



HAL
open science

Integrating heterogeneous knowledges for understanding biological behaviors: a probabilistic approach

Jérémie Bourdon, Damien Eveillard, Samuel Gabillard, Theo Merle

► To cite this version:

Jérémie Bourdon, Damien Eveillard, Samuel Gabillard, Theo Merle. Integrating heterogeneous knowledges for understanding biological behaviors: a probabilistic approach. 2007. hal-00199362

HAL Id: hal-00199362

<https://hal.science/hal-00199362v1>

Preprint submitted on 18 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Integrating heterogeneous knowledges for understanding biological behaviors: a probabilistic approach

J eremie Bourdon^{1*}, Damien Eveillard¹, Samuel Gabillard¹, Theo Merle²

1. LINA - CNRS and University of Nantes, Nantes, France

2. Ecole Normale Sup erieure (ENS) - Cachan (Bretagne), Bruz, France

*email: jeremie.bourdon@univ-nantes.fr

Abstract

Despite recent molecular technique improvements, biological knowledge remains incomplete. Reasoning on living systems hence implies to integrate heterogeneous and partial informations. Although current investigations successfully focus on qualitative behaviors of macromolecular networks, others approaches show partial quantitative informations like protein concentration variations over times. We consider that both informations, qualitative and quantitative, have to be combined into a modeling method to provide a better understanding of the biological system. We propose here such a method using a probabilistic-like approach. After its exhaustive description, we illustrate its advantages by modeling the carbon starvation response in *Escherichia coli*. In this purpose, we build an original qualitative model based on available observations. After the formal verification of its qualitative properties, the probabilistic model shows quantitative results corresponding to biological expectations which confirm the interest of our probabilistic approach.

1 Introduction

The last decade has seen great successes in macromolecular network modeling. In particular, qualitative methods appear today as well-adapted for reasoning on biological systems, despite the current lack of quantitative informations (de Jong, 2002). Thus most of interesting and investigated knowledges concern local informations such as gene-gene or gene-protein interactions. They allow to build networks like on Figure 1 (A), that model the global qualitative behavior of a biological system. However, other experiments illustrated Figure 1 (B) give insights about various partial quantitative knowledges. They emphasize both molecular concentration variations and time-series. These two related kinds of partial quantitative information, *i.e.*, time and concentration, are well studied by other experiments (Wolfe, 2005) and reflect as well the overall system behavior. Both informations, qualitative and quantitative, have hence to be combined into a modeling method for giving a better understanding of the biological system. Due to the lack of quantitative informations, we propose a modeling approach that (i) spreads partial local informations through the qualitative network and (ii) gives insights about global behaviors. Probabilistic approaches are well adapted for bringing complementary quantitative or semi quantitative knowledges into a qualitative model. Among them, we suggest an original toll based approach that predicts various molecular productions combining both qualitative and partial quantitative knowledges. After an overview of our probabilistic approach (Sec. 2), we propose here to apply it on gene regulatory model of the carbon starvation response in *Escherichia coli*. In this purpose, we (Sec. 3.1) build a model based on a novel qualitative abstraction, validate its behavior using a formal verification approach, which (Sec. 3.2) allows us to accurately apply our probabilistic method. Such a protocol emphasizes several biological insights of interest.

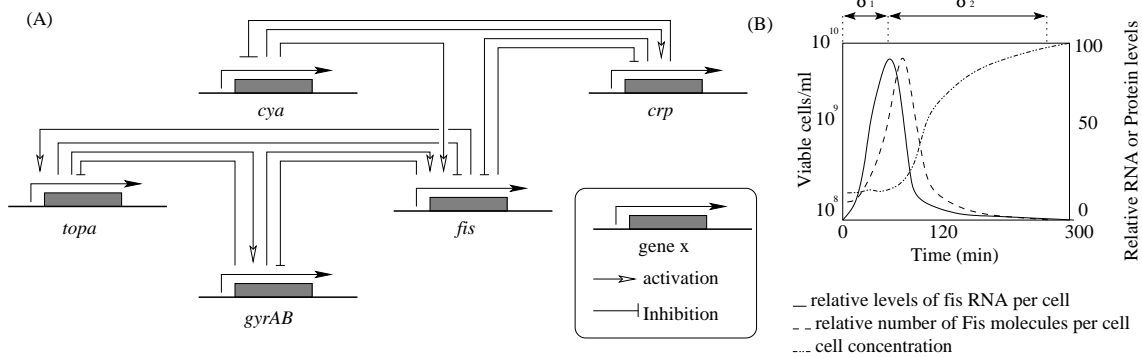


Figure 1: Biological informations concerning *Escherichia coli* carbon starvation system. (A) represents interactions between genes involved in the regulatory network (adapted from (Ropers et al., 2006)). (B) shows quantitative variations of macromolecules of interest (based on (Ball et al., 1992)). Note the linear relationship between *fis* RNA and Fis protein productions.

2 Method

2.1 Biological system formalization

We consider biological networks as graphs that show transitions between various components of the system. Each transition is related to variations of characteristic quantities of the system and produces its own impact on the whole system behavior. In a gene regulatory network, a qualitative graph arrow is associated with a production or consumption of the corresponding protein.

In order to abstract qualitative biological behaviors, we represent a gene regulatory network by a qualitative graph where each state stands for a qualitative variation of a gene activity. We focus on the macromolecular transformation derivative, which is more tractable to model detailed macromolecular concentration variations. As illustration, following interactions describe the fact that (i) gene x activates gene y and (ii) x represses y :

$$(i) \quad x \longrightarrow y^+ \quad (ii) \quad x \longrightarrow y^-$$

Such a representation implies that gene x produces protein X that activates gene y . Thus (i) and (ii) represent respectively an overall increase of Y protein production and an overall decrease of Y . Note that such an abstraction neglects post-transcriptional regulations which is particularly unappropriated for modeling eukaryote gene regulatory network.

This biological abstraction allows us to model various qualitative interactions. Considering that a gene x activity is summarized by two qualitative states x^+ and x^- , y activation by x might be described by the set of rules and its corresponding transitions:

$$\{x^+ \Longrightarrow y^+\} \wedge \{x^- \Longrightarrow y^+\}$$

A peak of gene x activity that activates y is represented by:

$$\{x^+ \Longrightarrow y^+\} \wedge \{x^- \Longrightarrow \emptyset\}$$

A minimal activity of gene x that activates y is symbolized by:

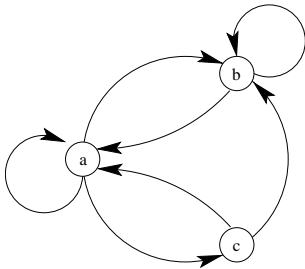
$$\{x^+ \implies \emptyset\} \wedge \{x^- \implies y^-\}$$

Gene y repressions by a gene x activity are modeled using similar rules that imply transitions toward y^- . Such an abstraction gives the opportunity to focus on qualitative behaviors. Reasoning on quantities associated with qualitative rules allows us to emphasize quantitative states of the system despite concurrent qualitative rules.

2.2 Graph model and quantities

We make the assumption that the biological system is associated with several *quantities* q_1, \dots, q_k that represent the current state of the system. For illustration, these quantities represent protein concentrations, or other non trivial quantities such as the number of times a particular pathway is taken by the living system. Studying the behavior of biological systems hence consists in understanding the evolution of these quantities. Note here that such quantities may not be experimentally measurable. Since the last decade, biological behaviors have been often described by qualitative graphs that abstract different component variations within the system. In our model, we consider that each transition of this qualitative graph implies a potential variation on each quantity. Here we propose a method that focusses on these quantities.

We consider two types of quantities. Some quantity variations are additive whereas others are multiplicative. (i) Each transition from i to j is associated with a real number $\delta_{(i,j)}$, the quantity q is additive if the quantity $q = x$ before the transition becomes $x + \delta_{(i,j)}$ after the transition. (ii) Each transition from i to j is associated with a strictly positive real number $\lambda_{(i,j)}$, the quantity q is multiplicative if the quantity $q = x$ before the transition becomes $x\lambda_{(i,j)}$ after the transition. Each quantity q is thus associated with a matrix C_q in which the element at position (i, j) is the contribution of the transition from i to j . We are looking for understanding the typical behavior of given additive or multiplicative quantities after a given time. These behaviors are controlled by an accumulation of small contributions. For illustration, we consider the following graph.



$$C_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 0.9 & 1.2 & 0.9 \\ 0.9 & 1.2 & 1 \\ 0.9 & 1.2 & 1 \end{pmatrix}$$

We consider two distinct quantities q_1 and q_2 . Their associated cost matrices are respectively C_1 and C_2 . q_1 counts the number of times that transition $c \rightarrow b$ is taken. q_2 models the concentration of a product. It increases by 20% for every transitions pointing to b and decreases by 10% for all other transitions (*i.e.*, abstraction of the product natural degradation). Note here that, by convention, a cost of 0 (or respectively 1 for multiplicative quantities) has been assigned to the non existing transitions $b \rightarrow c$ and $c \rightarrow c$. Thus, as illustration, given initial quantities $q_1 = 0$, $q_2 = 1$ and a trajectory, **abacbacacba**, their values become $q_1 = 2$ and $q_2 = 0.826$.

2.3 Probabilistic model

On applicative purpose, we are interested in the values of all quantities for a random trajectory. Transitions impact differently to the global system behavior. We assume here that each transition possesses its own probability. Thus, at each step, one chooses randomly between all the transitions that leave the current state. The sum of the probabilities associated with all edges that leave a given state is 1.

For fixed probabilities at all steps, this model is a weighted Markov chain. Nevertheless, probabilities may vary, showing a behavior controlled by a dynamical system (see (Vallée, 2001) for a further details about dynamical sources). This model is hence quite general and particularly accurate for theoretical studies since it includes simple probabilistic models such as Markov chains, Hidden Markov chains or trickiest models that handle unbounded correlations (*i.e.*, the choice made at one step influes on all next choices). In this last case, generating operators play the role of transition probabilities. For this reason, we assume our model as a *graph with dynamical sources* (or GDS model). The GDS is called *nice* if it satisfies some classical conditions of the theory of Markov chains and dynamical sources (namely, the graph is strongly connected and aperiodic and all the dynamical systems are topologically mixing and possess expansive branches). We consider the transition matrix $T = (t_{i,j})$ of the qualitative graph in which the element (i, j) is the generating operator relative to the transition from i to j . Reasoning on system properties implies to focus on quantities asymptotic properties. These mathematical properties are well studied in both theories of Markov chains and dynamical sources (Bourdon and Vallée, 2006).

2.4 Typical behaviors

Previous theoretical assumptions allow us to emphasize typical characteristics of quantities. More precisely, for a given GDS model, we provide results for the mean, the variance and the limit distribution. The following theorem synthesizes our results.

Theorem 1 *Let \mathcal{M} by a nice GDS model with transition matrix T and q a quantity with cost matrix C . Let Q_n be the random variable equal to the quantity q after n steps of the GDS model \mathcal{M} .*

- (i) *if q is an additive quantity, Q_n follows asymptotically (when n tends to ∞) a Normal law with mean and variance*

$$\mathbb{E}[Q_n] = \alpha_1 n + O(1) \quad \text{Var}[Q_n] = \alpha_2 n + O(1),$$

where $\alpha_1 = \lambda'(1)$ and $\alpha_2 = \lambda''(1) + \lambda'(1) - \lambda''(1)^2$ express by means of derivatives of the dominant eigenvalue of the matrix $\mathbb{A}(u)$ defined by $\mathbb{A}_{i,j}(u) = T_{i,j}u^{C_{i,j}}$.

- (ii) *if q is a multiplicative quantity, Q_n follows asymptotically (when n tends to ∞) a log-Normal law with mean and variance*

$$\mathbb{E}[Q_n] = \beta_1 \gamma_1^n + o(\Lambda_1^n) \quad \text{Var}[Q_n] = \beta_2 \gamma_2^n + o(\Lambda_2^n),$$

where $\gamma_1 = \lambda(e)$ and $\gamma_2 = \max(\lambda(e^2), \lambda(e)^2)$ express by means of the dominant eigenvalue of the matrix $\mathbb{A}(u)$ defined by $\mathbb{A}_{i,j}(u) = T_{i,j}u^{\ln C_{i,j}}$. β_1 and β_2 are constants corresponding to the dominant eigenvectors of $\mathbb{A}(e)$ and $\mathbb{A}(e^2)$. The error terms Λ_1 and Λ_2 verify $\Lambda_1 < \gamma_1$ and $\Lambda_2 < \gamma_2$.

Sketch of proof. See (Bourdon and Eveillard, 2007) for a complete proof of this theorem. Considering the additive case is sufficient, if q is a multiplicative quantity, then $\log q$ is an additive one and it is easy to obtain the results of (ii). The study involves several classical elements on the average-case analysis theory such as generating functions. Let m be the number of states of the GDS model and P_0 be a probability vector whose element i is the probability that initial state is state i . Since we consider asymptotic cases, this initial vector does not have any influence on the result. The generating function $Q(z, u)$ defined as

$$Q(z, u) = \sum_{n \geq 0} P_0 z^n \mathbb{A}(u)^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

permits to study the quantities of interest. Indeed,

$$\mathbb{E}[Q_n] = \frac{\partial}{\partial u} [z^n] Q(z, u) |_{u=1}.$$

For a nice GDS model, the matrix $\mathbb{A}(u)$ admits a dominant eigenvalue in a neighbourhood of $u = 1$ and decomposes as $\mathbb{A}(u) = \lambda(u)\mathbb{P}(u) + \mathbb{N}(u)$, where $\lambda(u)$ is the dominant eigenvalue, $\mathbb{P}(u)$ is the dominant eigenvector and $\mathbb{N}(u)$ is associated to the remainder of the spectrum (and is thus orthogonal to $\mathbb{P}(u)$). Consequently, for large n , one has

$$\mathbb{A}(u)^n \approx \lambda(u)^n \mathbb{P}(u).$$

It is easy to obtain a formula for the mean. The study of the variance follows similar assumptions and involves the second derivative of $Q(z, u)$. Finally, the limit law is obtained by applying Hwang's (Hwang, 1996) general result on bivariate generating functions. \square

Supplementary results have been obtained but they are not detailed here. Among others, we calculate the probability for a quantity to attain a given threshold before a given time t (it generalizes the hitting probability, common in the Markov chain theory) and the joint law of several quantities. Most on our computations extends in same cases when the graph is not strongly connected or aperiodic.

2.5 A typical biological study

Previous theories allow us to reason on system quantitative properties but provide as well the core of a dedicated software¹. This software works on GDS models with fixed probabilities and represents an accurate tool for simulating macro-molecular networks. As inputs, it needs a graph (or

¹POGG: Probabilities On Genetic Graphs is available at <http://www.sciences.univ-nantes.fr/lina/bioserv/POGG/>

a qualitative graph in a better case) and the cost matrix of quantities of interest. GDS probabilities are unknown or partially unknown which make almost impossible to predict the quantitative impact of an interaction on the system behavior. Nevertheless, experimental results of quantity behaviors, like protein concentrations, are known. POGG uses such an information and adopts a reverse engineering point of view. Previous theoretical results give some (in)equalities that relate unknown significance probabilities to experimental measures (of part or all of the quantities). POGG uses general techniques of local search theory, such as Tabu search, for estimating the impact of a local interaction in the whole biological system. The determined model gives us the opportunity to predict the behavior of others quantities. Note that the software also provides supplementary informations such as an approximation of the hyper-volume of models that are consistent with the measures. This information helps to decide whether a new measure is informative or not, using a simple comparison between different volumes.

3 Results

(Ropers et al., 2006) models the growth phase transition of a bacteria after a nutritional stress. In particular, the model shows the abandon of exponential growth state to a more stationary growth during a carbon starvation stage. Their qualitative results are relevant with experimental knowledges, which allows us to consider the model as an appropriate benchmark for our modeling approach. Furthermore, macromolecules that interact within the model are well studied. It gives us various partial quantitative informations that have to be introduced into the qualitative model.

3.1 Carbon starvation response in *Escherichia coli*: gene regulatory network and qualitative rules validation

We consider similar hypotheses to those exposed in (Ropers et al., 2006) and propose a new graph that represents identical qualitative behaviors of bacterial responses after a nutritional stress. For illustration and using abstractions described in Sec. 2.1, we detail in Figure 2 one particular biological component: *crp* gene. The gene *crp* is controlled by two promoters that are both repressed by Fis protein (González-Gil et al., 1998). Following assumptions from (Ropers et al., 2006), we omit the negative control of *crp* and summarize the impact of cAMP metabolite using rules that imply Cya and Crp protein and carbon starvation signal as well (Harman, 2001).

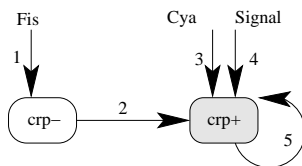


Figure 2: Qualitative representation of *crp* interaction with others genes and carbon starvation signal. 1 represents the repression of *crp* by Fis protein production. 2 and 5 are transitions for the basal synthesis rate which plays an important role during the exponential growth phase. Combination of concurrent rules 3 and 4 synthesizes the *crp* activation via cAMP metabolite.

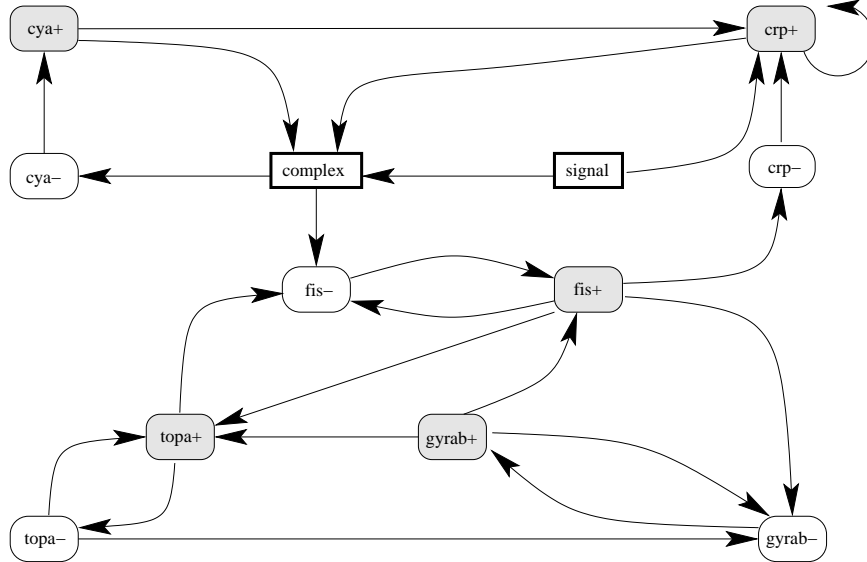


Figure 3: Qualitative graph representing genes regulatory network of carbon starvation response in *Escherichia coli*. Signal represents the input module that indicates carbon starvation condition.

We use a similar approach for describing each biological component of the gene regulatory network. Figure 3 represents the corresponding qualitative graph. Our aim is to demonstrate advantages of our probability approach. Therefore we will not detail here biological assumptions that have been used for building the model. See (Ropers et al., 2006) for exhaustive hypotheses. Before further *in silico* investigations, the model has to be validated. It is the *sine qua non* condition for applying a probabilistic approach. Although probabilities can be estimated using an appropriate optimization, our confidence in such parameters is related to the ability of the model to reproduce appropriate qualitative behaviors of the biological system (*i.e.*, various kinds of qualitative models can produce similar quantitative results). Using the symbolic model-checker of BIOCHAM (Calzone et al., 2006, Fages et al., 2004), we thus check qualitative rules in order to verify their consistency with experimental understandings. In particular, (Browning et al., 2004) shows an antagonistic relationship between *fis* and *crp* activities. For validating the model, we are able to ask positive queries (*i.e.* queries where the expecting answer is true) such as $(fis^+ \wedge \neg fis^-) \implies (crp^- \wedge \neg crp^+)$. We are as well able to ask negative queries (*i.e.* queries where the expecting answer is false), such as $(fis^+ \wedge \neg fis^-) \equiv (crp^- \wedge \neg crp^+)$. Using this formal verification on the qualitative model, we successfully check other biological properties like the relationship between the carbon starvation signal and *crp* expression (Ishizuka et al., 1994) as well with *cya* activities (Ball et al., 1992).

3.2 Probabilistic results

Therefore, we have at our disposal an accurate qualitative graph (Figures 3) and quantitative informations (Figure 4 (A)) that belong to the same bacterial system. Our modeling approach exploits such informations and predicts probabilities on graph transitions using a local search algorithm. In practice, we take into account the fact that Fis concentration is multiplied by 10 in

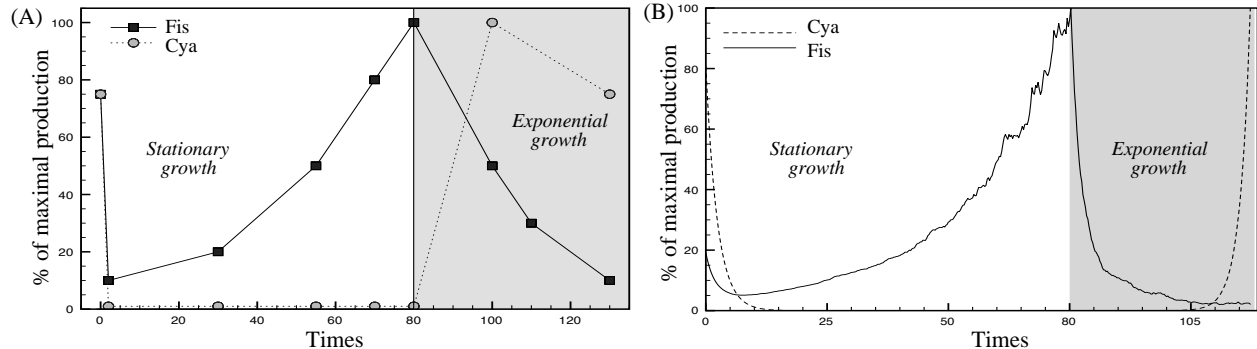


Figure 4: Summary of informations used and produced by the probabilistic approach. (A) shows variations of Fis and Cya protein concentrations in function of growth phases (Ball et al., 1992, Notley-McRobb et al., 1997). Two Fis variations during stationary growth have been used for estimating probabilities associated with qualitative transitions from Figure 3. It allows to reproduce quantitative behaviors of Cya and Fis during both growth phases (B).

80 minutes during the stationary growth phase. We assume Fis concentration q_{Fis} as a multiplicative quantity (see Sec. 2.2). Therefore it increases by 20% for each transition pointing to fis^+ , decreases by 20% for all transitions pointing to fis^- and decreases by 10% for natural degradation passing through all other transitions. We estimate the Fis quantity at time 80, $q_{Fis} \approx 10 \cdot \gamma^{1600}$ with $\gamma \approx 1.001$ (1600 corresponding to the number of steps performed by the model during 80 minutes, this number is established by considering Cya natural degradation during the first 2 minutes). Comparing this numerical value with constants from Theorem 1, we get a constraint that relates probabilities with a measure on the system. Local search methods allow to find a suitable probability matrix used for simulations.

Figure 4 (B) shows the estimated variation of Cya and Fis protein in function of growth phases. During the stationary growth, our model accurately predicts a decrease followed by an increase of Fis protein production (Ball et al., 1992). It emphasizes the ability of our approach to spread partial quantitative knowledges through the qualitative network. Despite a quantitative estimation using two measures during the stationary phase, interestingly, our model predicts efficiently the Fis concentration decrease during the exponential phase. This model artifact represents a quantitative emerging property of the biological system which gives insights about global behaviors.

Estimative Cya protein variations are as well consistent with experiments during stationary phase. However, despite an appropriate increase during the beginning of the exponential phase, the Cya production does not follow an expected peak (Notley-McRobb et al., 1997). It might reflect a shortcoming or a missing qualitative transition that represses the *cya* gene. We consider such an information as a guidance for future models or further experiments that might focus on *cya* gene regulations.

A close attention to estimated probabilities gives results that are related with the quantitative sensitivity of the model. More precisely, an estimation of the hyper-volume associated with the model emphasizes whether a new measure is informative or not. Our model shows that the probability associated with $topa^+$ and fis^- transition is highly constrained in order to maintain an

overall consistency between heterogeneous informations. This transition is a shortcut adapted from (Ropers et al., 2006) for representing DNA supercoiling effect on *fis* gene expression. Experiments suggest that *fis* is involved in fine tuning of the homeostatic control of DNA supercoiling (Schneider et al., 2000). A small change in DNA supercoiling drastically affects the *fis* expression. This information is accurate with our estimative impact of this transition on the global system behavior.

4 Discussion

Recent fruitful probabilistic approaches has been developed for studying gene regulatory networks (Shmulevich et al., 2002, Zhou et al., 2004, Kim et al., 2002). These approaches add probabilities to an already defined deterministic model. It gives the opportunity to study probability variation impacts and eventually to determine probability sets that accurately represent experiments. Knowing the transition probability graph, the major issue of these approaches is to compute the asymptotic (stationary) distribution and to reason on it.

Our original method appears as a complementary approach that adds new natural informations in a general probabilistic graph. It gives the opportunity to reason on emerging system properties by focusing on asymptotic properties of the probabilistic model. We prove that their asymptotics are related to natural constants on a weighted transition matrix. The proposed method allows to design constraints between probabilities and observations, which gives the opportunity to deal with unknown transition probabilities. Therefore our results are adapted to a large class of probabilistic models and their integration within a more general framework such as PBN and Bayesian networks seems promising.

The number of biological details at disposal defines the model abstraction level which conduces to choose an accurate biological abstraction. It is more or less discrete in function of the number of qualitative states. Our probabilistic-like technique is able to combine quantitative informations with various qualitative abstractions of biological systems, *i.e.*, from boolean to PDE network (de Jong, 2002). Therefore, our method emphasizes a convenient flexibility for analyzing biological systems because it presents major advantages for integrating heterogeneous knowledges such as those that constitute the *Escherichia coli* starvation system.

During this study, various biological models were elaborated. After probability optimization, most of them give relevant quantitative simulation results. Nevertheless, they remain inconsistent with their ability to reproduce the whole set of expecting experimental behaviors. It hence confirms the support of reasoning rather than just simulating that prevents to validate the model using few simulations. Furthermore, it emphasizes the need for an appropriate qualitative validation of model behaviors prior to apply our probabilistic technique. In this purpose, the biological system has been described using a set of original qualitative rules. It allows us to use a formal verification technique in a qualitative validation requirement. Therefore our technique appears as a natural extension of regular qualitative modeling approaches for extending robust qualitative models toward quantitative properties.

References

- Ball, C. A., Osuna, R., Ferguson, K. C., and Johnson, R. C., 1992. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J Bacteriol*, **174**(24):8043–56.
- Bourdon, J. and Eveillard, D., 2007. Toll based measures for dynamical graphs. Technical Report q-bio/0702060, arXiv.
- Bourdon, J. and Vallée, B., 2006. Pattern matching statistics on correlated sources. In *Proc. of LATIN'06*, number 3887 in LNCS, pages 224–237.
- Browning, D. F., Beatty, C. M., Sanstad, E. A., Gunn, K. E., Busby, S. J. W., and Wolfe, A. J., 2004. Modulation of CRP-dependent transcription at the *Escherichia coli* acsP2 promoter by nucleoprotein complexes: anti-activation by the nucleoid proteins fis and ihf. *Mol Microbiol*, **51**(1):241–54.
- Calzone, L., Fages, F., and Soliman, S., 2006. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, **22**(14):1805–7.
- de Jong, H., 2002. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, **9**(1):67–103.
- Fages, F., Soliman, S., and Chabrier-Rivier, N., 2004. Modelling and querying interactions networks in the biochemical abstract machine BIOCHAM. *Journal of Biological Physics and Chemistry*, **4**:64–73.
- González-Gil, G., Kahmann, R., and Muskhelishvili, G., 1998. Regulation of crp transcription by oscillation between distinct nucleoprotein complexes. *EMBO J*, **17**(10):2877–85.
- Harman, J. G., 2001. Allosteric regulation of the cAMP receptor protein. *Biochim Biophys Acta*, **1547**(1):1–17.
- Hwang, H. K., 1996. Large deviations for combinatorial distributions. I. central limit theorems. *Annals of Applied Probability*, **6**(1):297–319.
- Ishizuka, H., Hanamura, A., Inada, T., and Aiba, H., 1994. Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the crp gene. *EMBO J*, **13**(13):3077–82.
- Kim, S., Li, H., Dougherty, R., Cao, N., Chen, Y., Bittner, M., and Suh, E. B., 2002. Can markov chain models mimic biological regulation? *Journal of Biological Systems*, **10**(4):431–445.
- Notley-McRobb, L., Death, A., and Ferenci, T., 1997. The relationship between external glucose concentration and cAMP levels inside *Escherichia coli*: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology*, **143**:1909–1918.
- Ropers, D., de Jong, H., Page, M., Schneider, D., and Geiselman, J., 2006. Qualitative simulation of the carbon starvation response in escherichia coli. *BioSystems*, **84**(2):124–52.
- Schneider, R., Travers, A., and Muskhelishvili, G., 2000. The expression of the *Escherichia coli* fis gene is strongly dependent on the superhelical density of DNA. *Mol Microbiol*, **38**(1):167–75.
- Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W., 2002. Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, **18**(2):261–274.
- Vallée, B., 2001. Dynamical sources in information theory: fundamental intervals and word prefixes. *Algorithmica*, **29**(1-2):262–306.
- Wolfe, A. J., 2005. The acetate switch. *Microbiol Mol Biol Rev*, **69**(1):12–50.
- Zhou, X., Wang, X., Pal, R., Ivanov, I., Bittner, M., and Dougherty, E. R., 2004. A bayesian connectivity-based approach to constructing probabilistic gene regulatory networks. *Bioinformatics*, **20**(17):2918–2927.