

A Dynamic Model to Simulate the Genetic Regulatory Circuit Controlling the Mercury Ion Uptake by *E. coli* cells

GHEORGHE MARIA*

University Politehnica of Bucharest, Laboratory of Chemical & Biochemical Reaction Engineering, P.O. 35-107, Bucharest, Romania

A structured dynamic model has been proposed to simulate the mercury ion transport and reduction in E. Coli cells, as well as the stationary and dynamic characteristics of the genetic regulatory circuit (GRC) controlling the involved mer operon expression responsible for the whole mercury uptake process. Starting from a simpler variant, the current model extension accounts for a larger number of stationary data recorded under a wide range of environmental mercury levels and using cultures of cloned cells with various levels of plasmid encoded mer operon. After screening among regulatory loops alternatives, the model key parameters are identified from data, while subsequent model refinements are focus on the flexibility in adjusting the dynamic properties by imposing optimal regulatory efficiency criteria of the whole GRC. The model is written in the variable cell volume and isotonic condition framework. A whole-cell modelling approach allows placing the studied GRC in an E. coli cell of known characteristics, thus mimicking the GRC behaviour in connection with the whole-cell content replication during the equilibrated cell growth, or under stationary or dynamic perturbations in the environment.

Keywords: mercury ion uptake, genetic regulatory circuit modelling

Metallic ion and organometallic compound uptake, utilization or neutralization in bacteria is an intense investigated subject over the last decades, due to tremendous environmental implications but also due to emergent industrial or medical applications. Very good reviews have been published on bacterial resistance to mercury, silver and other heavy metals, or on the metabolic control of zinc, copper, cobalt, iron, nickel, and manganese utilization in bacteria (FEMS Microbiology Reviews special issue on June 2003).

Among them, an interesting subject concerns the specialized genetic control developed by some bacteria to cope with toxic mercury compounds from environment. Instead of "neutralizing" the toxic ions by building complex chelate compounds, thus consuming lot of cell energy and metabolites to maintain chelate homeostasis, bacteria developed an efficient defense system by simply reducing the ions to volatile metal, less toxic and easily eliminable from the cell by membrane diffusion. The whole uptake process is closely controlled by a GRC involving the *mer* operon expression. At least seven genes are involved in this process, the ratio of their expression being controlled according to the cytosolic content of mercury. The resulted cross- and self-control of the *mer* operon expression is aiming at eventually shrinking the import of large amounts of mercury, which might lead to the blockage of cell RSH species, exhausting the cell resources, and eventually compromising the whole cell metabolism.

To kinetically model the mercury uptake process in Gram-negative bacteria, several trials have been published using unstructured models based on experimental observations (Michaelis-Menten or Haldane type; [1-4]). If such a model can suffice for scaling-up the process, for a deep understanding of the regulatory pathway and of the GRC dynamic properties, more information are necessary to build-up a structured model. The construction of a GRC model is however not an easy

task, due to the well-known high degree of freedom and low estimability of the cell processes. This is why, a systematic approach is often used, comprising a preliminary step in detecting inference among genes, the structure of the gene regulatory network, and its main regulatory loops using a large variety of methods and -omics data (the so-called 'reverse engineering' approach; reviews [5-7]). After disassembling the whole system in parts (modules) then, by performing tests and applying a suitable numerical/sensitivity analysis, one defines rules that allow recreation of the whole and its characteristics, thus reproducing by simulation the real system (the so-called 'integrative understanding'; [8-10]). One approach consists in combining lumped regulation modules of individual gene expression, separately studied. Thus, the model complexity is considerably reduced 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.

Dynamic simulation of GRC can be realized in various ways, by using binary (Boolean) variables, continuous, stochastic, or mixed type of variables (reviews [11-12]). Each approach presents advantages and disadvantages well pointed out in the literature. The continuous variable models continue to be popular in spite of predicting only the average behaviour of a cell population, and not being able to well covering stochastic processes (such as gene mutation, cell signalling, random faulty in GRC, etc.), or to predict fluctuations in different cells and multimodal population distributions. Among advantages, it is to mention the possibility to represent the metabolic kinetics under various continuous perturbations, binding thermodynamics, cell-division cycle, oscillatory processes, or molecular diffusion. Due to the large size of the identification problems and the time-course data type (ca. 10^3 - 10^4 number of states, and 10^3 gene transcription factors

* email: gmaria99m@hotmail.com

TFs), co-regulated genes are clustered together, and usually a mixed structured representation with some lumped terms (Michaelis-Menten, Haldane, or Hill type) is derived. Successful trials to also account for stochastic behavior of gene expression in deterministic models have been reported [13,14].

Concerning the mercury ion uptake in bacteria, systematic experiments have been done in the literature to study the process (review of Barkay et al. [19]), leading to determine the main reaction steps in *E. coli* and to propose cell modifications to increase the source of *mer* operon responsible for controlling the whole mercury reduction process [2-4,15]. Process apparent kinetics (unstructured, Michaelis-Menten type) has been studied on various micro-organisms, while large-scale applications are already reported, such as efficient mercury removal from industrial wastewaters [1,16,17]. Such reduced kinetics models can fairly represent the overall mercury uptake in cells and, after identification of rate constants from batch experiments, can be successfully used to scale-up the process. However, the overall models can not accurately represent the cell response to stationary or dynamic perturbations in the environment, self-regulation of the process under various conditions, *mer* gene expression amplification and the GRC characteristics in connection to the cell volume growth and inner-cell content replication. This is why a whole-cell modelling framework is to be considered, by including individual gene expression regulatory modules but also, at a certain level of generality, the genome and proteome replication.

It is to mention that available literature includes extended qualitative information on the process characteristics, the structure and role of the involved species, some regulatory features, and quantitative information on only the apparent reduction kinetics under stationary conditions, and on the homeostasis level of expression of some genes in modified cells of increased *mer* operon content. Unfortunately, no structured / quantitative information exists on the regulatory loops dynamics, time-scale windows of parallel-consecutive involved reactions, level of transcription factors, or details on the regulatory loops.

Recently, it was presented [18] a modular structured dynamic model of the mercury ion uptake in *E. Coli* cells including the controlling GRC, by fitting stationary recorded data from batch experiments [2], corresponding to an external average level of mercury of $[Hg_{env}^{2+}] = 40 \mu M$, and using cells cloned with plasmids leading to a *mer* operon level of ca. 3 nM. The scope of this paper is to extend such a model validity by performing some modifications, by accounting for a larger number of data sets and information from cell cultures cloned with various levels of *mer* operon producing plasmids, and by checking various alternatives of GRC controlling the stationary and dynamic characteristics of the whole mercury uptake process. Due to the previously mentioned reasons, the study will use a step-by-step approach on developing the regulatory loops, ranging first the stationary characteristics of the model by fitting the available data to identify the key parameters. The subsequent model refinement will be focused on adjusting the dynamic properties of the GRC model by using little available information on gene expression dynamics, and by applying a similarity analysis of the gene expression vs. already studied regulatory modules, aiming at optimizing individual but also the holistic regulatory efficiency of the GRC.

Available process information and modelling possibilities

General information available on mercury ion reduction in Gram-negative bacteria

Mercury resistance in bacteria is one of the most studied metallic-ion uptake and release process [19]. Instead of building carbon- and energy-intensive disposal devices into the cell (like metallothionein or hydroxamate chelators which maintain homeostasis of other transition metals), that might consume lots of cell resources, the bacteria developed a simpler and a more efficient defense system against toxic mercury ions/compounds, consisting in the catalytic reduction to volatile metal, less toxic and easily eliminable from the cell by simple membrane diffusion. Such a process is favored by the large content of low molecular-mass thiol redox buffers (RSH, in millimolar concentrations), able to coordinate ionic Hg^{2+} , and of reductants such as NAD(P)H, necessary to convert it to the neutral metal.

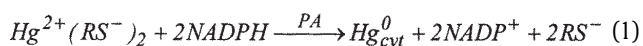
To prevent deactivation and/or blockage of cysteine-rich proteins involved in various metabolic processes, a dedicated *mer* operon has evolved into Gram-negative bacteria to develop a specific, high-affinity, genetic regulatory circuit controlling the mercury membranar and cytosolic transport as well as its reduction. The operon includes at least seven genes *GmerX* ($X = R, T, P, C, A, B, D$), of which expression leads to production of *PmerX* proteins. The *mer* operon expression happens even in the absence of the cytosolic mercury, being self-maintained by the presence of small amounts of *PmerD* protein (in the present study $G \bullet$ denotes genes, while $P \bullet$ denotes proteins). However, *GmerR* gene expression is quickly amplified by the presence of cytosolic $[Hg_{env}^{2+}]$, even in small amounts ($> 100 \text{ nM}$), inducing the quantitative production of *PmerR* mercury ion reductase in ca. 2-3 min [19]. An increase in the *PmerR* concentration will trigger the production of the other *mer* proteins in an apparent consecutive synthesis schema. If the organic mercury compounds are not present in the environment, the protein *PmerB* is not produced. The membrane transport of Hg^{2+} in Gram-negative bacteria is a quite complex process itself, requiring production of at least three enzymes, i.e. the periplasmic *PmerP*, and the inner membrane proteins *PmerT* and *PmerC* (even if two other proteins *PmerF* and *PmerE* can also be involved in the transport; [3]). As more Hg^{2+} is imported into cytosol, the accelerated expression of *GmerR* will trigger production of more transport and reductase proteins until a certain saturation level is reached, when the limited cell resources will impose flattening of the levels of *mer* proteins leading to a quasi-constant mercury reduction rate, irrespectively to the increased amount of mercury which might be present in the environment.

While the role of each *mer* gene and protein in the mercury ion reduction process is generally known, not all the regulatory loops are perfectly understood. A schematic representation of the main process steps is proposed in figure 1 (PT lump includes *PmerT*, *PmerP*, *PmerC* permesases, while *mer* genes and proteins are simply denoted by GR,GT,GA,GD and PR,PT,PA,PD respectively). The *mer* operon expression starts with the expression of *GmerR* gene (GR), encoding the protein *PmerR* (or PR) that induces the consecutive expression of the subsequent structural genes *GmerT*, *GmerP*, *GmerC* (encoding the permease lump PT), then of the gene GA (encoding the reductase PA), and of the gene GD (encoding the control protein PD). The promoter/operator regions for *GmerR* and *GmerTPCAD* are located between genes *GmerR* and *GmerT*. The whole process is induced but also inhibited by

large amounts of cytosolic Hg^{2+} (usually maintained below 3500-5000 nM), being controlled by the presence of PD (the last produced *mer* protein), and self regulated at every gene expression level. When organic R-Hg compounds are also present, expression of a gene *GmerB* is triggered before *GmerD* expression, thus being produced the *PmerB* organomercurial lyase that degrades the R-Hg complex with releasing the neutral Hg^0 and methane.

The Hg^{2+} import into the cell across the outer membrane is mediated by *PmerP*, while across the inner membrane by *PmerT* and *PmerC*, in a succession of reactions involving Hg-protein complexes ([3,4] for details). The permeation is intensively consuming the cell energy and, due to the successive coupling-decoupling steps, it is the slowest step of the whole mercury uptake process, being several times slower than the mercury reduction. Such a result has been experimentally proved [2] in separate experiments measuring the mercury reduction rate of intact *E. Coli* cells, comparatively to the mercury uptake by the permeabilized *E. Coli* cells presenting a negligible transport resistance created by means of certain membrane treatments. The whole process is inhibited by large concentrations of environmental mercury, eventually limiting its access into the cytosol and thus avoiding the blockage of thiolic compounds involved in various metabolic processes.

The key protein for mercury reduction is the reductase PA (Hg:NADP⁺ oxidoreductase E.C. 1.16.1.1) that catalyses transformation of cytosolic $[\text{Hg}_{\text{cyt}}^{2+}]$ in atomic $[\text{Hg}_{\text{cyt}}^0]$, easily removable by simple membranar diffusion with any efflux system:



The large excess of regenerable NADPH and RSH satisfies the reaction requirements. As observed [3], normal cells can not tolerate higher cytosolic mercury content, so the whole process is highly regulated by positive and negative regulatory loops at every *GmerX* gene expression level (fig. 1 an attempt of reaction pathway). While the GR expression is both induced and limited by the cytosolic mercury concentration, the expression of GR is triggering those of GT, which in turn will trigger the GA expression, and eventually the expression of GD gene. While a self-control of expression is expected by similarity with most of studied gene expression regulatory modules [11,12,34], a cascade control is managing the whole expression pathway. An additional control is exerted by the last synthesized protein PD, allowing the expression of GR in the absence of mercury ions (maintaining low levels of *PmerX* proteins into the cell), and helping PR protein in quickly repressing the GA expression when substrate $\text{Hg}_{\text{env}}^{2+}$ is exhausted (because PA reductase has also an oxidase activity which results in production of toxic H_2O_2). Consequently, the *mer* operon expression control is more complex, including not only self but also a cross-control in a cascade schema in which the involved proteins are the control nodes. As remarked [19], despite of being the longest studied bacterial toxic metal resistance loci, the *mer* operon continues to bring new insights to gene regulation and involved enzymatic metabolic processes of the bacteria, as long as significant parts of the process control mechanism remain still unknown.

Recovering experimental data collected at various mer operon levels

The available data [2,3] basically consist in the measured apparent reaction rates for mercury membrane transport and its cytosolic reduction in *E. Coli* cells, derived from batch cultures experiments carried out at environmental mercury concentrations in the range of $[\text{Hg}_{\text{env}}^{2+}] = 0-120 \mu\text{M}$. To separately determine a kinetic expression of the mercury ions transport and reduction rates, experiments have been repeated using intact *E. Coli* cells (thus fixing the membrane transport rate), but also using permeabilized cells that present a negligible transport resistance (thus fixing the mercury reduction rate). Membrane permeabilization was realized by using a brief treatment with ethyl ether on ice (see original papers for experimental details). Based on these measurements, unstructured Michaelis-Menten rate expressions have been proposed for both metabolic steps, and the rate constants have been estimated and checked under various environmental conditions. These parameters will be used in further structured model developments.

To also study the influence of the *mer* operon content on the gene expression level and whole process efficiency, the same experiment has been repeated by using cloned *E. coli* cells with different amounts of plasmids that realize a *mer* gene content amplification. The plasmids are maintained at the following copy numbers per cell: 3, 67, 78, 124, and 140. For every plasmid level, the level of some *mer* genes (*GmerT* and *GmerA*) in stationary conditions have been determined, as well as the amplification of the expression levels resulting for the encoded proteins *PmerT* and *PmerA*. One important conclusion derived is that amplification of the entire *mer* operon in the intact *E. coli* cells does not result in a considerable increase of the reduction activity of the cell. As long as the membrane transport of the mercury is the limiting step, an increase in the *mer* gene content with the same factor does not lead to the same increase of the corresponding permease and reductase levels, due to different turnover rates of the gene expression in the operon. Another conclusion concerns the realized level of *mer* genes for a continuously increased amount of plasmids as sources of the *mer* operon. It was measured that a 47-fold increase in plasmid copy number (from 3 to 140 per cell) will result in only a 10.8-fold increase in the *GmerA* gene copy number, and in only 5.4-fold increase in the *GmerT* gene level. Consequently, to account for such an effect in further modelling developments, an average damping factor β of the *mer* plasmid level was introduced, the same for all *mer* genes. The β -factor, ranging between β_{min} and β_{max} , can be evaluated for every plasmid level by using the linear interpolation formula:

$$\beta = \beta_{\text{min}} + \frac{\beta_{\text{max}} - \beta_{\text{min}}}{\text{plGmer}_{\text{max}} - \text{plGmer}_{\text{min}}} (\text{plGmer} - \text{plGmer}_{\text{min}});$$

$$\frac{(\sum [\text{GmerX}])}{(\sum [\text{GmerX}])_{\text{min}}} = \frac{[\text{plGmer}]}{[\text{plGmer}]_{\text{min}}} / \beta, \quad (2)$$

where *plGmer* denoted the level of the plasmid, while 'min' or 'max' indices denotes minimum and respective maximum levels used in the experiments; the sum is including all active and inactive forms of *GmerX* gene, X = R, T, P, C, A, B, D). As an example, by using the values of $\text{plGmer}_{\text{min}} = 3$ and $\text{plGmer}_{\text{max}} = 140$, and adopting $\beta_{\text{min}} = 1.5$ and $\beta_{\text{max}} = 3$, a value of $\beta = 1.8$ results for a

plasmid level of pGmer = 30. That means that a step increase of 30/3 times in pGmer plasmid will result in a step increase in the GmerX of only 10/1.8 = 5.5 times. In fact, due to optimal regulatory efficiency requirements imposed to the gene expression (to be further detailed), the active concentration of the genes is only half of thus computed gene copy numbers. The parameters β_{\min} and β_{\max} remain to be fitted together with other model parameters when adjusting model predictions to match the available experimental data sets.

Concerning the GmerX genes, there are no systematic data on the expression level dynamics. It was however observed that, at low levels of environmental mercury of $[\text{Hg}_{\text{env}}^{2+}] = 0.1\text{-}0.2 \mu\text{M}$, the expression of GR gene occurs quantitatively in few minutes after stimulation [19]. Such an observation and a similarity analysis with information from literature will be used to range the dynamic characteristics of the involved GRC. At the same time [2] offered important information on the stationary levels of some GmerX expression. It was recorded that a step increase in the pGmer plasmid will result not only in a dumped increase of the GmerX genes, but also in an even more reduced increase of the stationary levels of PmerX proteins. Also, the realised ratio of different GmerX genes is not reflected in the same ratio of the increased protein PmerX levels. For instance, a 47-fold increase in plasmid copy numbers will result in a 10.8-fold increase in the GmerA gene copy numbers, and in an only 5-fold increase in the PA reductase content in the cell. Also a 5.4-fold increase in the GmerT gene level will result in an only 2.5-fold increase in the PmerT stationary level. By closely controlling the mercury permease level, the cell prevents a too much increase of the mercury content of the cytosol, thus preventing a too much consumption of cell energy and involvement of a large number of RSH molecules in coping with the mercury removal.

To elaborate the structured model of the mercury uptake, nominal values have been adopted from literature for the *E. coli* cell main characteristics (cell volume, cell cycle, cell concentration and density of the culture), presented in table 1 together with the information source. Major species concentration, calculated with the footnote (a) formula based on the copy number content of the cell, correspond to the K-12 strain of *E. coli* described by the EcoCyc [20] databank.

The proteome concentration, of $\sum_j^{all} c_{Pj,cyt} = 10^7 \text{ nM}$ was evaluated by considering the cell as including 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 [21]. The genome includes ca. 4500 genes (of one copy). The lumped metabolite concentration of $[\text{MetP}]_s = 3 \times 10^8 \text{ nM}$ and those of external nutrients were adopted from literature [12,22]. Only the concentration of $[\text{MetG}]_s \approx 2.0017 \times 10^7 \text{ nM}$ (table 1) resulted from fulfilment of the state-law constraint for an isotonic cell system, i.e.:

$$\sum_j^{all} c_{j,cyt} = \sum_j^{all} c_{j,env}, \Rightarrow \sum_j^{all} c_{MetGj,cyt} = \sum_j^{all} c_{j,env} - \sum_j^{all} c_{MetPj,cyt} - \sum_j^{all} c_{Gj,cyt} - \sum_j^{all} c_{Pj,cyt} \quad (3)$$

where c_j = species j concentration. The mercuric ion concentration range in the environment $\text{Hg}_{\text{env}}^{2+}$ and those of the *mer* plasmid cover the experiments [2]. A reference value for the mercury ion concentration in

the cytosol $[\text{Hg}_{\text{env}}^{2+}]$ was obtained by averaging the experimental data.

The saturation level of PmerX proteins was considered for large concentrations of mercury in the environment, i.e. for $[\text{Hg}_{\text{env}}^{2+}] = 120 \mu\text{M}$. Corresponding values for $[\text{PR}]_s$ and $[\text{PD}]_s$ have been determined by fitting the model with the stationary data sets. The same rule has been applied to determine the stationary saturation concentration of $[\text{PA}]_s$, the fitted value being of the same order of magnitude to its Michaelis-Menten constant (3700 nM). The lumped permease concentration $[\text{PT}]_s = [\text{PmerP}]_s + [\text{PmerT}]_s + [\text{PmerC}]_s$ was evaluated with the approximate formula $1.5 \{[\text{PA}]_s + 15\} + [\text{PR}]_s$, based on the observations that $[\text{PmerP}]_s = [\text{PA}]_s + 10\text{-}20 \text{ nM}$, $[\text{PmerT}]_s = 0.5[\text{PmerP}]_s$, and $[\text{PmerC}]_s = [\text{PR}]_s$ (one copy number correspond to ca. 1 nM for a born cell volume of 10^{-15} L) [19].

The used reaction schema for the quick self- and cross- control of the gene expression remains an open question. Literature information indicates a certain number of quick buffering reactions, of type $G + \text{TF} \Leftrightarrow \text{GTF}$, by which the transcription factors TFs adjust the gene G activity through catalytically inactive GTF species. A cascade control of the transcription and translation is proved as leading to better regulatory performances [10].

The whole cell modelling framework

The classical modelling framework of dynamic cell processes using continues variables is that of a constant volume, the differential mass balance of species $dc_j/dt = g_j(\mathbf{c}, \mathbf{k})$ being written in terms of concentrations, accounting for the cell growing rate as a 'decay' rate of key species (often lumped with the degrading rate) in a so-called 'diluting' rate D . Such a representation might suffice for many applications, but not for accurate modelling of cell regulatory GRC under perturbed conditions, often distorting the prediction quality. The variable-volume whole-cell (VVWC) modelling framework, intensively promoted [22, 10-12, 34], and adopted in this study, presents the advantage of explicitly linking the volume growth, external conditions, osmotic pressure, cell content ballast, and net reaction rates for all cell-components. Basically, the model mass balance equations and conservation relationships of the form:

$$\frac{dc_j}{dt} = \frac{1}{V} \frac{dn_j}{dt} - D c_j = r_j(\mathbf{c}, \mathbf{k}) - D c_j = g_j(\mathbf{c}, \mathbf{k})$$

[at steady-state $g_j(\mathbf{c}_s, \mathbf{k}) = 0$]; $j = 1, \dots, n_s$;

$$\pi V(t) = RT \sum_{j=1}^{n_s} n_j(t); \quad (\text{Pfeffer's law in diluted solutions})$$

$$D = \frac{1}{V} \frac{dV}{dt} = \left(\frac{RT}{\pi} \right) \sum_j^{n_s} \left(\frac{1}{V} \frac{dn_j}{dt} \right), \quad (4)$$

(where V = cell volume; c_j = species j concentration; n_j = species j number of moles; D = cell content dilution rate, i.e. the cell volume logarithmic growing rate; π = osmotic pressure; T = temperature; R = universal gas constant; n_s = number of species inside the cell; t = time). The model hypotheses correspond to an open cell system, of uniform content and negligible inner gradients, being separated by the environment through a semi-permeable membrane, of negligible volume and resistance to nutrient diffusion and following the cell growing dynamics. The constant osmotic pressure of the system, $\pi_{\text{cyt}} = \pi_{\text{env}} = \text{constant}$, leads to the same total concentration inside and outside the cell,

$$\left(\sum_j^{all} c_j \right)_{\text{cyt}} = \left(\sum_j^{all} c_j \right)_{\text{env}}$$

Variable	Value	Observations
<i>Cell overall characteristics:</i>		
- initial cell volume (cytoplasm, $V_{cyt,o}$)	1.1×10^{-15} L	Adopted from various sources
- cell cycle time (t_c)	138.6 min	Ibidem
- dilution rate ($D_s = \ln(2)/t_c$)	0.3 h^{-1}	
- cell concentration in the culture medium (c_{cell})	10^{10} cells (L env) $^{-1}$	[2]
- cell density in the culture medium (ρ_{cell})	10^6 mg protein (L cell) $^{-1}$	[2]
<i>Species concentrations (nM)</i>		
- mercuric ion in the environment, [Hg_{env}^{2+}]	0-120 μM	[2-3]
- mercury in the cytoplasm, [Hg_{cyt}^{2+}], [Hg_{cyt}^0]	0-3430	[4]
- average concentration of mercury in the cytoplasm, $c_{Hg2cy,ref}$	1866	Ibidem
- lumped nutrients used for genome synthesis, [$NutG$] (ref. to the environmental volume)	3×10^7	[12,22]
- lumped nutrients used for proteome synthesis, [$NutP$] (Ibidem)	3×10^8	Ibidem
- lumped metabolites used for proteome synthesis, [$MetP$]	3×10^8	Ibidem
- lumped metabolites used for genome synthesis, [$MetG$]	2.0017×10^7	Footnote (b)
- lumped genome (active part), [G]	4500/2	Footnotes (c,d)
- lumped proteome, [P]	1×10^7	Footnote (e)
- various <i>mer</i> operon plasmid vectors, [$plGmer$], [$Gmer$]	3-140	[2-3]
- genes expressing the <i>mer</i> proteins, [GR],[GT],[GA],[GD]	$\frac{[GX]_{min}}{2\beta} \cdot \frac{[plGmer]}{[plGmer]_{min}}$	Footnotes (d,f); X= R,T,A,D
- catalytic inactive forms of <i>mer</i> genes, [$GRPRPR$], [$GTPTPT$], [$GAPAPA$], [$GDPDPD$],	Ibidem	Ibidem
- proteins initiating and controlling the <i>mer</i> operon expression, [PR], [PD]	100	Fitted from data; see also obs. of Barkay et al. [19]
- lumped permease (PT= PmerT + PmerC) for membranar transport of Hg_{env}^{2+}	1000-7000	Footnote (g)
- reductase for cytosol mercury ions, [PA]	1000-4000	Fitted from data; see also Footnote (h)
- proteic dimers of <i>mer</i> proteins with TF [$PXPX$] with X= R,T,A,D.	1-100	Evaluated by optimizing the GRC holistic properties [11-12]

Footnotes:

(a) Inner cell concentrations are evaluated with the formula [11]:

$c_j = (\text{copynumbers of species } j \text{ per cell}) / (N_A V_{cyt})$, where $N_A = 6.022 \times 10^{23}$ is the Avogadro number, V_{cyt} = average volume of the cell.

(b) calculated from the state-law constraint for an isotonic and isothermal cell system:

$$\sum_j c_{j, cyt} = \sum_j c_{j, env} \quad , \quad \sum_j c_{MetGj, cyt} = \sum_j c_{j, env} - \sum_j c_{MetPj, cyt} - \sum_j c_{Gj, cyt} - \sum_j c_{Pj, cyt} .$$

(c) The considered K-12 strain of *E. coli* genome includes ca. 4500 genes (EcoCyc [20]).

(d) The maximum regulatory effectiveness of expression takes place for equal active and inactive G-forms at steady-state [10-12], i.e. $[Gj]_s = [GjTF]_s$, where TF denotes the transcription factor adjusting the gene activity.

(e) The considered K-12 strain of *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 [20,21].

(f) As reported by Philippidis et al. [2], the level of *mer* genes is less than those of the introduced plasmid vectors (of $plGmer$ level), diminished with a β factor ranging between β_{min} and β_{max} . A linear interpolation formula is:

$$\beta = \beta_{min} + \frac{\beta_{max} - \beta_{min}}{plGmer_{max} - plGmer_{min}} (plGmer - plGmer_{min}) .$$

(g) protein lump $PT = PmerP + PmerT + PmerC$ concentration is calculated according to experimental observations reviewed by Barkay et al. (2003), with the approximate formula (in nM):

$$[PT] = [PmerP] + [PmerT] + [PmerC] \approx [PmerP] + 0.5 [PmerP] + [PmerR] = 1.5 [PmerP] + [PmerR] = 1.5 \{ [PA] + 15 \} + [PR]$$

(h) Concentration of reductase PA is comparable with those of the permease PmerP at homeostasis [19], being of the order of magnitude of its Michaelis-Menten constant (3700 nM), and close to its substrate Hg_{cyt}^{2+} limitation constant (of ca. 1800 nM),

Table 1
E. COLI CELL MAIN CHARACTERISTICS AND STATIONARY CONCENTRATIONS OF OUTER AND INNER SPECIES CONSIDERED FOR A REFERENCE CELL 'SATURATED' IN PmerX PROTEINS (X= R,T,P,C,A,B,D) CORRESPONDING TO HIGH LEVELS OF ENVIRONMENTAL MERCURY OF $[Hg_{env}^{2+}] = 120 \mu\text{M}$. CONCENTRATIONS ARE EXPRESSED IN NM REFERRING TO THE CELL VOLUME OR ENVIRONMENTAL SPACE ^(a)

In fact, the supplementary constraint introduced by the Pfeffers' state equation leads, for an isotonic and isothermal cell, to an increase in the species connectivity: inner species interact directly in common reactions and metabolic steps, but also indirectly via the common cell volume to which all components contribute. Thus, a perturbation in some species level or their net reaction rates will influence the volume growth, which in turn will perturb the other cell component concentrations by means of the so-called 'secondary perturbations'. In the VVWC model the rates of individual reactions are constrained by the periodicity of the cell cycle and by the requirement that molar amounts of all components and volume must double in exactly one cell cycle. To be consistent with the mentioned hypotheses, the models require that each cell process be included at some level of detail, i.e. as an individual or lumped species and reaction.

The VVWC approach is not only increasing the system estimability, but it is proved as being more promising in predicting local and holistic regulatory characteristics of the metabolic network, such as [10-12]: simulation of secondary perturbations into the cell; simulations of interactions of the studied metabolic path with the proteome and genome by placing it in a whole-cell and mimicking the equilibrated cell growth; a more realistic representation of the regulatory units/modules performances; simulation of the cell overall content 'ballast' effect in smoothing perturbations, etc. For instance, variable environment will produce smaller perturbations

but longer transient times for cells with a large content compared to cells with a 'sparing' content. The VVWC disadvantages consist in an increased model complexity and requirement to include information not only on the studied metabolic pathway but also on all the cell key species (even treated as lumps) and on the holistic properties. In spite of its increased complexity, the VVWC approach can be very advantageous in fairly representing complex cell processes such as GRC.

In the GRC modular (lumped) representation, a module is formed by a certain number of interconnected regulatory units somehow related to a certain gene expression, each one controlling the rate of components' synthesis or consumption [11,12,23,24]. Due to frequent lack of structured information to estimate all kinetic parameters, the known low GRC estimability is recovered by optimizing the holistic properties in terms of regulatory efficiency [12,25-27]. The regulatory efficiency is appreciated by means of a quite large number of performance indices took from the theory of automated systems, such as [11]: i) *Stationary efficiency*, defined as low values of normalized sensitivity coefficients $S(c_j; c_i) = (\partial \ln(c_j) / \partial \ln(c_i))$; ii) *Responsiveness* to exo- or endogeneous stimuli, in terms of small transient times τ_j necessary to a species j stationary level to reach a new steady-state (with a certain tolerance); iii) *Dynamic efficiency*, expressed by the species j fast recovering time ($\tau_{rec,j}$) of the steady-state (with a certain tolerance) after an

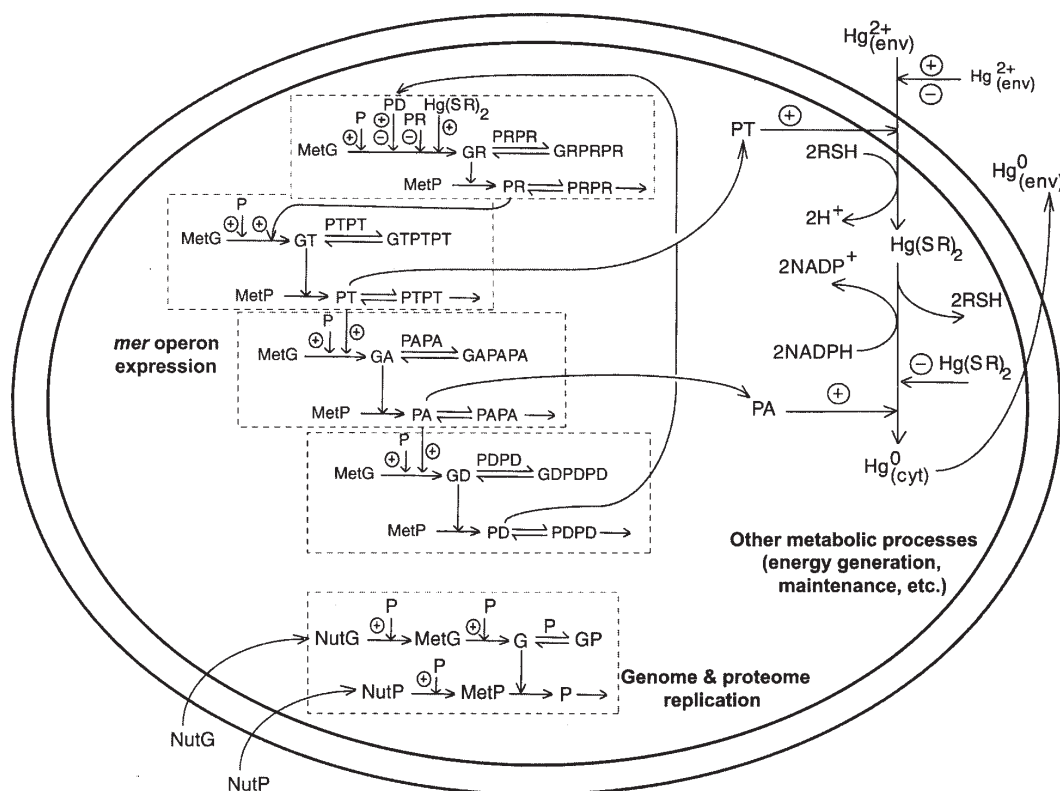


Fig. 1. The whole-cell model for reducing the Hg^{2+} ions from environment to volatile Hg^0 in *E. coli* bacteria. The simplified reaction path includes: two modules for mediated transport of Hg^{2+} into cytosol (catalysed by PT) and its reduction (catalysed by PA); five regulatory modules of *mer* operon expression including successive synthesis of PR (the transcriptional activator of other protein synthesis), lumped PT permease, PA reductase, and of the control protein P and genome G replication. The regulatory system 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 \rightarrow = proteins; G \rightarrow = 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; \oplus / \ominus positive or negative feedback regulatory loops

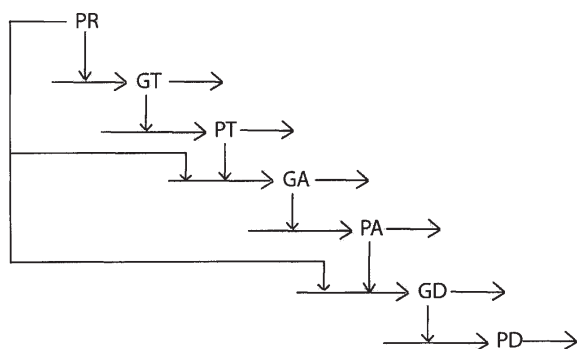
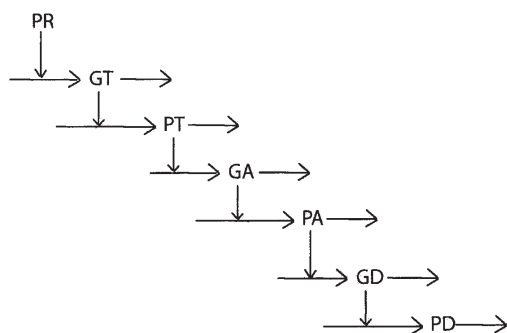
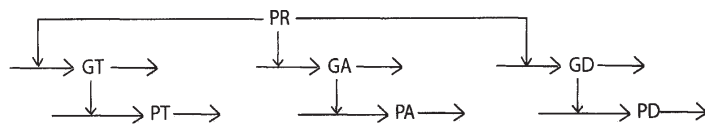


Fig. 2. Various alternatives of the PR-controlled paths considered for the *mer* operon expression: (up) Parallel expression of GT, GA, and GD; (middle) Consecutive expression in the succession GT, GA, and GD (each expression level controlling the next step); (down) Parallel-consecutive expression with a mixed control

impulse-like perturbation into the cell; iv) *Overall responsiveness* of the GRC to stimuli, approximated by the average transient time $AVG(\tau_i)$ of species; v) *Species connectivity* expressed by the synchronised response to a certain perturbation, in terms of standard deviation $STD(\tau_i)$; vi) *System stability, stability region and strength*, evaluated by using the $J = dg/dc$ system Jacobian matrix (e.g. the necessary condition for stability is that $Re(\lambda_j) < 0$ for all j , where λ_j are the eigenvalues of the J matrix). Such indices can also be related to individual regulatory elements / effectors, thus better pointing out their role.

The structured modular representation of a GRC is computationally very tractable because the regulatory properties can quantitatively be studied in connection to certain types of regulatory units, and a similarity analysis among modules can often be applied. For instance, a linear dependence between the regulatory unit performance index and the number of the allosteric control steps can be easily established for every TF type [11,28].

Dynamic model construction using a simplified metabolic pathway

The modular gene expression regulatory circuit and pathway

The main two reaction steps of the mercury uptake in *E. coli* cells consist in the mercury transport into the cell and its enzymatic reduction (fig. 1). The Michaelis-Menten kinetics proposed [2-4] was adopted, together with the fitted saturation/inhibition constants, K_{m_7} , K_{m_8} , K_{i_7} , K_{i_8} (the 7th and 8th reactions in table 2). Strong variation of these constants with the Gmer plasmid level reveals in fact the complexity of these steps. The $r_{max,7} = k_7 c_{PT}$ and $r_{max,8} = k_8 c_{PA}$ terms have been considered variable in the present study, requiring a separate identification of the turnover numbers k_7 and k_8 . As k_7 is two-orders of

magnitude smaller than k_8 , it clearly appears that mercury transport is the limiting step of the process. In fact, normal cells can not tolerate higher cytosolic mercury content, so the cell permeases do not allow importing large amounts of mercury overstepping $[Hg_{cyl}^{2+}] > 5 \mu M$ [3-4]. Thus, the mercury import plays the role of the 'bottle-neck' for the whole process, providing the advantage to the survival of the cell. Even if the *mer* operon concentration into the cell is increased, the expression level of PmerX proteins is maintained different, being closely adjusted according to the environmental conditions. Both mercury transport and reduction reactions involve important amounts of RSH compounds (such as glutathione) and regenerable NADPH, requiring a close control of the mercury import and *mer* gene expression. The large excess of RSH and NADPH, usually existing in the normal *E. coli* cells (ca. 140 μM), satisfies the reaction requirements, as long as the Michaelis-Menten constant for NADPH in the reduction rate expression is 10-times smaller (ca. 14 μM).

When building-up a structured dynamic model of the mercury uptake in bacteria, the first question concerns the type of GRC schema to be adopted, once the production of PR was induced by the cytosolic mercury. Figure 2 presents three alternatives: i) a parallel expression of GT, GA, and GD, catalysed by PR (adopted by the simplified version of the model, [18]); ii) a consecutive expression of the GmerX genes, the PR catalysing only expression of the first gene; iii) a parallel-consecutive expression schema with a closely control exerted by the PR protein. Preliminary calculations using the whole-cell model indicate the consecutive schema as being more suitable for describing the ratio of the expression levels under various environmental and inner *mer* operon content conditions (comparative results not

Table 2

IDENTIFIED RATE CONSTANTS OF THE *E. COLI* WHOLE-CELL MODEL USING THE AVAILABLE DATA SETS [2-4,15] AT VARIOUS *MER* PLASMID LEVELS (GMER) AND STATIONARY REDUCTION CONDITIONS ($[TF]_s = 4 \text{ nM}$; STATIONARY LEVELS OF PROTEINS AT THE REFERENCE CELL CONDITIONS ARE THOSE OF TABLE 1; REACTION RATES ARE REFERRED TO THE CELL VOLUME; THE SYMBOL ‘~’ INDICATE APPROXIMATE VALUES TO AVOID DISPLAYING 16-DIGITS MANTISSA NUMBERS)

Reaction	Rate expression, $\frac{1}{V} \frac{dn_j}{dt}$, (nM min ⁻¹)	Rate constants (units in min and nM)
<i>Genome & proteome replication</i>		
$NutG \xrightarrow{P} MetG$	$k_1 c_{NutG} c_P$	3.3526×10^{-10}
$NutP \xrightarrow{P} MetP$	$k_2 c_{NutP} c_P$	5.1678×10^{-10}
$MetG \xrightarrow{P} G$	$k_3 c_{MetG} c_P$	1.1192×10^{-13}
$MetP \xrightarrow{G} P$	$k_4 c_{MetP} c_G$	7.4106×10^{-8}
$G + P \rightarrow GP$	$k_5 c_G c_P$	$\sim 10^{-2}$
$GP \rightarrow G + P$	$k_6 c_{GP}$	10^5 (adopted)
<i>Mercury ion import into cytosol</i> ^(a)		
$Hg_{env}^{2+} \xrightleftharpoons[-2H^+]{+2RSH, (PT)} Hg_{cyt}^{2+}$	$\frac{k_7 c_{PT} c_{Hg_{env}^{2+}}}{K_{mt} + c_{Hg_{env}^{2+}}}$	[Gmer] $r_{max,t}$ K_{mt} 3 nM 8.2×10^6 3800 67 nM 13.4×10^6 4700 78 nM 17.5×10^6 3600 124 nM 20.6×10^6 5100 140 nM 19.8×10^6 6000
<i>Mercury ion reduction</i> ^(a)		
$Hg(SR)_2 \xrightleftharpoons[-2NADP^+]{+2NADPH, (PA)} Hg_{cyt}^0 + 2RSH$	$\frac{k_8 c_{PA} c_{Hg_{cyt}^{2+}}}{K_{mP} + c_{Hg_{cyt}^{2+}} + c_{Hg_{cyt}^{2+}}^2 / K_{iP}}$	[Gmer] K_{mP} K_{iP} 3 nM 12600 96700 67 nM 15900 91500 78 nM 22800 87400 124 nM 15500 75200 140 nM 19500 93100 $k_8 = 1.2016 \times 10^5$
<i>Volatile elementary mercury removal</i>		
$Hg_{cyt}^0 \rightarrow Hg_{env}^0$	$k_9 c_{Hg_{cyt}^0}$	1.0105×10^4
<i>Gene GR expression module</i>		
$MetG \xrightarrow{Hg_{cyt}^{2+}, P, PD, PR} GR$	$k_{10} c_{MetG} c_{PD}^{n_{PD}} c_{PR}^{n_{PR}} \left[\frac{1 + b \left(\frac{c_{Hg_{cyt}^{2+}}}{c_{Hg_{2cy,ref}} + c_{Hg_{cyt}^{2+}}} \right)^{n_H}}{c_{Hg_{2cy,ref}} + c_{Hg_{cyt}^{2+}}} \right]$	7.7457×10^{-12} $b = 2a^4$; $a = 2.5$ $c_{Hg_{2cy,ref}} = 1866$ $n_{PD} = 1$; $n_{PR} = -0.5$; $n_H = 2$
$MetP \xrightarrow{GR} PR$	$k_{11} c_{MetP} c_{GR}$	4.6200×10^{-11}
$GR + PRPR \rightleftharpoons GRPRPR$	$k_{12} c_{GR} c_{PRPR}$; $k_{13} c_{GRPRPR}$	2.5000×10^4 ; 10^5 (adopted)
$2PR \rightleftharpoons PRPR$	$k_{14} c_{PR}^2$; $k_{15} c_{PRPR}$	4.0000×10^3 ; 10^5 (adopted)
<i>Gene GT expression module</i>		
$MetG \xrightarrow{P, PR} GT$	$k_{16} c_{MetG} c_P c_{PR}$	1.1607×10^{-16}
$MetP \xrightarrow{GT} PT$	$k_{17} c_{MetP} c_{GT}$	1.9912×10^{-9}
$GT + PTPT \rightleftharpoons GTPTPT$	$k_{18} c_{GT} c_{PTPT}$; $k_{19} c_{GTPTPT}$	2.5000×10^4 ; 10^5 (adopted)
$2PT \rightleftharpoons PTPT$	$k_{20} c_{PT}^2$; $k_{21} c_{PTPT}$	5.3572×10^{-2} ; 10^5 (adopted)
<i>Gene GA expression module</i>		
$MetG \xrightarrow{P, PT} GA$	$k_{22} c_{MetG} c_P c_{PT}$	4.2477×10^{-19}
$MetP \xrightarrow{GA} PA$	$k_{23} c_{MetP} c_{GA}$	1.3250×10^{-9}
$GA + PAPA \rightleftharpoons GAPAPA$	$k_{24} c_{GA} c_{PAPA}$; $k_{25} c_{GAPAPA}$	2.5000×10^4 ; 10^5 (adopted)
$2PA \rightleftharpoons PAPA$	$k_{26} c_{PA}^2$; $k_{27} c_{PAPA}$	1.2345×10^{-1} ; 10^5 (adopted)
<i>Gene GD expression module</i>		
$MetG \xrightarrow{P, PA} GD$	$k_{28} c_{MetG} c_P c_{PA}$	6.4483×10^{-19}
$MetP \xrightarrow{GD} PD$	$k_{29} c_{MetP} c_{GD}$	1.1049×10^{-10}
$GD + PDPD \rightleftharpoons GDPDPD$	$k_{30} c_{GD} c_{PDPD}$; $k_{31} c_{GDPDPD}$	2.5000×10^4 ; 10^5 (adopted)
$2PD \rightleftharpoons PDPD$	$k_{32} c_{PD}^2$; $k_{33} c_{PDPD}$	4.0000×10^1 ; 10^5 (adopted)

Footnote:

(a) Parameters $r_{max,t}$, K_{mt} , K_{mP} , K_{iP} are adopted at the values indicated by Philippidis et al. [2].

presented here). Such a schema was also suggested by the preliminary qualitative observations from literature [19].

Once the modular regulation is adopted for the GRC schema, the next modelling step is to decide on the level of detail adopted for every gene expression

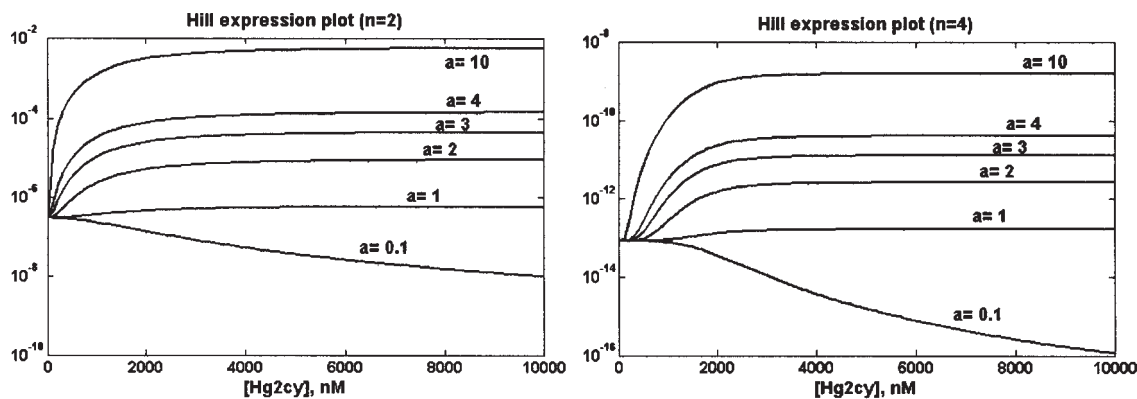


Fig. 3. Logarithmic plot of the Hill-type algebraic function E (eq. 5) dependence on the cytosolic mercury ion concentration, for the Hill constants $a = 0.1-10$, and $n_H=2$ (left) and $n_H=4$ (right), using the average $c_{\text{Hg2cy,ref}} = 1866$ nM. The sigmoidal-like increase of E is dumped as the Hill constants (a, n_H) are getting closer to 1

regulatory module. To fill not-up the model with too many parameters, difficult to estimate, and lacking of structured dynamic data on the *mer* gene expression, a quick screen among regulatory modules from literature [11,12,28] indicates the simplest alternative those with one step backward control of lumped transcription and translation, and a gene activity adjustment via rapid buffering reactions of type $G + \text{TF} \rightleftharpoons \text{GTF}$.

To be effective, these buffering reactions must be more rapid than the catalyzed synthesis itself, while the mass conservation law must be all time fulfilled, $[G] + [\text{GTF}] = \text{constant}$. As indicated in the literature, the time-window for such processes is of micro-seconds [29], so the dissociation constant of the buffering reactions was adopted every time at a value of 10^5 min^{-1} (table 2). As proved, the maximum regulatory efficiency of the protein synthesis reaction at steady-state corresponds to $[G]_s / [G]_{\text{total}} = 1/2$, when the maximum regulation sensitivity vs. perturbations is reached (see [10-12,25] for a more detailed discussion). Consequently, concentrations of active and inactive G-forms have to be taken equal at the steady-state, i.e. $[G]_s = [\text{GP}]_s$, or $[\text{GmerX}]_s = [\text{GmerX}::\text{TF}]_s$ (fig. 1; index 's' denotes the steady-state). In fact, dimeric TFs are proved to be more effective than single molecules, in accordance to the observations [30] and simulations [10,12]. Thus, dimeric $\text{PmerX}::\text{PmerX}$ species have been used to confer optimal regulatory properties to the GRC. Generally, the allosteric control of a gene G activity is performed by successive rapid buffering reactions $G + \text{TF} \rightleftharpoons \text{GTF} + \text{TF} \rightleftharpoons \text{GTF}_2 + \dots + \text{TF} \rightleftharpoons \text{GTF}_n$, developing successive inactive species of type $[\text{GTF}_n]$. Such a consecutive schema can amplify the regulatory efficiency of the whole expression module, with a 1.3–2 multiplicative factor for every new added buffer reaction (e.g. shortening the recovering rates of steady-state or transition times after a perturbation; [11]).

Such regulatory modules with self-control via $\text{TF} = \text{PXPX}$ proteins have been proposed for the four expression modules of GR, GT, GA, GD genes, and also for the expression of the lumped genome G (see the dashed rectangles in fig. 1). The consecutive expression schemes also include the catalytic role of the lumped proteome P, aiming at accounting for the influence of the whole-cell content on smoothing the environmental perturbations. The TF dimer levels for *mer*-gene expression (PP or PXPX in fig. 1), are to be chosen to confer maximum efficiency to the regulatory modules and some dynamic properties to the whole GRC, but also to avoid exaggerate consumption of cell energy (i.e. avoiding overshoots in the repressing/de-repressing enzyme levels, [31]).

A special module G/P (fig. 1) has been included into the whole-cell construction to mimic the genome (G-lump) and proteome (P-lump) replication and, due to their large concentrations, the cell content 'ballast' and 'inertial' effect when coping with external or internal perturbations. As investigated in previous works, this module is necessary not only due to the VVWC framework specific requirements, but also to explicitly account for lumped interactions of individual gene expression with the genome and proteome under the isotonic osmolarity constraint.

The kinetic model of table 2 (attached to the reaction schema of fig.1) basically includes elementary reactions, excepting for the mercury import and reduction steps, and also for the GR expression induction, adopted of Hill type. This GRC crucial step is involving a large number of controlling loops, accounted for by means of PD and PR species reaction orders n_{PD} and n_{PR} to be fitted from data. The GR expression amplification at low cytosolic mercury concentrations c_{Hg2cy} is represented by the following Hill type term:

$$E = \frac{1}{c_{\text{Hg2cy,ref}}^{n_H}} \cdot \frac{1 + b (c_{\text{Hg2cy}} / c_{\text{Hg2cy,ref}})^{n_H}}{1 + (c_{\text{Hg2cy}} / c_{\text{Hg2cy,ref}})^{n_H}}, \quad b = 2a^4 \quad (5)$$

The logarithmic plot of the Hill expression E in figure 3 reveals a close dependence of the shape on the inducer level (c_{Hg2cy}) but also on the Hill parameters (a, n_H). The additive terms in the numerator allow maintaining a certain level of the reaction rate in the absence of inducer, while smaller values of constant a (close to 1 in our application) allows flattening the GR synthesis (reaction no. 10 in table 2) irrespectively of the inducer level. Such a kinetics accounts for n_H molecules of inducer $\text{Hg}(\text{RS})_2$ allosterically binding the promoter site. The value of $c_{\text{Hg2cy,ref}}$ is taken at the average level of the cytosolic mercury (experimentally determined).

Ranging the stationary data predictions

A recent proposed structured model for mercury uptake in *E. coli* [18] used a modular control of the *mer* operon expression, based on a parallel expression of GA, GT, and GD catalysed by PR. The model, estimated from stationary data [2] for $[\text{Hg}^{2+}]_e = 0-120 \mu\text{M}$ and *mer* plasmid level of $\text{Gmer} = 3 \text{ nM}$, has been found to satisfactory represent the overall reduction rate ($r_{\text{Hg,s}}$), but also the ratio of the *mer* gene expression levels under various environmental mercury conditions. The Hill-type induction of the process was very pushful

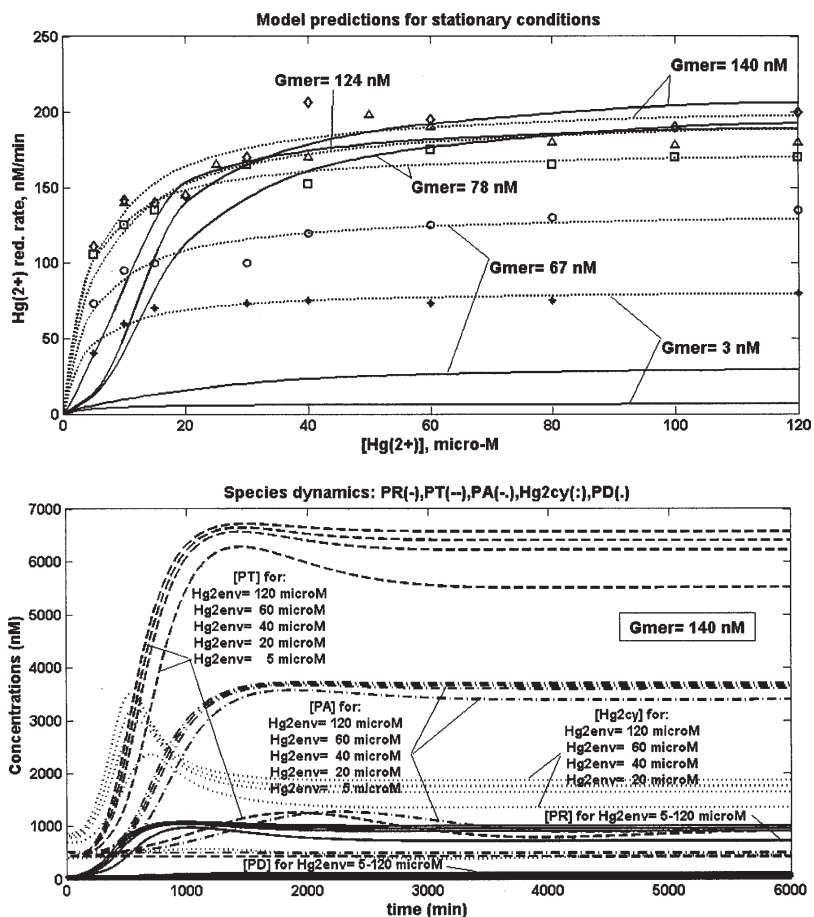


Fig. 4. (up) Experimentally recorded stationary reduction rates of mercury ions at various environmental $[Hg^{2+}]$ realised by the *E. coli* cells cloned with *mer* operon plasmid pGmer in concentrations of 3 nM (*), 67 nM (O), 78 nM (\square), 124 nM (\diamond), and 140 nM (Δ). The predicted reduction rates are given by Philippidis et al. [2] using an overall unstructured model (:), and by the present structured model (-) identified using $p[Gmer] = 140$ nM, $[PR]_s = 1000$ nM, $[PA]_s = 3700$ nM, $[PD]_s = 100$ nM, $[TF]_s n_H = 4$ nM, $a=4$, $n_H=4$, $\beta=1$, $n_{PD}=1$, $n_{PR}=-0.5$, and $c_{Hg2cy,ref} = 1866$ nM parameters (Gmer denotes plasmid concentration). (down) Long term dynamic evolution of the PmerX protein concentrations in *E. coli* cells cloned with $[p[Gmer]] = 140$ nM plasmid, for step perturbations in the environment from $[Hg_{env}^{2+}]_s = 100$ nM to various mercury ion levels varying from 5 μ M to 120 μ M

($a=10, n_H=4, c_{Hg2cy,ref} = 3430$ nM), while the transition times among various steady-state conditions (of the order of 200 min), have been realized using only one buffering reaction in every gene expression regulatory schema. Extrapolation of model predictions for higher Gmer levels encounters however difficulties.

In contrast, the current model variant adopts a consecutive schema for the *mer* gene expression, and a larger number of parameters to detail the process induction, and mercury transport and reduction steps (fig. 1 and table 2). The first attempt with this model was to fit all stationary data sets [2], in terms of Hg(II) reduction rates for $[Hg_{env}^{2+}]_s = 0-120$ μ M and plasmid levels of Gmer= 3-140 nM (fig. 4), by maintaining an accentuated Hill induction of the *mer* operon expression ($a=4, n_H=4, c_{Hg2cy,ref} = 1866$ nM), with the same first-order PD catalysis, and a self-PR repression of order $n_{PR} = -0.5$ [32]. The Hg(II) transport and reduction kinetic parameters have been adapted for every Gmer level (table 2), while the others parameters have been estimated from stationary data at saturation conditions of $[Hg_{env}^{2+}]_s = 120$ μ M and Gmer= 140 nM. The resulted model predictions, presented in figure 4 (up), fail to reproduce the stationary reduction rates $r_{Hg,s}$ at low $[Hg_{env}^{2+}]_s$ and Gmer levels, and any other trial to find a better estimate also fail. The shape of the $r_{Hg,s}$ curves at low Hg_{env}^{2+} concentrations is inadequate, suggesting the existence of significant *mer* proteins at low mercury

concentrations of $[Hg_{env}^{2+}] < 0.1$ μ M, that can ensure larger $r_{Hg,s}$ levels. It is however to observe in figure 4 (down) that the net Hill induction ensures a quite satisfactory increase of PmerX protein levels over few minutes for a step increase of $[Hg_{env}^{2+}]_s$ from 0.1 μ M to various higher levels, with a maximum 13-fold increase for PT and PR, 7-fold for PA, and 2-fold for PD.

A second trial with the proposed model was to fit all available experimental data sets by adopting a 'milder' induction of the GR expression, using lower values for Hill parameters of ($a=1, n_H=4$). The results presented in figure 5 (up) reveal a perfect match with the stationary data, even better than those reported [2] using the unstructured global kinetic model (dashed lines). However, the results from figure 5 (down) reveal a near flat evolution of all PmerX proteins, even in the absence of the external mercury. Such a result, indicating an already adapted microorganism to high loads of mercury, contradict the experimental observations [2] that reported significant variations of the PmerX protein levels in time and with the Gmer plasmid content into the cloned cells.

Consequently, a compromise between these two extremes have been realised by fitting the model vs. all data sets but adopting moderate Hill parameters ($a=2.5, n_H=2$) and the same β damping factor when calculating the *mer* gene levels from the plasmid content (eq. 2). The identified parameters are presented

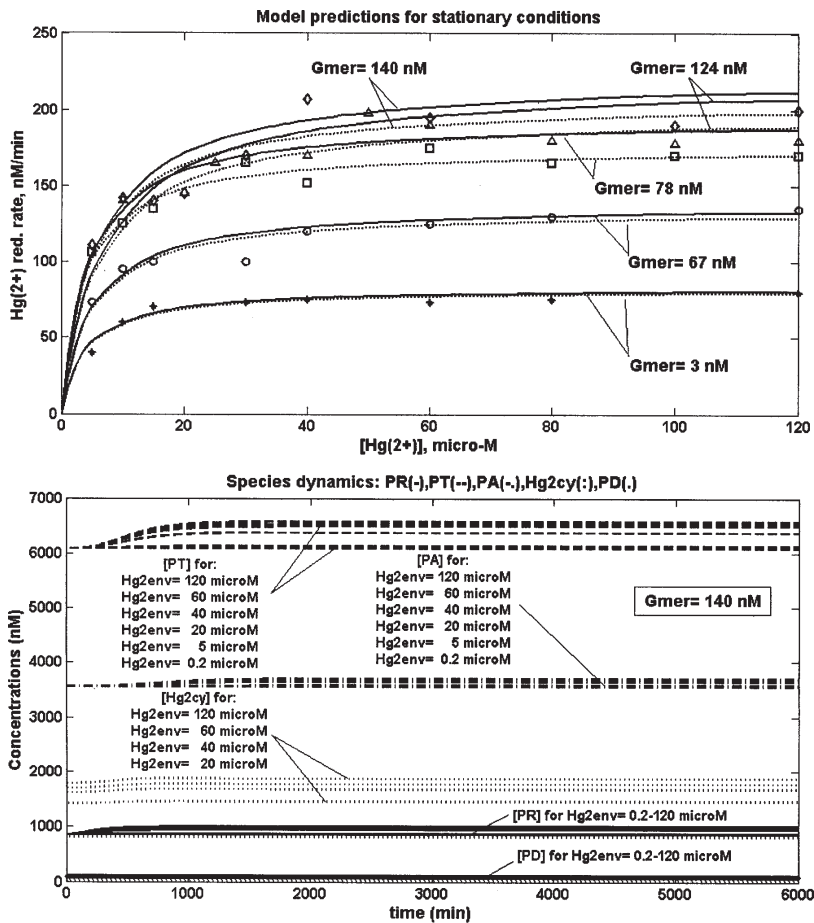


Fig. 5. (up) Experimentally recorded stationary reduction rates of Hg^{2+} at various pGmer concentrations (*,O,•,◇,△), and predictions of the unstructured model of Philippidis et al. [2] (:), and of the present structured model (-) identified using pGmer= 140 nM, [PR]_s = 1000 nM, [PA]_s = 3700 nM, [PD]_s = 100 nM, [TF]_s = 4 nM, $a=1$, $n_H=4$, $\beta=1$, $n_{PD}=1$, $n_{PR}=-0.5$, and $c_{Hg2cy,ref} = 1866$ nM parameters (Gmer denotes plasmid concentration). (down) Long term dynamic evolution of the PmerX protein concentrations in *E. coli* cells cloned with [pGmer]= 140 nM plasmid, for step perturbations in the environment from $[Hg_{env}^{2+}] = 100$ nM to various mercury ion levels varying from 0.2 μ M to 120 μ M

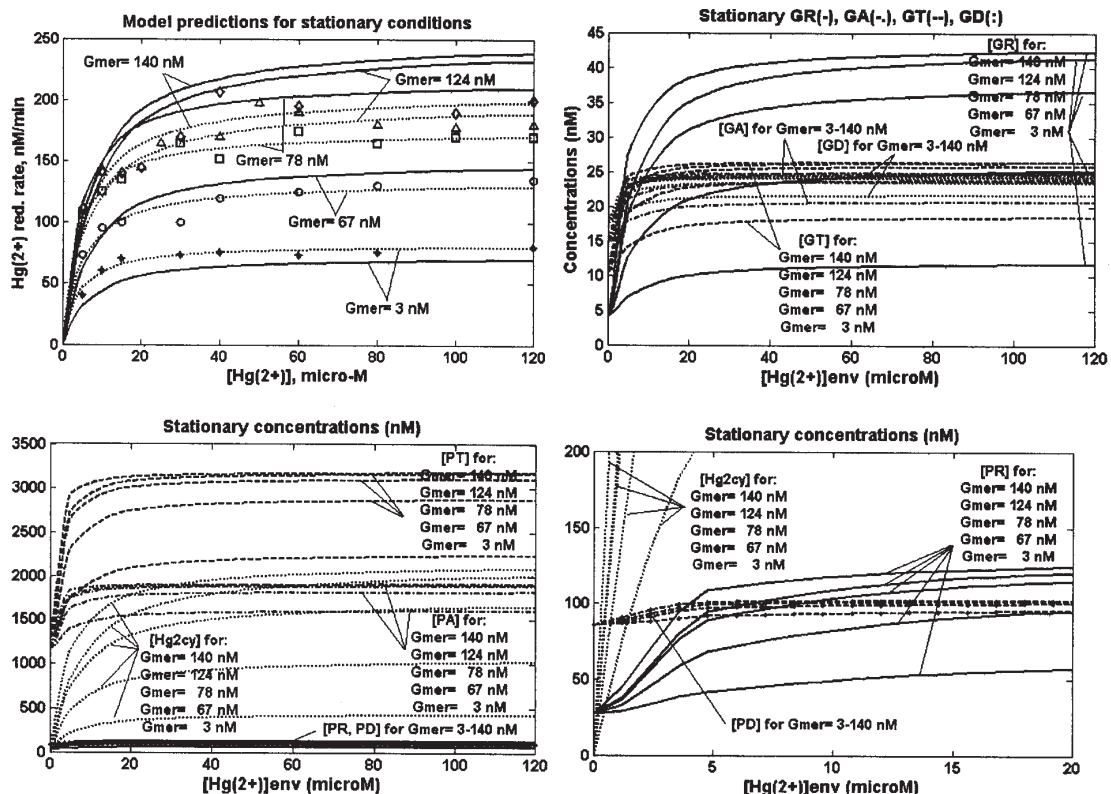


Fig. 6. (up-left) Experimental stationary reduction rates of Hg^{2+} at various pGmer concentrations (*,O,•,◇,△), and predictions of the unstructured model of Philippidis et al. [2] (:), and of the present structured model (-) identified using [pGmer]= 140 nM, [PR]_s = 100 nM, [PA]_s = 1800 nM, [PD]_s = 100 nM, [TF]_s = 4 nM, $a=2.5$, $n_H=2$, $\beta_{min}=1.5$, $\beta_{max}=3$, $n_{PD}=1$, $n_{PR}=-0.5$, and $c_{Hg2cy,ref} = 1866$ nM parameters (Gmer denotes plasmid concentration). (up-right) Predicted mer genes stationary concentrations in the *E. Coli* cells cloned with [pGmer]= 3-140 nM plasmid concentrations, for stationary $[Hg_{env}^{2+}] = 10^{-4}$ -120 μ M. (Down) Predicted PmerX proteins and $[Hg^{2+}]_{cyt}$ stationary concentrations in the *E. coli* cells cloned with [pGmer]= 3-140 nM plasmid concentrations, for stationary $[Hg_{env}^{2+}] = 10^{-4}$ -120 μ M

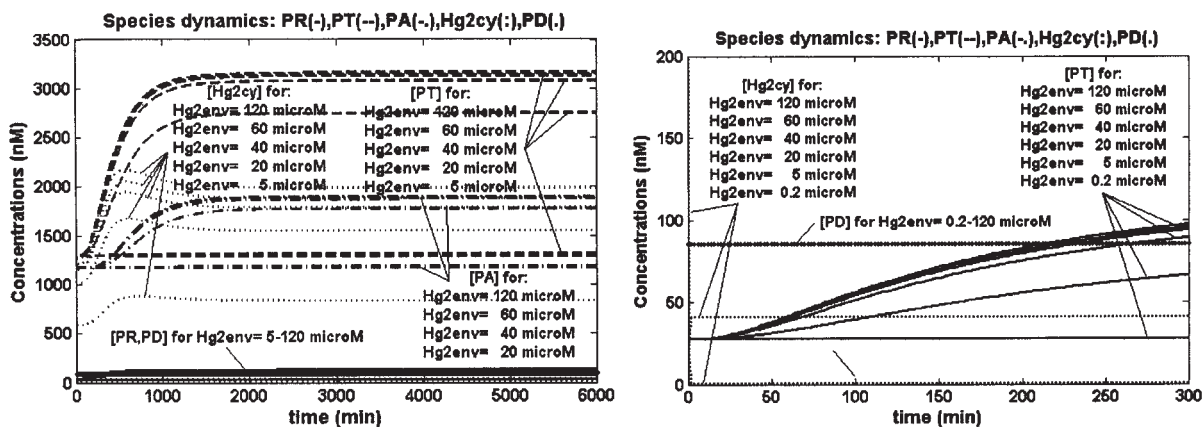
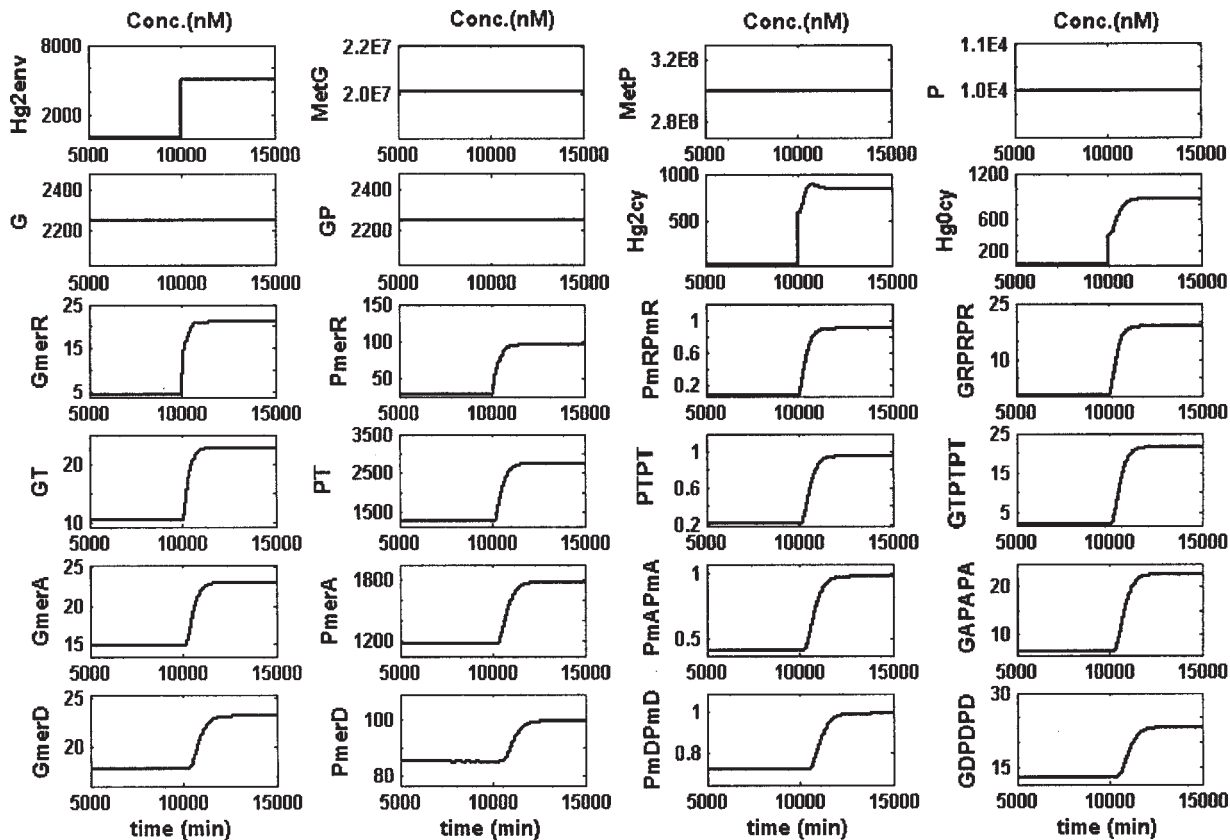


Fig. 7. (up) Typical evolution of species concentrations predicted by the cell model, after a step-perturbation in the environment from $[Hg_{env}^{2+}]_s = 0.1 \mu M$ to $5 \mu M$, for $[plGmer] = 140 \text{ nM}$, $[TF]_s = 1 \text{ nM}$. (down) Dynamic evolution of the PmerX protein concentrations in *E. coli* cells cloned with $[plGmer] = 140 \text{ nM}$ plasmid, for step perturbations in the environment from $[Hg_{env}^{2+}]_s = 100 \text{ nM}$ to various levels varying from $0.2 \mu M$ to $120 \mu M$

in table 2. The results, displayed in figure 6 (up-left), show a quite satisfactory agreement with the data especially for low and moderate Gmer plasmid levels, and a slightly over-estimation of maximum 20-25% of stationary reduction rates $r_{Hg,s}$ at higher Gmer levels. In contrast, the GmerX genes rise with the Hg_{env}^{2+} (fig. 6, up-right) is of ca. 8.5-fold for GR, 2.5-fold for GT, 1.7-fold for GA and GD, roughly matching the experimental observations (excepting for GA of which concentration increase is moderate). Also the PmerX protein content in the cell is increasing significantly with the external mercury levels (fig. 6, down), that is ca. 6-fold for PR, 3-fold for PT, 1.5-fold for PA and PD, which are in concordance with the experimental observations (excepting for PA of a moderate increase). As displayed in the figure 6 (down), the increase of the PmerX content occurs at even very low levels of $[Hg_{env}^{2+}]_s$.

Ranging the dynamic properties of the GRC model

As previous studies proved, the advantage of using a modular GRC control schema with self-adjustable expression for each gene is the possibility to easily range the dynamic responsiveness (i.e. transient times τ_i between steady-states) and dynamic efficiency (i.e. recovering times $\tau_{rec,i}$ of a steady-state) of the whole network via suitable levels chosen for the TFs intermediates [11-12].

The species transient times have been evaluated, usually with a 1% tolerance [11], after applying a step-like perturbation in the environmental $[Hg_{env}^{2+}]_s$ level. Such responses to a step perturbation from $0.1 \mu M$ to 5 nM are displayed in figure 7 (up). The dynamics of PmerX proteins after step perturbations of various magnitudes are represented in figure 7 (down) for $[TF]_s = 1 \text{ nM}$. It is to remark a satisfactory fast initiation of the *mer* operon expression, the rising time for quantitative production of PR being of order of few minutes, as indicated in the literature [19].

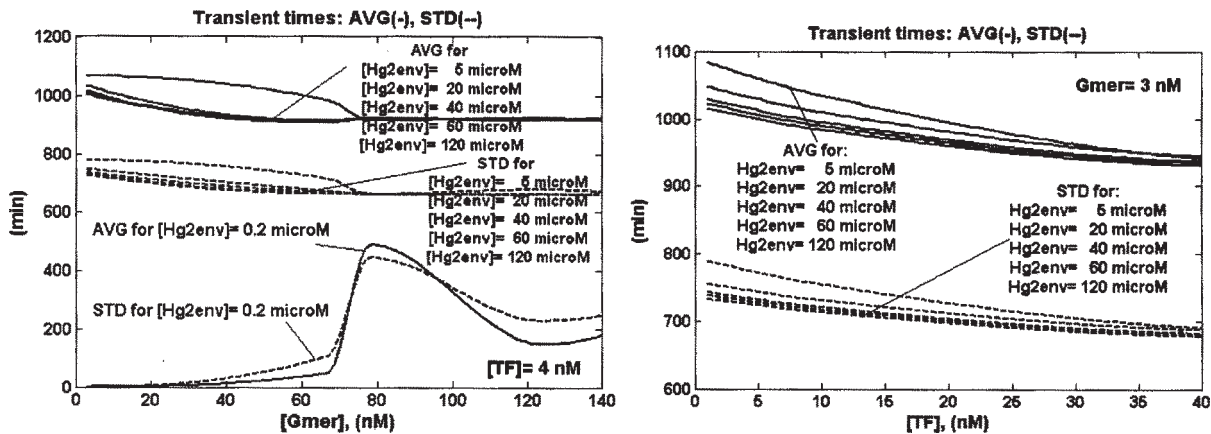


Fig. 8. Predicted average of species transient times $AVG = \text{avg}(\tau)$ (with 1% tolerance), and their standard deviation $STD = \text{st.dev}(\tau)$, after a step-perturbation in the environmental mercury from $[Hg_{env}^{2+}]_s = 100 \text{ nM}$ to various levels varying from $0.2 \mu\text{M}$ to $120 \mu\text{M}$ (for $[TF]_s = 4 \text{ nM}$). Variation of AVG and STD with the plGmer plasmid level (left), and with the transcription factor TF level (right) (Gmer denotes plasmid concentration).

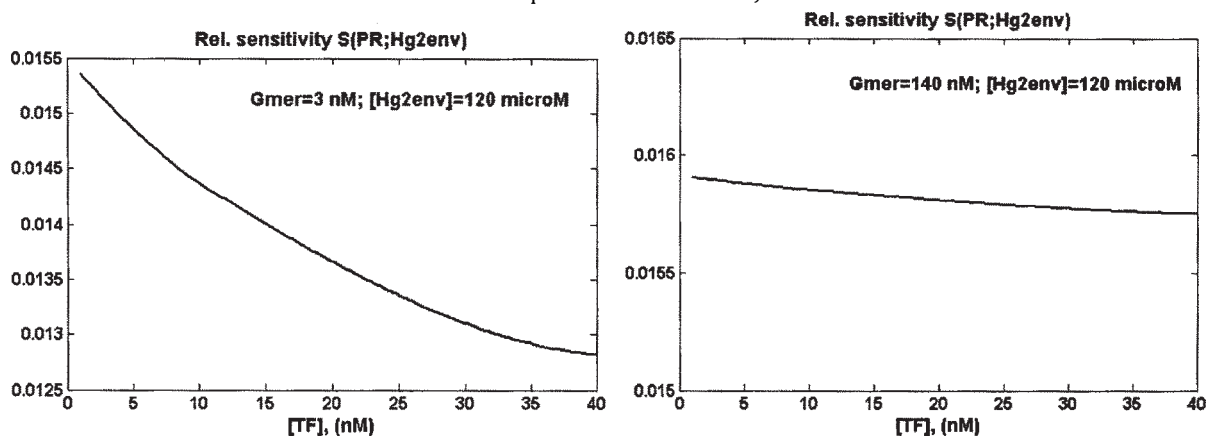


Fig. 9. Predicted PR species relative sensitivity to external mercury concentration, $S = (\partial \ln(c_{PR}) / \partial \ln[Hg_{env}^{2+}])_s$ for various $[TF]_s$ levels, for $[Hg_{env}^{2+}]_s = 120 \mu\text{M}$, and mer plasmid $[p\text{IGmer}] = 3 \text{ nM}$ (left) and $[p\text{IGmer}] = 140 \text{ nM}$ (right) (Gmer denotes plasmid concentration)

Lacking of any systematic measurements on the mer gene expression dynamics, an unique value has been adopted for the $[TF]_s$ irrespectively of gene. A plot of the average (AVG) and standard deviation (STD) of τ in figure 8 (right) reveals their continuous decrease, with the $[TF]_s$ level used in the buffering reactions controlling each gene expression. The same result was obtained for the recovering times after a dynamic impulse-like perturbation in the environment (not presented here). Variation of AVG and STD with the level of Gmer plasmid and $[Hg_{env}^{2+}]_s$ is presented in figure 8 (left), revealing a quite flat performance at higher Gmer-levels.

In fact, large values of $[TF]_s$ intermediates (e.g. more than 20-40 nM as indicated in the literature) are not favourable, because large overshoots require lot of energy to degrade the excess of protein. The milder Hill-inductive schema of the current model leads to predicting a 3-4 times slower average response to stimuli comparatively to the parallel expression schema with an accentuated Hill-induction step proposed [18] for Gmer = 3 nM. However, these performance indices can be easily adjusted by changing the allosteric control of every gene expression, by adopting a larger numbers of buffering reactions (e.g. $n=2-4$). It was proved in this case that the realised AVG decrease was with a 1.3-2 de-multiplicative factor for every new buffering reaction added to the allosteric control schema [10-11]. Even if the transition or species recovering times after a perturbation are larger than of the cell cycle (139 min here), such a situation is not uncommon. As observed [33] in the oscillatory GRC, the transition times in the

inductive/co-repressed GRC systems can be much larger than the cell-division cycle, the state of the oscillator, transient switch or amplifier being transmitted from generation to generation of cells.

Results and discussion

As resulted from the stationary and dynamic characteristics, the proposed kinetic model for mercury uptake and GRC control in *E. coli* cells can satisfactorily represent the cell behavior in a wide range of concentration intervals for the environmental mercury and for cloned cells with an increasingly content of mer operon (Figs. 6-7). Given the available data on stationary rates for mercury reduction, and on the expression level for some proteins at various Gmer levels, but very few on the gene expression dynamics, the model can be seen as satisfactory. However, the degree of detail of the control schema is susceptible to improvements at every mer gene level, thus improving not only the ratio of expression among various mer genes to closer fit the experimental data, but also predictions of stationary reduction rates at higher Gmer levels.

One possibility to get improvements in model predictions is to consider separate β -dumping factors for every gene to calculate its real copy number when the bacteria cell is cloned with mer plasmids. Of course, that will require complete measurements on gene content under various conditions, and will complicate the parameter identification step.

Another possibility is to detail the allosteric schema that control the expression of each mer gene, and to separately calculate the optimal $[TF]_s$ intermediate levels conferring

the regulatory characteristics that better reproduce dynamic measurements. Such a degree of detail will however require gene transcript measurements using high-time-resolution time-series microarray technique, with a sampling frequency of seconds to minutes [5]. Such a requirement is based on the observation that the signal transduction is very fast, from few seconds to minutes from binding the signal molecule(s) to the specific binding site on DNA until the transcription factor activation.

In all presented simulations, the cell system has been proved as presenting a stable homeostasis for the key-species, the real part of the eigenvalues of the differential model Jacobian matrix being all negative $\text{Re}(\lambda_i(\mathbf{J})) < 0$ (not presented here). The system was proved as being quite responsive to stimuli at low $[\text{Hg}_{\text{env}}^{2+}]_s$ levels, and less responsive at higher levels, as revealed by the plots of figure 9 on the normalized sensitivity coefficients of $[\text{PR}]_s$ vs. the external mercury, $S(\text{PR}; \text{Hg}_{\text{env}}^{2+}) = (\partial \ln[\text{PR}] / \partial \ln[\text{Hg}_{\text{env}}^{2+}])_s$.

The model is quite flexible, the Hill parameters (a, n_H) and the partial orders n_{PD} and n_{PR} to which PD and PR control the GR expression being kept as identifiable constants, that is the key characteristics of the process inductive step. It is to observe that the current amplification feature of the Hill-type induction step is milder than those identified by Maria [18] at low Gmer levels, which can be explained by the much broader validity of the current model. However, such a result might suggest a change of Hill type induction parameters with the level of *mer* operon, which remains a theoretical and experimental issue to be further disclosed.

Starting from such an observation, an interesting idea to improve the model adequacy, by covering broad operating intervals, is to consider the possibility to vary the parameters of the Hill-type inductive step of the *mer* operon transcription according to the $[\text{Hg}_{\text{env}}^{2+}]_s$ and *mer* plasmid level into the cell. The Hill-induction can become more exacerbated at low levels of Gmer plasmid and closer to a saturated Michaelis-Menten kinetics for high levels. However, such an approach will introduce new parameters into the model, difficult to be checked individually. Also the partial orders of reaction of PD might take different values according to the mercury amount into the cell, thus better blocking the overproduction of PA in the absence of $\text{Hg}_{\text{env}}^{2+}$, e.g. by simply changing the sign of n_{PD} . The dynamic characteristics of the GRC can be relatively easily adjusted through the expandable allosteric schema adopted for controlling each gene expression, and the adjustable levels of TFs, as soon as consistent dynamic data will be available.

In fact, the obtained results with such a reduced structured model reveal the difficulties encountered when representing the complexity of the expression control process only using a few number of parameters identified from stationary data and few dynamic information or from a similarity-based analysis.

Conclusions

An attempt to model the kinetics of the mercury uptake process in *E. coli* cells, by also including the GRC controlling the expression of the involved *mer* genes, reveals the advantages of using a modular pathway approach with the possibility of transferring information from similar studied systems. Even if a rough model adequacy has been obtained, the study reveals the difficulties encountered when the model is covering a broad range of variable

behaviors, in terms of stimuli levels and of the *mer* operon content (for the cloned cell cases), by using a reduced pathway structure and incomplete data sets especially on the gene expression dynamics.

Structured cell models can however present certain advantages, in spite of a much larger experimental effort necessary to check the dynamics of the involved metabolic steps together with the homeostatic levels of key-species and reactions, and of computational steps necessary to establish the best compromise between the detailing degree of pathway representation and the model predictive capabilities. While an unstructured overall (Michaelis-Menten type) kinetic model can suffice for scaling-up the process and for equipment design, for a deep understanding of the process aiming at modifying the micro-organism and at transferring the knowledge to study similar cell systems, a structured model is crucial. Modelling the inner cell process dynamics will help in understanding and predicting the cell response to stationary or dynamic perturbations in the environment, self-regulation of the process under various conditions, gene expression amplification, or other GRC characteristics in connection to the cell volume growth and inner-cell content replication.

The present case study on modelling the mercuric ion reduction in *E. coli* cells suggests that, according to the available information, a suitable combination of elementary and lumped kinetic terms (Michaelis-Menten, Haldane, or Hill type) can be a good choice to simulate the concerned GRC characteristics. For instance, the current model allows to simulate: mercury transport and reduction rate inhibition with the substrate; the quick and efficient control exerted by PR protein on *mer* gene expression; the whole GRC dynamics and its regulatory properties linked to the adjustable level of intermediates and the chosen allosteric control schema; linkage of the concerned GRC to the lumped genome/proteome; whole-cell response and key-species sensitivity and synchronization in coping to various external perturbations. Further developments of the whole-cell model are seeking for supplementary details of the cellular process as soon as quantitative observations on the individual cell components' dynamics will become available.

The study confirms the conclusions [2] that simple cell cloning with increased amounts of plasmid, aiming at raising the *mer* operon content, does not improve much the reduction process overall rate. This is because the cell developed a self-protecting system aiming to shrink the import of large amounts of mercury, which might lead to blockage of RSH species, exhausting the cell energetic resources, and eventually compromising the whole metabolism. A possibility is to design mutant cells where expression of the transport *mer* genes to be controlled by a separate GRC, apart from those controlling the reductase gene expression. In such a manner, larger amount of mercury can be processed with the expense of a modified metabolism directed toward metal ion uptake (up to a critical level of tolerability to be further determined), which eventually will raise the industrial application feasibility.

The used VVWC modelling framework appears to be a promising alternative to evaluate the GRC properties during simulation of a certain metabolic process. Thus, by placing the *mer* operon regulatory modules in an *E. coli* growing cell, the bacterial mercury resistance can be better studied and the whole-cell behaviour can be mimicked under simulated stationary or perturbed environmental conditions. Special effects, such as the

cell content 'inertial' effect in treating perturbations, and the effect of the indirect or secondary perturbations transmitted via the cell-volume under isotonic osmolarity conditions, can also be roughly simulated.

Acknowledgment. Funding support from the DAAD Project A/09/02572 (2009) at TUHH (Germany) is acknowledged. The author is also grateful to Professor An-Ping Zeng (Institute of Bioprocess and Biosystems Engineering, Technische Universität TUHH Hamburg-Harburg, Germany) for the constructive discussions on modelling GRC.

Notations

a, b - rate constants in the Hill-type kinetic expression
 c_j - species (lump, or 'pool') concentration
 D - cell content dilution rate (i.e. cell-volume logarithmic growing rate)-
 g - kinetic model function vector
 $J = dg / dc$ - kinetic model Jacobian matrix
 k - kinetic constant vector
 K - equilibrium or kinetic constants
 n_H - Hill-coefficient
 n_{PD}, n_{PR} - partial orders of reaction
 n_j - species j number of moles
 n_s - no. of species
 N_A - Avogadro number
 r_j - species j reaction rate
 R - universal gas constant
 $S(y; x) = \partial \ln(y) / \partial \ln(x)$ - relative sensitivity of y vs. x
 t - time
 t_c - cell-cycle time
 T - temperature
 V - cell volume

Greeks

β - damping factor used to calculate the mer gene content from the concerned plasmid content
 $\lambda(J)$ - eigenvalues of the dynamic model Jacobian
 π - osmotic pressure
 ρ - density
 $\tau_{rec,j}$ - species j recovering time of the steady-state
 τ_j - species j transition time from one steady-state to another

Index

cyt - cytoplasm
 env - environment
 max - maximum
 min - minimum
 o - initial
 ref- reference
 s-steady-state

Abbreviations

AVG (τ_j)-average of τ_j
 G_{\bullet} - gene
 Gmer, pGmer-plasmids generating mer operons into the cell
 GmerX-mer genes
 GRC-genetic regulatory circuit
 MetG, MetP-metabolites
 NutG, NutP-nutrients
 P_{\bullet} -protein
 PmerX-mer proteins
 Re()-real part
 RSH-compounds including thiol redox groups
 STD(τ_j) - standard deviation of τ_j
 TF-transcription factor
 VVWC-variable volume whole-cell
 [.] - concentration

References

1. LEONHÄUSER, J., RÖHRICHT, M., WAGNER-DÖBLER, I., DECKWER, W.D., Eng. Life Sci., **6**, 2006, p. 139

2. PHILIPPIDIS, G.P., MALMBERG, L.H., HU, W.S., SCHOTTEL, J.L., Applied & Environmental Microbiology, **57**, 1991a, p. 3558
3. PHILIPPIDIS, G.P., SCHOTTEL, J.L., HU, W.S., Mathematical modelling and optimization of complex biocatalysis. A case study of mercuric reduction by Escherichia coli. Expression Systems & Processes for DNA Products. National Science Foundation Report ECE-8552670, University of Minnesota; 1991b.
4. PHILIPPIDIS, G.P., SCHOTTEL, J.L., HU, W.S., Biotech. Bioeng., **37**, 1991c, p. 47
5. HE, F., BALLING, R., ZENG, A.P., Reverse engineering and verification of gene networks, JI. Biotechnology, 2009 (under revision).
6. BANSAL, M., GATTA, G. D., DI BERNARDO, D., Bioinformatics, **22**, 2006, p. 815
7. CAMACHO, D., VERA LICONA, P., MENDES, P., LAUBENBACHER, R., Ann. N.Y. Acad. Sci., **1115**, 2007, p. 73-
8. STYCZYNSKI, M.P., STEPHANOPOULOS, G., Comp. & Chem. Eng., **29**, 2005, p. 519
9. ZAK, D.E., VADIGEPALLI, R., GONYE, G.E., DOYLE III, F.J., SCHWABER, J.S., OGUNNAIKE, B.A., Comp. & Chem. Eng., **29**, 2005, p. 547
10. MARIA, G., Chemical and Biochemical Engineering Quarterly, **20**, 2006, p. 353
11. MARIA, G., Chemical and Biochemical Engineering Quarterly, **19**, 2005, p. 213
12. MARIA, G., Chemical and Biochemical Engineering Quarterly, **21**, 2007, p. 417
13. KHANIN, R., VINCIOTTI, V., WIT, E., Proc. Natl. Acad. Sci. USA, **103**, 2006, p. 18592
14. TIAN, T., BURRAGE, K., Proc. Natl. Acad. Sci. USA, **103**, 2006, p. 8372
15. PHILIPPIDIS, G.P., SCHOTTEL, J.L., HU, W.S., Enzyme Microb. Technol., **12**, 1990, p. 854
16. DÖBLER, I.W., CANSTEIN, H., LI, Y., DECKWER, W.D., Env. Sci. & Technol., **34**, 2000, p. 4628
17. DECKWER, W.D., BECKER, F.U., LEDAKOWICZ, S., DÖBLER, I.W., Environ. Sci. Technol., **38**, 2004, p. 1858
18. MARIA, G., Chemical and Biochemical Engineering Quarterly, **23**, 2009, p. 323
19. BARKAY, T., MILLER, S.M., SUMMERS, A.O., FEMS Microbiology Reviews, **27**, 2003, p. 355
20. 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/>
21. ALLEN, G.C. Jr., KORNBERG, A., JI. Biological Chemistry, **266**, 1991, p. 11610
22. MORGAN, J.J., SUROVTSEV, I.V., LINDAHL, P.A., J. theor. Biology, **231**, 2004, p. 581
23. HLAVACEK, W.S., SAVAGEAU, M.A., J. Mol. Biol., **266**, 1997, p. 538
24. SAVAGEAU, M.A., Math. Biosciences, **180**, 2002, p. 237
25. SEWELL, C., MORGAN, J., LINDAHL, P., J. theor. Biol., **215**, 2002, p. 151
26. Van SOMEREN, E.P., WESSELS, L.F.A., BACKER, E., REINDERS, M.J.T., Signal Processing, **83**, 2003, p. 763
27. SCHWACKE, J.H., VOIT, E.O., Theoretical Biology and Medical Modelling, **1**, 2004, p. 1
28. YANG, Q., LINDAHL, P., MORGAN, J., J. theor. Biol., **222**, 2003, p. 407
29. SUN, B., CHIU, D.T., J. Am. Chem. Soc., **125**, 2003, p. 125:3702.
30. SALIS, H., KAZNESSIS, Y., Comp. & Chem. Eng., **577**, 2005, p. 577
31. ROSENFELD, N., ELOWITZ, M.B., ALON, U., J. Mol. Biol., **323**, 2002, p. 785
32. VOIT, E.O., IEE Proc. Syst. Biol., **152**, 2005, p. 207
33. ELOWITZ, M.B., LEIBLER, S., Nature, **403**, 2000, p. 335.
34. MARIA, G., Rev. Chim. (Bucuresti), **59**, no. 3, 2008, p. 318

Manuscript received: 28.10.2009