilized cells: Basics and applications*, Elsevier, Amsterdam, Netherlands.
253. Moser A (1988) *Bioprocess Technology: Kinetics and Reactors*. SpringerVerlag, New York, USA.
254. Trambouze P, Van Landeghem H, Wauquier JP (1988) *Chemical reactors: Design, engineering, operation*. Edition Technip, Paris, France.
255. Narasimhan B, Mallapragada SK, Peppas NA (1999) Release kinetics, data interpretation. In: Mathiowitz E (Ed.), *Encyclopedia of controlled drug delivery*, Wiley, New York, USA, pp. 921-935.

256. Doraiswamy LK, Sharma MM (1984) Heterogeneous reactions: Analysis, examples, and reactor design, Wiley, New York, USA, pp. 1-2.
257. Rojanschi V, Ognean T (1989) Cartea operatorului din statii de tratare si epurare a apelor (Operator's manual working in the wastewater treatment and purification plants), Ed. Tehnica, USA.
258. Green-Ruiz C (2006) Mercury(II) removal from aqueous solutions by nonviable *Bacillus* sp. from a tropical estuary. *Bioresource Technology* 97(15): 1907-1911.
259. Trun NJ, Gottesman S (1990) On the bacterial cell cycle: *Escherichia coli* mutants with altered ploidy. *Genes Dev* 4(12A): 2036-2047.
260. Mackay D (2001) Multimedia environmental models - The fugacity approach, Boca Raton, Lewis Publishers, USA.
261. Whitham AG, Sparks RSJ (1986) Pumice. *Bulletin of Volcanology* 48: 209-223.
262. Lambert DP (1996) Ventilation criteria for IDMS facility, DOE contract DE-AC09-89SR18035 (Report WSRC-RP-96-0282), Westinghouse Savannah River Company, Aiken, USA.
263. Shen L, Chen Z (2007) Critical review of the impact of tortuosity on diffusion. *Chem Eng Sci* 62(14): 3748-3755.
264. Alvarez E, Cancela MA, Navaza JM, Taboas R (2002) Mass transfer coefficients in batch and continuous regime in a bubble column, In: Proc. Intl. Conference on Distillation and Absorption (Baden-Baden, Germany, Sept. 2002), Düsseldorf: GVC - VDI Society of Chemical and Process Engineering.
265. Laurent M, Charvin G, Guespin-Michel J (2005) Bistability and hysteresis in epigenetic regulation of the lactose operon. *Cellular and Molecular Biology* 51(7): 583-594.
266. Gonze D (2010) Coupling oscillations and switches in genetic networks. *BioSystems* 99: 60-69.
267. Rassiss D, Nussinovitch A, Saguy IS (2002) Collapse, shrinkage and structural changes in dried alginate gels containing fillers. *Food Hydrocolloids* 16(2): 139-151.
268. Chen M (2020) Novel approaches for in vivo evolution, screening and characterization of enzymes for metabolic engineering of *Escherichia coli* as hyper L-tryptophan producer, PhD thesis, TU Hamburg, Germany.
269. Ghose TK, Fiechter A, Blakebrough N (1977-1978) *Advances in Biochemical Engineering*, Springer Verlag, Berlin volumes 7-10.
270. Liese A, Seelbach K, Wandrey C (2006) *Industrial biotransformations*, Wiley-VCH, Weinheim.
271. Buchholz K, Hempel DC (2006) From gene to product (editorial), *Eng Life Sci* 6(5): 437-437.
272. Chaudhuri J, Al-Rubeai M (Eds.) (2005) *Bioreactors for Tissue Engineering Principles, Design and Operation*, Springer Verlag, Berlin, Germany.
273. Nedovic V, Willaert R (2005) *Applications of cell immobilisation technology*, Springer verlag, Amsterdam, Netherlands.
274. Abel O, Marquardt W (2003) Scenario-integrated on-line optimisation of batch reactors. *J Process Control* 13(8): 703-715.
275. Lee J, Lee KS, Lee JH, Park S (2001) An on-line batch span minimization and quality control strategy for batch and semi-batch processes. *Control Eng Pract* 9(8): 901-909.
276. Ruppen D, Bonvin D, Rippin DWT (1998) Implementation of adaptive optimal operation for a semi-batch reaction system. *Comput Chem Eng* 22(1-2): 185-199.
277. Agrawal P, Koshy G, Ramseier M (1989) An algorithm for operating a fed-batch fermentator at optimum specific-growth rate. *Biotechnol Bioeng* 33(1): 115-125.
278. Shuler ML (ed.), (1989) *Chemical engineering problems in biotechnology*. American Institute of Chemical Engineers, New York, USA.
279. Roubos JA (2002) *Bioprocess modeling and optimization-Fed-batch clavulanic acid production by Streptomyces clavuligerus*, PhD Thesis, TU Delft, USA.
280. Zak DE, Vadigepalli R, Gonye GE, Doyle III FJ, Schwaber JS, et al. (2005) Unconventional systems analysis problems in molecular biology: A case study in gene regulatory network modelling. *Comput Chem Eng* 29(3): 547-563.
281. Slominski A, Semak I, Pisarchik A, Sweatman T, Szczesniewski A, et al. (2002) Conversion of L-tryptophan to serotonin and melatonin in human melanoma cells. *FEBS Letters* 511(1-3): 102-106.
282. Chen M, Ma C, Chen L, Zeng AP (2021) Integrated laboratory evolution and rational engineering of GalP/Glk-dependent *Escherichia coli* for higher yield and productivity of L-tryptophan biosynthesis. *Metabolic Engineering Communications* 12: e00167.
283. Chen L (2016) *Rational metabolic engineering and systematic analysis of Escherichia coli for L-tryptophan bioproduction*, PhD thesis, TU Hamburg, Germany.
284. Chen L, Zeng AP (2017) Rational design and metabolic analysis of *Escherichia coli* for effective production of L-tryptophan at high concentration. *Appl Microbiol Biotechnol* 101: 559-568.
285. Chen M, Chen L, Zeng AP (2019) CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided in vivo screening of enzyme variants for a more efficient chorismate pathway in *E. coli*, *Metabolic Engineering Communications* 9: e00094.
286. Li Z, Wang H, Ding D, Liu Y, Fang H, et al. (2020) Metabolic engineering of *Escherichia coli* for production of chemicals derived from the shikimate pathway. *Journal of Industrial Microbiology and Biotechnology* 47(6-7): 525-535.
287. Niu H, Li R, Liang Q, Qi Q, Li Q, et al. (2019) Metabolic engineering for improving L-tryptophan production in *Escherichia coli*, *Journal of Industrial Microbiology and Biotechnology* 46(1): 55-65.

288. Carmona SB, Flores N, Martínez-Romero, E Gosset G, Bolívar F, et al. (2020) Evolution of an *Escherichia coli* PTS<sup>-</sup> strain: a study of reproducibility and dynamics of an adaptive evolutive process. *Applied Microbiology and Biotechnology* 104: 9309-9325.
289. Bhartiya S, Chaudhary N, Venkatesh KV & Doyle III FJ (2006) Multiple feedback loop design in the tryptophan regulatory network of *E. coli* suggests a paradigm for robust regulation of processes in series. *J R Soc Interface* 3: 383-391.
290. Xiu ZL, Zeng AP, Deckwer WD (1997) Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. *Journal of Biotechnology* 58(2): 125-140.
291. Xiu ZL, Chang ZY, Zeng AP (2002) Nonlinear dynamics of regulation of bacterial *trp* operon: Model analysis of integrated effects of repression, feedback inhibition, and attenuation. *Biotechnol Prog* 18(4): 686-693.
292. Chen L, Chen M, Ma C, Zeng AP (2018) Discovery of feed-forward regulation in *L*-tryptophan biosynthesis and its use in metabolic engineering of *E. coli* for efficient tryptophan bioproduction, *Metabolic Engineering* 47: 434-444.
293. Carlsson B, Zambrano J (2014) Analysis of simple bioreactor models - a comparison between Monod and Contois kinetics, IWA Special International Conference: Activated Sludge - 100 Years and Counting. Essen, Germany.
294. Mathews CK, van Holde KE, Ahem KG (1999) *Biochemistry*. Prentice Hall, New Jersey, USA.
295. Stephanopoulos G, Simpson TW (1997) Flux amplification in complex metabolic networks. *Chem Eng Sci* 52(15): 2607-2627.
296. Calhoun KA, Swartz JR (2006) Total amino acid stabilization during cell-free protein synthesis reactions. *Journal of Biotechnology* 123(2): 193-203.
297. Noor E, Eden E, Milo R, Alon U (2010) Central Carbon Metabolism as a minimal biochemical walk between precursors for biomass and energy. *Molecular Cell* 39(5): 809-820.
298. Ruby EG, Neelson KH (1977) Pyruvate production and excretion by the luminous marine bacteria. *Applied and Environmental Microbiology* 34(2): 164-169.
299. Kreth J, Lengeler JW, Jahreis K (2013) Characterization of pyruvate uptake in *Escherichia coli* K-12. *Plos One* 8(6): E67125.
300. Bier M, Teusink B, Kholodenko BN, Westerhoff, H.V., 1996, Control analysis of glycolytic oscillations. *Biophysical Chemistry* 62(1-3): 15-24.
301. Silva AS, Yunes JA (2006) Conservation of glycolytic oscillations in *Saccharomyces cerevisiae*, *Genetics and Molecular Research*, 3(5): 525-535.
302. Loeblein C, Perkins J, Srinivasan B, Bonvin D (1997) Performance analysis of on-line batch optimization systems, *Comput Chem Eng* 21: S867-S872.
303. Maria G, Dan A (2011) Derivation of optimal operating policies under safety and technological constraints for the acetoacetylation of pyrrole in a semi-batch catalytic reactor. *Comput Chem Eng* 35(1): 177-189.
304. Bharat A (2013) Process analytical technology (PAT), Msc. Diss., P.D.V.V.P.F.S. College of pharmacy, Ahmed Nagar, India.
305. Avili MG, Fazaelpoor MH, Jafari SA, Ataei SA (2012) Comparison between batch and fed-batch production of rhamnolipid by *Pseudomonas aeruginosa*, *Iranian journal of Biotechnology* 10(4): 263-269.
306. Mihalachi M, Maria G (2019) Influence of PEP glycolytic precursor on tryptophan synthesis dynamics in *E coli* cells. *UPB sci bull series B* 81(2): 29-36.
307. Termonia Y, Ross J (1981a) Oscillations and control features in glycolysis: Numerical analysis of a comprehensive model. *Proc Natl Acad Sci USA* 78(5): 2952-2956.
308. Termonia Y, Ross J (1981b) Oscillations and control features in glycolysis: Analysis of resonance effects. *Proc Natl Acad Sci USA* 78(6): 3563-3566.
309. Rao SS (2009) *Engineering optimization - Theory and practice*. Wiley, New York, USA.
310. Burgard AP, Pharkya P, Maranas CD (2003) OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84(6): 647-657.
311. Ma HW, Zhao XM, Yuan YJ, Zeng AP (2004) Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph *Bioinformatics* 20(12): 1870-1876.
312. He F, Balling R, Zeng AP (2009) Reverse engineering and verification of gene networks: Principles, assumptions and limitations of present methods and future perspectives. *J Biotechnol* 144(3): 190-203.
313. Jandt U, Shao S, Wirth M, Zeng AP (2011) Spatio-temporal modeling and analysis of transient gene delivery. *Biotechnol Bioeng* 108(9): 2205-2217.
314. Kacser H, Burns JA (1973) *The control of flux*. Proc. Symp Soc Exper Biol, Cambridge Univ Press, Cambridge, UK, 27: 65-104.
315. Papin JA, Price ND, Palsson BO (2002) Extreme pathway lengths and reaction participation in genome-scale metabolic networks *Genome Res* 12(12): 1889-1900.
316. Arkin A, Ross J (1995) Statistical construction of chemical reaction mechanisms from measured time-series. *J Phys Chem* 99: 970-979.
317. Arkin A, Shen PD, Ross J (1997) A test case of correlation metric construction of a reaction pathway from measurements. *Science* 277(5330): 1275- 1279.
318. Kauffman KJ, Pajerowski JD, Jamshidi N, Palsson BO, Edwards JS (2002) Description and analysis of metabolic connectivity and dynamics in the human red blood cell. *Biophys J* 83(2): 646-662.
319. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 83(3): 1331-1334.
320. Vallino JJ, Stephanopoulos, G (1990) Flux determination in cellular bioreaction networks: Applications to lysine fermentations, In: *Frontiers in bioprocessing*

(Sikdar SK, Bier M, Todd P, Eds.), CRC Press, Boca Raton, USA.

321. Vallino JJ, Stephanopoulos G (1993) Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol Bioeng* 41(6): 633-646.
322. Raghunathan AU, Perez-Correa JR, Biegler LT (2003) Data reconciliation and parameter estimation in flux-balance analysis. *Biotechnol Bioeng* 84(6): 700-709.
323. Schmidt K, Carlsen M, Nielsen J, Villadsen J (1997) Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol Bioeng* 55(6): 831-840.
324. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, et al. (2007) Metabolic flux analysis at ultra-short time scale: Isotopically nonstationary <sup>13</sup>C labeling experiments. *J Biotechnol* 129(2): 249-267.
325. Deshpande RR, Yang TH, Heinzle E (2009) Towards metabolic and isotopic steady state in CHO batch cultures for reliable isotope-based metabolic profiling. *J Biotechnol* 4(2): 247-263.
326. Yuan Y, Yang TH, Heinzle E (2010) <sup>13</sup>C Metabolic flux analysis for larger-scale cultivation using gas chromatography-combustion-isotope ratio mass spectrometry. *Metab Eng* 12(4): 392-400.
327. Maier K, Hofmann U, Reuss M, Mauch K (2008) Identification of metabolic fluxes in hepatic cells from transient <sup>13</sup>C-labeling experiments: Part II. Flux estimation. *Biotechnol Bioeng* 100(2): 355-370.
328. Modak J, Deckwer WD, Zeng AP (2002) Metabolic control analysis of eucaryotic pyruvate dehydrogenase multienzyme complex. *Biotechnol Prog* 18(6): 1157-1169.
329. Selvarasu S, Lee DY, Karimi IA (2007) Identifying synergistically switching pathways for multi-product strain improvement using multiobjective flux balance analysis. *Comput Aided Chem Eng* 24: 1007-1012.
330. Saha R, Suresh S, Park W, Lee DY, Karimi IA (2007) Strain improvement and mediator selection for microbial fuel cell by genome scale in silico model, Proc. 17<sup>th</sup> European Symposium on Computer Aided Process Engineering (Plesu V, Agachi PS Eds.) Elsevier, Amsterdam, Europe.
331. Nagrath D, Avila-Elchiver M, Berthiaume F, Tilles AW, Messac A, et al. (2010) Soft constraints-based multiobjective framework for Flux Balance Analysis. *Metab Eng* 12(5): 429-445.
332. Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic network. *ONAS* 99(23): 15112-15117.
333. Sainz J, Pizarro F, Perez-Correa JR, Agosin E (2003) Modeling of yeast metabolism and process dynamics in batch fermentation. *Biotechnol Bioeng* 81(7): 818-828.
334. Vera J, Atauri P, Cascante M, Torres NV (2003) Multicriteria optimization of biochemical systems by linear programming: Application to production of ethanol by *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 83(3): 335-343.
335. Link H, Vera J, Weuster-Botz D, Darias NT, Franco-Lara E (2008) Multi-objective steady state optimization of biochemical reaction networks using a constrained genetic algorithm. *Comput. Chem. Eng* 32(1): 1707-1713.
336. Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol* 19: 125-130.
337. Wunderlich Z, Mirny LA (2006) Using the topology of metabolic networks to predict viability of mutant strains. *Biophys J* 91(6): 2304-2311.
338. Kayser A, Weber J, Hecht V, Rinas U (2005) Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-ratedependent metabolic efficiency at steady state. *Microbiology* 151(Pt 3): 693-706.
339. Orth JD, Fleming RMT, Palsson BO (2008) Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *Eco Sal Web: Escherichia coli and Salmonella: Cellular and molecular biology*. ASM Press Module 10.2.1.
340. Palsson B (2003) Flux-balance analysis: Basic concepts, Lecture #6, Systems Biology Research Group, University of California, San Diego.
341. Varma A, Palsson BO (1993) Metabolic capabilities of *Escherichia coli*. II. Optimal growth patterns. *J Theor Biol* 165(4): 503-522.
342. Bard JF (2005) Practical bilevel optimization Kluwer Academic Publ Boston.
343. Gumus ZH, Floudas CA (2001) Global optimization of nonlinear bilevel programming problems. *J Global Opt* 20(1): 1-31.
344. Gumus ZH, Floudas CA (2005) Global optimization of mixed integer bilevel programming problems. *Comput. Manag. Sci* 2(2): 163-251.
345. Mitsos A, Lemonidis P, Barton PI (2008) Global solution of bilevel programs with a nonconvex inner program. *J Global Opt* 42(3): 475-513.
346. Floudas CA (1995) Nonlinear and mixed-integer optimization. Oxford University Press Oxford, UK.
347. Coleman T, Branch MA, Grace A (1999) Optimization toolbox for use with MATLAB. User's Guide (ver. 2), MathWorks Inc, Natick.
348. Sbalzarini IF, Mueller S, Koumoutsakos P (2000) Multiobjective optimization using evolutionary algorithms, Proc. Summer Program, Center for Turbulence Research. NASA Ames Stanford University, USA, pp. 63-74.
349. Suna M, Kahraman R (2001) A comparative study of multi objective optimization methods in structural design. *Turk J Engin Environ Sci* 25: 69-78.
350. Tang KZ, Sun TK, Yang JY (2011) An improved genetic algorithm based on a novel selection strategy for nonlinear programming problems. *Comput Chem*

Eng 35(4): 615-621.

351. Dan A, Maria G (2012) Pareto optimal operating solutions for a semibatch reactor based on failure probability indices. *Chemical Engineering & Technology* 35(6): 1098-1103.
352. Fletcher R (1980) *Practical methods of optimization: Constrained optimization*, Wiley, New York, volume 2.
353. Zwillinger D, Krantz SG, Rosen KH (1996) *Standard Mathematical Tables and Formulae*. CRC Press, Boca Raton, USA.
354. Heinzle E, Dunn IJ, Furukawa K, Tanner RD (1982) Modelling of sustained oscillations observed in continuous culture of *Saccharomyces Cerevisiae*, In: Aarne H (Ed.) *Proc. Modelling & Control of Biotechnical Process IFAC Conference*. Helsinki, Finland Aug 17-19.
355. Tyson JJ, Novak B (2001) Regulation of the Eukaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions. *J theor Biol* 210(2): 249-263.
356. Tyson JJ (2002) Biochemical Oscillation. In: Fall CP, Marland ES, Wagner JM, Tyson JJ (Eds.) *Computational Cell Biology*. Springer verlag, Berlin chap 9, pp 230-260.
357. Franck UF (1980) Feedback kinetics in physicochemical oscillators *Ber. Bunsenges Phys Chem* 84(4): 334-341.
358. Kiparissides A, Koutinas M, Kontoravdi C, Mantalaris A, Pistikopoulos EN (2011) 'Closing the loop' in biological system modeling-From the in vitro. *Automatica* 47(6): 1147-1155.














AUTHOR NAME  
**GHEORGHE MARIA**