Extended Repression Mechanisms in Modelling Bistable Genetic Switches of Adjustable Characteristics within a Variable Cell Volume Modelling Framework

G. Maria*
Department of Chemical & Biochemical Engineering,
University Politehnica of Bucharest, P.O. 35–107 Bucharest, Romania

Original scientific paper
Received: May 22, 2013
Accepted: August 30, 2013

Genetic switches (GS) are a category of genetic regulatory circuits (GRC) responsible for the control and adaptation of cell metabolism to environmental changes. When mechanistic dynamic models of continuous variables are employed, a large variety of extended/reduced mathematical formulations have been proposed over the last decades, trying to simplify the regulatory network representation without risking missing the essential features of the real system. These models include the main GS species, which are: the DNA, mRNA, protein (enzyme) product, repressor, inducer / activator, RNA polymerase, and various individual or lumped metabolites. This study presents a comparison between Hill-type reduced models from literature and proposed non-Hill type extensions that replace the Hill coefficients by complex repression mechanisms. These complexes include elementary steps of the cross-repression control in inducible GS by means of rapid reversible reactions involving DNA/mRNA and dimeric transcription factors (TFs). Investigation is made under a variable-volume whole-cell (VVWC) holistic modelling framework, by placing the GS in a growing *E. coli* cell of known characteristics, and by mimicking the GS response to exogenous/endogenous gene expression activators/repressors. Comparison is made in terms of GS regulation indices, by analysing the influence of some system parameters (level of inducer, transcription factor, target protein, and Hill-type induction/repression nonlinearity) on the GS efficiency, and the close relationship between GS complexity and its stability.

Key words:
Bistable genetic switch, dynamic modelling, variable volume, whole-cell, repression complexes, Hill kinetics

Introduction

Balanced cell growth under favourable external conditions is characterized by the existence of key-species homeostasis, which corresponds to quasi-invariance of key species concentrations (enzymes, proteins, metabolites) ensuring content replication and metabolism maintenance despite external perturbations (in nutrients and metabolites) or internal cell changes. Such an internal regulation of synthesis/degradation reactions is realized by a complex genetic regulatory network that links a large number of GRCs used to control individual gene expression, such as:[1–5] toggle-switch (mutual repression control in two gene expression modules, that creates decision-making branch points between on/off states according to the presence of certain inducers); hysteretic / history-dependent behaviour (in the presence of certain exogenous inducers); oscillators (producing regular fluctuations of cell species levels); specific treatment of external signals by controlled gene expression such as amplitude filters, noise filters, or signal/ stimuli amplifiers; signalling circuits and cell–cell communicators. As the cell regulatory systems are module-based organized,[6] complex feed-back and feed-forward loops are employed for self- or cross-activation / repression of interconnected gene expression, leading to different interaction alternatives (directly/inversely, perfect/incomplete coupled/uncoupled connections[1,7]) of a gene with up to 23–25 other genes.[8] Consequently, the cell metabolism can be changed by modifying/designing GRCs conferring new properties/functions to the mutant cells (i.e. desired 'motifs'), while engineered/synthetic gene circuits can be designed by using the Synthetic Biology tools.[4,9–14]

A key step in the GRC analysis is the possibility to *in-silico* (model-based) test their properties, and to design modified circuits by simulating their regulatory effectiveness.[15] Structured mechanistic-based kinetic models, including hundreds to thousands biochemical cell reactions, continue to be

*Corresponding author: e-mail: gmaria99m@hotmail.com

developed, leading to considerable improvement in the predictive power at a cellular but also at a macroscopic (bioreactor) level. These models are suitable for accurately predicting the cell response/adaptation to environmental perturbations by modelling GRCs controlling the cell metabolism, and the behaviour of cloned/modified microorganisms of industrial or medical interest.[16–21] Thus, the modern bioengineering concept "from gene to product"[22] focused on improving the industrial bioprocess performance by designing modified cells, can be efficiently assisted by adequate GRC simulators. The *in-silico* analysis of a GRC uses a reductionistic approach by decoupling simple networks from the complex cell system to be studied individually, followed by rules to "recreate" the real system.[23–24]

In particular, modelling and designing genetic switches in modified cells continues to be a very attractive subject due to its tremendous importance in adapting the cell metabolism to certain objectives.[42] Starting from the classical Jacob-Monod[25] GS mechanism and from the simple model published by Griffith in 1968,[26] a large number of models continues to be reported, their complexity depending on the approached system. Examples include GS from *E. coli*, such as: lactose(*lac*)-operon (including three mutually repressed gene expression regulatory modules acting as a 'multiple'-switch[27]); galactose(*gal*)-operon (including three structural genes, one dimeric repressor and two promoters[28,29]); tryptophan(*trp)*-operon (including five structural genes, one promoter and one operator[30]), the uncoupled GRCs of asparagine (*asn*), purine (*pur*), and guanine (*gua*) operators, etc. (see Hasty *et al*.[23] for other GSs).

Modelling GRCs responsible for gene expression/GS control requires steady experimental and computational efforts to decipher the regulatory loops/motifs at a quantitative level.[4] The number of involved species in GRCs (genes, mRNA, proteins) is of $O(10^3–10^4)$, the number of gene TFs of $O(10^3)$, while the number of parameters of gene-TF interactions theoretically corresponds to the number of states multiplied by the number of TFs.[31] For instance, in *E. coli* there are 350 proteins working as TFs able to bind to DNA, 43 % of proteins act as repressors, 35 % as activators, and 22 % as dual regulators,[30] their activity being modulated by small molecules of effector species (amino acids or metabolites). However, the genetic networks in a cell are known as being sparsely interconnected, a gene expression being related to a small number (up to 23–25) of other gene expressions, and thus the majority of the parameters of such a "complete" gene network model will be zero.[31] Alternatively, the modular-based representation of such a complexity becomes more suitable in reproducing the GRC characteristics,[10,23] in an 'expandable building-blocks'-like construction able to represent the tight control of gene expression of optimized properties, a quick dynamic response of high sensitivity to specific inducers, and robustness of the genetic circuit (i.e. low sensitivity vs. external noise).

Griffith[26] proved that in a gene expression the mRNA level can easily display bistable (low or high) expression levels depending on the demand for the encoding protein, being simultaneously regulated by positive and negative feedback loops involving several regulatory components (allostery) of high nonlinearity. As these GRC motifs constitute the building blocks of large gene regulatory networks, the modular modelling approach appears to be justified.[6] One semi-autonomous Gene Expression Regulatory Module (GERM) includes all individual and lumped cell components and reactions that participate in the control of a certain gene expression.[18] Linking GERM in regulatory networks leads to the control of the whole cell metabolism, ensuring metabolic efficiency (minimum energy and substrate consumption), individual or associative component functions, hierarchical organization, system homeostasis, balanced cell growth, and regulatory efficiency at various network levels.[32] Moreover, the main steps in a GERM can be modulated by varying the TF level, i.e. either an inhibitor protein binding the operator to prevent a certain gene expression (negative regulation), or a protein binding the promoter to enable RNA polymerase to initiate the gene transcription (positive regulation), or RNA processing, RNA translation, post-translational protein modification, enzyme inactivation or degradation, etc.[15]

To dynamically represent the GERM/GS characteristics in a quantitative way, a large variety of models using continuous, Boolean, or stochastic variables have been proposed, in a structured approach accounting for individual or lumped TFs, intermediates, metabolites, enzyme complexes, etc.[24] The continuous variable mechanistic models, using ordinary differential equations (ODE) and requiring more structured information, are still largely used as being computationally very tractable to adequately represent continuous cell processes, response to perturbations, cell adaptation, and system complex nonlinear behaviour, by accounting for mechanistically-based interactions among individual or lumped species and reactions (aggregate pools), of physical significance.[33] The Boolean approach, with variables taking only discrete values, is suitable only for a rough representation of the GRC, becoming unrealistic for large networks, being unable to reproduce molecular interactions in detail, such as slow and continuous responses to perturbations.[15] The stochastic variable approach, replacing the 'average' solution of continuous-vari-

able ODE models by a detailed random-based simulator, is very suitable for representing stochastic events such as cell signalling, gene mutation, faulty switches, random interactions of species present in small amounts, etc. Inherent stochastic fluctuations in the species concentrations can introduce significant variability into the dynamics of genetic regulation, and can even "mask" some GRC steady-states.[33] Because stochastic models are very laborious, hybrid models (e.g. stochastic-continuous variables) have often been reported as being feasible to represent inducible GRCs (review of Maria[18,34]).

This paper is focused on developing mechanistic ODE models of GSs based on the reported modelling experience. Recently, Maria[5,34] tested some simplified representations of GS, by including a Hill-type gene expression activation (of fourth order), and single cross- and self-repression buffering reactions to control the gene transcription using dimeric TF-s, all being formulated in a VVWC modelling framework.

The aim of this paper is to extend the GS modelling investigation of Maria,[5] by comparing reduced Hill-type kinetic models from literature with proposed non-Hill type extensions that replace the Hill coefficients with complex repression mechanisms. The proposed models include multiple and adjustable elementary steps of the nonlinear cross-repression control in inducible GS by means of rapid reversible elementary reactions involving DNA/mRNA. The GS properties are tuned by means of several parameters, i.e. the number of repression steps, TF-level, and adjustable self-repression strength according to the stability and effectiveness targets. Model comparison is made by employing several GS regulation indices, such as: switch certainty, steady-state stability and sensitivity to exogenous inducers, response rate and transient or recovering times after a stationary or dynamic perturbation in the external inducers or GS target proteins. The designed GS is placed in an *E. coli* cell (K-12 strain of known characteristics[35]), by analysing the influence of some system parameters (level of inducer, TFs, target protein, and induction/repression nonlinearity) on the GS efficiency. The analysis points out the close relationship between regulatory system complexity, its efficiency and GS stability. Model versatility suggests further extensions and use for *in-silico* design of engineered gene circuits of practical interest.

Continuous variable GERM/GS models in literature

Literature includes a large number of attempts in representing GSs, by including the main species involved in the gene expression, that are (Fig. 1):

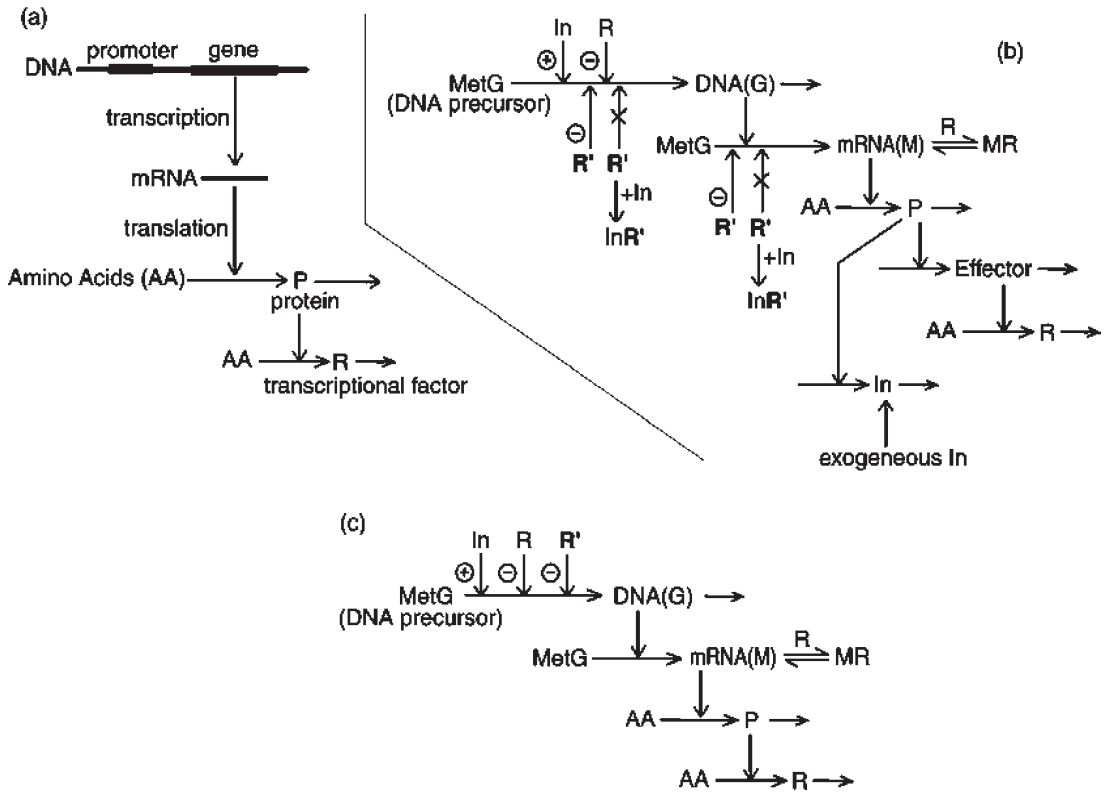

Fig. 1 – *Simplified representations (a-c) of a gene G expression regulatory module. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions; G = gene encoding protein P; M = mRNA; R, R' = transcriptional factors (repressors); In = inducer; MetG = DNA precursor metabolites.*

DNA (gene G), mRNA (intermediate M), end product/enzyme (P), repressor (R), inducer/activator (A,I,In), RNAP (RNA polymerase binding to free promoter), operon (O), and various individual or lumped metabolites. Some models also account for the delay time before re-initiation of the transcription, or to synthetize mRNA. As reviewed by Smolen *et al*.,[33] the time delays in GRC loops and multi-stability lead to a history-dependent response to perturbations. Also, TFs are reported acting in dimeric or even tetrameric forms involving feedback loops with several intermediates, while highly-nonlinear species interactions are essential in getting a large variety of complex dynamic behaviours.[3,23,36] However, extended kinetic models incorporate too many species mass balances, with parameters difficult to estimate from often incomplete data and, consequently, difficult to use for practical applications. An alternative is to use reduced or "multi-scale" models that combine unstructured with structured process characteristics to generate more precise predictions.[19,24,31,37,38]

Depending on the structure and manner of representing the kinetic terms, the ODE models of GERM/GS can be classified as follows:[15]

i) *Extended ODE* dynamic models trying to mechanistically include a significant number of elementary steps involved in the gene expressions. As an example, Salis & Kaznessis[3,36] designed a bistable switch from the *lac* and *ara* operons of *E. coli*, in which the transcriptional regulation is modelled by using a stochastic approach accounting for 40 elementary reactions and 27 species (reduced version) or 70 reactions and 50 species (extended version), with repressors in dimeric or tetrameric form. Several reviews on extended mechanistic GRC models including a large array of experimental data, kinetic parameters, regulatory indices (sensitivity, robustness, efficiency, responsiveness, stability strength, species connectivity) are given in literature.[5,24,26,34]

ii) *Reduced ODE* models include lumped species and reactions, with both elementary and Michaelis-Menten/Hill-type kinetic expressions based on quasi-steady-state assumptions for mRNA, TFs (activators, repressors), or intermediates.[39] The empirical Hill-like activation $\left(k_o + [A]^n / (K + [A]^n)\right)$,[39] or inhibition/repression $\left([R]^n / (K^n + [R]^n)\right)$[30] terms of transcriptional kinetics account for the existence of a saturating phenomenon, "a passive diffusion of the inducer throughout the cell membrane" (parameter k_o), and a non-linearity of the gene expression response to the activation/repression action of TFs.[15,39] To get a GS, $n > 1$ should be set to avoid existence of only one unique stable steady-state.[33,40] In fact, $n \geq 2$ Hill-type kinetics is a lumped representation of an allosteric control (i.e. at least two successive rapid reversible buffering reactions) at various levels:[30] operator site or mRNA binding the repressor, RNA polymerase binding the activator/inducer. The Hill exponents also determine the co-operative characteristics of repressor-operator and repressor-inducer binding.[41] There are different ways of simplifying the GERM representation in a GS, as follows:

– include only the two mRNA (M1,M2), or only the two product proteins synthesis (P1,P2) with Hill type cross-repression terms of $n \geq 2$ exponents (usually $n = 2$–4);[30,40,42]

– include mRNA (M), protein P and repressor R synthesis (Jacob & Monod[25] model), with mutual Hill repression of mRNA formation in the two or three GERMs of {M,P,R}-type [27,43] of Hill exponent $n \geq 2$. Many times the repressors of M1-M2 synthesis are considered being the protein products P2-P1 respectively, with strong Hill cross-repression exponents $n > \sec^p(\pi / p) > 1$, depending on the concentration of the activator or inhibitor (usually $n = 2$–4; sec = reciprocal to the cosine; p = no. of successive repression steps of the negative feedback loop).[26,44] The self-repression of gene expression can also be included with a Hill exponent $n = 2$.[45] Polynikis *et al*.[15] studied the influence of the two Hill repression coefficients (n_1, n_2 for the repression of M1, M2 synthesis) on the GS properties and found that for n_1, $n_2 > 2$ the system exhibits multi-stationarity with a Hopf bifurcation point, while for $n_1 n_2 > 4$ the system exhibits oscillatory behaviour in certain regions of the parametric space. Voit[46] pointed out for a {G,M,P,I} reduced GERM that Hill induction coefficient $n > 1$ (usually $n = 4$) leads to multi-stationarity of different stability type depending on the initial species concentrations, by recommending the Hill self-repression exponent of value 0.5. Widder *et al*.[47] considered a {G,M,P} system without self-regulation (repression), with activation and cross-repression Hill exponents $n \geq 3$ inducing Hopf multi-stability. Smolen *et al*.[33] and Tabaka[30] detailed the importance of time delays (of tenths of seconds) for the GS syntheses (due to species diffusion reasons), and of the degradation rate of activator/repressor in adjusting the GS protein levels.

iii) *S-type* ODE models, with an apparent power-law representation of the nonlinear GERM/GS kinetics, are simple and computationally very convenient.[48] Even if the apparent rate constants lack of any physical meaning and fictitious auxiliary variables are necessary for non-power-law terms, such a modelling approach was proved to be versatile and effective in representing complex behaviour of GRC, such as: saturation or sigmoidal response, multi-stability, bifurcations, oscillatory behaviour, and hysteresis.[1,10] However, lacking the essential information on key-intermediates and elementary

steps dynamics, such representations cannot accurately reproduce some GRC properties, such as steady-state stability strength, or responsivity to nonlinear perturbations. Also, most of the model parameters' lack of significance, and min/max thresholds are usually imposed on the activators/repressors.[34,46,49]

iv) Piece-wise-linear approximations of Hill functions, or discretized ODE models, use simple algebraic relationships among species concentrations to overcome the incomplete quantitative data cases, without using kinetic parameters required by the detailed mechanistic models.[15,49,50]

In spite of tremendous progress made in the cell process analysis and "-omics" databanks, a consensus concerning a mathematical framework to decide the best modelling approach does not exist. However, an adequate, even simple GRC model needs to include all relevant steps and intermediates of the reaction pathway, but also to reproduce connections of the GRC with the rest of the cell responsible for the holistic response of bacteria to different environmental perturbations under a continuous cell-volume growth and content replication.

When continuous variables are used, the default-modelling framework is that using species concentrations, by accounting for the cell-growing rate as a 'constant decay' rate of key-species (often lumped with the degrading rate) in a so-called 'diluting' rate. Such a representation might be satisfactory for many applications, but could distort the predictions of an accurate modelling of cell regulatory processes under perturbed conditions, as long as connection with the whole-cell content dynamics is not accounted for.[18,32] By contrast, the variable-volume whole-cell (VVWC) modelling framework formulated by Grainger *et al*.,[51] and adopted in this study in the variant of Maria,[18] explicitly links the volume growth, external conditions, osmotic pressure, cell content ballast and net reaction rates for all cell-components.[52] As underlined by Maria,[18,32] in a VVWC representation the dilution rate is not constant, the large cell ballast tending to stabilize the system by smoothing the perturbations. Thus, the GRC responsiveness is slightly lower comparatively to an isolate GRC and, consequently, some of the regulatory properties are not overestimated.

The proposed bistable genetic switch model

Isotonic variable volume modelling framework

The VVWC modelling framework promoted by Maria,[18] basically accounts for the cell volume dynamics and cell content dilution rate by linking the osmotic pressure, system temperature and the sum of all species reaction rates (individual or lumped; see the simplificatory hypotheses in Table 1). Such a supplementary constraint equation increases species connectivity, and better reproduces some of the cell holistic properties, such as the cell large content inertial effect in treating perturbations, the effect of the indirect or secondary perturbations of cell species concentrations transmitted via cell-volume variation under isotonic conditions (ca. 80 % of the cell cycle). It should also be mentioned that a VVWC model must include all cell species at some level of detail (i.e. individual or lumped species and reactions), in order to be consistent with the mentioned hypotheses because all components contribute to the common volume dynamics. In such a manner, the number of model rate constants increases, leading to a corresponding increase in the identification effort. The GRC model rate constants are estimated from experimental concentrations of key-species at homeostasis (i.e. fulfilment of the steady-state condition under a balanced cell growth), from experimental species dynamics information (kinetic data, if any), and from imposing optimum regulatory properties in terms of maximum responsiveness and efficiency (i.e. minimum transition or steady-state recovering times after a stationary or dynamic perturbation). Some physical meaning constraints, such as bounded rate constants and species concentration levels are also included.

The dynamics of the individual species but also of the lumped content can be mimicked under simulated stationary or perturbed environmental conditions, by "placing" the GERM chain in a growing cell. Reported tests with simple GRC models under a VVWC formulation are very promising in predicting local and holistic features of the metabolic network vs. classical formulation that tends to overestimate some of the regulatory dynamic properties.[18,32,53,54] The main VVWC modelling advantages come from the possibility of linking the dynamics of the studied metabolic pathway to the rest of the (lumped) cell, thus offering a holistic evaluation of the involved GRC in the cell context. Even if the VVWC models work in parallel with species copynumbers and cytosolic concentrations (see the linking formula in Table 1 – last column), one disadvantage of using continuous variable formulations is the possibility of translating fractional concentrations to fraction of copynumbers. For instance, for a born *E. coli* cell volume of $V_{cyt,o} = 1.66 \cdot 10^{-15}$ L, one gene G copynumber translates to 1 nmol L^{-1} concentration, while a concentration of [G] = 0.5 nmol L^{-1} must be interpreted either as an average of time-invariant in a population of cells (e.g. half of all cells containing 1 copynumber of G), or as a time-dependent average for a single cell (e.g. that cell contains 1 copynumber of G half of the time).

Table 1 – *Dynamic variable cell-volume modelling framework (after Maria[18]).*

$\frac{dc_j}{dt} = \frac{1}{V}\frac{dn_j}{dt} - D\,c_j = g_j(\boldsymbol{c},\boldsymbol{k})$ $\frac{1}{V}\frac{dn_j}{dt} = r_j(\boldsymbol{c},\boldsymbol{k})\ ;\ \ j = 1,...,n_s$	Species mass balance (continuous state variables)
$D = \frac{1}{V}\frac{dV}{dt} = \left(\frac{RT}{\pi}\right)\sum_j^{n_s}\left(\frac{1}{V}\frac{dn_j}{dt}\right)$	Cell content dilution rate (from Pfeffer's law in diluted isotonic solutions)
$\overline{RT} = \frac{}{V}\sum_{j=1}^{n_s} n_j = \sum_{j=1}^{n_s} c_j = \sum_{j=1}^{n_s} c_{jo}$	Constant osmotic pressure constraint
$\sum_j^{all} c_{j,cyt} = \sum_j^{all} c_{j,env}$	Isotonic osmolarity constraint

Hypotheses:

– open cell system of uniform content (negligible diffusion resistance);

– semi-permeable membrane, of negligible volume and resistance to nutrient diffusion, following the cell growing dynamics;

– constant osmotic pressure, ensuring membrane integrity ($\pi_{cyt} = \pi_{env}$ = constant);

– average logarithmic growing rate: $D_s = \ln(V/V_o) = \ln(2)/t_c$;

– balanced cell growth with a constant growing rate over the cell-cycle under quasi-constant environmental conditions;

– homeostatic stationary growth for: $\left(dc_j/dt\right)_s = g_j(\boldsymbol{c}_s,\boldsymbol{k}) = 0$;

– isotonic perturbations in cell volume result from variations in the species copynumber;

– species concentration formula: (*no. of copies per cell*)/ $(N_A \times V_{cyt,o})$.

The two gene expression regulatory modules of the switch

Detailed representation of a GERM includes a multi-level control of intermediate and product synthesis (mRNA, enzymes, regulators) by means of positive/negative feedback loops and (dimeric) TFs of variable concentration. Such adjustable effector species range the catalytic activity at each cascade level to cope with internal and external perturbations.[5,32] Negative self-regulation is proved to speed up the response time of the GRC to external stimuli, and promotes robustness to fluctuations in production rate.[4,5] In contrast, positive self-regulation slows responses and can lead to bi-stability of the expression module.

The proposed GS reaction pathway of Fig. 2 (in a VVWC formulation) includes two interconnected GERMs, of identical structure, with mutually (cross-)repressed gene G2 and G3 expression. An additional GERM is added to mimic the cell ballast, inertial effect to perturbations, and the controlled replication of the lumped cell content (genome G1 and proteome P1). The G1/P1 cell content replication module is necessary to make the cell model consistent with the isotonic assumption, the VVWC holistic formulation requiring inclusion of all cell species at a certain detailing degree (individual or lumped). Thus, the dilution constant D in Table 1, kept constant over the cell cycle in the classical ki-

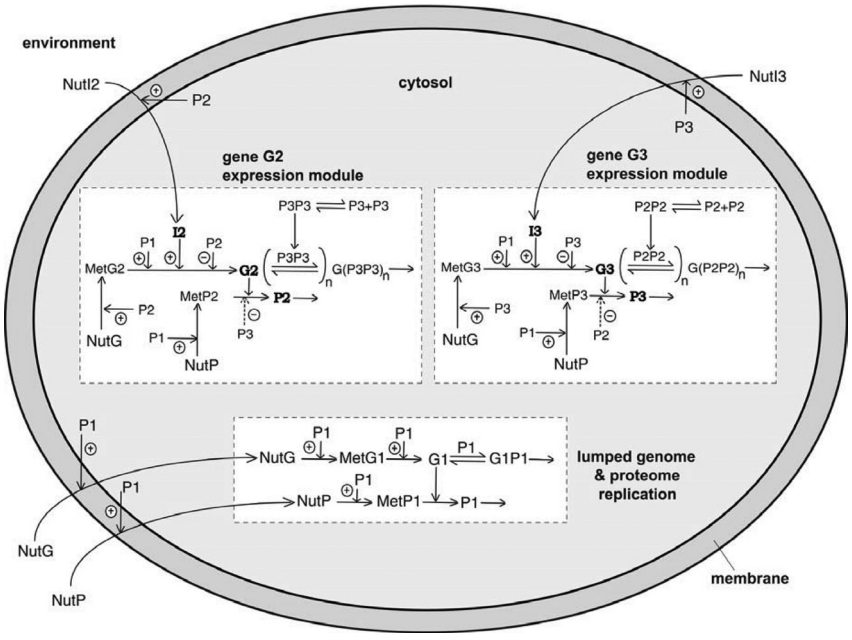

Fig. 2 – *Genetic switch model including two gene G2 and G3 expression (self- and cross-) regulatory modules. The GRC is placed in a growing E. coli cell, by mimicking the homeostasis and cell response to stationary and dynamic perturbations in environmental NutI2 and NutI3 inducers with a VVWC model. The cell content ('ballast') influence is mimicked by including the lumped proteome P1 and genome G1 replication module. Notations: G1/P1 = lumped genome/proteome; MetG1/MetP1 = lumped metabolome; P2, P3 = genetic switch target proteins; MetG2, MetG3, MetP2, MetP3 = individual metabolites; I2, I3 = inducers; NutI2, NutI3 = external stimuli; NutG, and MetG (lumped external nutrients) are the precursors of DNA, mRNA, and amino-acids respectively; ⊕/Θ = positive/negative regulatory loops. Horizontal arrows indicate reactions; vertical arrows indicate catalytic actions (dashed vertical arrows indicate effects included in only some of the presented GS models).*

Table 2 – *Estimated kinetic parameters of tested GS models for stationary E. coli cell growth conditions of Table 2 (TF level of* $[P2P2]_s = [P3P3]_s = 5$ *nmol* L^{-1}*). Rate constant units correspond to minutes and nmol* L^{-1}*, while rates* $(dn_j/dt)/V$ *are formulated for variable volume conditions.*

	Reaction	Rate expression (nmol L^{-1} min^{-1})	Rate constants		
			pure Hill model	buffering reactions of cross-repression	
			WC-B0	WC-B1	WC-B4
1	$NutG + P1 \rightarrow MetG1 + P1$	$k_1[NutG][P1]$	$4.61 \cdot 10^{-10}$	$4.61 \cdot 10^{-10}$	$4.61 \cdot 10^{-10}$
2	$NutP + P1 \rightarrow MetP1 + P1$	$k_2[NutP][P1]$	$7.11 \cdot 10^{-10}$	$7.11 \cdot 10^{-10}$	$7.11 \cdot 10^{-10}$
3	$MetG1 + P1 \rightarrow G1 + P1$	$k_3[MetG1][P1]$	$1.56 \cdot 10^{-13}$	$1.56 \cdot 10^{-13}$	$1.56 \cdot 10^{-13}$
4	$MetP1 + G1 \rightarrow P1 + G1$	$k_4[MetP1][G1]$	$1.03 \cdot 10^{-7}$	$1.03 \cdot 10^{-7}$	$1.03 \cdot 10^{-7}$
5	$G1 + P1 \rightarrow G1P1$	$k_5[G1][P1]$	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$
6	$G1P1 \rightarrow G1 + P1$	$k_6[G1P1]$	10^5 (a)	10^5 (a)	10^5 (a)
7	$NutI2 + P2 \rightarrow I2 + P2$	$k_7[NutI2][P2]$	$1.38 \cdot 10^{-3}$	$1.38 \cdot 10^{-3}$	$1.38 \cdot 10^{-3}$
8	$NutI3 + P3 \rightarrow I3 + P3$	$k_8[NutI3][P3]$	$1.38 \cdot 10^{-3}$	$1.38 \cdot 10^{-3}$	$1.38 \cdot 10^{-3}$
9	$NutG + P2 \rightarrow MetG2 + P2$	$k_9[NutG][P2]$	$4.62 \cdot 10^{-7}$	$4.62 \cdot 10^{-7}$	$4.62 \cdot 10^{-7}$
10	$NutP + P1 \rightarrow MetP2 + P1$	$k_{10}[NutP][P1]$	$2.31 \cdot 10^{-12}$	$2.31 \cdot 10^{-12}$	$2.31 \cdot 10^{-12}$
11	$MetG2\ (+P1+P2+I2) \rightarrow G2$	$\frac{k_{11}[MetG2][P1](1+B[I2]^4)}{(K_{G2}+[I2]^4)[P2]^{nR}}$	$k_{11} = 3.46 \cdot 10^{-13}$ $B = 2$ $nR = 1$ K_{G2} (b)	$3.46 \cdot 10^{-13}$ $B = 2$ $nR = 1$ K_{G2} (b)	$8.66 \cdot 10^{-12}$ $B = 2$ $nR = 3$ K_{G2} (b)
12	$MetP2 + G2 \rightarrow P2 + G2$	$k_{12}\frac{[MetP2][G2]}{K_{P2}+[P3]^{nH}}$	$k_{12} = 4.33 \cdot 10^{-6}$ $nH = 2$ K_{P2} (c)	$9.70 \cdot 10^{-8}$ $nH = 0$ $K_{P2} = 1$	$6.58 \cdot 10^{-7}$ $nH = 0$ $K_{P2} = 1$
13	$NutG + P3 \rightarrow MetG3 + P3$	$k_{13}[NutG][P3]$	$4.62 \cdot 10^{-7}$	$4.62 \cdot 10^{-7}$	$4.62 \cdot 10^{-7}$
14	$NutP + P1 \rightarrow MetP3 + P1$	$k_{14}[NutP][P1]$	$2.31 \cdot 10^{-12}$	$2.31 \cdot 10^{-12}$	$2.31 \cdot 10^{-12}$
15	$MetG3\ (+P1+P3+I3) \rightarrow G3$	$\frac{k_{15}[MetG3][P1](1+B[I3]^4)}{(K_{G3}+[I3]^4)[P3]^{nR}}$	$k_{15} = 3.46 \cdot 10^{-13}$ $B = 2$ $nR = 1$ K_{G3} (b)	$3.46 \cdot 10^{-13}$ $B = 2$ $nR = 1$ K_{G3} (b)	$8.66 \cdot 10^{-12}$ $B = 2$ $nR = 3$ K_{G3} (b)
16	$MetP3 + G3 \rightarrow P3 + G3$	$k_{16}\frac{[MetP3][G3]}{K_{P3}+[P2]^{nH}}$	$k_{16} = 4.33 \cdot 10^{-6}$ $nH = 2$ K_{P3} (c)	$9.70 \cdot 10^{-8}$ $nH = 0$ $K_{P3} = 1$	$6.58 \cdot 10^{-7}$ $nH = 0$ $K_{P3} = 1$
17	$P2 + P2 \rightarrow P2P2$	$k_{17}[P2][P2]$	–	$2.00 \cdot 10^3$	$2.00 \cdot 10^4$
18	$P2P2 \rightarrow P2 + P2$	$k_{18}[P2P2]$	–	10^5 (a)	10^5 (a)
19	$P3 + P3 \rightarrow P3P3$	$k_{19}[P3][P3]$	–	$2.00 \cdot 10^3$	$2.00 \cdot 10^4$
20	$P3P3 \rightarrow P3 + P3$	$k_{20}[P3P3]$	–	10^5 (a)	10^5 (a)
21	$G2 + P3P3 \rightarrow G2(P3P3)$	$k_{21}[G2][P3P3]$	–	$2.00 \cdot 10^5$	$2.00 \cdot 10^4$
22	$G2(P3P3) \rightarrow G2 + P3P3$	$k_{22}[G2(P3P3)]$	–	10^5 (a)	10^5 (a)
23	$G2(P3P3) + P3P3 \rightarrow G2(P3P3)_2$	$k_{23}[G2(P3P3)][P3P3]$	–	–	$2.00 \cdot 10^4$
24	$G2(P3P3)_2 \rightarrow G2(P3P3) + P3P3$	$k_{24}[G2(P3P3)_2]$	–	–	10^5 (a)
25	$G2(P3P3)_2 + P3P3 \rightarrow G2(P3P3)_3$	$k_{25}[G2(P3P3)_2][P3P3]$	–	–	$2.00 \cdot 10^4$
26	$G2(P3P3)_3 \rightarrow G2(P3P3)_2 + P3P3$	$k_{26}[G2(P3P3)_3]$	–	–	10^5 (a)
27	$G2(P3P3)_3 + P3P3 \rightarrow G2(P3P3)_4$	$k_{27}[G2(P3P3)_3][P3P3]$	–	–	$2.00 \cdot 10^4$
28	$G2(P3P3)_4 \rightarrow G2(P3P3)_3 + P3P3$	$k_{28}[G2(P3P3)_4]$	–	–	10^5 (a)
29	$G3 + P2P2 \rightarrow G3(P2P2)$	$k_{29}[G3][P2P2]$	–	$2.00 \cdot 10^5$	$2.00 \cdot 10^4$
30	$G3(P2P2) \rightarrow G3 + P2P2$	$k_{30}[G3(P2P2)]$	–	10^5 (a)	10^5 (a)
31	$G3(P2P2) + P2P2 \rightarrow G3(P2P2)_2$	$k_{31}[G3(P2P2)][P2P2]$	–	–	$2.00 \cdot 10^4$
32	$G3(P2P2)_2 \rightarrow G3(P2P2) + P2P2$	$k_{32}[G3(P2P2)_2]$	–	–	10^5 (a)
33	$G3(P2P2)_2 + P2P2 \rightarrow G3(P2P2)_3$	$k_{33}[G3(P2P2)_2][P2P2]$	–	–	$2.00 \cdot 10^4$
34	$G3(P2P2)_3 \rightarrow G3(P2P2)_2 + P2P2$	$k_{34}[G3(P2P2)_3]$	–	–	10^5 (a)
35	$G3(P2P2)_3 + P2P2 \rightarrow G3(P2P2)_4$	$k_{35}[G3(P2P2)_3][P2P2]$	–	–	$2.00 \cdot 10^4$
36	$G3(P2P2)_4 \rightarrow G3(P2P2)_3 + P2P2$	$k_{36}[G3(P2P2)_4]$	–	–	10^5 (a)

(a) adopted value for rate constant of ca. $10^7 D_s$.[12,18,20,56]
(b) $K_{G2} = [I2_{ref}]^4$; $K_{G3} = [I3_{ref}]^4$; $[I2_{ref}] = [I3_{ref}] = 1$ nmol L^{-1};
(c) $K_{P2} = [P3_{ref}]^{nH}$; $K_{P3} = [P2_{ref}]^{nH}$; $[P2_{ref}] = [P3_{ref}] = 10$ nmol L^{-1}.

netic formulations, becomes slightly variable with an evolution depending on the internal reactions as response to external perturbations.

The GS reaction rate expressions are presented in Table 2, model formulation being made in two alternatives:

i) *model WC-B0* (reactions no. 7–16 in Table 2), used as a reference model, is inspired from literature.[40,46] The model includes only reduced Hill-type kinetic expressions for genes {G2,G3} and target proteins {P2,P3} synthesis, similar to the reported models of Voit,[46] Widder *et al*.,[47] and Tyson *et al*.[26] The Hill-exponent for gene expression activation with {I2,I3} is adopted at a value of n = 4, as recommended by Voit[46] and Widder *et al*.,[47] corresponding to a cooperative binding of four inducer molecules to the promoter, thus resulting in a highly nonlinear amplification of the gene expression. The Hill-type cross-repression of protein synthesis (P2 by P3, and vice-versa[40]) was applied directly to the {mRNA,DNA} lump (reactions 12 and 16, similar to Griffiths' formulation;[26] Fig 2). As suggested in literature, different Hill exponents $nH \geq 2$ (for multi-stability reasons) and self-repression coefficients $nR \geq 1$ will be investigated in relation to other GS properties. The apparent dissociation constants K_G and K_P are adopted at the average value of inducer and repressor concentrations (taken at the powers 4 and nH respectively). Low values for the self-repression exponent $nR << 1$ (that is less than one effector molecule per gene operator, as suggested by Voit[46]) are avoided, being considered unrealistic in most of reported GERMs.[10] The WC-B0 model is completed with the nutrient and inducer import reactions into the cell, and formation of gene and protein metabolic precursors {MetG2,MetG3,MetP2,MetP3}.

ii) *models WC-Bn* are original extensions of the WC-B1 model proposed by Maria,[5] and include Hill-type rate expressions only for genes {G2,G3} synthesis, by replacing the Hill-expression for {P2,P3} synthesis (reactions 12 and 16) by a non-Hill type complex repression mechanism. This complex includes a certain number n_{buff} of explicit rapid reversible (buffering) elementary reactions cross-binding dimeric repressors TF = P2P2 and TF = P3P3 to the gene G3 and G2 operator, respectively. The "catalytic" gene activity can be adjusted in such a manner by employing successive buffering reactions of type Gi + PjPj ⇔ GiPjPj ⇔ … ⇔ $Gi(PjPj)_n$. Consequently, models WC-Bn include the core reactions no. 1–16, plus reactions 17–20 for reversible synthesis of dimeric TF = {P2P2,P3P3}, and the reactions 21–36 corresponding to n_{buff} = 1–4 buffering steps. The TFs were adopted in a dimeric form, as reported by most of the experimental studies (even if tetrameric repressors are also possible[3,36]). It should be mentioned that WC-Bn models include extensions of the WC-B1 model proposed by Maria[5] in detailing the cross-repression mechanism and self-repression strength by means of a number of tunable model parameters (nR, n_{buff}, [TF], repression scheme).

The proposed models are quite flexible, the GS regulatory properties being adjusted by varying the level of external inducers, TFs, target proteins, and the strength of cross-/self-repressing elements (nH, n_{buff}, nR). In this lumped representation, the proteins play the role of permeases and metabolases for exogenous stimuli {NutI2, NutI3} import and inducer {I2,I3} production. There are also some other simplificatory hypotheses adopted, which are as follows: the dilution rate is uniform for all species, the degradation steps (of repressor, expressed protein, activator) are neglected, and no delay time constants have been included in kinetic expressions to account for successive protein synthesis steps of ribosome binding, peptide elongation, protein folding, dimeric complexes formation, and their diffusion to the DNA-binding site. In contrast to other simplificatory GS models,[48,49,55] no min/max thresholds have been imposed to the inducers or repressors to "artificially" limit the DNA/mRNA synthesis rate.

Mimicking the cell content 'ballast' replication

To mime the GS behaviour in the VVWC environment, the two GERMs of the GS have been placed in an *E. coli* cell of known characteristics (the K-12 strain[35]). To also mime the whole content replication and volume growth under stationary or perturbed environmental conditions, a GERM describing the replication of the lumped genome (G1) and proteome (P1) (reactions 1–6 in Table 2) has also been added. For simplicity, the P1/G1 synthesis is assumed to be controlled by a rapid buffering reaction G1 + P1 ⇔ G1P1, close to its equilibrium, with a dissociation constant much larger than those of the core synthesis.[18] The lumped proteome P1 plays the role of permease for nutrients NutG, NutP import, of a metabolase for metabolites MetG1, MetP1, MetP2, MetP3 synthesis, and of a polymerase for the lumped genome G1 and target genes G2 and G3 production (Fig. 2). Thus, some holistic cell properties can be modelled, such as the cell content inertial/smoothing effect in treating perturbations, the effect of the indirect or secondary perturbations transmitted via the cell-volume under isotonic osmolarity conditions.[18]

Genetic switch performance indices

When environmental concentration of an exogenous inducer (e.g. NutI2) changes, its cytosolic level (I2) varies, leading to activation of G2 expres-

sion and, as a consequence, to the repression of the other gene of the switch (G3). The cell GS reacts vice-versa when raising NutI3 inducer level instead of NutI2. The resulting GS is either "On" when [P2] > [P3], or "Off" otherwise. The short transient times, high sensitivity to specific inducers, robust dynamic response, and tight control of gene expression are the common goals when designing an optimised GS.[1] To quantitatively characterize the GS efficiency, several indices have been defined, some of them in a similar way to those used by the nonlinear system control theory:[5]

– *switch certainty (SC),* which is high when the stationary ratio of the two co-expressed proteins $[P2]_s/[P3]_s$ is high following an induction with NutI2. A "good" switch must usually realize SC >5–10 (depending on the inducer level).[36,40]

– steady-state *stability strength*, related to the firmness of GS steady-states (of local stability) obtained for certain inducer level. For the ODE model of general form $d\boldsymbol{c}/dt = \boldsymbol{g}(\boldsymbol{c},\boldsymbol{k},t) \approx \boldsymbol{J}^T\boldsymbol{c}$ (Table 1), the stability condition requires that $\mathrm{Re}(\lambda_j) < 0$ for all j (where λ_j are the eigenvalues of Jacobian $\boldsymbol{J} = d\boldsymbol{g}/d\boldsymbol{c}$ matrix evaluated at the steady-state). One proposed alternative in this paper is to relate the *stability strength* to how much is the smallest absolute eigenvalue $|\lambda_j|$ different than D_s (it should be mentioned that in a VVWC model formulation, the minimum of $|\lambda_j(\boldsymbol{J})|$ is equal to D_s at all times due to the imposed isotonic conditions).[56] Various other alternatives, such as the stability region,[1] or the smallest absolute eigenvalues of the monodromy matrix[54] might be used instead.

– relative *sensitivity coefficients* of species steady-state concentrations (vector $\boldsymbol{c}$) vs. stationary perturbations of external inducers/nutrients (vector ***Nut***), numerically evaluated by performing the cell model differentiation:[18]

$$\left[\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{c}}\right]_s \left[\frac{\partial \boldsymbol{c}}{\partial \boldsymbol{c}_{\boldsymbol{Nut}}}\right]_s + \left[\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{c}_{\boldsymbol{Nut}}}\right]_s = \boldsymbol{0}\,. \qquad (1)$$

– *stationary responsiveness* of homeostatic species levels to stationary environmental perturbations, expressed by the small transient times τ_j necessary for a reference species *j* stationary-level (e.g. the GS target proteins P2,P3) to reach a new steady-state (with a tolerance of 1 %[18]) after applying a stationary ("step-like") perturbation in the external stimulus. In some reported models, explicit thresholds are imposed on the expressed enzymes, which will eventually determine activation or repression limits for the GS gene expression.[46,57]

– *dynamic responsiveness* (efficiency) of homeostatic species levels vs. dynamic perturbations, expressed by the small recovering times $\tau_{rec,j}$ necessary for a reference species *j* (e.g. the GS target proteins P2,P3) to return to the initial steady-state (with a tolerance of 1 %[18]) after applying a dynamic ("impulse-like") perturbation in a cytosolic or external species.

– *overall responsiveness* of the modular GRC approximated by the average of the transient times $\mathrm{AVG}(\tau_j)$ of all cell species after applying a stationary ("step-like") perturbation in the external stimulus. Moreover, a global measure of the species *connectivity* (synchronisation) during the transition can be done by the standard deviation of species transient times τ_j.[18]

Comparing several ODE models for a GS design in *E. coli*

The two GERMs in Fig. 2 are placed in an *E. coli* cell of nominal characteristics presented in Table 3 (K-12 strain,[35] the lumped genome/proteome = $[P1]_s$ / $[G1]_s$ = 10^7 nmol L^{-1} / 4500 nmol L^{-1} being known) in order to exemplify the GS properties predicted by models WC-B0 and WC-Bn under a VVWC framework. The accounted high levels of P1, G1, MetGj, MetPj will mimic the cell 'ballast' effect when small perturbations occur. A generic GS including proteins {P2,P3} of initial levels given in Table 3 (in the absence of GS external inducers) has been formulated for an easy comparative analysis. Other intermediate species, such as TFs and inducers, are taken at different levels to study their influence on the GS efficiency.

The rate constants of every tested GS-model have been identified by solving the stationary model equations (Table 1) with substituted nominal concentrations of species from Table 3 (locally stable system), and for average $[TF]_s$ = 5 nmol L^{-1} (in the inducer absence). As proved by several authors, the dissociation constants of the buffering reactions used in the cross-repression for adjusting the gene activity, of type Gi + PjPj ⇔ GiPjPj ⇔ … ⇔ $Gi(PjPj)_n$, must be adopted at a much higher value than the dilution rate ($10^7 D_s$ here, see Table 2 footnote-a). Thus, the control reactions of G + P ⇔ GP type are kept near equilibrium, because $K = k_{bind}/k_{diss} = [GP](1+D/k_{diss})/[G][P]$.[18] The n_{buff} = 1–4 reversible buffer reactions employed by models {WC-B1, WC-B2, WC-B3, WC-B4} use dimeric TF = {P2P2,P3P3}. Equal concentrations of catalytically active/inactive forms $[Gj]_s = [GjTF_n]_s$ are adopted at steady-state to ensure maximum regulatory efficiency vs. perturbations.[53,54] The Hill constants in {G2,G3} and {P2,P3} syntheses (reactions 11,12,15,16 of Table 2), are adopted at values which follow the average concentration of inducers {I2,I3} and target proteins {P2,P3}, which means B = 2 (similar to Voit[46]), and $K_{G2} = [I2_{ref}]^4$ = 1 (nmol L^{-1})^4,

Table 3 – *E. coli cell characteristics and nominal stationary concentrations of key-species considered in the GS model (inner cell concentrations are evaluated with formula in Table 1)*

Variable	Value	Observations
– initial cell volume (cytoplasma, $V_{cyt,o}$)	$1.66 \cdot 10^{-15}$ L	Volkmer & Heinemann[59]
– cell cycle time (t_c)	100 min	Trun & Gottesman[60]
– lumped nutrients used for genome synthesis, $[NutG]_s$	$3 \cdot 10^7$ nmol L^{-1}	referring to the environmental volume[34,56]
– lumped nutrients used for proteome synthesis, $[NutP]_s$	$3 \cdot 10^8$ nmol L^{-1}	referring to the environmental volume[34,56]
– lumped metabolites used for proteome synthesis, $[MetP1]_s$	$\sim 3 \cdot 10^8$ nmol L^{-1}	Morgan *et al.*[56]; Maria[34]
– lumped metabolites used for target {P2,P3} synthesis, $[MetP2]_s$, $[MetP3]_s$	10^6 nmol L^{-1}	adopted
– lumped metabolites used for genome synthesis, $[MetG1]_s$	$\sim 2 \cdot 10^7$ nmol L^{-1}	footnote (a)
– lumped metabolites used for target {G2,G3} synthesis, $[MetG2]_s$, $[MetG3]_s$	10^4 nmol L^{-1}	adopted (by keeping comparable ratios vs. MetP1/MetG1)
– lumped genome (active part), $[G1]_s$	4500/2 nmol L^{-1}	footnotes (b,c)
– lumped proteome, $[P1]_s$	$1 \cdot 10^7$ nmol L^{-1}	footnote (b)
– target proteins $[P2]_s$, $[P3]_s$	5 nmol L^{-1}	adopted
– gene $[G2]_{tot}$ or $[G3]_{tot}$ expressing the P2 or P3 proteins, respectively	1 nmol L^{-1}	footnote (c)
– inactive catalytic forms of target genes, $[Gi(PjPj)_n]_s$	$[Gi]_{tot}/(n+1)$ nmol L^{-1}	footnote (c)
– environmental species $[NutI2]_s$, $[NutI3]_s$ inducing target genes G2, G3 expression	0–10 nmol L^{-1}	referring to the environmental volume
– cytosolic species $[I2]_s$, $[I3]_s$ inducing target genes G2, G3 expression	0–10 nmol L^{-1}	adopted

(a) calculated from the state-law constraint for an isotonic and isothermal cell system:
$$\sum_j^{all} c_{j,cyt} = \sum_j^{all} c_{j,env} \Rightarrow \sum_j^{all} c_{MetGj,cyt} = \sum_j^{all} c_{j,env} - \sum_j^{all} c_{MetPj,cyt} - \sum_j^{all} c_{Gj,cyt} - \sum_j^{all} c_{Pj,cyt} .$$

(b) The considered K-12 strain of E. coli genome includes ca. 4500 genes, 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.[35]

(c) Maximum regulatory expression effectiveness takes place for active and inactive equal G-forms at steady-state, i.e. $[Gj]_s = [GjTFj]_s$, where TF denotes the transcription factor adjusting the gene activity.[18] For instance, in a system including G2 inactivation through two (n_{buff} = 2) successive buffering reactions $G2 \overset{+P3P3}{\rightleftarrows} G2(P3P3) \overset{+P3P3}{\rightleftarrows} G2(P3P3)_2$, the active G2 concentration results from the steady-state equality: $[G2]_s = [G2(PjPj)]_s = [G2(P3P3)_2]_s = [G2]_{tot}/3$.

$K_{G3} = [I3_{ref}]^4 = 1$ (nmol L^{-1})^4, $K_{P2} = [P3_{ref}]^{nH} = 10^{nH}$ (nmol L^{-1})^{nH}, $K_{P3} = [P2_{ref}]^{nH} = 10^{nH}$ (nmol L^{-1})^{nH}. These constants can be easily ranged for every GS case according to the average {P2,P3} and {I2,I3} level of "sensitivity". The estimated rate constants for the nominal cell conditions are presented in Table 2 for models WC-B0, WC-B1 (n_{buff} = 1), and WC-B4 (n_{buff} = 4).

A typical simulation of the individual/lumped GS species response in *E. coli* is plotted in Fig. 3 after a "step" perturbation in the environmental stimulus NutI2 from 0 to 1 nmol L^{-1}. Predictions are generated by simulation over tenths of cell cycles by using the model WC-B4 with four cross-repressing reactions of {G2,G3} expression, for nR = 3 self-repression exponent ([TF] = 5 nmol L^{-1}). As expected, species present in large amounts (of order 10^5–10^8 nmol L^{-1}) display a negligible response to the NutI2 small perturbation (of 1 nmol L^{-1}), while the cellular species directly connected to the NutI2/I2 inducer pathway are very strongly affected. Even if the plots in Fig. 3 are represented for a large time-scale (thousands of minutes), the species transient times for recovering their homeostasis are from the order of minutes up to few cell-cycles (hundreds of minutes, see Table 4, last column; see also the Elowitz & Leibler[27] discussion on transmission of the effect of certain perturbations from generation to generation as an expression of cell adaptation to the environment).

Various values have been checked for the Hill cross-repression exponent nH = 2–8 when employing the model WC-B0 (with only Hill-type kinetics) in order to compare the GS-model characteristics from the perspective of regulatory efficiency indices. To keep similar terms of comparison, models WC-Bn have been considered with the number of TFs per gene operon corresponding to $n_{buff} = nH/2$ successive repression reactions. For every studied case, the self-repression exponent nR, and [TF] level have been varied, by computing every time the GS regulatory indices after applying a stationary

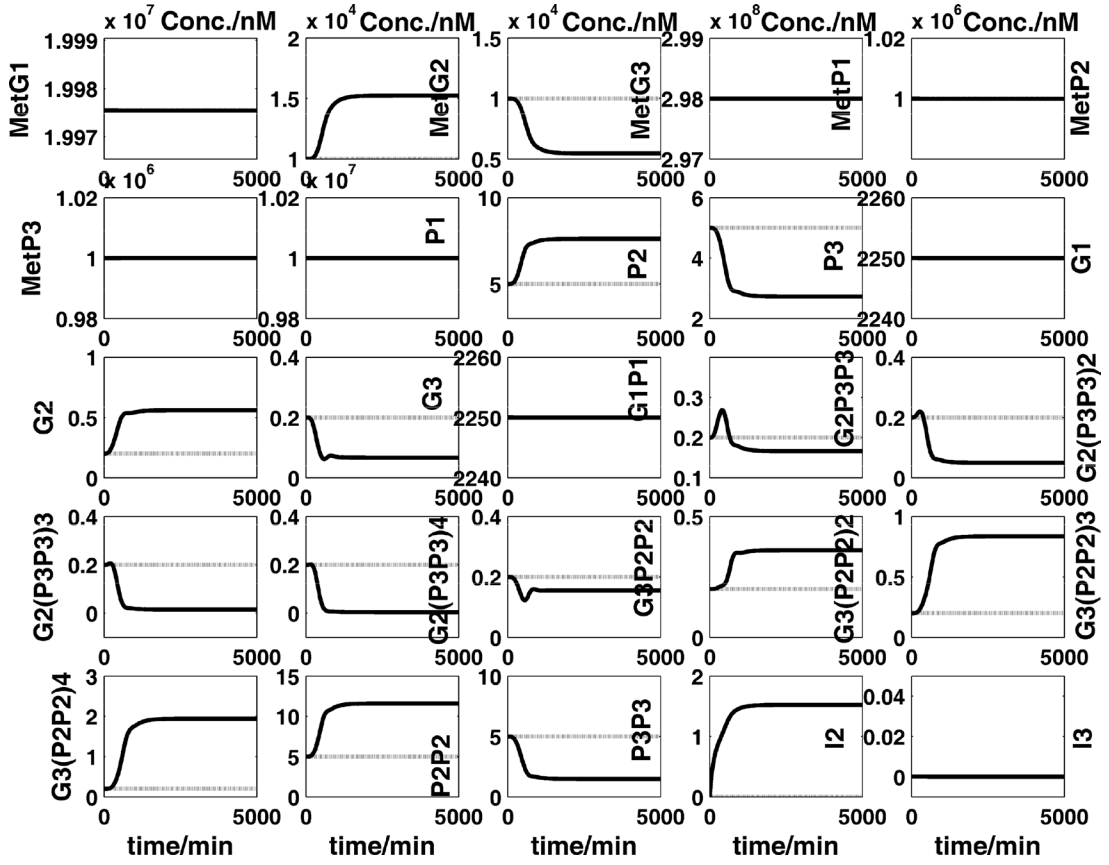

Fig. 3 – *Individual or lumped cellular species dynamics after a "step"-like stationary perturbation in the external stimulus [NutI2] from 0 to 1 nmol L^{-1} (applied at time t = 0). Predictions generated by model WC-B4 under nominal conditions in Table 2 (TF level of* $[P2P2]_s = [P3P3]_s$ *= 5 nmol L^{-1};* $n_{buff} = 4$; *nR = 3).*

perturbation in external stimulus NutI2, or a dynamic perturbation in the target protein P2 (with the initial cell condition in Table 3). The results, summarized in Table 4 (for all models), Fig. 4 (model WC-B0), and Fig. 5 (model WC-B1), lead to the following conclusions:

i) The *SC* increases with the decrease of self-repression strength (*nR*), with the inducer level {NutI2, NutI3} up to a certain 'saturation' (depending on the imposed *B,* K_{G2}, K_{G3} constants), and slightly with the initial target protein {P2,P3} level. For the WC-B0 model, the switch certainty SC increases very sharply with the induction exponent *nH* for weak (*nR* = 0.5) self-repression (from SC = 373 for *nH* = 2 to SC = 2031 for *nH* = 4), but presents much lower values for a stronger self-repression (*nR* = 1) in the two GERMs (SC ≈ 4, Fig. 4 – top row). A similar SC behaviour is reported for WC-Bn models (Table 4), but with a much more moderate variation (from SC = 18 for *nR* = 0.5 and n_{buff} = 1, until SC ≈ 3 for *nR* = 3 and n_{buff} = 4; Table 4). In contrast, WC-B0 model predictions seem to be too optimistic as long as the gene 'catalytic' activity moderation due to the successive buffering reactions synchronization is neglected.

ii) GS-steady-state local stability (i.e. all $\mathrm{Re}(\lambda_j) < 0$) is a sensitive issue when comparing different models (Table 4). While stable steady-states are predicted by WC-B0 for Hill coefficients nH ≥ 2 (as previously reviewed), for the WC-Bn model cases the GS stability is closely related to the number of repression steps (i.e. n_{buff} and nR parameters). While the WC-B1 model (with n_{buff} = 1 cross-repression reaction) predicts two stable steady-states for nR ≥ 0.5, the model WC-B2 (with n_{buff} = 2) offers similar predictions for nR ≥ 2, while WC-B4 model (n_{buff} = 4) for nR ≥ 3 (with a reduced stability strength). It appears that the increase in the complexity of the repression mechanism, with involving more synchronized reactions and intermediates, "works" against the system stability.

iii) The *GS sensitivity* to external stimuli, expressed here by the relative sensitivity $S([P2];[NutI2])$ at nominal state (Table 4), indicates the large WC-Bn constructions as being less sensitive to external inducers than the simple WC-B0 formulations (from 1.5 to 5 times, depending on the *nR* and TF level; Figs. 4–5).

iv) The *stationary responsiveness* to external stimuli {NutI2,NutI3}, quantified here by the

Table 4 – *Genetic-switch performances designed in an E. coli cell, predicted by models WC-B0 (Hill-type cross-repression), and WC-Bn (P2,P3 synthesis cross-repression by TFs over n_{buff} buffering reactions). Notations: TF = {P2P2, P3P3}; AVG = average; NG = negligible (less than 1 min); nH = Hill cross-repression exponent of P2,P3 synthesis; nR = self-repression exponent of G2,G3 synthesis rate; n_{buff} = no. of successive repression reactions used in WC-Bn models for P2,P3 synthesis; Index 's' = stationary-state.*

	Model (cross-repression parameter)	*nR* (self-repression parameter)	[TF]	Switch certainty[(a)] $[P2]_s/[P3]_s$	Stability index[(c)] $\lvert\lambda_{min}\rvert \cdot 10^3$	Sensitivity index[(c)] $\lvert S([P2];[NutI2])\rvert \cdot 10^{13}$	Response to perturbations	
							stationary $[NutI2]_s$	dynamic ± 0.1$[P2]_s$
							transient τ_{P3}/AVG (min/min)[(d)]	recovering $\tau_{rec,P2}$/AVG (min/min)[(e)]
two TFs per gene operator site	WC-B0 ($nH = 2$)	0.5	–	373.7	0.26	2.7	4918 / 2195	1122 / 247
		1	–	4.21	1.56	1.7	3738 / 1595	258 / 96
	WC-B1[(f)] ($n_{buff} = 1$)	0.5	5	18.1	8.23	0.64	4650 / 2758	43 / 15
		1	0.5	13.8	0.41	0.14	3808 / 2206	393 / 397
		1	5	3.31	1.33	0.31	3676 / 2206	58 / 18
		1	10	2.57	1.53	0.35	3448 / 2128	NG / 4
		2	5	1.43	3.32	0.11	1449 / 892	46 / 20
four TFs per gene operator site	WC-B0 ($nH = 4$)	0.5	–	2031	9.25	3.35	4938 / 2169	998 / 235
		1	–	4.12	1.85	2.00	3685 / 1502	225 / 77
		2	–	1.41	3.65	1.10	1087 / 610	125 / 50
	WC-B2[(f)] ($n_{buff} = 2$)	2	0.5	3.85	1.63	0.41	2122 / 1389	245 / 230
		2	5	2.56	2.12	0.31	2274 / 1529	55 / 56
		2	10	2.28	2.24	0.29	2284 / 1550	NG / 22
		3	5	1.39	4.13	0.49	1000 / 762	41 / 42
eight TFs per gene operator site	WC-B0 ($nH = 8$)	1	–	3.45	2.15	2.4	3793 / 1500	211 / 74
		3	–	1.18	4.69	0.81	NC / 22	130 / 42
	WC-B4[(f)] ($n_{buff} = 4$)	3	0.5	2.52	1.78	0.55	1300 / 928	428 / 364
		3	5	2.79	1.07	0.61	1359 / 959	410 / 519
		3	10	2.92	0.72	0.64	1382 / 977	NG / 436
		4	5	1.58	4.54	0.68	576 / 630	50 / 177

(a) Switch certainty index expressed as $[P2]_s/[P3]_s$ ratio in the presence of an exogenous inducer of $[NutI2]_s$ = 1 nmol L^{-1}.

(b) Stability strength expressed as the minimum of $\left|\lambda_j(\boldsymbol{J})\right|$ different from D_s (for a stable cell in a VVWC model formulation, the minimum of $\left|\lambda_j(\boldsymbol{J})\right|$ is equal all the time to D_s.[56]

(c) Relative sensitivity of P2 level vs. the environmental level of its synthesis inducer NutI2 (under the same nominal conditions).

(d) P3 species transient time (τ_{P3}), and average AVG transient times of all cell species to reach the steady-state (with a tolerance of 1 %[18]) after a "step-like" perturbation in $[NutI2]_s$ from 0 to 1 nmol L^{-1}.

(e) P2 species recovering time ($\tau_{rec,P2}$), and average AVG recovering times of all cell species to reach the steady-state (with a tolerance of 1 %[18]) after an "impulse-like" perturbation of ± 10 $\%[P2]_s$.

(f) WC-B1 stable for $nR > 0.5$, WC-B2 stable for $nR \geq 2$, WC-B4 stable for $nR \geq 3$.

{P3,P3} rise-time τ_j, and the cell average $\mathrm{AVG}(\tau_j)$ to reach the new steady-state after a stationary perturbation in stimuli {NutI2,NutI3}, indicates comparable high values for all models, the perturbation effect being transmitted over 10–20 cell generations (similar to experimental observations of $(5 \div 22)$ reported by Voit[46]). As underlined by Elowitz & Leibler,[27] the transition times in the inductive/co-repressed GS can be much higher than the cell-division cycle, the state of the transient switch or amplifier being transmitted from generation to generation of cells. A sharp reduction in the transient time can be given by the self-repression level, larger $nR \geq 3$ values reducing the transition period and increasing the stability index (to the detriment of SC).

v) The *dynamic responsiveness efficiency* of WC-Bn models is incomparably better (tenths of times) than those predicted by the Hill-WC-B0 models, the small (± 10 %) "impulse"-perturba-

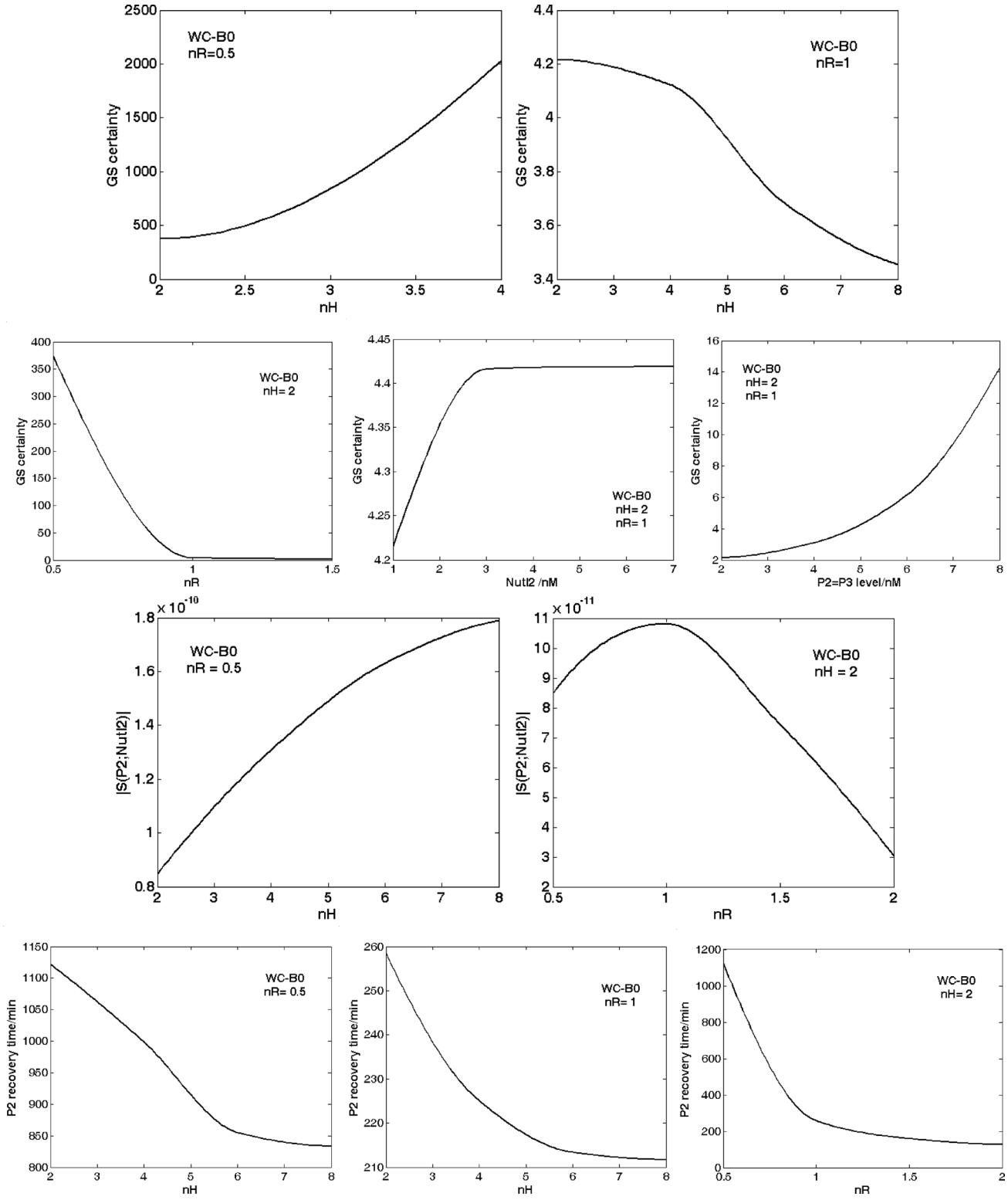

Fig. 4 – *Model WC-B0 predictions of genetic-switch certainty (1st and 2nd rows), sensitivity index* $S([P2];[Nutl2])$ *(3rd row), and P2 recovery time (4th row) as a function of model parameters nH, nR,* $[Nutl2]_s$*, initial* $[P2]_s = [P3]_s$ *level (nominal* $[P2]_s = [P3]_s = 5$ *nmol* L^{-1}*).*

tions in the key species being "extinguished" over a few minutes during the cell cycle (e.g. $AVG(\tau_{rec,j})$ = 4–18 minutes for WC-B1 with [TF] = 5–10 nmol L^{-1}, $nR = 1$ in Table 4). The dynamic efficiency increases with the TF-level and stability index (to the detriment of SC index), and decreases with the strength/complexity of the repression control (i.e. with the increase of nH, nR, n_{buff} parameters) (Fig. 4–5). This GS efficiency index better reflects the superiority of more elaborated GS models of

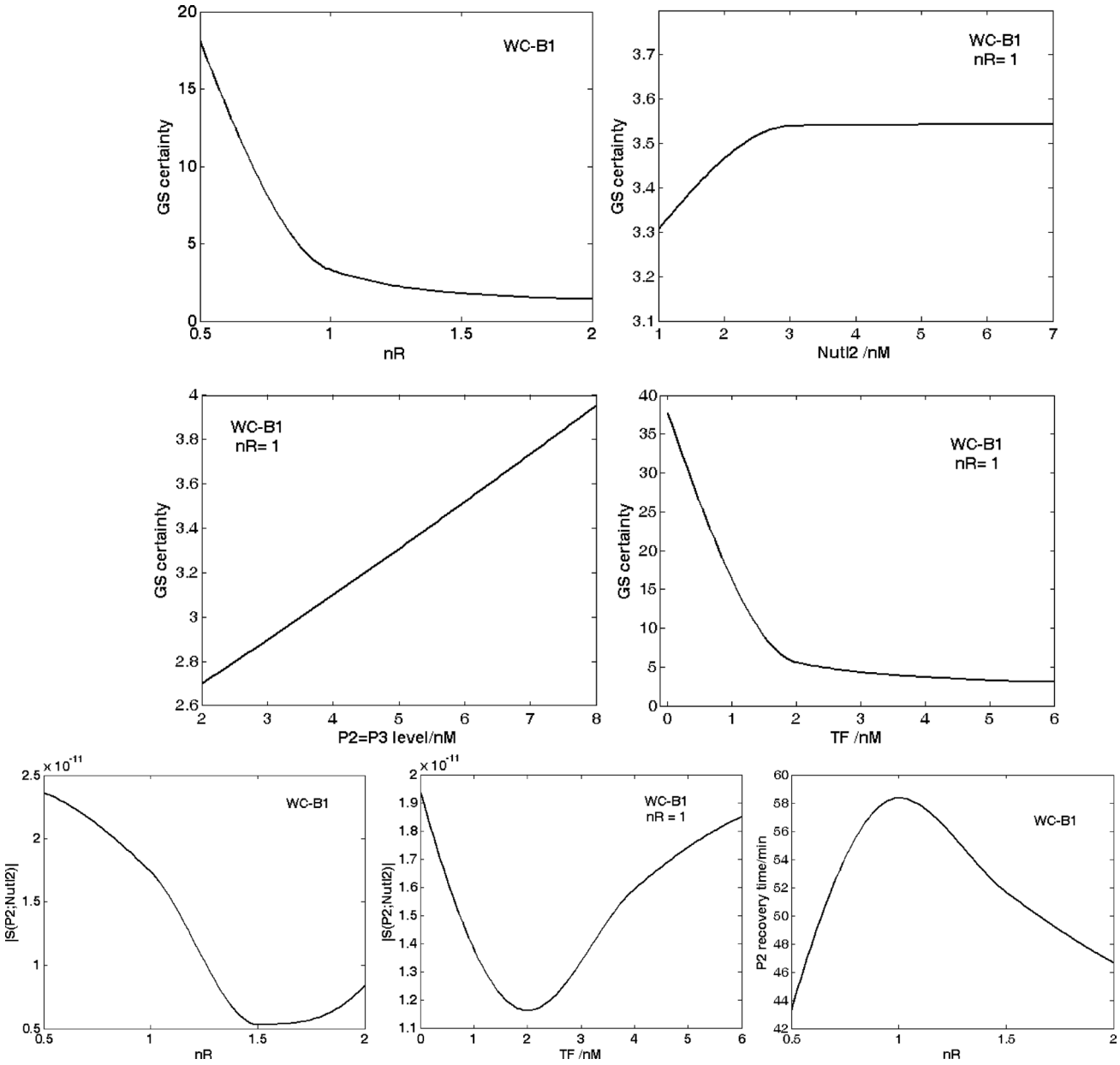

Fig. 5 – *Model WC-B1 predictions of genetic-switch certainty (1st and 2nd rows), sensitivity index* $S([P2];[NutI2])$ *, and P2 recovery time (3rd row) as a function of model parameters nR,* $[NutI2]_s$*, initial* $[P2]_s = [P3]_s$*, and* $[TF]_s$ *level (nominal* $[P2]_s = [P3]_s = 5$ *nmol* L^{-1}*, and* $[TF]_s = 5$ *nmol* L^{-1}*).*

WC-Bn type that explicitly include intermediate steps of the key regulatory loops.

vi) The *GS properties* depend on the employed model structure and parameters. For solving practical GS applications, the most flexible model here seems to be the WC-B0 with $nH = 2$ (lumped Hill-type), or its corresponding model WC-B1 (with $n_{buff} = 1$ explicit cross-repression step), being easily adaptable to a certain case study (of known P2,P3,G2,G3 and inducer levels) by means of several tuning parameters (nH, nR, n_{buff}, B, K_{G2}, K_{G3}, K_{P2}, K_{P3}, $[TF]_o$), using 'wild'- or cloned cells with known {G2,G3} plasmid levels. The WC-Bn models seem to be more flexible, by replacing the adjustment of nH with the adjustment of two parameters $\{n_{buff}, [TF]_o\}$.

vii) The *estimated rate constants* of the buffering reactions of {G2,G3} activity in WC-Bn models (Table 2) are comparable to those reported in the literature. For instance, the repressor monomer dimerization/dissociation, as well as the repressor binding to operator/repressor–operator complex dissociation constants of ca. (10^2 L nmol^{-1} min^{-1} / 1–10^3 min^{-1})[36] are comparable to (10^3 L nmol^{-1} min^{-1} /10^5 min^{-1}) in WC-B1 model, the differences coming from the adopted large dissociation constant (10^5 min^{-1}) for all reactions involving TFs. The apparent rate constants in WC-Bn reactions that in-

clude large lumps (i.e. P1, MetP1, MetP2, MetP3, MetG1, MetG2, MetG3) are obviously much smaller than those reported in literature. However, these constants become comparable when including the lumped cell 'ballast' contribution. For instance, the mRNA (genes G2,G3) synthesis rate constants of ca. 10^{-6} L nmol^{-1} min^{-1} (including [P1]) is comparable to the reported 10^{-4} L nmol^{-1} min^{-1} values,[58] the differences coming from the considered large lumped {MetG2,MetG3} precursors (of 10^4 nmol L^{-1} levels). However, such adopted high levels of metabolites do not reduce the generality of the analysis, trying to highlight the influence of the "rest" of the cell on the GS behaviour.

As a general observation, the GS efficiency depends not only on the model type, but also on the levels of external nutrients and inducers (NutG, NutP, NutI2, NutI3), the cell cycle period, and the levels of key-intermediates involved in the reaction pathway (TF represented by P2P2 and P3P3 here). Even if GS modelling analysis is kept at a generic level, comparison of hybrid WC-Bn models with Hill-type reduced models (of GERMs similar to those of Voit[46]) under the VVWC framework is relevant enough to promote such constructions as a worthy alternative to the classical representations.

The explicit models of some allosteric regulatory steps (like cross-repression in WC-Bn models) seems to be a promising alternative for the better reproduction of GS dynamic responsiveness and stability properties, and better reflection of the diminished SC due to the cell 'ballast' effect in smoothing all perturbations, but especially those of the existing species in small amounts. Moreover, the reduced Hill-type models (e.g. WC-B0) seem to offer too 'optimistic' predictions, such as very high SC or large regions of stability independent of the *nR* exponent or TF level (Table 4). In contrast, the WC-Bn models appear to be more suitable for pointing out complex relationships between GS bi-stability strength and the negative regulatory loops of self- (via *nR* exponent) and cross-repression (via n_{buff}, TF level), implying more intermediate species.

Conclusions

As underlined in literature and pointed out by the present numerical analysis, the detailed GS models that include a complex repression mechanism with lumped terms (reactions, species, rate expressions) but also explicit intermediate control steps (concerning induction, cross- and self- repression) can offer a more flexible and mechanistic-based GS model structure of higher tunability (due to the larger number of parameters), even if the required estimation/tuning effort increases considerably. An acceptable trade-off between the model simplicity, its estimability vs. available information (by including a reasonable number of parameters of physical significance), computing tractability, and predictive quality should be realized every time when simulating the dynamic properties of the complex genetic regulatory networks.

A combination of Hill-type activation steps completed with rapid buffering reactions (using dimeric TFs) adjusting the genes activity, and a modular GERM construction (like WC-Bn models here) seem to be promising for obtaining increased accuracy in the evaluated regulatory GS indices compared to the lumped power-law/Hill-type kinetic models. The local/holistic GS regulatory properties can be easily ranged according to available dynamic data by tuning the model parameters and TF-levels controlling the repressing enzymes, for known average levels of key-species. For instance, our study demonstrates how an increase in the complexity of the repression mechanism, with involving more synchronized reactions and intermediates, can negatively affect the GS system state stability.

Moreover, the modular approach is proved to be more appropriate to study GRCs, by offering the advantage of an expandable simulation framework in accordance with the modelling purposes. The local but also the holistic GRC regulatory properties can be studied in such a manner, allowing a more realistic characterisation of the system (in terms of stability, flexibility, multiplicity, efficiency, robustness).

The VVWC modelling framework employed in this study can be viewed as another level of complexity used in developing dynamic cell models, when the GRC dynamic properties should be evaluated from another perspective. In particular, by placing the two gene expression regulatory modules of a GS in a simulated cell of known characteristics, the GS properties can be studied by mimicking stationary or perturbed environmental conditions, by accounting for direct but also indirect GS interactions with the genome and proteome transmitted via the common cell-volume under isotonic osmolarity conditions.

Being formulated at a generic level, such reasonable simple GS models can be included in larger metabolic network models by simply tuning the model parameters, being able to simulate some observable effects, such as: i) quick and efficient cross- and self-control of the gene expressions by means of adjustable dimeric TFs; ii) gene expression amplification at low levels of exo/endogenous inducers; iii) quantitative characterization of the GS response to various perturbations; iv) 'inertial'/smoothing cell content effect in treating perturbations under isotonic conditions; v) GS synchronization in adjusting the target gene expression. Desir-

able properties of a GS, such as SC, stability strength, stationary and dynamic efficiency, switch responsiveness, robustness and selectivity to external stimuli, can be checked quite easily by adjusting the GS model parameters, leading to the use of GS models for cell design purposes.

Nomenclature

B – rate constant in Hill expressions
c_j – species (individual, lump, or 'pool') concentration
D – cell content dilution rate
$\boldsymbol{g}$ – kinetic model function vector
$\boldsymbol{J} = d\boldsymbol{g} / d\boldsymbol{c}$ – dynamic model Jacobian matrix
k, K – kinetic constants
n, nH, nR– Hill-kinetics exponents
n_{buff} – number of repressing reversible reactions binding dimeric TFs to gene operator
n_j – species j number of moles
n_s – number of species
N_A – Avogadro number
r_j – species j reaction rate
R – universal gas constant
$\tilde{S}(y;x) = \partial \ln(y) / \partial \ln(x)$ – relative sensitivity of y vs. x
t – time
t_c – cell-cycle time
T – temperature
V – cell volume (cytoplasma)

Greeks

$\lambda(\boldsymbol{J})$ – eigenvalues of the Jacobian matrix
π – osmotic pressure
$\tau_{rec,j}$ – species j recovering time of the steady-state
τ_j – species j transition time from one steady-state to another

Index

cyt – cytoplasm
diss – dissociation
env – environment
o – initial
ref – reference (nominal)
s – steady-state
tot – total

Abbreviations

A – activator
AVG () – average of ()
G – gene
GERM – gene expression regulatory module
GRC – genetic regulatory circuit
GS – genetic switch
I – inducer
M – mRNA
Met – metabolite
Nut – nutrient
ODE – ordinary differential equations
P – protein
R – repressor
SC – switch certainty
TF – transcription factor
VVWC – variable-volume whole-cell
WC – whole-cell

References

1. *Hlavacek, W. S., Savageau, M. A.,* J. Mol. Biol. **266** (1997) 538.
2. *Atkinson, M. R., Savageau, M. A., Myers, J. T., Ninfa, A. J.,* Cell **113** (2003) 597.
3. *Kaznessis, Y. N.,* Chem. Eng. Sci. **61** (2006) 940.
4. *Alon, U.,* An introduction to system biology. Design principles of biological circuits, Chapman & Hall / CRC, Boca Raton, 2007.
5. *Maria, G.,* Asia-Pac. J. Chem. Eng. **4** (2009) 916.
6. *Kholodenko, B. N., Kiyatkin, A., Bruggeman, F. J., Sontag, E., Westerhoff, H. V., Hoek, J. B.,* Proceedings of the National Academy of Sciences of USA **99** (2002) 12841.
7. *Hlavacek, W. S., Savageau, M. A.,* J. Mol. Biol. **255** (1996) 121.
8. *Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T. S., Cantor, C. R., Collins, J. J.,* Proc. Natl. Acad. Sci. U. S.A. **101** (2004) 8414.
9. *Hasty, J., McMillen, D., Collins, J. J.,* Nature **420** (2002) 224.
10. *Wall, M. E., Hlavacek, W. S., Savageau, M. A.,* J. Mol. Biol. **332** (2003) 861.
11. *Benner, S. A., Sismour, A. M.,* Nat. Rev. Genet. **6** (2005) 533.
12. *Heinemann, M., Panke, S.,* Bioinformatics **22** (2006) 2790.
13. *Schümperli, M., Pellaux, R., Panke, S., Knoll, M., Pleiss, J., Calles, B., de Lorenzo, V., Davidescu, F. P., Jørgensen S. B.,* A modular platform for biosynthesis of complex molecules, Synthetic Biology 3.0 Conference, Zürich (CH), 24–27 June, 2007.
14. *He, F., Balling, R., Zeng, A. P.,* Biotechnology **144** (2009) 190.
15. *Polynikis, A., Hogan, S. J., di Bernardo, M.,* J. Theor. Biol. **261** (2009) 511.
16. *Endy, D., Brent, R.,* Nature **409** (2001) 391.
17. *De Jong, H.,* J. Comput. Biol. **9** (2002) 67.
18. *Maria, G.,* Chem. Biochem. Eng. Q. **19** (2005) 213.
19. *Roeva, O., Pencheva, T., Tzonkov, S., Arndt, M., Hitzmann, B., Kleist, S., Miksch, G., Friehs, K.,* J. Biotechnol. **10** (2007) 592.
20. *Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., Herwig, R.,* Systems biology. A textbook, Wiley-VCH, Weinheim, 2009.
21. *Chen, Z., Wilmanns, M., Zeng, A. P.,* Trends Biotechnol. **28** (2010) 534.

22. *Buchholz, K., Hempel, D. C.,* Eng. Life Sci. **6** (2006) 437.
23. *Hasty, J., Isaacs, F., Dolnik, M., McMillen, D., Collins, J. J.,* Chaos **11** (2001) 207.
24. *Styczynski, M. P., Stephanopoulos, G.,* Comput. Chem. Eng. **29** (2005) 519.
25. *Jacob, F., Monod, J.,* J. Mol. Biol. **3** (1961) 318.
26. *Tyson, J. J., Albert, R., Goldbeter, A., Ruoff, P., Sible, J.,* J. R. Soc. Interface **5** (2008) S1.
27. *Elowitz, M. B., Leibler, S.,* Nature **403** (2000) 335.
28. *Hua, S. S., Markovitz, A.,* J. Bacteriol. **110** (1972) 1089.
29. *Ponnambalam, S., Spassky, A., Busby, S.,* FEBS Letters **219** (1987) 189.
30. *Tabaka, M.,* Regulation of trp operon in E. coli, PhD Thesis, Institute of Physical Chemistry of the Polish Academy of Sciences, Warsaw, 2009.
31. *Zak, D. E., Vadigepalli, R., Gonye, G. E., Doyle III, F. J., Schwaber, J. S., Ogunnaike, B. A.,* Comput. Chem. Eng. **29** (2005) 547.
32. *Maria, G.,* Chem. Biochem. Eng. Q. **20** (2006) 353.
33. *Smolen, P., Baxter, D. A., Byrne, J. H.,* Bull. Math. Biol. **62** (2000) 247.
34. *Maria, G.,* Chem. Biochem. Eng. Q. **21** (2007) 417.
35. *EcoCyc*, Encyclopedia of Escherichia coli K-12 genes and metabolism, SRI Intl., The Institute for Genomic Research, Univ. of California at San Diego, 2005. (http://ecocyc.org/).
36. *Salis, H., Kaznessis, Y.,* Comput. Chem. Eng. **29** (2005) 577.
37. *Dondo, R., Marques, D.,* Lat. Am. Appl. Res. **32** (2002) 195.
38. *Roubos, J. A.,* Bioprocess modelling and optimization – Fed-batch clavulanic acid production by Steptomyces clavuligerus, PhD Thesis, TU Delft, 2002.
39. *Laurent, M., Charvin, G., Guespin-Michel, J.,* Cell. Mol. Biol. **51** (2005) 583.
40. *Cherry, J. L., Adler, F. R.,* J. Theor. Biol. **203** (2000) 117.
41. *Kaern, M., Blake, W. J., Collins, J. J.,* Annu. Rev. Biomed. Eng. **5** (2003) 179.
42. *Gonze, D.,* BioSystems **99** (2010) 60.
43. *Goodwin, B. C.,* Temporal organization in cells, Academic Press, New York, 1963.
44. *Tyson, J. J., Othmer, H. G.,* Prog. Theor. Biol. **5** (1978) 1.
45. *Del Vecchio, D.,* Design and analysis of an activator-repressor clock in E. Coli, Proc. American Control Conference, July 2007, New York. pp. 1589–1594.
46. *Voit, E. O.,* IEE Proc. – Systems Biology **152** (2005) 207.
47. *Widder, S., Schicho, J., Schuster, P.,* J. Theor. Biol. **246** (2007) 395.
48. *Savageau, M. A.,* Math. Biosci. **180** (2002) 237.
49. *Grognard, F., de Jong, H., Gouzé, J. L.,* Piecewise-linear models of genetic regulatory networks: theory and example, In: *Queinnec, I., Tarbouriech, S., Garcia, G., Niculescu, S.* (Eds.), Biology and control theory: current challenges, Lecture Notes in Control and Information Sciences **357**, 137–159, Springer-Verlag, Berlin, 2007.
50. *Casey, R., de Jong, H., Gouze, J. L.,* J. Math. Biol. **52** (2006) 27.
51. *Grainger, J. N. R., Gaffney, P. E., West, T. T.,* J. Theor. Biol. **21** (1968) 123.
52. *Kinoshita, A., Nakayama, Y., Tomita, M.,* In silico analysis of human erythrocyte using E-Cell system, 2nd Intl. Conference on Systems Biology, California Institute of Technology, Pasadena (USA), November 4–7, 2001.
53. *Yang, Q., Lindahl, P., Morgan, J.,* J. Theor. Biol. **222** (2003) 407.
54. *Maria, G.,* Chem. Biochem. Eng. Q. **17** (2003) 99.
55. *Savageau, M. A.,* Chaos **11** (2001) 142.
56. *Morgan, J. J., Surovtsev, I. V., Lindahl, P. A.,* J. Theor. Biol. **231** (2004) 581.
57. *Rosenfeld, N., Elowitz, M. B., Alon, U.,* J. Mol. Biol. **323** (2002) 785.
58. *Wang, X., Li, Y., Xu, X., Wang, Y.,* Biosystems **100** (2010) 31.
59. *Volkmer, B., Heinemann, M.,* PLoS One **6** (2011) 1.
60. *Trun, N. J., Gottesman, S.,* Gene. Dev. **4** (1990) 2036.