

# DEMSIM: a discrete event based mechanistic simulation platform for gene expression and regulation dynamics

Madhukar S. Dasika, Anshuman Gupta, Costas D. Maranas\*

Department of Chemical Engineering, The Pennsylvania State University, 112A Fenske Laboratory, University Park, PA 16802, USA

Received 30 April 2004; received in revised form 20 July 2004; accepted 22 July 2004

Available online 18 September 2004

## Abstract

In this paper, a discrete event based mechanistic simulation platform DEMSIM is developed for testing and validating putative regulatory interactions. The proposed framework models the main processes in gene expression, which are transcription, translation and decay processes, as stand-alone modules while superimposing the regulatory circuitry to obtain an accurate time evolution of the system. The stochasticity inherent to gene expression and regulation processes is captured using Monte Carlo based sampling. The proposed framework is applied to the extensively studied *lac* operon system, the SOS response system and the *araBAD* operon system of *Escherichia coli*. The results for the *lac* gene system demonstrate the simulation framework's ability to capture the dynamics of gene regulation, whereas the results for the SOS response system indicate that the framework is able to make accurate predictions about system behavior in response to perturbations. Finally, simulation studies for the *araBAD* system suggest that the developed framework is able to distinguish between different plausible regulatory mechanisms postulated to explain observed gene expression profiles. Overall, the obtained results highlight the effectiveness of DEMSIM at describing the underlying biological processes involved in gene regulation for querying alternative regulatory hypotheses.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** Gene networks; Stochasticity; Discrete event simulation

## 1. Introduction

Gene expression is the primary method through which a living organism processes the information stored in its DNA to form all functional cellular components. Elucidation of regulation mechanisms has been an important challenge for understanding the fundamental organization and functioning of biological systems. To date, many data-driven approaches have been developed that use DNA microarray data to unravel the underlying network of genetic interactions. These broadly include clustering approaches (Spellman, 1998; Ang et al., 2001; Helmann et al., 2003), Boolean networks (Akutsu and Miyano, 2000; Ideker et al., 2000), differential equations (Chen et al., 1999; D'haeseleer

et al., 1999; Hoon et al., 2003; Dasika et al., 2004), Bayesian networks (Friedman et al., 2000) and neural networks (Vohradsky, 2001). We refer to these class of methods as “top-to-bottom” approaches as they attempt to elucidate the complex web of DNA, protein and metabolite interactions by using “snap-shot” data (top layer) to infer the inner workings (bottom layer). Alternatively, as illustrated in Fig. 1, “bottom-to-top” approaches rely on detailed mechanistic descriptions of the underlying molecular processes to construct a predictive model of interaction parameterized to comply with experimental observations. In this paper, we introduce such a “bottom-to-top” simulation platform that accounts for the mechanistic detail of various processes underlying gene expression and regulation.

The fundamental processes that govern the flow of information from the DNA to a working component (proteins, ribosomes, etc.) in a cell are transcription and

\*Corresponding author. Tel.: +814-863-9958; fax: +814-865-7846.  
E-mail address: [costas@psu.edu](mailto:costas@psu.edu) (C.D. Maranas).

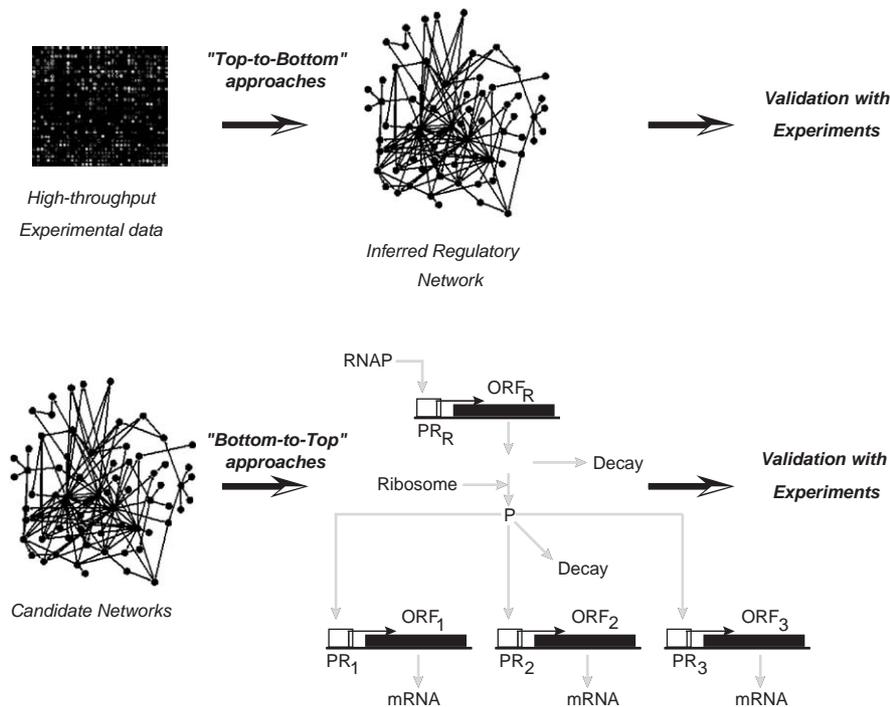


Fig. 1. Alternative approaches employed to investigate regulatory networks. The “top-to-bottom” approach uses snap-shot experimental data such as those obtained from microarrays to explain the inner workings of the regulatory networks. On the other hand, “bottom-to-top” approaches take into account the molecular mechanisms of the underlying processes to develop a predictive model.

translation. These processes, coupled with decay mechanisms and various regulatory interactions, largely control the level of gene expression in a cell. Many researchers have attempted to model the gene regulation process by abstracting these underlying processes using ordinary differential equations. Specifically, Agger and Nielsen (1999) modeled the regulation dynamics of a genetic system using equilibrium kinetics, Cheng et al. (1999, 2000) developed a model to describe the inhibition of *lac* operon by triplex forming oligos. Shea and Ackers (1985) developed a model for the O<sub>R</sub> control system of bacteriophage  $\lambda$ . Other differential equation models include efforts by Goutsias and Kim (2004) and Hatzimanikatis and Lee (1999).

The research cited above utilizes differential equations to represent systems that are essentially discrete in nature. Dynamics of gene expression and regulation in many cases involve interactions between relatively small numbers of molecules. For example, the number of available RNA polymerase molecules is estimated to be approximately 35 in *E. coli*, while the number of available ribosomes is estimated to be approximately 350 (Kierzek et al., 2001). In such discrete systems, rates of reaction are no longer deterministic; the reactions occur in a stochastic and discontinuous fashion, rendering the differential equation representation only a coarse approximation (Carrier and Keasling, 1997). Under these conditions, stochastic fluctuations become

important resulting in significant variability in the number of molecules of the species around their average value. Many experimentally verified instances of stochastic variability of genetic systems have been reported in literature. For example, the expression of plasmids containing *araBAD* promoter at subsaturating levels of inducer revealed the existence of both induced and uninduced cells in the population (Siegele and Hu, 1997). Elowitz and Leibler (2000) have reported that the expression of a synthetically constructed oscillating network exhibits noisy behavior. On the theoretical/computational front, Monte Carlo based simulation methods have been employed by a number of researchers for studying the stochastic evolution of genetic systems (McAdams and Arkin, 1997; Arkin et al., 1998; Kepler and Elston, 2001; Kastner et al., 2002; Kurata et al., 2003). These methods largely employ the stochastic simulation algorithm developed by Gillespie (1976, 1977; Gibson and Bruck, 2000). Alternatively, Carrier and Keasling proposed a Monte Carlo based algorithm to study the expression of prokaryotic systems (Carrier and Keasling, 1997, 1999).

A systems engineering view reveals that gene expression dynamics are governed by processes that are essentially event driven, i.e. many events have to take place in a predetermined order with uncertain start and execution times to accomplish a certain task. Fig. 2 highlights the many parallels between gene expression

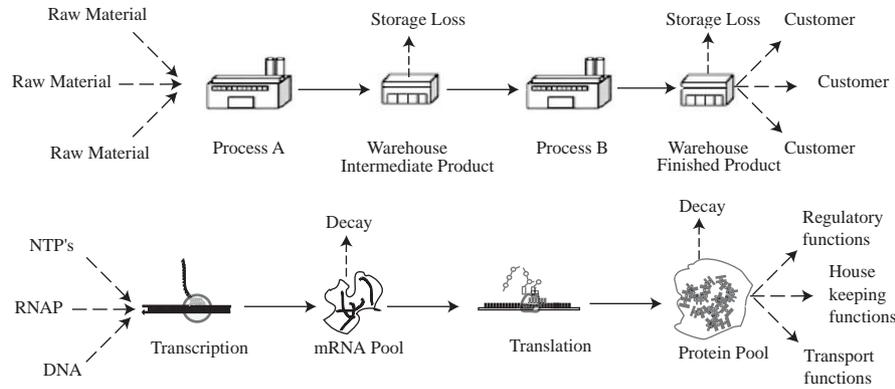


Fig. 2. As in manufacturing processes, gene expression is also event driven implying that many events have to take place in a predetermined order to accomplish a certain task.

and manufacturing systems. In analogy to a manufacturing facility which produces a certain amount of finished product at a particular time with a certain probability, the transcription process produces mRNA transcripts with probability determined by the cellular environment and availability of required components. Similarly, accumulating mRNA and protein levels in the cell are akin to product inventory held in warehouses in a manufacturing system. Motivated by the numerous parallels between these two seemingly different settings, we propose the use of discrete event simulation, which is a powerful tool employed to model and simulate supply chains and manufacturing systems, to model and simulate gene expression systems.

To this end, in this work we describe the *discrete event* based *mechanistic simulation* platform DEMSIM that we have developed for testing and hypothesizing putative regulatory interactions. The key feature of the DEMSIM platform is the event-based modeling and integration of the fundamental processes underlying gene expression (such as transcription, translation and species decay) with system-specific regulatory circuitry. In the next section, we outline the level of mechanistic detail that is accounted for in the various biological processes followed by a description of the computational and algorithmic issues that arise while implementing the simulation framework. Subsequently, the scope of the simulation framework to answer biologically relevant questions is investigated through three examples. The extensively studied *lac* operon system is simulated for verifying that the developed tool can indeed be trained to generate the experimentally obtained biological response of a genetic system. Then, the predictive capabilities of DEMSIM are probed by applying it to simulate the SOS response in *E. coli*. Finally, the sensitivity of the proposed approach to discriminate between alternative regulatory hypotheses is examined using the *araBAD* system of *E. coli* as a benchmark.

## 2. Methods

Effective simulation of gene expression and regulation dynamics entails the detailed modeling and integration of the underlying biochemical processes with the regulatory machinery. To this end, we have modeled each of the underlying transcription, translation and decay processes as stand-alone *modules*. Each module is characterized by *physical* and *model* parameters. Physical parameters correspond to parameters which are known *a priori* from literature sources and are fixed within the simulation framework (e.g. length of gene, transcription rate, etc.). In contrast, model parameters are regression parameters that are fitted using the available experimental data. Subsequently, the simulation is driven by communication between these modules in accordance with the specifics of the regulatory circuitry of the biological system being investigated. Furthermore, the mechanistic detail of the underlying processes is represented as a sequence of discrete events within the modules. The sequence of events that govern a given module and the associated parameters are described below.

### 2.1. Description of discrete event modules

#### 2.1.1. Transcription module

The mechanism of transcription is fairly well understood compared to other biological processes (Alberts et al., 1994; Hardinson, 2002a, b). The physical parameters required for this module include the length of the open reading frame (ORF)  $L_{ORF}^i$  [nucleotides] for each gene  $i$ , the foot print size of the RNAP enzyme  $L_{RNAP}$  [nucleotides] and the rate of transcriptional elongation  $\alpha_{Tp}$  [nucleotides/s]. The foot print size  $L_{RNAP}$  is the number of nucleotides that the RNAP has to transcribe before it clears the promoter for the subsequent transcription process. The model parameter associated with this module is the gene specific RNAP binding

parameter ( $K_{RNAP}^{bi}$ ) which quantifies the probability of the RNAP successfully binding to the promoter site. The discrete events constituting the transcription module are schematically shown in Fig. 3A. The transcription module begins with the transcription initiation event. A Monte Carlo based description is used to account for the inherent randomness associated with *all* stochastic events, including the binding events. Specifically, a uniformly distributed random number between 0 and 1 is generated and compared to the binding parameter associated with the event. If the magnitude of the generated random number is less than the binding parameter, then successful binding is assumed to have taken place otherwise the binding is assumed to have failed. If binding is successful, then the elongation phase is initiated, otherwise, promoter binding is reattempted as shown in Fig. 3A. The elongation phase consists of sequential elongation events whereby the mRNA transcript is produced one nucleotide at a time. Once the RNAP has transcribed  $L_{RNAP}$  nucleotides, the promoter is declared to be cleared and made available for additional transcription initiation events. This allows for the possibility of multiple RNAP molecules simulta-

neously transcribing a gene. We also account for the concurrent translation of an incomplete transcript, which is a well-known characteristic of prokaryotic systems, by checking for the formation of the nascent ribosome-binding site (RBS). This is achieved by comparing the length of the elongating mRNA to the ribosome footprint size  $L_{Rib}$  [nucleotides]. If the length of elongating mRNA is equal to  $L_{Rib}$ , then the newly formed RBS is made available for either initiation of translation or mRNA decay.

### 2.1.2. mRNA decay module

The complete mechanism of mRNA decay is still unresolved and many theories have been put forward to explain it (Marianne, 1999). However, it is largely accepted that mRNA decay is initiated when the enzyme RNase E endonuclease (RNase E) binds to the transcript (Carrier and Keasling, 1997). In view of this relatively well established hypothesis, we have modeled the decay process as a competitive binding event where the RNase E and the ribosomal assembly both compete for the free RBS on the elongating or complete mRNA transcript (see Fig. 3B). The gene specific RNase E

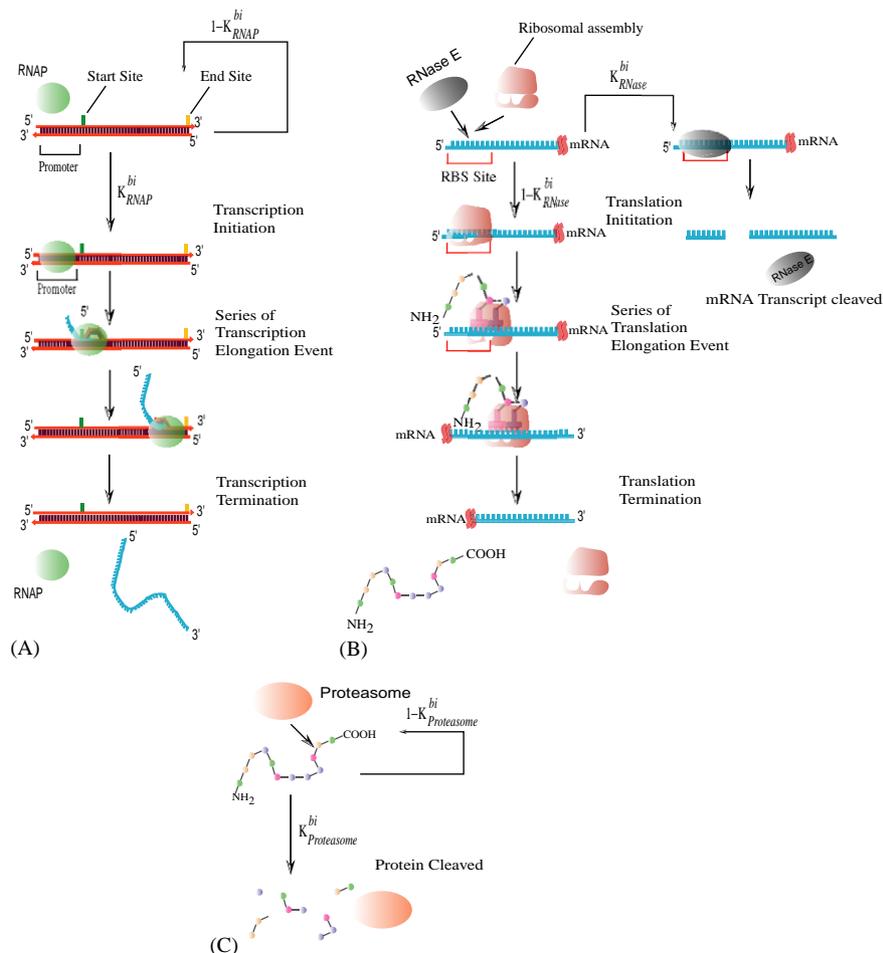


Fig. 3. Sequence of events governing: (A) Transcription module; (B) mRNA decay and Translation module; (C) Protein decay module.

binding parameter ( $K_{RNase}^{bi}$ ) quantifies the probability of successful binding of the RNase E to an mRNA transcript. If RNase E binds to the RBS, then the mRNA transcript is cleaved, otherwise the ribosomal assembly binds to the RBS and translation is initiated.

### 2.1.3. Translation module

Upon successful initiation by ribosome binding, a series of elongation events is executed through which the protein polypeptide chain is formed through the discrete addition of amino acid molecules (Fig. 3B) at the rate determined by the translation elongation rate parameter  $\alpha_{Tr}$  [codons/s]. RBS clearance is checked after each elongation event by comparing the number of nucleotides translated by the ribosome to  $L_{Rib}$ . If the ribosome has cleared the RBS, then the RBS is made available for the competitive binding of the RNase E and ribosomal assembly.

### 2.1.4. Protein decay module

Protein decay is modeled by the binding of the proteasomal assembly to the fully translated protein molecule (Alberts et al., 1994) as shown in Fig. 3C. The gene specific proteasome binding parameter  $K_{Proteasome}^{bi}$  determines the frequency with which the proteasomal assembly binds to a protein molecule and cleaves it into its constituent amino acids. Table 1 summarizes all the modules described above along with the associated physical and model parameters.

Table 1  
Modules and associated parameters

Module	Value
<i>Transcription</i>	
Physical	
$L_{RNAP}$ (nt)	60 nt (Kierzek et al., 2001)
$L_{Rib}$ (nt)	33 nt (Carrier and Keasling, 1997)
$L_{ORF}^i$ (nt)	KEGG Database
$\alpha_{Tr}$ (nt s <sup>-1</sup> )	50 nt/s (Hardinson, 2002a, b)
Model	
$K_{RNAP}^{bi}$	Fitted
<i>mRNA decay</i>	
Model	
$K_{RNase}^{bi}$	Fitted
<i>Translation</i>	
Physical	
$L_{Rib}$ (nt)	33 nt (Carrier and Keasling, 1997)
$\alpha_{Tr}$ (codons s <sup>-1</sup> )	10 codons/s (Hardinson, 2002a, b)
<i>Protein decay</i>	
Model	
$K_{Proteasome}^{bi}$	Fitted

## 2.2. Modeling of gene regulation

Regulation of gene expression occurs at varying degrees at all steps of the transcription through translation cascade. In DEMSIM, we assume that transcriptional initiation is the key step in gene regulation. This hypothesis has been put forth by a number of other researchers and supported by both experimental (Helmann et al., 2003) and computational investigations (Shen-Orr et al., 2002). The regulatory logic thus directly or indirectly alters the binding interactions of the RNAP with the promoter region of the DNA. In the context of our modeling framework, this is captured as the effect of the regulatory machinery on the probability of successful RNAP binding to the promoter region. Note that here the term regulatory logic is employed to describe a wide range of regulatory mechanisms which can be readily accounted for in our simulation framework. For example, a regulatory protein might regulate a target gene only if the concentration of the regulatory protein is beyond a threshold. In that case, the implementation of the regulatory criterion would entail checking if the concentration of the regulatory protein is above the specified threshold and subsequently making  $K_{RNAP}^{bi}$  dependent on the output of the regulatory logic. Separate RNAP binding parameters are assigned to binding events that represent alternative outcomes of the regulatory logic. The relative magnitude of these parameters quantifies the nature and strength of regulation (upregulation/down regulation). The regulatory logic employed for the test systems considered in this study are discussed in the results section.

## 2.3. Implementation of simulation framework

The DEMSIM software implementation consists of the following three key components: (i) an *event list* that contains all the events that need to be executed along with their respective execution times, (ii) a *global simulation clock* that records the progress of simulation time as events are sequentially executed, and (iii) a set of *state variables* that characterize the system and which are updated every time an event is executed. At every time step, events corresponding to all active (non-terminated) modules in the system are included in the event list. Subsequently, the event list is sorted and the event having the smallest execution time is executed. The simulation clock is advanced and the execution time of all other events is updated. Such a sequential procedure prevents the occurrence of “causality errors” by ensuring that an event with a later time stamps is not executed before an event with an earlier time stamp (Tropper, 2002). Furthermore, since the execution of certain events leads to the creation of new modules and the termination of existing ones, the number of active modules in

Table 2  
Events and execution times

Module	Execution time (s)	Value
<i>Transcription</i>		
Initiation event	$t_{bind}$	0.1 s
Elongation event	$1/\alpha_{Tp}$	0.02 s
<i>mRNA decay</i>		
Initiation event	$t_{bind}$	0.1 s
<i>Translation</i>		
Elongation event	$1/\alpha_{Tr}$	0.1 s
<i>Protein decay</i>		
Initiation event	$t_{bind}$	0.1 s

the system is updated and new events are included in the event list. This procedure is then repeated for the duration of the simulation horizon and state variables such as number of mRNA and protein molecules are recorded.

We use a fixed-time step of 0.10 s for stepping forward in time. This time interval corresponds to the duration between two translation elongation events (since  $\alpha_{Tr} = 10$  codons/s) and five transcription elongation events (since  $\alpha_{Tp} = 50$  nt/s). Table 2 lists the events associated with each module and the associated execution times. This time step, which results in the lumping of 5 transcription elongation events into a compound “pseudo” transcription event, is chosen to balance computational accuracy and CPU time requirements. Other assumptions include: (i) transcription and translation machinery are present in excess so that dilution by cell growth and gene expression can be neglected (Carrier and Keasling, 1997); and (ii) post transcriptional and post translational modifications take place instantaneously (Albert and Othmer, 2003; Goutsias and Kim, 2004). The DEMSIM framework is implemented using the C programming language on a 16 node linux cluster with dual Intel 3.4 Ghz Xeon processors.

### 3. Results

To highlight and probe its capabilities, the DEMSIM framework is applied to three different test systems. Given the stochastic nature of the underlying processes, multiple simulation runs are needed to glean a statistically complete picture of the temporal evolution of the system. The simulation runs are averaged out to extract the mean trajectory and the standard deviation is estimated at each time point. The results of the simulations are presented by plotting the mean trajectory and the  $\pm 1\sigma$  regions, where  $\sigma$  denotes the standard deviation.

#### 3.1. Example I—*lac* operon system of *E. coli*

The *lac* operon of *E. coli* has been extensively studied as a model system for understanding prokaryotic gene regulation (Kennell and Riezman, 1977; Wong et al., 1997; Vilar et al., 2003). We use this relatively simple genetic system to verify that the various model parameters embedded within DEMSIM can indeed be tuned using experimental data. In particular, we focus our attention on the expression of *lacZ* gene, the first within the operon which also includes genes *lacY* and *lacA*. Transcription from the *lac* operon is inhibited by the product of *lacI* gene located upstream of the operon. However, in the presence of lactose, the gene product of *lacI* combines with lactose to form an inactive product, thus turning the operon ON. This enables transcription of the *lacY* gene which encodes the protein responsible for transport of lactose into the cell.

In addition to the basic modules described earlier, the simulation of the *lac* operon system requires a model for transport of lactose into the cell. To this end, the kinetic model developed by Wong et al. (1997) is used. This model relates the rate of change of intracellular lactose to the amount of extracellular lactose and the amount of *lacY* protein. The mathematical form of the model is described in the appendix. All simulation runs begin with no lactose present inside the cell and the copy number/cell of mRNA and protein of all the genes is assumed to be zero (i.e. cold start). The regulatory logic is modeled by making the RNAP binding parameter for the *lac* operon conditionally dependent on the relative amounts of the inducer (lactose) and repressor (*lacI* protein) in the cell. This is achieved by utilizing the following rule based representation within the simulation framework.

$$K_{RNAP}^b = \begin{cases} \alpha & \text{if } [lacI] \leq [Lactose], \\ \beta & \text{if } [lacI] > [Lactose], \end{cases}$$

where  $[lacI]$  and  $[Lactose]$  are the number of *lacI* protein and lactose molecules respectively and  $\alpha > \beta$  in accordance with the inducer/repressor role of lactose/*lacI*. In addition to these RNAP binding parameters, two other model parameters that need to be tuned are the RNase E and proteasome binding parameter for *lacZ*. These parameters are estimated by applying DEMSIM within a predictive–corrective loop whereby the parameters are tuned such that the simulation results match experimentally reported data. Specifically, we use the following experimental data for fitting (Kennell and Riezman, 1977): *lacZ* mRNA half-life (1.3 min); average rate of production of *lacZ* protein (20 molecules/s); steady-state number of *lacZ* mRNA transcripts (62 molecules/cell). The values of the fitted model parameters are listed in Table 3. Figs. 4A and B show the simulated profiles for the number of *lacZ* mRNA and protein molecules, respectively. The simulated values for the

three quantities used for fitting are: *lacZ* mRNA half-life (1.5 min); average rate of production of *lacZ* protein ( $29 \pm 2$  molecules/s); steady-state number of *lacZ* mRNA transcripts ( $60 \pm 6$  molecules/cell). These results for the *lac* operon system clearly suggest that the DEMSIM framework is able to reproduce the dynamics of gene expression using appropriately tuned model parameters.

### 3.2. Example II—SOS response system of *E. coli*

In this example, we expand both the scale of the system under consideration, in terms of the number of genes whose expression is simulated, as well as the scope

Table 3  
Fitted parameter values for *lacZ* gene

Parameter	Condition	Value
$K_{RNAP}^{bi}$	$[lacI] \geq [Lactose]$	$1.0 \times 10^{-3}$
$K_{RNAP}^{bi}$	$[lacI] \leq [Lactose]$	$7.125 \times 10^{-1}$
$K_{RNase}^{bi}$	—	$8.0 \times 10^{-3}$
$K_{Proteasome}^{bi}$	—	$9.0 \times 10^{-5}$

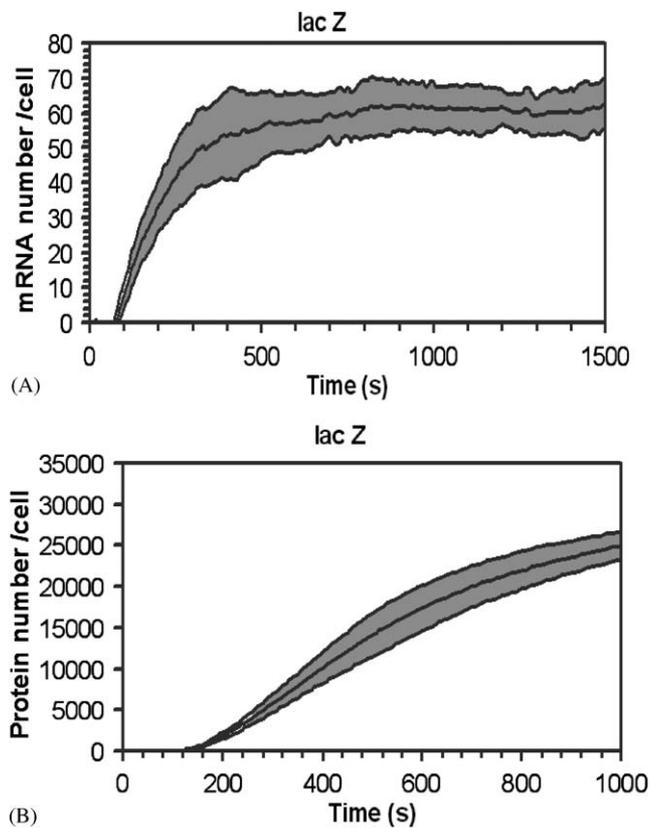


Fig. 4. (A) The fitted profile for the *lacZ* mRNA copy number. (B) The fitted profile for *lacZ* protein copy number. The center solid line shows the mean profile of 50 simulation runs and the shaded region represents the  $\pm 1\sigma$  regions.

of issues addressed using DEMSIM. We explore the capabilities of DEMSIM to not only reproduce experimental data with which it was trained but also its ability to predict the de novo response of the system to an externally imposed perturbation. To this end, the specific system that we investigate is the SOS response of *E. coli*. Irradiation of cells with UV light produces DNA lesions that transiently block the process of replication. It is now known that cells respond to this stress by upregulating the expression of several genes that function to repair the DNA lesions (Kuzminov, 1999; Henestroza et al., 2000; Janion, 2001). This response is termed as the SOS response (see Fig. 5). Many of the genes involved in the repair of DNA damage are negatively regulated by the *lexA* repressor protein, which binds to a consensus sequence located upstream of the promoter. Upregulation of these genes occurs when the *recA* protein binds to the single stranded DNA created at replication forks. This introduces a conformational change in the *recA* protein, turning it into a coprotease that cleaves the *lexA* repressor. As soon as the cellular concentration of *lexA* diminishes, the genes suppressed by *lexA* are more frequently transcribed. Following repair of DNA damage, the coproteolytic activity of *recA* diminishes leading to an increase in the *lexA* concentration and thus returning the cell to its original state as shown in Fig. 5 (Brent and Ptashne, 1981; Bertrand-Burggraf et al., 1987; Sassanfar and Roberts, 1990; Rehrauer et al., 1996). From the larger set of about 30 genes which are known to be regulated by the *lexA* repressor, we selected a subset of six genes to simulate (Courcelle et al., 2001; Khil and Camerini-Otero, 2002). In addition to *lexA*, the genes that we considered are: *polB* (production of DNA polymerase II); *uvrA*, *uvrB* (nucleotide excision repair); *ruvA* (recombination process); and *dinI* (inhibitor of *umuD*).

#### 3.2.1. Modeling of gene regulation

The regulatory logic for this system is formulated as follows. The probability of successful binding of the *lexA* protein to the protein-binding region of a gene is postulated to be given by

$$K_{lexA}^{bi} = 1 - \frac{2e^{(-\Phi(i) \cdot [lexA])}}{1 + e^{(-\Phi(i) \cdot [lexA])}}$$

for  $i = lexA, dinI, polB, uvrA, uvrB, ruvA$ .

Here,  $\Phi(i)$  is a gene specific regulatory constant and  $[lexA]$  is the number of molecules of *lexA* protein. Fig. 6 shows the dependence of  $K_{lexA}^{bi}$  on  $[lexA]$  for different values of  $\Phi(i)$ . Parameter  $\Phi(i)$  quantifies the relative binding strength of the *lexA* repressor to a particular gene  $i$  with a higher value of  $\Phi(i)$  implying a higher magnitude for  $K_{lexA}^{bi}$  (and hence higher probability of repression). Note that the above formulation ensures

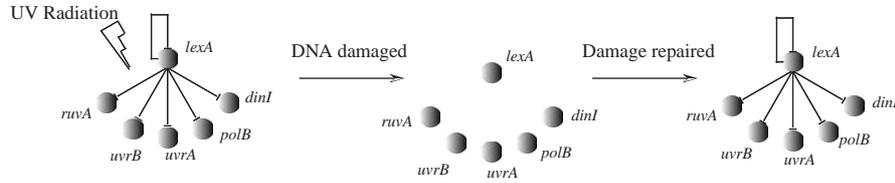


Fig. 5. UV radiation damages the DNA duplex. The damage to DNA acts as a signal to de-repress the genes normally repressed by the *lexA* repressor. Consequently, these genes are more frequently expressed. After the damage to DNA has been repaired, the repressor activity of *lexA* is reestablished thus returning the cell to its original state.

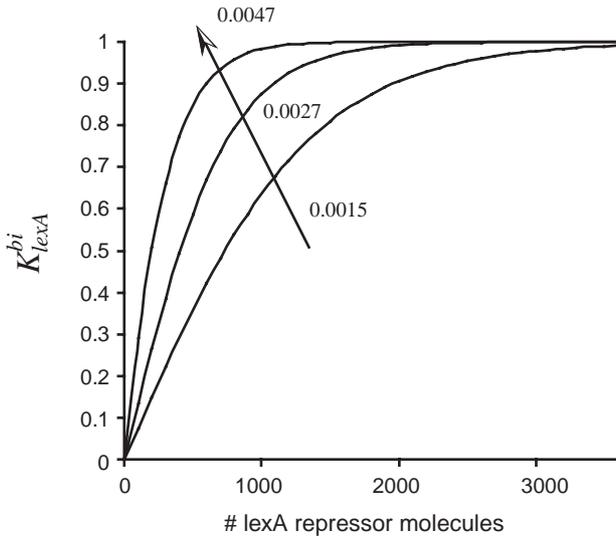


Fig. 6. Typical profiles for the probability of repression as a function of the amount of *lexA* repressor (copy number). The probability of repression is a monotonically increasing function of the repressor level with diminishing returns. The larger the value of  $\Phi(i)$ , the higher the probability of repression as indicated by the black arrow.

that the probability of repression given by  $K_{lexA}^{bi}$  is between 0 and 1 for all values of  $\Phi(i)$  and  $[lexA]$  with  $K_{lexA}^{bi} \rightarrow 0$  as  $[lexA] \rightarrow 0$  and  $K_{lexA}^{bi} \rightarrow 1$  as  $[lexA] \rightarrow \infty$ . Fig. 7 pictorially depicts the regulatory logic for the SOS response system. The *lexA* repressor binds to the operator region of the genes with a probability given by  $K_{lexA}^{bi}$ . If the repressor binds, then the gene is repressed otherwise the gene is unrepressed. The magnitude of  $K_{RNAP}^{bi}$  is made contingent on the outcome of the regulatory logic as illustrated in Fig. 7 with the relative magnitudes of  $(K_{RNAP}^{bi})_{Repressed}$  and  $(K_{RNAP}^{bi})_{Unrepressed}$  quantifying the strength of repression for each of the genes. Enhanced *lexA* cleavage under irradiated conditions is simulated by increasing  $K_{Proteasome}^{bi}$  for *lexA* gene by a factor of  $X_{lexA} (> 1)$ :

$$(K_{Proteasome}^{bi})^{Irradiated} = X_{lexA} (K_{Proteasome}^{bi})^{Unirradiated}$$

As a result of enhanced cleavage, the number of *lexA* molecules in the cell decrease reducing the magnitude of  $K_{lexA}^{bi}$ . This decreases the probability of repression of the genes in the system and the genes are more frequently

transcribed. After the repair time ( $T_{Repair}$ ) has elapsed, the value of  $K_{Proteasome}^{bi}$  for *lexA* is restored to its initial value thus gradually returning the cell to its original state.

### 3.2.2. Parameter estimation

The gene specific mRNA decay parameter  $K_{RNase}^{bi}$  is estimated by matching the simulated decay of mRNA level in the absence of transcription to the experimentally observed mRNA half-life. For a given value of the decay parameter, the simulations are run by ‘‘arresting’’ the processes of transcription. In the context of the simulation framework, this is accomplished by setting the value of  $K_{RNAP}^{bi}$  to zero. The simulated value of half-life corresponding to the assumed decay parameter is then estimated by measuring the time needed for the initial mRNA level to drop by half.  $K_{Proteasome}^{bi}$  is fitted similarly by ‘‘arresting’’ both the transcription and translation processes. Fig. 8 shows the average values of the simulated mRNA (Fig. 8A) and protein (Fig. 8B) half-lives as a function of the  $K_{RNase}^{bi}$  and  $K_{Proteasome}^{bi}$ , respectively. Subsequently, the factor  $X_{LexA}$ , which accounts for the enhanced *lexA* cleavage post irradiation, is similarly fitted by adjusting its value to reproduce the experimentally observed post-irradiation half-life of approximately 1–2 min (Sassanfar and Roberts, 1990). The time required to repair the damage to DNA is set at 45 min based on the observations of Courcelle et al. (2001).

The remaining parameters are gene specific RNAP binding parameter under repressed state  $(K_{RNAP}^{bi})_{Repressed}$ , RNAP binding parameter under unrepressed state  $(K_{RNAP}^{bi})_{Unrepressed}$  and the gene specific regulatory constant  $\Phi(i)$ . Since these parameters account for the generation of the mRNA transcripts and protein molecules in the cell, we refer to this set of parameters as generation parameters. The generation parameters are fitted by simultaneously adjusting their values to match the experimentally observed mRNA fold changes in both irradiated and unirradiated cells and the protein levels in the unirradiated cells. This procedure relies on the assumption that a direct correspondence exists between the mRNA transcript level and the fluorescence intensity measured in the microarray experiments. Fig. 9 highlights the procedure employed to estimate the

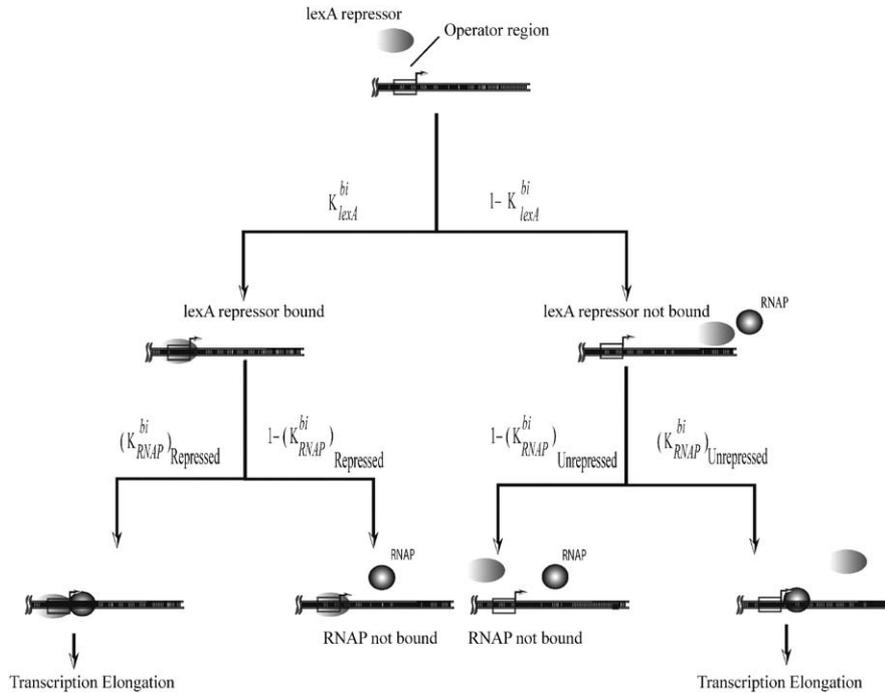


Fig. 7. The regulatory logic employed to simulate the SOS response system.

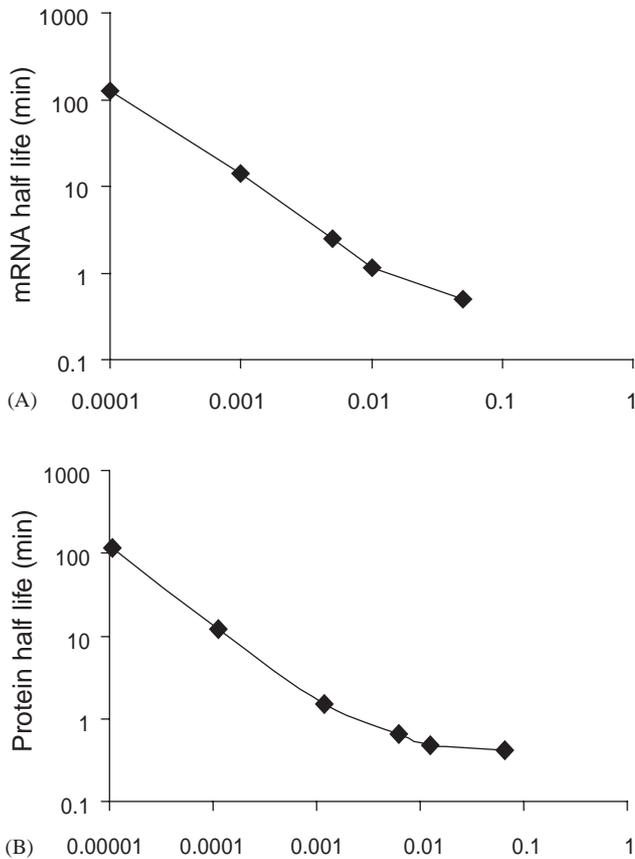


Fig. 8. The average half-life as a function of the governing decay parameter. (A) mRNA half-life as a function of  $K_{RNase}^{bi}$ . (B) Protein half-life as a function of  $K_{Proteasome}^{bi}$ .

generation parameters. Beginning with an initial guess for the values of the generation parameters, simulations are run using the previously estimated values for the decay parameters. After the simulation equilibrates (simulation warm-up time), the mRNA and protein levels in the cells are recorded for 5000s. These measurements correspond to the mRNA and protein levels under *unirradiated* conditions. Subsequently, the cleavage of *lexA* repressor is enhanced for a duration of  $T_{Repair}$  seconds, by multiplying  $K_{Proteasome}^{bLexA}$  with the previously estimated factor  $X_{LexA}$ , and the mRNA levels are recorded for a period of 5000s as shown in Fig. 9. These measurements correspond to the mRNA levels under *irradiated* conditions. The *unirradiated* and *irradiated* mRNA levels are compared to experimentally observed mRNA fold changes reported by Courcelle et al. (2001). Also, the recorded protein levels are compared to experimentally reported protein levels in *unirradiated* cell cultures (Kuzminov, 1999). The generation parameters are adjusted until the simulated measurements are in reasonable agreement with experimental observations. Table 4 summarizes the parameters for the SOS response system and the experimental data used to estimate the parameters. The values for the estimated parameter values are provided in Table 5. The simulated mRNA fold changes under *unirradiated* conditions are plotted in Fig. 10. The experimentally reported values are also plotted for comparison. Fig. 11 shows similar comparisons for the *irradiated* conditions. Similarly, the simulated and the

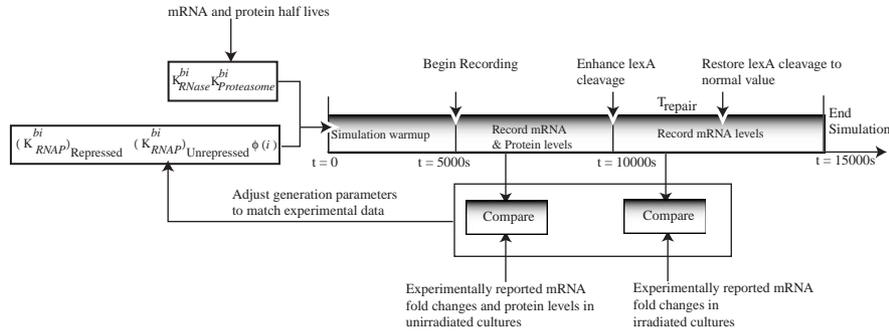


Fig. 9. This figure illustrates how the parameters of the SOS response system are fitted to reproduce experimental data. The gene specific mRNA and protein decay parameters are estimated from the experimental mRNA and protein half lives, respectively. Subsequently, the generation parameters are adjusted until the simulation results match experimental data for mRNA fold changes in *unirradiated* and *irradiated* cultures and the protein levels in *unirradiated* cultures.

Table 4  
Parameters for SOS response system

Parameter	Reference
$(K_{RNAP}^{bi})_{Unrepressed}$ , $(K_{RNAP}^{bi})_{Repressed}$ , $\Phi(i)$	Adjusted to match (i) mRNA fold changes in <i>unirradiated</i> cells (Courcelle et al., 2001) (ii) mRNA fold changes in <i>irradiated</i> cells (Courcelle et al., 2001) (iii) Protein levels in <i>unirradiated</i> cells (Kuzminov, 1999)
$K_{RNase}^{bi}$	Selected to reproduce experimentally observed mRNA half-life (Bernstein et al., 2002)
$K_{Proteasome}^{bi}$	Selected to reproduce experimentally observed protein half-life; 60 min for <i>lexA</i> (Sassanfar and Roberts, 1990) 10–30 min for other genes (Typical Value)
$X_{LexA}$	Selected to reproduce the <i>lexA</i> protein half-life of about 1–2 min post irradiation (Sassanfar and Roberts, 1990)
$T_{Repair}$	Set at 45 min (Courcelle et al., 2001)

Table 5  
Fitted parameter values for SOS response system

Gene	$(K_{RNAP}^{bi})_{Repressed}$	$(K_{RNAP}^{bi})_{Unrepressed}$	$\Phi(i)$	$K_{RNase}^{bi}$	$K_{Proteasome}^{bi}$	$X_{LexA}$
<i>lexA</i>	$9.5 \times 10^{-5}$	$4.75 \times 10^{-4}$	$3.7 \times 10^{-3}$	$3.5 \times 10^{-3}$	$2.4 \times 10^{-5}$	30.0
<i>uvrA</i>	$9.5 \times 10^{-5}$	$2.85 \times 10^{-4}$	$2.7 \times 10^{-3}$	$3.5 \times 10^{-3}$	$5.0 \times 10^{-4}$	
<i>dinI</i>	$9.5 \times 10^{-5}$	$4.75 \times 10^{-4}$	$5.7 \times 10^{-3}$	$3.8 \times 10^{-3}$	$6.0 \times 10^{-5}$	
<i>polB</i>	$9.5 \times 10^{-5}$	$4.75 \times 10^{-4}$	$1.5 \times 10^{-3}$	$3.8 \times 10^{-3}$	$7.0 \times 10^{-4}$	
<i>uvrB</i>	$9.5 \times 10^{-5}$	$9.50 \times 10^{-4}$	$4.7 \times 10^{-3}$	$4.8 \times 10^{-3}$	$7.0 \times 10^{-5}$	
<i>ruvA</i>	$9.5 \times 10^{-5}$	$9.50 \times 10^{-4}$	$1.7 \times 10^{-3}$	$3.2 \times 10^{-3}$	$1.0 \times 10^{-4}$	

experimentally estimated values for the protein levels in *unirradiated* cell cultures are listed in Table 6. In line with the observations for the *lac* system, the fitted parameter values are able to accurately reproduce the experimental data used to train the model.

### 3.2.3. Model validation

Next, the trained model is validated by comparing its predictions of protein levels in *irradiated* cultures to experimentally reported values. Table 7 lists the simulated peak protein levels estimated from the average

of 120 simulation runs and the corresponding experimentally obtained values. While good agreement with experimental estimates is observed for *uvrB*, *polB* and *uvrA* genes, some deviation is observed for *dinI* and *ruvA* genes. One possible reason for these deviations could be that the simulation framework might not account for all regulatory interactions involving these genes. In addition to the peak protein levels, the dynamics of the temporal response of *lexA* protein on induction of SOS response are also found to be in good agreement with experimental observations of Sassanfar and Roberts

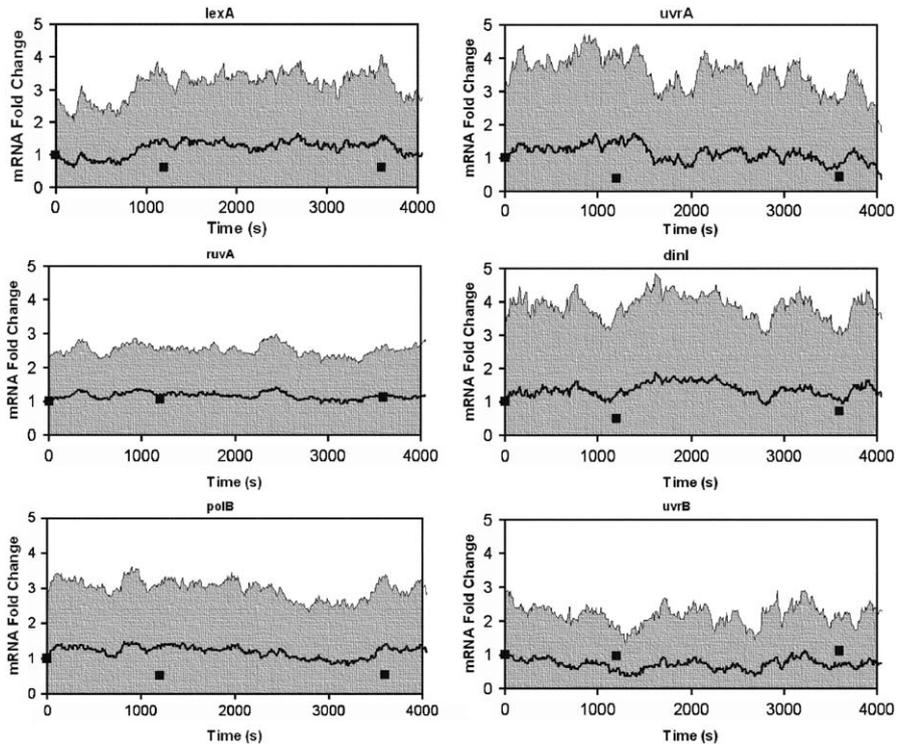


Fig. 10. Fitted mRNA profiles under *unirradiated* conditions: The simulation results are the average of 120 simulation realizations. Both the mean trajectory and the  $\pm 1\sigma$  regions are plotted, where  $\sigma$  denotes the standard deviation. The black squares are experimentally reported values (Courcelle et al., 2001).

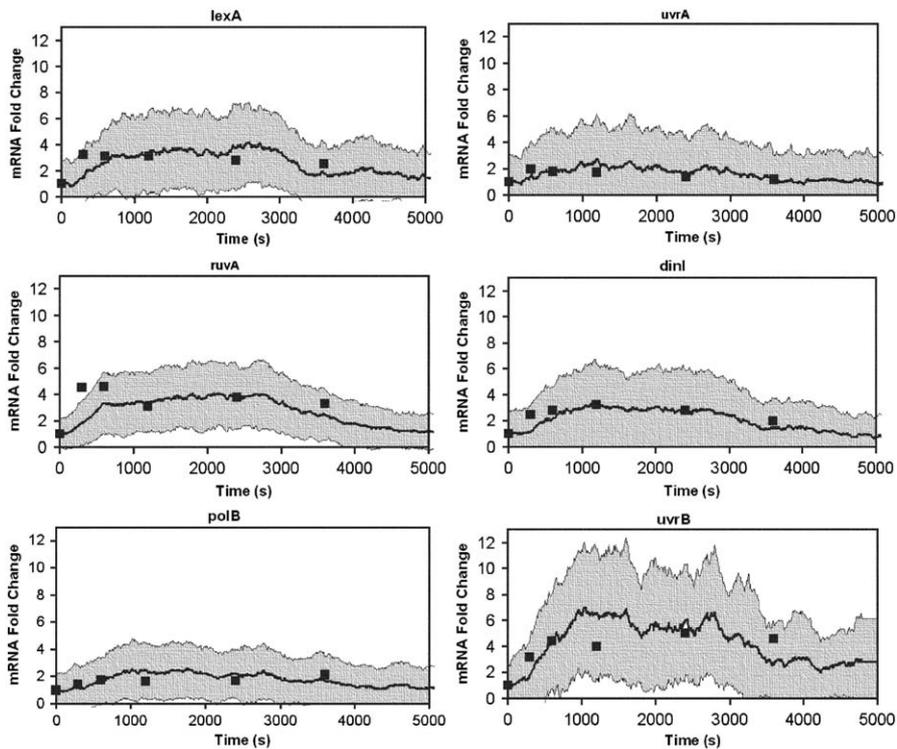


Fig. 11. Fitted mRNA profiles *irradiated* conditions: The simulation results are the average of 120 simulation realizations. Both the mean trajectory and the  $\pm 1\sigma$  regions are plotted, where  $\sigma$  denotes the standard deviation. The black squares are experimentally reported values (Courcelle et al., 2001).

Table 6  
Comparison between experimental and the fitted values of the protein levels under *unirradiated* conditions

Gene	No of copies/cell	
	Fitted	Experimental <sup>a</sup>
<i>lexA</i>	1306	1300
<i>uvrA</i>	49	20
<i>dinI</i>	384	500
<i>polB</i>	72	40
<i>uvrB</i>	243	250
<i>ruvA</i>	669	700

The protein numbers are represented as number of copies of the protein per cell.

<sup>a</sup>Based on Kuzminov (1999).

Table 7  
Comparison between experimental and the simulation predictions for the protein levels under *irradiated* conditions

Gene	No of copies/cell	
	Predicted	Experimental <sup>a</sup>
<i>lexA</i>	143	130
<i>uvrA</i>	112.5	250
<i>dinI</i>	1120	2300
<i>polB</i>	175	300
<i>uvrB</i>	1421	1200
<i>ruvA</i>	2158	5600

The protein numbers are represented as number of copies of the protein per cell.

<sup>a</sup>Based on Kuzminov (1999).

(1990) as shown in Fig. 12. These results highlight how, given adequate experimental data, the DEMSIM framework can first be trained and then be used as a predictive tool for generating responses of genetic systems.

### 3.3. Example III—Induction dynamics of *araBAD* operon of *E. coli*

The ability of the simulation framework to discriminate between alternative regulatory hypotheses is probed by applying it to the *araBAD* system in *E. coli*. The *araBAD* operon has been extensively studied as it serves as an excellent model for the feed forward loop motif (Seabold and Schleif, 1998; Schleif, 2000, 2003; Wu and Schleif, 2001). The *crp* gene activates both the *araBAD* operon and the *araC* gene in presence of inducer cAMP. The *araC* gene product transcriptionally activates the *araBAD* operon in presence of inducer L-arabinose resulting in a feed forward loop motif (see Fig. 13A). In addition, since the nature of regulation (i.e. activation) by the *crp* gene is the same for both the operon and the *araC* gene, the motif is termed as a

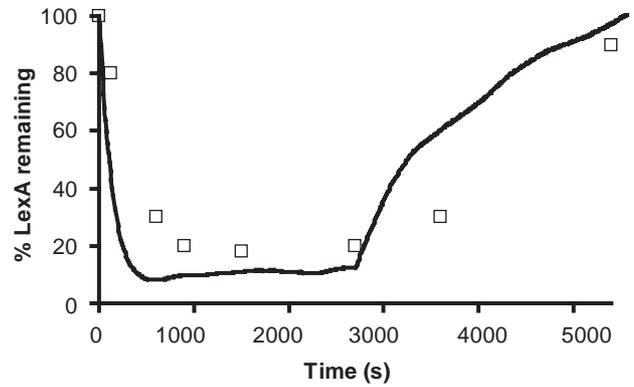


Fig. 12. The temporal response of *lexA* gene in terms of change in the protein level following the induction of SOS response. The simulation profile is the average of 120 simulation runs. The squares are experimental measured values (Sassanfar and Roberts, 1990).

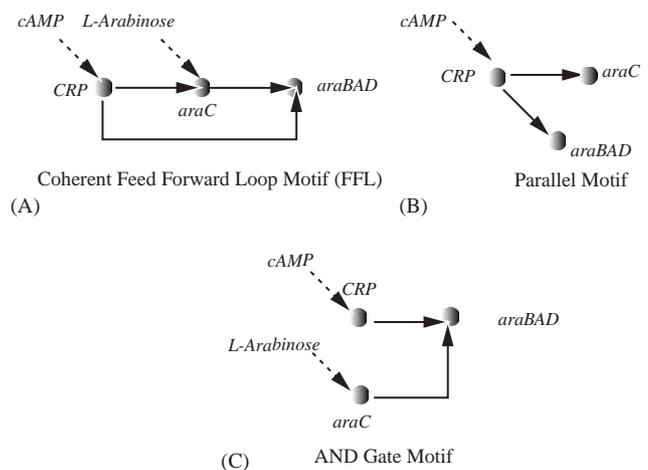


Fig. 13. Alternative regulatory mechanisms considered in this study, (A) FFL motif, (B) parallel motif, (C) simple AND gate motif.

*coherent feed forward loop* (FFL). Theoretical studies (Shen-Orr et al., 2002) have suggested that this system acts as a sign sensitive delay element. This implies that while the motif delays the cells response to an ON step in the stimulus, no delay in response is observed in the case of the complementary OFF step. In addition, Mangan et al. (2003) have investigated the responses of the *araBAD* FFL motif to cAMP ON and cAMP OFF steps. By comparing the response of the motif to that of *lac* promoter, which is a model for the simple AND gate motif, the authors have concluded that the *araBAD* system exhibits sign sensitive delay kinetics.

We used the DEMSIM framework to simulate two different regulatory mechanisms which both support the experimentally observed enhanced expression of the *araBAD* operon and the *araC* gene on addition of cAMP to a system saturated with L-arabinose. The first motif corresponds to a FFL (Fig. 13A) and the second motif represents a parallel motif (Fig. 13B). A simple

AND gate motif is also considered where both *crp* and *araC* enhance the expression of the *araBAD* operon as shown in Fig. 13C. Identical values are assigned to gene specific decay parameters for all the three mechanisms so that the decay dynamics exhibited by the motifs are the same. Furthermore, the gene specific generation parameters are fitted such that all three motifs exhibit similar *araBAD* expression in systems which are saturated and starved of the inducer cAMP. Subsequently, the response of the motifs to cAMP ON and OFF steps is generated and the responses of the FFL and parallel motif are compared to the response of the simple AND gate response. Simulation results shown in Fig. 14 indicate that the parallel motif model for gene regulation fails to capture the sign sensitive delay nature of the operon. In contrast, the FFL motif correctly exhibits a delayed response to a cAMP ON step (Fig. 14A) while no delay is observed in response to cAMP OFF step (Fig. 14B), suggesting that FFL is indeed the most plausible regulatory mechanism. These results highlight the ability of the DEMSIM framework to effectively discriminate between alternative regulatory mechanisms.

#### 4. Discussion

In this paper, we introduced a discrete event based mechanistic simulation platform (DEMSIM) and used it for testing and hypothesizing putative regulatory

interactions. The key feature of the developed simulation framework was the modeling of underlying biological processes, such as transcription, translation and decay, using stand-alone *modules*. Each module was characterized by a sequence of discrete events in accordance with the level of mechanistic detail considered. A rule based Monte Carlo procedure was employed for capturing the randomness inherent to the molecular binding events. Subsequently, communication within the modules was driven by taking into account system specific regulatory information. A distinction was made between physical and model parameters, with the former determined either from literature or online databases and the latter determined by fitting simulation results to experimental data.

The developed tool was benchmarked by applying it to three biological systems with different levels of complexity. The relatively simple *lac* operon was used to verify that parameters embedded in DEMSIM can indeed be trained using experimental data. Subsequently, the more complex SOS response system was used to probe the predictive capabilities of the developed framework. Simulation results indicated that the tool was able to make fairly accurate predictions regarding data that was *not* used for training the model parameters. Finally, the *araBAD* system was used to highlight the developed tool's sensitivity to discriminate between relatively "close" regulatory hypotheses.

The versatility of the DEMSIM framework allows us to conduct numerous *in silico* experiments. For example, the framework employed for SOS response system can be used to make predictions regarding the gene expression dynamics in a *lexAdef* genetic context, where the genes are expressed constitutively (Quillardet et al., 2003). If the model predictions are correct, then the developed model can be used to ask more complex questions regarding the biological system. For example, one could investigate the timing of induction of SOS response or the effect of single stranded DNA (ssDNA). If the model predictions are incorrect, then the experimental data can be used to refine the current model to prepare a more accurate representation of the underlying physical interactions. This exercise can provide valuable insights into the workings of the gene expression and regulatory interactions at a molecular level.

Many "top-to-bottom" computational frameworks employ high-throughput biological data to infer plausible regulatory hypotheses. For example, the GRAM algorithm proposed by Joseph et al. (2003), utilizes gene expression data and genome-wide location analysis for DNA-binding regulators, to predict putative regulatory interactions. In contrast the DEMSIM framework takes into account the underlying mechanistic detail of the gene expression and regulation processes to construct a predictive model. Furthermore, the simulation results demonstrate the ability of the framework to verify and

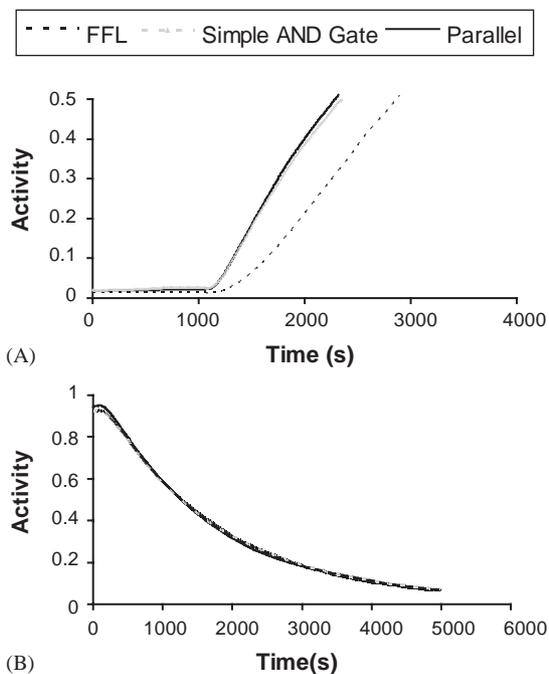


Fig. 14. The temporal responses of the alternative regulatory motifs compared to a simple AND gate motif. (A) cAMP ON step (B) cAMP OFF step.

also discriminate between relatively “close” regulatory hypotheses. These observations suggest that DEMSIM, which adopts a “bottom-to-top” approach, can be employed in tandem with “top-to-bottom” computational frameworks such as GRAM to verify and complete the candidate regulatory hypotheses generated by the latter approaches. However, unlike “top-to-bottom” approaches, extending the simulation framework to simulate large-scale gene networks requires enormous computational resources. One possible way of addressing this problem is to exploit the modular structure of large-scale regulatory networks. Recent studies have indicated that the regulatory networks can be decomposed into clusters of motifs (Shen-Orr et al., 2002; Alon, 2003). Hence, the regulatory hypotheses generated by the “top-to-bottom” approaches can be investigated for their modularity and the generated sub-networks/motifs can be simulated using the proposed framework. Comparison of simulation predictions with experimental data would then serve to verify, correct and complete the inferred hypotheses.

Due to the underlying stochastic nature of the simulation framework, extending the framework to model systems with larger copy numbers of species involved is difficult as the number of events increases by many folds. In such systems we envision a hybrid simulation framework that uses both differential equation based and stochastic methods in tandem (Kiehl et al., 2004). While differential equations can be used to model species with high copy number, DEMSIM can be used selectively for only low copy number species. We are currently working towards developing an integrated computational framework that brings to bear both “top-to-bottom” and “bottom-to-top” approaches to identify and verify candidate regulatory networks.

## Acknowledgements

Financial support by NSF Award BES0120277 is gratefully acknowledged. The authors would also like to thank Dr. Antonios Armaou for helpful discussions and suggestions.

## Appendix

In the kinetic model for transport of inducer (lactose) developed by Wong et al. (1997) the rate of transport,  $V$  of inducer into the cell is given by

$$V = \left( k_{in} \frac{[Lactose]_{ext}}{[Lactose]_{ext} + K_T} - k_{out} \frac{[Lactose]_{in}}{[Lactose]_{in} + K_T} \right) \times [lacY].$$

Here,  $[lacY]$  is the available amount of protein generated by the  $lacY$  gene (permease);  $k_{in}$  is the specific rate constant for transport of lactose into the cell and has a value of 35.8 mol lactose/mol permease/s;  $k_{out}$  is the specific rate constant for transport of lactose out of the cell and has a value of 1.19 mol lactose/mol permease/s;  $K_T$  is the saturation constant for lactose transport and has a value of  $2.6 \times 10^{-4}$  M; and  $[Lactose]_{ext}$  is the external lactose concentration set at 0.001 M.

## References

- Agger, T., Nielsen, J., 1999. Genetically Structured Modeling of Protein Production in Filamentous Fungi. *Biotechnol. Bioeng.* 66, 164–170.
- Akutsu, T., Miyano, S., 2000. Algorithms for inferring qualitative models of biological networks. *Pac. Symp. Biocomput.* 5, 290–301.
- Albert, R., Othmer, H.G., 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223, 1–18.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D., 1994. *Molecular Biology of THE CELL*. Garland Publishing Inc., New York & London.
- Alon, U., 2003. Biological networks: the tinkerer as an engineer. *Science* 301, 1866–1867.
- Ang, S., Lee, C.Z., Peck, K., Sindici, M., Matrubhutam, U., Gleeson, M.A., Wang, J.T., 2001. Acid-induced gene expression in *Helicobacter pylori*: study in genomic scale by microarray. *Infect. Immun.* 69 (3), 1679–1686.
- Arkin, A., Ross, J., McAdams, H.H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *E. coli* Cells. *Genetics* 149, 1633–1648.
- Bernstein, J.A., Khodursky, A.B., Lin, P.H., Chao, S.L., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *E. coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA* 99 (15), 9697–9702.
- Betrand-Burggraf, E., Hurstel, S., Daune, M., Schnarr, M., 1987. Promoter properties and negative regulation of the *uvrA* gene by the LexA repressor and its amino-terminal DNA binding domain. *J. Mol. Biol.* 193, 293–302.
- Brent, R., Ptashne, M., 1981. Mechanism of action of the *lexA* gene product. *Proc. Natl Acad. Sci. USA* 78 (7), 4204–4208.
- Carrier, T.A., Keasling, J.D., 1997. Mechanistic modeling of prokaryotic mRNA decay. *J. Theor. Biol.* 189, 195–209.
- Carrier, T.A., Keasling, J.D., 1999. Investigating autocatalytic gene expression systems through mechanistic modeling. *J. Theor. Biol.* 201, 25–36.
- Chen, T., He, H.G.L., Church, G.M., 1999. Modeling gene expression with differential equations. *Pac. Symp. Biocomput.* 4, 102–111.
- Cheng, B., Fournier, R.L., Relue, P.A., 1999. The inhibition of *E. coli lac* operon gene expression by antigene oligonucleotides-mathematical modeling. *Biotechnol. Bioeng.* 70 (4), 467–472.
- Cheng, B., Fournier, R.L., Relue, P.A., Schisler, J., 2000. An experimental and theoretical study of *E. coli lac* operon gene expression by antigene oligonucleotides. *Biotechnol. Bioeng.* 74 (3), 220–229.
- Courcelle, J., Khodursky, A., Peter, B., Brown, P.O., Hanawalt, P.C., 2001. Comparative gene expression profiles following exposure in wild-type and SOS-Deficient *E. coli*. *Genetics* 158, 41–64.
- Dasika, M.S., Gupta, A., Maranas, C.D., 2004. A mixed integer linear programming (MILP) framework for inferring time delay in gene regulatory networks. *Pac. Symp. Biocomput.* 9, 474–485.

- D'haeseleer, P., Shoudan, L., Somogyi, R., 1999. Linear modeling of mRNA expression levels during CNS development and injury. *Pac. Symp. Biocomput.* 4, 41.
- Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nat. Genet.* 403, 335–338.
- Friedman, N., Linial, M., Nachman, I., Pe'er, D., 2000. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620.
- Gibson, M.A., Bruck, J., 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.* 104, 1876–1889.
- Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403–434.
- Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81 (25), 2340–2361.
- Goutsias, J., Kim, S., 2004. A nonlinear discrete dynamical model for transcriptional regulation: construction and properties. *Biophys. J.* 86, 1922–1945.
- Hardinon, R.C., 2002a. *Molecular Genetics—vol. I.* McGraw-Hill Primis Custom Publishing, New York.
- Hardinon, R.C., 2002b. *Molecular Genetics—vol. II.* McGraw-Hill Primis Custom Publishing, New York.
- Hatzimanikatis, V., Lee, K.H., 1999. Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metab. Eng.* 1, 275–281.
- Helmann, J.D., Wu, M.F., Gaballa, A., Kobel, P.A., Morshedi, M.M., Fawcett, P., Paddon, C., 2003. The global transcriptional response of *Bacillus subtilis* to peroxide stress is coordinated by three transcription factors. *J. Bacteriol.* 185, 243–253.
- Henestroza, A.R.F.d., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J.J., Ohmori, H., Woodgate, R., 2000. Identification of additional genes belonging to the LexA regulon in *E. coli*. *Mol. Microbiol.* 35 (6), 1560–1572.
- Hoon, M.J.L., Imoto, S., Kobayashi, K., Ogasawara, N., Miyano, S., 2003. Inferring gene regulatory networks from time-ordered gene expression data of *Bacillus subtilis* using differential equations. *Pac. Symp. Biocomput.* 8, 17–28.
- Ideker, T.E., Thorsson, V., Karp, R.M., 2000. Discovery of regulatory interactions through perturbations: inference and experimental design. *Pac. Symp. Biocomput.* 5, 302–313.
- Janion, C., 2001. Some aspects of the SOS response system—a critical survey. *Acta Biochim. Pol.* 48 (3), 599–610.
- Joseph, Z.B., Gerber, G.K., Lee, T.I., Yoo, J.Y., Robert, F., Gordon, D.B., Fraenkel, E., Jaakkola, T.S., Young, R.A., Gifford, D.K., 2003. Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.* 21 (11), 1337–1342.
- Kastner, K., Solomon, J., Fraser, S., 2002. Modeling a *HOX* gene network *in Silico* using a stochastic simulation algorithm. *Dev. Biol.* 246, 122–131.
- Kennell, D., Riezman, H., 1977. Transcription and translation initiation frequencies of the *E. coli lac* operon. *J. Mol. Biol.* 114, 1–21.
- Kepler, T.B., Elston, T.C., 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81, 3116–3136.
- Khil, P.P., Camerini-Otero, P.D., 2002. Over 1000 genes are involved in the DNA damage response of *E. coli*. *Mol. Microbiol.* 44 (1), 89–105.
- Kiehl, T.R., Mattheysses, R.M., Simmons, M.K., 2004. Hybrid simulation of cellular behavior. *Bioinformatics* 20 (3), 316–322.
- Kierzek, A.M., Zaim, J., Zielenkiewicz, P., 2001. The effect of transcription and translation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J. Biol. Chem.* 276 (11), 8165–8172.
- Kurata, H., Matoba, N., Shimizu, N., 2003. CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle. *Nucleic Acids Res.* 31 (14), 4071–4084.
- Kuzminov, A., 1999. Recombinatorial repair of DNA damage in *E. coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* 63 (4), 751–813.
- Mangan, S., Zaslaver, A., Alon, U., 2003. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.* 334, 197–204.
- Marianne, G.M., 1999. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu. Rev. Genet.* 33, 193–227.
- McAdams, H.H., Arkin, A., 1997. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* 94, 814–819.
- Quillardet, P., Rouffaud, M.A., Bouige, P., 2003. DNA array analysis of gene expression in response to UV radiation in *E. coli*. *Res. Microbiol.* 154, 559–572.
- Rehrauer, W.M., Lavery, P.E., Palmer, E.L., Singh, R.N., Kowalczykowski, S.C., 1996. Interaction of *E. coli* RecA protein with the LexA repressor. *J. Biol. Chem.* 271 (39), 23865–23873.
- Sassanfar, M., Roberts, J.W., 1990. Nature of the SOS-inducing signal in *E. coli* The involvement of DNA replication. *J. Mol. Biol.* 212, 79–96.
- Schleif, R., 2000. Regulation of the L-arabinose operon of *E. coli*. *Trends Genet.* 16 (12), 559–565.
- Schleif, R., 2003. AraC protein: a love–hate relationship. *BioEssays* 25, 274–282.
- Seabold, R.R., Schleif, R., 1998. Apo-Arac actively seeks to loop. *J. Mol. Biol.* 278, 529–538.
- Shea, M.A., Ackers, G.K., 1985. The O<sub>R</sub> control system of bacteriophage lambda a physical–chemical model for gene regulation. *J. Mol. Biol.* 181, 211–230.
- Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *E. coli*. *Nat. Genet.* 31, 64–68.
- Siegele, D.A., Hu, J.C., 1997. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl Acad. Sci. USA*, 94, 8168–8172.
- Spellman, 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9, 3273–3297.
- Tropper, C., 2002. Parallel discrete-event simulation applications. *J. Parallel Distrib. Comput.* 62, 327–335.
- Vilar, J.M.G., Guet, C.C., Leibler, S., 2003. Modeling network dynamics: the *lac* operon, a case study. *J. Cell Biol.* 161 (3), 471–476.
- Vohradsky, J., 2001. Neural model of the genetic network. *J. Biol. Chem.* 276, 36168–36173.
- Wong, P., Galdney, S., Keasling, J.D., 1997. Mathematical model of the *lac* operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Prog.* 13, 132–143.
- Wu, M., Schleif, R., 2001. Mapping Arm-DNA-binding domain interactions in AraC. *J. Mol. Biol.* 307, 1001–1009.