Learning Robust Dynamic Networks in Prokaryotes by Gene Expression Networks Iterative Explorer (GENIE)

Oscar Harari, Cristina Rubio-Escudero, Patricio Traverso, Marcelo Santos, and Igor Zwir*

Abstract Genetic and genomic approaches have been used successfully to assign genes to distinct regulatory networks, but the uncertainty concerning the connections between genes, the ambiguity inherent to the biological processes, and the impossibility of experimentally determining the underlying biological properties only allow a rough prediction of the dynamics of genes. Here we describe the GENIE methodology that formulates alternative models of genetic regulatory networks based on the available literature and transcription factor binding site evidence. It also provides a framework for the analysis of these models optimized by genetic algorithms, inferring their optimal parameters, simulating their behavior, evaluating them by integrating robustness, realness and flexibility criteria, and contrasting the predictions to experimentally results obtained by Gene Fluorescence Protein analysis. The application of this method to the regulatory network of the bacterium *Salmonella enterica* uncovered new mechanisms that enable the inter-connection of the PhoP/PhoQ and the PmrA/PmrB two component systems. The predictions were experimentally verified to establish that both transcriptional and post-transcriptional mechanisms are employed to connect these two systems.

1 Introduction

One of the biggest challenges of the post genomic era is determining when, where and for how long genes are turned on or off [1]. Gene expression is determined

O. Harari and C. Rubio-Escudero
Dept. Computer Science and Artificial Intelligence, University of Granada, E-18071, Spain

M. Santos and I. Zwir
Dept. Computer Science, University of Buenos Aires, C1428EGA, Argentina

P. Traverso and I. Zwir
Howard Hughes Medical Institute, Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110-1093, USA

*Corresponding author: zwir@borcim.wustl.edu

by protein-protein interactions among regulatory proteins and with RNA polymerase(s), and protein-DNA interactions of these transacting factors with cis-acting DNA sequences in the promoters of regulated genes [2]. These interactions define complex genetic networks and the qualitative agreement between model and experiment in a series of studies depends both on the design of the network topology, which most of the times includes uncertain connections between genes, as well as on the dynamic behavior of the network, which is affected by the ambiguity inherent to the biological processes (e.g., monomer or dimmer binding of promoters, enzymes having kinase and/or phosphatase activities, etc.) and the mathematical models used to represent them. Moreover, the number of genes considered in the networks is usually large compared to the number of the available measurements (e.g., time-point expression), thus, more than one possible model may be consistent with the subjacent data. Finally, the data always contains a substantial amount of noise [3, 4] which in addition to previous problems, makes it difficult to deduce the implications of the underlying logic of genetic networks through experimental techniques alone.

We propose a methodology termed GENIE, for Gene Expression Networks Iterative Explorer, which embraces the uncertainty inherent to the biological problem and the imprecision of their underlined mathematical models by using an iterative approach. First, GENIE proposes a network topology based on DNA sequence analysis of transcription factor interactions, which, together with previous knowledge from the literature, constitute the raw material for the architecture design. Second, it transform the hypothesis provided by the network topology, by means of its possible chemical reactions and physical constraints, into a system of nonlinear ordinary differential equations [5]. Rather than advocating a single and definitive model of the genetic network, we describe a variety of optimal models learned by random walk [6] and improved by genetic algorithm techniques. Third, the network non-linear models are evaluated by testing their ability to reproduce the biological behavior observed in vivo. Fourth, the successful models are tested by considering different emergent properties, such as flexibility to reproduce all possible functional patterns, and robustness to changes in parameters and initial conditions.

We apply GENIE to uncover regulatory networks in the bacteria *Salmonella enterica serovar Typhimurium* by focusing on cross-link between the PhoP/PhoQ and PmrA/PmrB two-component systems, which govern virulence and the adaptation to low Mg^{2+} and high Fe^{3+} environments, respectively [7], and verified our predictions by measuring time-dependent gene expression using Gene Fluorescence Protein (GFP) techniques.

2 Problem: Computational and Biological Challenges

Modeling genetic networks: The scientific community has put a considerable amount of effort into designing approaches to model genetic networks [8]. Most of the models define species as nodes, and interaction between them as links of a graph. They differ in the values assigned to the nodes (i.e. initial concentration) and

links (i.e. the value of the interaction between the species), generating alternative models. Indeed, the interactions between elements can be considered as static or dynamic, and the entire model can be studied in a stochastic or deterministic context (e.g. Boolean, discrete or continues) [9].

The usage of continuous values to determine the level of gene expression and relationships among them results the most expressive model, because it allows capturing biological properties that can be experimentally observed. Ordinary Differential Equations (ODE's) are good approximation to continues models: ODE's capture the system by equations that calculate the difference of concentration of species (i.e. RNA, proteins) along the time. Statical ODE's [10] model the systems when they reach their steady state (i.e. the system has reached an equilibrium in which the difference of concentrations of species in function of time is equal to zero).

In contrast, dynamic models [5] do not necessarily consider this equilibrium, enabling the observation of the gene expression behavior over time. This important characteristic allows the temporal simulation of the system, results critical when studying biological systems, in which is possible to experimentally observe the dynamics for different sets of stimuli. An interesting concept of dynamic ODE's models is that the actual values of the parameters are not a priori estimated. Instead, the model can be evaluated by employing different sets of parameters to test if it follows certain macroscopic patterns previously known. As a result, the quality of the obtained network is not only determined by the chosen model, but also by the design of the inference method (i.e. learning strategy) that estimates the parameters of the network.

Two-component systems: In prokaryotes organisms, the "two-component systems" are small networks that control an important amount of cellular functions, constituting the main mechanism of signal transduction that allows the bacteria to modify its cellular behavior in response to environmental stimuli. These systems include a sensor protein that responds to specific signals and phosphorylates its cognate regulators. The response regulators are mostly transcription factors proteins (TF) that once they become phosphorylated, bind the DNA, and then, activate or repress their target genes. Although there are between 30 to 60 different two-component systems identified in bacterial genomes, they are not completely understood and some of them can be also preserved in eukaryotic genomes [11].

The PhoP/PhoQ two-component system constitutes a master regulator in *Salmonella enterica*, regulating the transcription of more than 2% of the genes in response of a low extra cellular Mg^{2+}. Another two component system present in this bacteria is the PmrA/PmrB system, which is related to the polymyxin B antibiotic inducted resistance; resistance to cell death mediated by Fe^{3+} among others. The target genes regulated by this system independently respond to two signals: high level of extra cellular Fe^{3+}, sensed by the PmrB protein; and low levels of Mg^{2+}, sensed by the PhoQ protein. This cross-talk between both two-component systems is mediated by the *pmrd* gene, which resulting protein PmrD can bind the PmrA protein probably in a posttranscriptional or posttranslational fashion. Curiously, *pmrD* harbors a PmrA binding site that results in a negative feedback that closes the regulatory loop. Although, this system has been widely studied [11], the exact mechanisms that defines the system dynamics is still unknown (see Fig. 1).

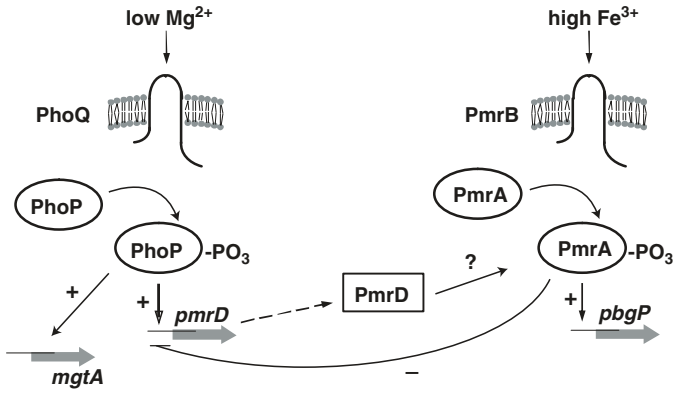

Fig. 1 The PhoP/PhoQ-PmrA/PmrB functional scheme in *Salmonella enterica serovar Typhimurium*. The PhoQ protein senses low Mg^{2+} and the PmrB protein high Fe^{3+} concentrations from the environment and both proteins phosphorylate their cognate response regulators PhoP and PmrA, respectively. Although each of these proteins control the expressions of their target genes in response to their own signal, an alternative cross-talk suggest that some genes regulated by the PmrA protein can be regulated by PhoP in low Mg^{2+} conditions via the PmrD protein. Indeed, a transcriptional negative feedback has been detected in the *pmrD* gene

3 Discovering Genetic Networks Using GENIE

GENIE is devoted to infer genetic regulatory networks. It consists of three main phases (see Fig. 2): (1) *discovery of the components of the studied system*, where we analyze the literature, databases and experimental evidence of cis-features (e.g., TF binding sites) to formulate alternative architectures for a genetic network and encode these models as continuous ODE's; (2) *identification of the desired system*, where we learn the parameters of the network, simulate its dynamics, and evaluate the performance of different models both by probabilistic measures and correlation with experimental results; and (3) *sensitivity analysis of the system parameters*, where we evaluate the robustness of the learned system and extract emergent properties from the evaluated architectures that may uncover biological significance (e.g., gene expression diversity).

3.1 System Components

GENIE discovers genetic regulatory networks by formulating hypothetical architectures, representing them as continuous models encoded as biochemical reactions that encode the dynamics of the system under different constrains.

In spite of the fact that continuous model require a proper parameter configuration, they offer the advantage that the parameterized components on which they are constructed can model a complete set of analogous gene dynamic (i.e. expression

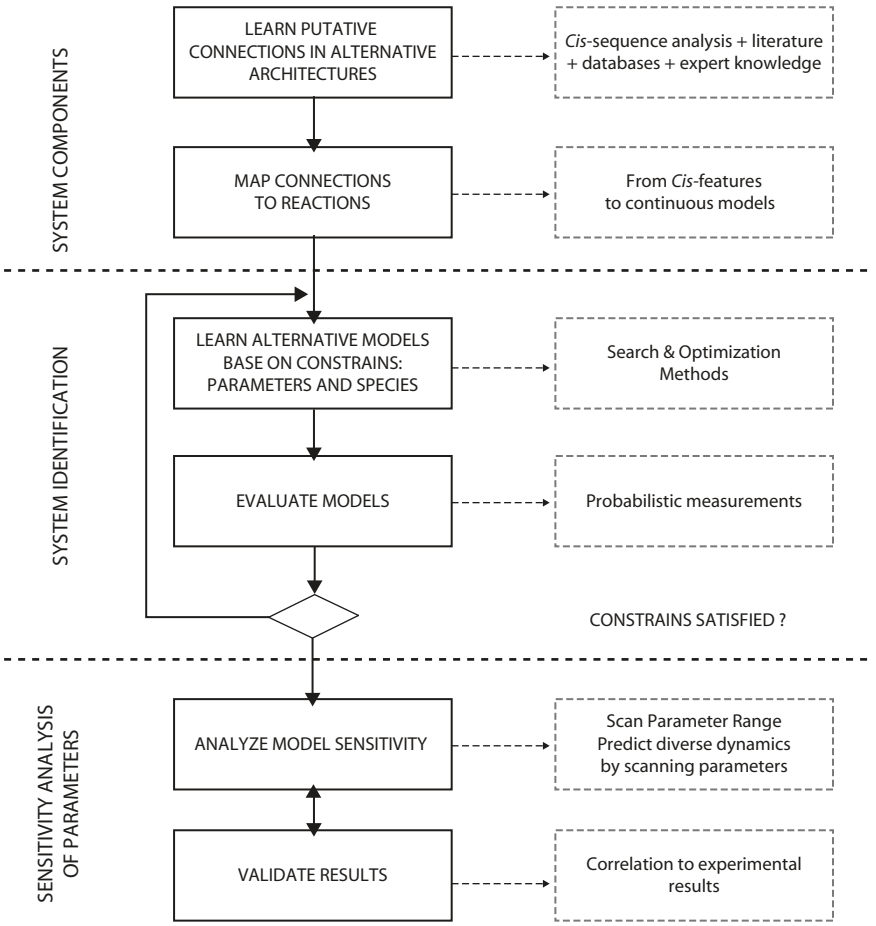

Fig. 2 Flowchart of the GENIE methodology. Each phase is decomposed into different task that are implemented in the methodology

intensity and rise time or order), thus these models can be customized to predict gene expression of a complete cluster of genes.

Network architecture: Transcriptional regulation evidence can be found in sequences (*Cis-sequence analysis methods*). We employ machine learning techniques [11] that analyze genome sequences and databases [12] to uncover initial hypothesis about architectures. Moreover, regulation evidence can be reinforced by microarrays experiments, however, the constraints in such analyses hitherto allow a relatively crude classification of gene expression patterns into a limited number of classes (e.g., up- and down-regulated genes).

Developing incremental mode: our methodology incrementally formulates network architectures to find the minimal one, according to the number of species and interactions, that exhibits the experimentally observed properties. It starts with a model that reflects the recovered information and postulates the most general

possible hypothesis for the unknown interactions. We express the rules that determine the behavior of genetic regulatory networks by decomposing the network into an aggregation of functional modules (e.g. negative/positive gene autoregulation, gene direct regulation, (des)phosphorylation of a protein) [2], which in turn are translated into a system of ODE's [5].

Mapping sequence-based circuits into continuous models: GENIE relays on Ingeneue software [5] which provides the Cash Karp method to integrate ODE's (see [5] for a comparative analysis of Cash Karp, SEBE, SEAPC, and Adams-Bashforth-Moulton alternatives applied to biological problems such ours).

Network constrains: The optimization of genetic networks has to consider at least two kind of constraints: Input/output constrains, where input signals activate the system and produce a desired output gene expression; and temporal constrains, which impose that the genes have to be ON and OFF at certain times with a specific order. [13].

3.2 System Identification

3.2.1 Learning Network Parameters and Species

GENIE employs both random walk and genetic algorithms (GA) strategies to search and optimize for parameters that identify the system. The random walk (RW) approach is a formalization of the intuitive idea of taking successive steps, each in a random direction (Meir et al report in [5] that RW approach obtains better estimation than linear optimization methods). GA provide a learning method motivated by an analogy to biological evolution [14]: it iteratively updates a pool of hypothesis, called population, to identify the best one. On each iteration, all members of the population (represented as chromosomes) are evaluated according to the fitness function. A new population is then generated by applying genetic operators (i.e. crossover and mutation) to the most fit individuals.

Chromosome representation: we encode the parameters of the ODE's as a vector of real numbers. *Fitness function:* It considers the value of every specie for each constrain at simulated time = 300 seconds (see section 3.2.2). *Selection:* we employ bit tournament to select the population that breed the new generation. *Crossover*: new individuals are generating by applying both two-point crossover and arithmetical crossover operators. *Mutation:* we employ the classical uniform mutation. *Elitism*: we retain the 3 solutions with best score in the elite set.

3.2.2 Evaluating Networks using Probability Measurements

We determine the capability of network architectures and their related parameters to reproduce the behavior of the living organism by applying a score function which evaluates the predicted concentration of distinguished species (equation (1))

$$score = \frac{\sum_i T(x_i)}{1+\sum_i T(x_i)} \qquad \begin{array}{l} T_{off} = \alpha_{\max}\left(\left(x_i/x_t\right)^3 \Big/ \left(1+\left(x_i/x_t\right)^3\right)\right) \\ T_{on} = \alpha_{\max}\left(1-\left(\left(x_i/x_t\right)^3 \Big/ \left(1+\left(x_i/x_t\right)^3\right)\right)\right) \end{array} \tag{1}$$

where i represents each specie; x_t represents the threshold for each specie; and $\alpha_{\max}$ is the worst possible value (i.e. 0.5). The functions T_{on} and T_{off} calculate a scalar value based on the half-maximal-activity threshold, according to the constrains (i.e. activated/repressed). A score value close to 0 indicates a high similitude with the constrained (we consider that a solution represents the expected pattern if its score is below 0.3).

Moreover, based on this score, we can compute the frequency of feasible solutions, and estimate the corresponding probability of randomly finding a configuration for a genetic network that fulfill the constrains (equation (2)). A high probability of finding configurations that reproduce the expected pattern can indicate that the functionality is more related to the network architecture itself than to the parameters:

$$p^n = f \quad p = 10^{\frac{\log f}{n}} \tag{2}$$

where p is the probability of randomly choosing a feasible solution (i.e. a configuration that allows the architecture to reproduce the expected patterns); f is the frequency of feasible configurations; and n is the number of parameters.

3.3 Sensitivity Analysis of Parameters

Different approaches have been proposed to evaluate the quality of the genetic regulatory network models (e.g. robustness; and flexibility) [15]. However these approaches partially evaluate the fidelity of model while representing a biological system. In this work, we propose a global quality measure based on: *Realism,* the model should be able to reproduce the experimentally observed behavior, relatively independent of its parameters; *Robustness*, network architectures should preserve the functional characteristic of the system when one or more parameters are perturbed. The models should tolerate variations without lost of realism of the link parameters (relations between species), because of the biological property of network resistance to subtle mutations of the participating genes; and node parameters (concentrations), because of the intrinsic noise of molecular systems. Finally, the *flexibility* criteria evaluates the capability of networks to simultaneously reproduce distinct patterns of behavior (i.e. constrains) of the system under study.

We evaluate the *Robustness* of a network architecture by randomly choosing a feasible solution and observe its behavior when we independently sample each parameter value, within a biological significant range, and fix the original configuration values for the other parameters (*scanning*). Thus, we determine a feasible solution range for the parameters, indicating possible alternatives to adapt the network to reflect the behavior of other genes.

Differential gene expression can be obtained from two distinct sources. A variety of network motifs integrated in the network architecture produce distinct expression in the target genes including the single-input (PhoP → *mgtA*), the chained (PhoP → *pmrD* → *pmrA* → *pbgP*) and the multi-component motif (PhoP → *pmrD* → *pmrA* → *pmrD*). Indeed, even within a particular network motif we can obtain differential expression in distinct target genes (e.g., PhoP → mgtA; PhoP → *mgtC*) by scanning the range of feasible solutions. Thus, allowing making predictions about diversity of unseen gene expression.

Furthermore, our methodology helps the evaluation of the biological significance of the results, by comparing the predictions to experimentally obtained results. We measure the promoter activity and growth kinetics for GFP reporter strains with high-temporal resolution [11]; smooth the activity signal by a polynomial fit (sixth order), and then we calculate the Pearson's coefficient to estimate the correlation of the experimental results and the values predicted by the model.

4 Application of GENIE to the PhoP/PhoQ-Pmra/PmrB Two-component System Cross-talk

Learning Phop/PhoQ-PmrA/PmrB putative architecture: We considered the crosslink of these two systems by a "forward" connection, from PhoP/PhoQ two component system to the PmrA/PmrB system and a "backward" connection in the opposite direction. Binding sites evidence (see Supplemental information Fig. 1 for the analysis of the TF evidence) and CHIP experiments show that PmrA represses the expression of *prmD* gene (i.e. "backward" connection). Given the fact that there is no binding site evidence of regulation of *pmrA* gene by the product of the *pmrD* gene we assume the "forward" connection of the systems is post-trancriptional (i.e.. PmrD protects the PmrA phosphatated form from the phosphating activity of PmrB) (see Fig. 3 for the final refined network architecture; and Supplemental information Fig. 2 for the initial reduced model).

Reactions and input concentrations of species: We translate architectures into a system of ODE's, by employing the Ingeneue library, which allows simulating the dynamic behavior of the network architecture (see Supplemental information Table 1 for the list of equations that model the final refined model).

The Fe^{3+} and Mg^{2+} concentration correspond to the "input" of the PhoP/PhoQ-PmrA/PmrB two component systems, while the values of *mgta*, *pbgP* and *pmrD* correspond to the "output" of the system. High values for the *mgta* and *pbgP* indicate the activation of the PhoP/PhoQ and PmrA/PmrB system respectively. A high value of *pmrD* shows the activation of the "forward" connection between the two systems, and a low one the activation of the "backward" connection. (see Supplemental information Table 2 for a list of expected patterns of behavior).

Learning parameters: We test the inference method by executing our GA using different configurations (i.e. population size, number of generations) and observe that both the population size and the maximum number of executions independently improve the quality of the results (see Table 1).

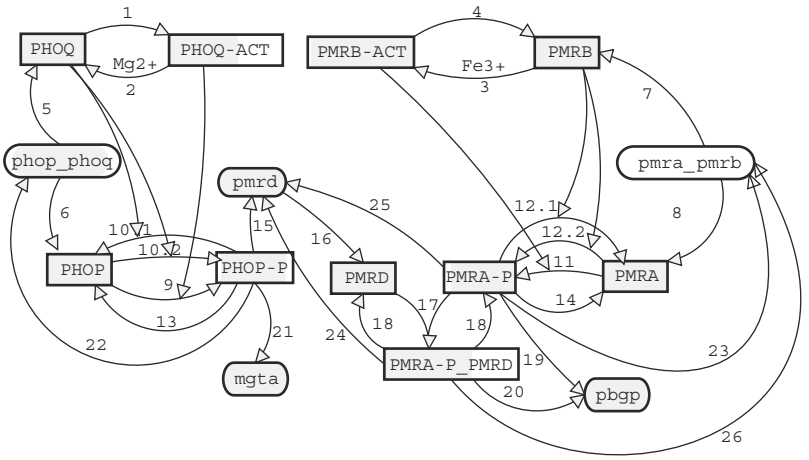

Fig. 3 Final refined model. The species interact as follows: 1/2- Low/High Mg^{2+} level favors the PHOP-ACT(ivated)/PHOP state in equilibrium. 3/4- High/Low Fe^{3+} level favors PMRB-ACT(ivated)/PMRB state in equilibrium. 5/6- phop_phoq is translated into PHOQ/PHOP proteins. 7/8- pmra_pmrb is translated into PMRB/PMRA proteins. 9- PHOP is phosphorilated (PHOP-P) by PHOQ-ACT kinase activity. 10.1- PHOP-P is desphosphorilated to PHOP by PHOQ phosphatase activity. 10.2- PHOP is phosphorilated to PHOP-P by PHOQ kinase activity. 11- PMRA is phosphorilated to PMRA-P by PMRB-ACT kinase activity. 12.1- PMRA-P is desphosphorilated to PMRA by PMRB phosphatase activity. 12.2- PMRA is phosphorilated to PMRA-P by PMRB kinase activity. 13/14- PHOP-P/PMRA-P is spontaneous desphosphorilated to PHOP/PMRA. 15- PHOP-P activates the *pmrD* transcription. 16- *pmrD* is translated into PMRD. 17- PMRD binds PMRA-P (constituting PMRD_PMRA-P) which activates *pbgP* and represses *pmrD* genes, but it is not affected by the phosphatase activity of PMRB-ACT. 18- PMRA-P_PMRD unbinds into PMRD and PMRA-P. 19/20- PMRA-P/ PMRA-P_PMRD activates the transcription of *pbgP* gene. 21/22- PHOP-P activates the transcription of *mgta*/*phoP_phoQ*. 23- PMRA-P activates the transcription of *pmrA_pmrB*. 24/25- PMRA-P_PMRD/PMRA-P represses the transcription of *pmrD*. 26- PMRA-P_PMRD activates the transcription of *pmrA_pmrB*

Table 1 Evaluation of the performance of the GA

Population size	Nbr. Generations	Evaluations	Best score	Best solution generation
50	100	5,000	0.1914	20
200	100	20,000	0.0522	9
50	250	12,500	0.0473	22

Table 2 Performance comparison (Random walk vs. GA)

Population size	Evaluations	Best score
Random Walk	100,000	>0.25
GA	1,100/12,500*	0.0473

* The GA obtained the best score after 1,100 evaluations. Heuristics like stall time can decrease the number of evaluations by indicating possible algorithm's stop condition.

Moreover, we compare the solutions obtained by the GA to the solutions obtained by the random walk approach, obtaining a score difference above 0,20 (see Table 2) [6].

Evaluating models: we initially propose a reduced model (see Supplemental information Fig. 2) designed as a test bed for our methodology: for sake of simplicity it lacks of the "forward connection" between the PhoP/PhoQ and PmrA/PmrB systems. We formalize this lack of realism by not specifying the expression of *pbgp* gene in a low Mg^{2+} and Fe^{3+} environment concentrations. The good probability measure obtained by this initial model ($p = 0.8341$) in a flexible configuration gives us a solid foundation to evolve it towards the final refined model (see Fig. 3), which reflect the "forward" connection. Along the process, we adapt the constrains to expect the expression of *pbgp* in the above conditions and relaxed the expression of *pmrD* in a low Mg^{2+} and Fe^{3+} environment concentrations (*pmrD* can be either activated by PhoP or repressed by PmrA). This final refined architecture that is more complex than the initial, thus requiring more parameters, actually obtains slightly better probability measure ($p = 0.8354$).

Furthermore, we measure the promoter activity and growth kinetics for GFP reporter strains with high-temporal resolution [11] forour distinguished genes *phoP*,*mgta* and*pmrD,* smooth the activity signal and then calculate the correlation to the predictions of the model. Pearson's coefficient indicates a correlation of 0.997 for *pmrD* gene; 0.983 for the *mgta* gene; and 0.991 for the *pbgp* gene, which reflects a highly correlated behavior between our predictions and the experimentally obtained values (see Fig. 4)

Sensitivity of the model: Our analysis of the sensitivity of the final refined network architecture for the PhoP/PhoQ-PrmA/PrmB system shows a tolerance of different magnitude order for distinct set of parameters. (see Supplemental information Fig. 3 for a detail description). Indeed, the final network architecture behaves according to the expected pattern when parameters (e.g nu_phop_mgta) take the entire biological meaningful range. Moreover, the architecture has only 3 parameters (i.e. a 4.5% of the 66 parameters) that can accept less than 25% of their entire range, what shows the robustness quality of our final refined network. (see Supplemental information Table 3 and Fig. 4)

Predicting by scanning ranges of feasible solutions: we hypothesize about the different kinetic behavior that genes co-regulated by PhoP might exhibit by scanning

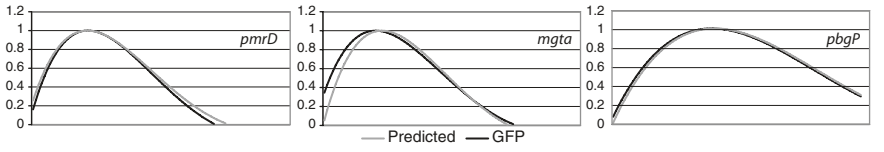

Fig. 4 Predicted and experimentally validated gene expression level. This charts reflects the high correlation between the predicted behavior (blue) and the experimentally obtained values (red) (i.e. promoter activity and growth kinetics for GFP reporter strains with high-temporal resolution)

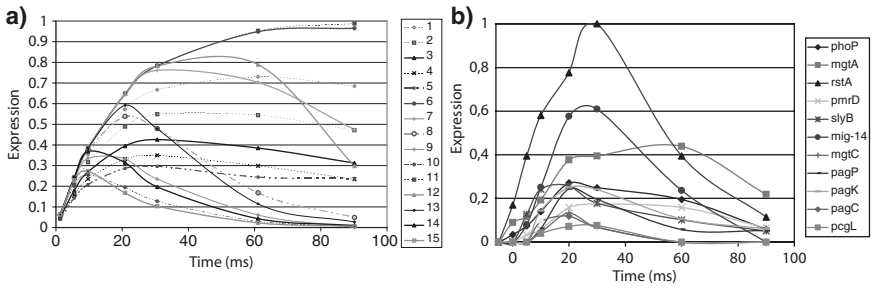

Fig. 5 Scanning parameters of feasible solutions. a) Predicted expression patterns. Multiple patterns result from scanning parameters of the single-input network motif controlled by the PhoP protein. Different symbols indicate distinct temporal order and intensity dynamics of the target genes. **b) Phop regulated genes growth kinetics for GFP.** The promoter activity and growth kinetics for GFP reporter strains with high-temporal resolution show different kinetic behavior of PhoP regulated genes

the parameters related to the *mgtA* specie (i.e. the distinguished specie that represents Phop regulated genes) in the previously learnt range of values. We observe that the simulation of the model can produce different patterns of rise time and level of expression, what is desirable for the operon of a master regulator like PhoP (see Fig. 5a and supplemental information Table 4 for the obtained results).

Validating results: we perform GFP experiments to evaluate the rise time and level of expression of PhoP regulated genes (see Fig. 5b), and calculate the correlation (c) between these experimentally obtained results to the patterns already predicted (See supplemental information Table 5 for a detail correlation results). Our analysis shows that pattern 12 predicts the dynamics of genes with early rise time and high level of transcription (i.e. $phop--c = 0.913$, $pmrD--c = 0.981$, and $mgtA--c = 0.975$); pattern 13 correlates to genes with a late rise time and low level of expression (i.e. $pagC--c = 0.917$ and $mgtC--c = 0.919$); and finally that pattern 8 predicts genes with an intermediate kinetic behavior (i.e. $rstA--c = 0.946$, $mig-14--c = 0.922$, $pcgL--c = 0.932$)

5 Concluding remarks

The experiments and simulations for the genetic regulatory network PhoP/PhoQ-PmrA/PmrB allowed us to extract several conclusions about the method shown in this work and the network under study: *(1) GENIE predicts interactions that explain experimentally observed behaviors*, the final refined architecture let predict the *in vivo* observed interaction between the two-component systems PhoP/PhoQ and PmrA/PmrB; *(2) PmrA/PmrB and PhoP/PhoQ constitute a robust and flexible genetic network*, our final refined model satisfies all of the constrains and only 4.5% of its parameters are limited to accept values that cover 25% of their biological meaningful range; *(3) GA approach is adequate for inferring regulatory genetic*

networks, the heuristic produces a better proportion of feasible solutions and better numerically evaluated results (i.e. scores) for the predictions which are highly correlated to experimental values.

Finally, we would like to remark that the decisions regards the architecture enhancement (i.e. adding or not new elements) are based on the definition of conditions (i.e. realness, robustness and flexibility) to be fulfilled by the models, all of them satisfied by our final refined model for the PhoP/PhoQ-PmrA/PmrB genetic regulatory network. Our approach to model regulatory genetic networks provides a framework to explore genetic regulatory networks, including biochemical elements (i.e. different equations to model the reactions), biological (i.e. constrains imposed to the networks), and computational (i.e. simulations and a learning strategy that tackles the high dimensional search space).

Tables and supplemental figures are available online at http://gps-tools2.wustl.edu/NICSO2007/Appendix_NICSO2007.pdf

Acknowledgments This work was partly supported by the Spanish Ministry of Science and Technology under Project BIO2004-0270-E, and I.Z. is also supported by and by Howard Hughes Medical Institute. O.H. acknowledges the doctoral MAEC- AECI fellowship.

References

1. Brenner, S., *Genomics. The end of the beginning.* Science, 2000. 287(5461): p. 2173–4.
2. Kærn, M., *Regulatory dynamics in engineered gene networks.*, in *4th International Systems Biology Conference*. 2003: Washington University, St. Louis.
3. Li, C. and W.H. Wong, *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.* Proc Natl Acad Sci U S A, 2001. 98(1): p. 31–6.
4. McAdams, H.H. and A. Arkin, *It's a noisy business! Genetic regulation at the nanomolar scale.* Trends Genet, 1999. 15(2): p. 65–9.
5. Meir, E., et al., *Ingeneue: a versatile tool for reconstituting genetic networks, with examples from the segment polarity network.* J Exp Zool, 2002. 294(3): p. 216-51.
6. Zwir, I., P. Traverso, and E.A. Groisman. *Semantic-oriented analysis of regulation: the PhoP regulon as a model network*. in *Proceedings of the 3rd International Conference on Systems Biology (ICSB)*. 2003. St. Louis, USA.
7. Hoch, J.A., *Two-component and phosphorelay signal transduction.* Curr Opin Microbiol, 2000. 3(2): p. 165–70.
8. Milo, R., et al., *Network motifs: simple building blocks of complex networks.* Science, 2002. 298(5594): p. 824–7.
9. Rubio-Escudero, C., et al. *Modeling Genetic Networks: Comparison of Static and Dynamic Models*. in *Evolutionary Computation,Machine Learning and Data Mining in Bioinformatics*. 2007. Valencia, Spain: Springer.
10. Batchelor, E. and M. Goulian, *Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system.* Proc Natl Acad Sci U S A, 2003. 100(2): p. 691–6.
11. Zwir, I., O. Harari, and E.A. Groisman, *Gene promoter scan methodology for identifying and classifying coregulated promoters.* Methods Enzymol, 2007. 422: p. 361–85.
12. Salgado, H., et al., *RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12.* Nucleic Acids Res, 2004. 32(Database issue): p. D303-6.

13. Alon, U., *An introduction to System Biology*. Mathematical and Computational Biology Series, ed. C. Hall/CRC. 2007, London: CRC Press, Taylor & Francis Group.
14. Mitchell, T.M., *Machine learning*. 1997, New York: McGraw-Hill. xvii, 414.
15. Meir, E., et al., *Robustness, flexibility, and the role of lateral inhibition in the neurogenic network.* Curr Biol, 2002. 12(10): p. 778–86.