Stochastic Modeling of Gene Positive Autoregulation Networks Involving Signal Molecules

Xin Fang,* William E. Bentley,† and Evanghelos Zafiriou*
*Department of Chemical and Biomolecular Engineering, and †Fischell Department of Bioengineering, University of Maryland, and Center for Biosystems Research, University of Maryland Biotechnology Institute, College Park, Maryland

ABSTRACT Positive autoregulation in gene regulation networks has been shown in the past to exhibit stochastic behavior, including stochastic bistability, in which an initially uniform cell population develops into two distinct subpopulations. However, positive autoregulation is often mediated by signal molecules, which have not been considered in prior stochastic analysis of these networks. Here we propose both a full model of such a network that includes a signal molecule, and a simplified model in which the signal molecules have been eliminated through the use of two simplifications. The simplified model is amenable to direct mathematical analysis that shows that stochastic bistability is possible. We use stochastic Petri networks for simulating both types of models. The simulation results show that 1), the stochastic behavior of the two models is similar; and 2), that the analytical steady-state distribution of the simplified model matches well the transient results at times equal to that of a cell generation. A discussion of the simplifications we used in the context of the results indicates the importance of the signal molecule number as a factor determining the presence of bistability. This is further supported from a deterministic steady-state analysis of the full model that is shown to be a useful indicator of potential stochastic bistability. We use the regulation of SdiA in *Escherichia coli* as an example, due to the importance of this protein and of the signal molecule, a bacterial autoinducer, that is involved. However, the use of kinetic parameter values representing typical cellular activities make the conclusions applicable to other signal-mediated positive autoregulation networks as well.

INTRODUCTION

Gene expression is subject to intrinsic noise (1), which may come from thermal fluctuations of molecular events that constitute the process of gene expression (2). When the numbers of some molecules in a cell are relatively low, the effect of fluctuations can be apparent and then even cloned cells in the same environments can evolve into different phenotypes (3). In such situations, gene regulation networks should be viewed as stochastic processes. Several gene regulation networks have been stochastically modeled. McAdams and Arkin (4) performed stochastic simulation for a simple regulation link of two genes. The expression of pyelonephritis-associated pili in uropathogenic *Escherichia coli* has been modeled as a Markov chain (5). A stochastic model for the quorum-sensing mechanism in *E. coli* was developed and used to obtain new insights on synthesis pathways of autoinducer-2 (AI-2) (6).

Gene regulation networks with positive autoregulation are strong candidates for consideration of stochastic behavior because positive autoregulation may cause bistability and lead an initially homogeneous cell population to develop into two distinct subpopulations. Arkin et al. (7) developed a stochastic model for the developmental pathway in phage λ-infected *E. coli* and simulation results predicted a bifurcation in the cell population, which was experimentally verified. In this model, although the λ-switch is essential for the selection for lysis/lysogeny pathways, the positive autoregulation of the gene expression for the CI protein is necessary for the bifurcation. Positive feedback modules are present in *E. coli* cells and experimental results indicate bistability of protein amounts, consistent with predictions from a stochastic model (8). In another experiment, stochastic bifurcation can even be directly observed under the microscope (9). Positive autoregulation networks have also been mathematically analyzed to obtain insight. Ferrell (10) discussed how cellular systems with positive feedback loops can exhibit bistability and reviewed some typical bistable systems. Kepler and Elston (11) analytically obtained the bistable probability distribution of protein number in a simple self-promotion network.

However, positive autoregulation in gene regulation networks is often mediated through signal molecules. Let us consider two examples in quorum sensing, a cell-to-cell communication mechanism, in which extracellular factors function as signal molecules. In *Vibrio fischeri*, when DNA sequences in operon$_R$ are deleted or autoinducers (AI) are at low levels, LsrR protein can have a positive regulation on *lsrR* gene in the presence of AI (12). In *E. coli*, the appearance of another kind of extracellular factor, autoinducer-2 (AI-2), can increase the expression of *sdiA* gene (13). Although it is still uncertain that the SdiA protein is related to this regulation, there are some supporting facts. First, SdiA has a LuxR-type sequence (14,15) and its N-terminal domain can bind to a factor (16), which suggests that SdiA may regulate genes by binding some factor. Second, AI-2 can enhance some SdiA-regulated gene expression (17), which suggests that AI-2 may be a factor. In this study, we have hypothesized that

Submitted December 10, 2007, and accepted for publication May 21, 2008.
Address reprint requests to Evanghelos Zafiriou, Tel.: 301-405-6625; E-mail: zafiriou@umd.edu.
Editor: Jennifer Linderman.

0006-3495/08/10/3137/09 $2.00
doi: 10.1529/biophysj.107.127175

SdiA protein upregulates its own expression by binding AI-2 molecules (18). Since the SdiA protein is involved in the regulation of cell division (15), the system we study may be used by extracellular signal molecules AI-2 to modulate cell division.

Work on the stochastic modeling of such signal-mediated networks has been limited. Through the study of the mathematical relation between the signal-mediated networks and the pure positive autoregulation networks that have been more extensively studied, conclusions from the latter networks, such as the property of bistability, may be carried to the signal-mediated networks. Here we present one network, in which positive autoregulation is mediated through signal molecules, as the full model, and another network, in which there is only pure positive autoregulation, as the simplified model. With some assumptions the simplified model can be obtained from the full model. Mathematical analysis and simulations for both models are performed stochastically and deterministically, so that the bistable properties of the two models can be studied in relation to each other, to explore under what circumstances the conclusions from the simplified model can be applied to the full model.

NETWORK MODELS

We have selected the regulation of SdiA in *E. coli* as our example network. As discussed in the previous section, there is evidence to suggest that the signal molecule AI-2 may be involved in SdiA protein upregulation of its own expression by binding to it. It should be noted, however, that the network structures that we study are more general and several kinetic parameters in the simulations are assigned values representing typical cellular activities. Therefore, we expect that conclusions reached from this study on the stochastic behavior of the network can be applicable to other signal-mediated positive autoregulation networks in *E. coli* and in other bacteria.

Full model

The signal-mediated positive autoregulation network, termed the full model, is shown in Fig. 1. The gene (DNA) has two states. Transcription rates under the two states are different, so mRNA can accumulate at different rates. After transcription, protein monomers (SdiA) are translated from mRNA and then form activated protein complexes (SdiAact) by binding signal molecules (AI2). Two activated complexes bind each other and become a dimer. The dimers determine which state the gene is in. When the dimer binds to the gene, the gene is in occupied state (DNA1) and the transcription is much faster than that of the unoccupied gene (DNA0). So we have a positive autoregulation which is dependent on signal molecules. Here we assume it is the dimer that binds to the gene because the regulatory proteins often function as dimers or higher-order oligomers (11,19).

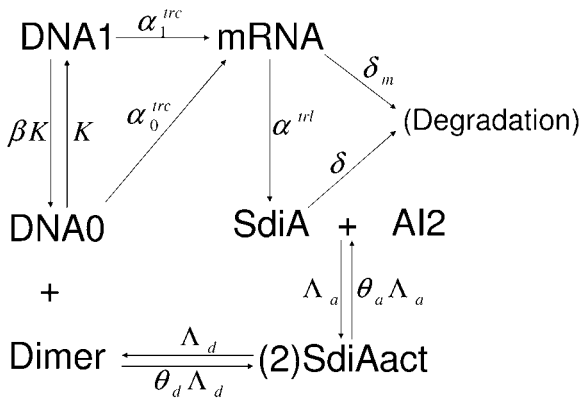

FIGURE 1 Full model with signal-mediated positive autoregulation.

The parameter values for Fig. 1, i.e., the rate constants of the reactions, are listed in Table 1. The transcription rate, α_0^{trc}, for the unoccupied gene is selected near the middle point on a logarithmic range for *E. coli* (20) and it corresponds to 200 transcripts per generation. The rate for the occupied gene, α_1^{trc}, is set to the maximum of the range. The mRNA decay rate, δ_{m}, corresponds to a half-life of 2 min, which is within the range of mRNA half-lives in *E. coli* (21). The protein decay rate, δ, corresponds to a half-life of 17 s (half-lives of proteins in cells vary widely (22) and we use a fast decay rate so that the bifurcation can be observed within 50 min). The translation rate, α^{trl}, is based on Chen et al. (23), and β, K on Kepler and Elston (11) with some adjustments within the same order of magnitude. The parameters for the formation of the complex and the dimerization were selected to result in parameter values for the simplified model, developed in the next section, close to those used in Kepler and Elston (11), while keeping the rate for the formation of the complex higher than the dimerization rate.

There are two physical meanings for these rate constants. In the deterministic model (ordinary differential equation, i.e., ODE, model), the constants are kinetic rate constants. In

TABLE 1 Parameters of full model

Parameters	Values	References
α_0^{trc}	1.2 min^{-1}	Based on transcription rate range: 10^{-4}–1 s^{-1} (20).
α_1^{trc}	60 min^{-1}	Set to maximum rate of one transcription per second (20).
δ_{m}	0.36 min^{-1}	Based on mRNA half-life range: 40 s–20 min (21).
α^{trl}	6 min^{-1}	Same order of magnitude as Chen et al. (23).
β	28	Same order of magnitude as Kepler and Elston (11).
K	1 min^{-1}	Same order of magnitude as Kepler and Elston (11).
δ	2.5 min^{-1}	Corresponds to half life of 17 s within the range for proteins (22).
θ_{a}	1000	Selected to be higher than θ_{d}.
Λ_{i}	1 min^{-1}	Adjusted based on Kepler and Elston (11) within same order of magnitude.
θ_{d}	250	Adjusted based on Kepler and Elston (11).
Λ_{d}	1 min^{-1}	Adjusted based on Kepler and Elston (11) within same order of magnitude.

the stochastic model, they are stochastic rate constants, which denote the probability of the reaction occurrence (2). When the deterministic rate constants are in terms of concentrations, the stochastic rate constants are equal to them for first-order reactions. For second-order reactions, the relation between stochastic rate and deterministic rate is (24)

$$\beta = \frac{k}{VN_A}. \tag{1}$$

Here β is the stochastic rate constant, k is the deterministic rate constant in terms of concentrations, V is the cell volume, and N_A is Avogadro's number. In the case where the deterministic rate constants are in terms of molecule numbers, for both the first-order and second-order reactions, the deterministic rate constants have the same values and units as the stochastic rate constants. Therefore the values in Table 1 can be used in both cases.

Simplified model

The simplified model is shown in Fig. 2. This model is a network in which positive autoregulation does not involve signal molecules. To obtain it from the full model, two simplifications are involved: First, transcription and translation, the two steps of gene expression in the full model, are merged into one reaction:

$$\text{DNA0} \xrightarrow{\alpha_0} \text{SdiA}, \tag{2}$$

$$\text{DNA1} \xrightarrow{\alpha_1} \text{SdiA}. \tag{3}$$

In the simplified model, mRNA is eliminated so we do not need to consider its activity. The second simplification is that activation and dimerization of protein monomers (SdiA) are simplified into a single reversible reaction:

$$2\text{SdiA} \underset{\theta^*\Lambda^*}{\overset{\Lambda^*}{\rightleftharpoons}} \text{Dimer}. \tag{4}$$

After this simplification, the activated protein complex (SdiAact) and signal molecules (AI2) are eliminated from the equations. The simplified model has the same structure as the general model examined by Kepler and Elston (11), where an analytical expression for the probability distribution was obtained.

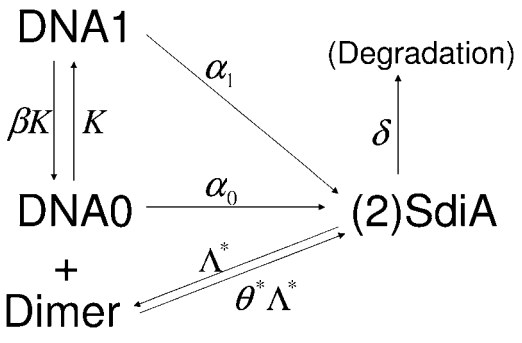

FIGURE 2 Simplified model with positive autoregulation.

These two simplifications are based on three assumptions: 1), The mRNA number is assumed at steady state; 2), the number of free signal molecules is considered constant; and 3), activation of protein is much faster than dimerization of the activated complex. Next, we show how to derive the model from the assumptions.

The first simplification is based on the first assumption. In the full model from Fig. 1, the translation kinetics are

$$r_{\text{SdiA}} = \alpha^{\text{trl}}[mRNA]. \tag{5}$$

The notation $[X]$ represents the molecule number of species X in the cell. With the assumption that mRNA is always at steady state, the mRNA number in the full model (Fig. 1) is obtained:

$$[mRNA] = \frac{\alpha_0^{\text{trc}}}{\delta_m}[DNA0] + \frac{\alpha_1^{\text{trc}}}{\delta_m}[DNA1]. \tag{6}$$

By using Eq. 6 in Eq. 5, a simplified rate function for protein monomer (SdiA) production is obtained as

$$r_{\text{SdiA}} = \alpha_0[DNA0] + \alpha_1[DNA1], \tag{7}$$

where

$$\alpha_0 = \alpha_0^{\text{trl}}\alpha^{\text{trc}}/\delta_m, \tag{8}$$

$$\alpha_1 = \alpha_1^{\text{trl}}\alpha^{\text{trc}}/\delta_m. \tag{9}$$

Equation 7 is just the kinetic rate function for reactions 2 and 3.

The second simplification is obtained from the last two assumptions. It is assumed that activation is much faster than dimerization. Hence, compared with dimerization, activation can be considered at quasi-equilibrium:

$$[SdiAact] = \frac{[SdiA][AI2]}{\theta_a}. \tag{10}$$

On the other hand, in those two steps dimerization is the determining step. Then for the overall reaction, the forward reaction rate is

$$r_f = \Lambda_d[SdiAact]^2 = \Lambda^*[SdiA]^2, \tag{11}$$

where

$$\Lambda^* = \frac{\Lambda_d[AI2]^2}{\theta_a^2}. \tag{12}$$

The overall reverse reaction rate is

$$r_r = \theta_d\Lambda_d[Dimer] = \theta^*\Lambda^*[Dimer], \tag{13}$$

where

$$\theta^* = \frac{\theta_a^2\theta_d}{[AI2]^2}. \tag{14}$$

Equations 11 and 13 describe the kinetics for the overall reaction 4 for the simplified model. Hence the second simplification is realized. From Eqs. 12 and 14 it can be concluded that θ^* and Λ^* can be used as effective rate constants because of the second assumption that the free signal molecule (AI2) number is constant.

Equations 8, 9, 12, and 14 indicate the relation between the parameters of the simplified and full models. With these equations and the constant signal molecule (AI2) number equal to 500, values for parameters of the simplified model are calculated and shown in Table 2.

ANALYSIS OF SIMPLIFIED MODEL

Analytical steady-state distribution

For networks with the structure of the simplified model, an analytical expression for the stochastic distribution under steady state has been obtained by Kepler and Elston (11). This distribution is based on two approximations. The first approximation is the small noise approximation. When the protein abundance is large, the protein monomer (SdiA) number can be expressed by defining a continuous variable instead of a discrete one. The second is the fast noise approximation, which means the transitions between the two gene states (DNA0 and DNA1) are fast but finite.

With these approximations, the steady-state density function of protein monomer number is

$$\overline{\rho}(x) = \frac{\lambda}{B(x)}\exp\left(2\int_0^x \frac{A(x')}{B(x')}\mathrm{d}x'\right), \tag{15}$$

where

$$A(x) = \frac{ba_0 + x^2}{b + x^2} - x - \frac{2xb(a_0 - 1)[((a_0 - 2) + x)x^2 + b(x - a_0)]}{\kappa(b + x^2)^4}, \tag{16}$$

$$B(x) = \frac{1}{m_o}\left(\frac{b(a_0 + x) + x^2(1 + x)}{b + x^2}\right) + \frac{2bx^2(a_0 - 1)^2}{\kappa(b + x^2)^3}. \tag{17}$$

(Note that there are a couple of typographical errors in the equations in Kepler and Elston (11) corresponding to Eqs. 15 and 17 above. Results shown in Fig. 6 in Kepler and Elston (11) match Eqs. 15 and 17.) There are four dimensionless rescaled parameters in the equations above: $a_0 = \alpha_0/\alpha_1$, $b = \beta\theta\delta^2/\alpha_1^2$, $\kappa = K\alpha_1^2/(\theta\delta^3)$, and $m_o = \alpha_1/\delta$. With the parameter values in Table 2, the values for the rescaled parameters are calculated as shown in Table 3. The value x is the stochastic variable denoting the number of protein monomer. The x is also dimensionless with the relation $x = [SdiA]/m_o$. The parameter λ is used to normalize the distribution $\overline{\rho}(x)$.

TABLE 2 Parameters of simplified model

Parameters	Values
α_1	1000 min^{-1}
α_0	20 min^{-1}
δ	2.5 min^{-1}
β	28
K	1 min^{-1}
θ^*	1000
Λ^*	0.25 min^{-1}

TABLE 3 Values for rescaled parameters

Parameters	Values
a_0	0.02
b	0.175
κ	64
m_o	400

The distribution of protein monomer number obtained from Eq. 15 is shown in Fig. 3. In this figure the plot of the density function (*solid curve*) is a two-peak curve, which means the pathway of protein production has a bifurcation. The bifurcation separates the bacterial cells into two subpopulations. In one subpopulation, cells produce protein at a slower rate. There is a lower number of proteins in the cells. The less the number of protein monomers is, the less the numbers of activated protein complexes and dimers are as well. Then, most of the time, the gene is in the unoccupied state and the protein production rate becomes even slower. As time goes on, this group of cells reaches steady state with a lower protein number. The other subpopulation goes in the opposite direction. Faster protein production causes higher number of dimers and puts the gene in the occupied state more frequently, resulting in much faster protein production rate. Hence cells in this group reach steady state with a higher protein number. The distribution discussed here is a steady-state distribution.

We can establish that the fluctuation between DNA0 and DNA1 plays an important role in the existence of bistability by trying different values for K. The value for K in Table 2 is 1 min^{-1}, which is a moderate rate for the binding of dimer to DNA0 and can cause a bistable distribution as shown in Fig. 3. Next we set $K = 0.1$ min^{-1}, in which case it is hard for the dimer to bind DNA0. Then the distribution has only one peak at a low protein number (Fig. 4). Finally we set $K = 10$ min^{-1}, for which the binding of dimer to DNA0 is very rapid.

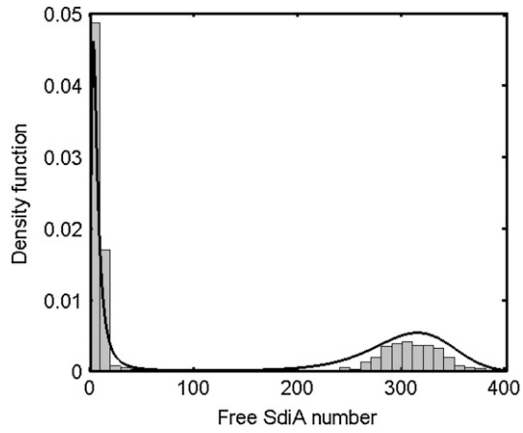

FIGURE 3 Distribution of protein monomer (free SdiA) number for the simplified model. (*Solid curve*) Analytical steady-state distribution of monomer number. (*Bar graph*) Distribution obtained from SPN simulation after 50 min.

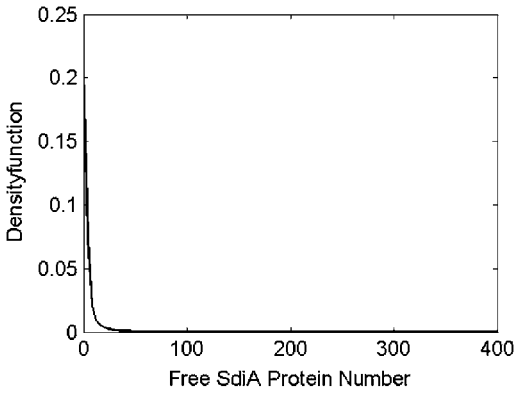

FIGURE 4 Distribution of protein monomer (free SdiA) number for the simplified model with low rate of binding of dimer to DNA0 ($K = 0.1\ \text{min}^{-1}$ so $\kappa = 6.4$).

The distribution is shown in Fig. 5 and it has only one peak at a high protein number.

Stochastic simulation

In this work, stochastic simulation is realized with stochastic Petri networks (SPN). An SPN is a mathematical formalism for representing and simulating events that are stochastic and discrete. Goss and Peccoud (24) showed that they are a very good tool for the stochastic simulation of molecular biological systems and since then they have been used more extensively for this purpose. The SPN for the simplified model is constructed and shown in Fig. 6. Cell activities like gene expression, molecular binding and unbinding, are all described as chemical reactions. In the SPN the tokens (*circles*) represent the chemical species including DNA, mRNA, proteins, and signal molecules. The transitions (*vertical lines*) represent the chemical reactions, with arrows pointing from reactants and into products. For second-order reactions, there are input and output gates (*triangles*) in Fig. 6, indicating preconditions and effects of the transition firing, because

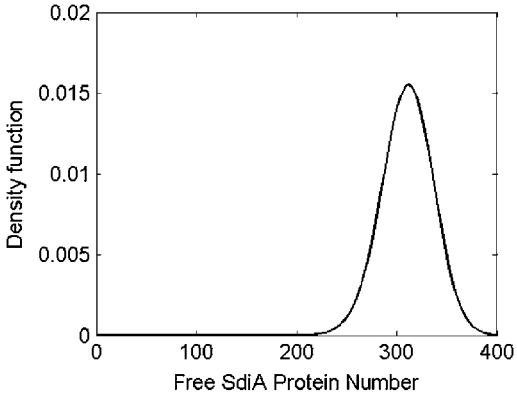

FIGURE 5 Distribution of protein monomer (free SdiA) number for the simplified model with high rate of binding of dimer to DNA0 ($K = 10\ \text{min}^{-1}$ so $\kappa = 640$).

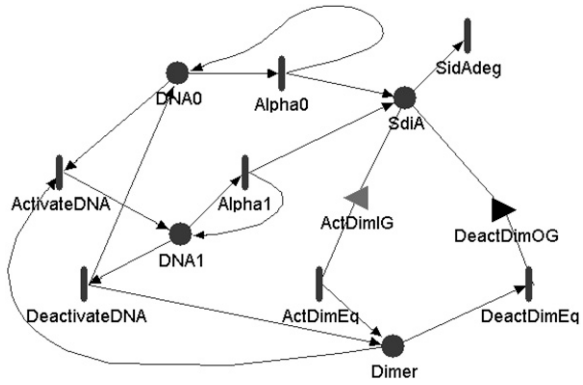

FIGURE 6 SPN for simplified model.

reactants and products of the corresponding reactions involve two molecules. The Mobius software package is used to simulate the SPN. This software is available from the Performability Engineering Research Group (PERFORM) at the University of Illinois at Urbana-Champaign (25). Fig. 6 and all other figures in this article showing SPN representations were created in Mobius.

The result of the SPN stochastic simulation after 50 min is the bar graphic in Fig. 3, which matches the analytically obtained steady-state distribution (*solid curve*) very well. The analytical steady-state distribution is derived from master equations with the two approximations mentioned above (11). Both master equations and SPN suppose that the reactions are Markov chains in which the probability of reaction depends on current molecule numbers (2,24). Hence, since the two methods are based on the same stochastic foundations, one would expect the results to match well.

The analytical steady-state distribution can predict the long-run trend of stochastic simulation. The distribution from the stochastic simulation is not a steady-state distribution. Unlike the steady-state distribution, which is independent of initial condition, stochastic simulations with different initial conditions may result in different distributions. However, as long as the system evolves for long enough time, stochastic simulation can reach the same steady-state distribution from any initial condition. In Fig. 3 the two distributions match very well because we choose a proper initial condition. From this initial condition the cell can almost reach steady state within a single generation time. (The time length of 50 min is an approximate generation time of *E. coli* cells.)

Deterministic simulation

With small-noise and fast-noise approximations, the deterministic ODE for the simplified model is (11)

$$\frac{\mathrm{d}x}{\mathrm{d}(\delta t)} = \frac{ba_0 + x^2}{b + x^2} - x. \tag{18}$$

Note that scaled time δt should be used here, while in Kepler and Elston (11) the corresponding equation uses t. The ODE solution is shown in Fig. 7.

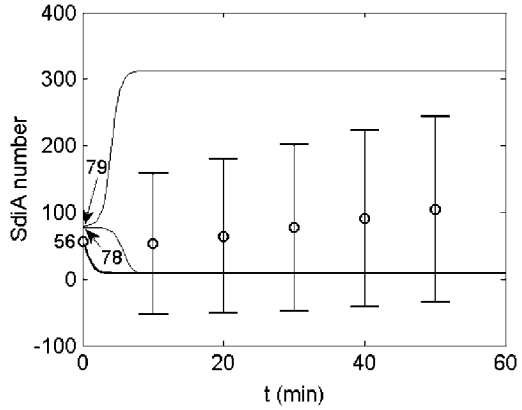

FIGURE 7 Protein monomer (free SdiA) number versus time for the simplified model. (*Circles*) Mean of stochastic simulation; bar graphs show the standard deviation. (*Bold solid curve*) Deterministic path from 56 initial protein monomers, which is the same initial condition as that of stochastic simulation. (*Thin solid curves*) Deterministic paths from the initial conditions above and below the switch point.

Comparison between the deterministic and the stochastic simulation indicates the existence of bifurcation. For monostable systems, the stochastic mean is close to the deterministic path, and as a result such systems can be adequately described in deterministic form. However, the system we are examining is a bistable system. Therefore, from the same initial condition with protein monomer number equal to 56, the stochastic mean (*circles* in Fig. 7) does not match the deterministic path (*bold solid curve* in Fig. 7). We can also see that the standard deviation for the stochastic simulation (*bar graph* in Fig. 7) is very large, which indicates the bifurcation of the cell population.

The deterministic model has a switch point between 78 and 79 molecules in Fig. 7. The two deterministic paths starting above and below the switch point (*thin solid curves*) verify the existence of bifurcation. Initial number of protein monomer higher than the switch point leads to the higher stable steady state. And lower initial protein monomer number leads to the lower stable steady state. Of course, for the stochastic simulation, although the initial condition is below the deterministic switch point, a cell subpopulation at high SdiA number is also obtained, as clearly shown in Fig. 3.

ANALYSIS OF FULL MODEL

Stochastic simulation

The SPN of the full model is shown in Fig. 8. As we mentioned before, gene expression is separated into transcription and translation. Activation and dimerization are not considered as one step here. The result shown in Fig. 9 is the distribution of protein monomer number in cells with 50-min age for the full model.

The distribution in Fig. 9 is very similar to that of the simplified model. First, as in the stochastic simulation result

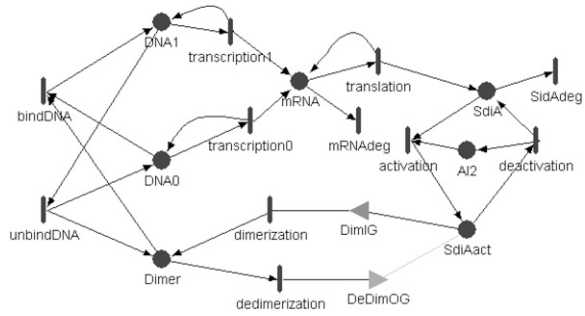

FIGURE 8 SPN for full model.

for the simplified model, there is bistability in the distribution for the full model. Hence the full model also predicts the stochastic bifurcation for the positive autoregulation network, which verifies the conclusions from the simplified model. Second, the shapes of the distributions from the two models are quite similar. For both the simplified and the full model, the peaks at the lower number of protein monomer are tall and thin, indicating that within this group of cells, the fluctuations of protein number are small. The peaks at the higher protein number are more flat, so the fluctuations in those cells are relatively large.

The higher-number peaks in Figs. 3 and 9 are not in exactly the same position. This is because of the second assumption for the simplification from the full model to the simplified model. In the simplified model, the free signal molecule number is assumed to be equal to 500. In the full model, we assume the total number of all forms of signal molecules, including free signal molecules and signal molecules in the form of activated protein complexes and dimers, to be a constant equal to 600. However the difference between the two peak values is not significant (at ~50). Therefore the simplification of the full model does not change the peak positions significantly. (It should be noted that the selection of 600 as the total number of signal molecules was also guided by the deterministic steady-state analysis discussed in the next section.)

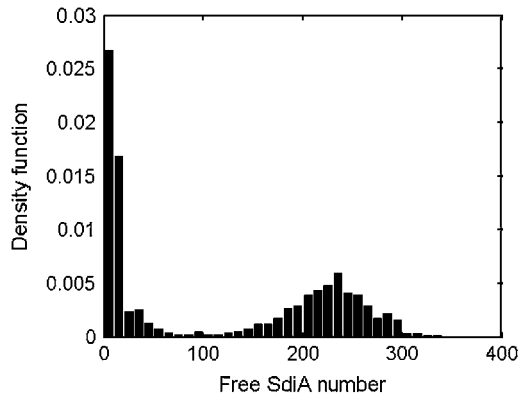

FIGURE 9 Distribution of protein monomer (free SdiA) number for the full model.

An important reason for using the model simplification is that for the full model we cannot obtain an analytical distribution as we did for the simplified model. So we have to simplify the full model first, and obtain a mathematical description for the simplified model. Such a description of the distribution allows a detailed analysis of the effect of different parameters on the presence or not of stochastic bistability (11).

As we discussed earlier, the analytical distribution matches the simulation distribution for the simplified model very well, and the stochastic simulation distributions of the full and the simplified models are very similar. Hence, the result obtained from mathematical analysis of the simplified model can be expected to be applicable to the full model.

Deterministic steady-state analysis

In the previous section, we determined that the analytical distribution for the simplified model can suggest the steady-state distribution for the full model. SPN simulation can further provide distributions for the full model at finite times. However, the direct mathematical analysis for the full model is very difficult. Often a deterministic analysis can indicate a pathway bifurcation that may also be present in the stochastic system. Here we examine the use of deterministic analysis as an indicator for the bistable distribution of the full model.

First, for simplicity of notation, a variable z is defined as equal to the number of dimers:

$$z = [\text{Dimer}]. \tag{19}$$

At the deterministic steady states, protein production and degradation in Fig. 1 are in balance:

$$\alpha_0[\text{DNA0}] + \alpha_1[\text{DNA1}] - \delta[\text{SdiA}] = 0. \tag{20}$$

In Eq. 20 the third term is the degradation rate of protein. The summation of the first two terms is the production rate of protein, in Eq. 7. Equation 7 is used for the simplified model, but at the deterministic steady states it can also be used for the full model, because the simplification was based on the assumption that [mRNA] is at steady state.

In Fig. 1, the transition of gene states of the full model is at equilibrium:

$$\frac{[DNA0]}{[DNA1]} = \frac{\beta}{z}. \tag{21}$$

Furthermore, the gene states can be normalized to a single gene. Then we have

$$[DNA0] + [DNA1] = 1. \tag{22}$$

From Eqs. 20–22, the steady-state protein monomer number is obtained as

$$[SdiA] = \frac{\alpha_0}{\delta}\frac{\beta}{\beta+z} + \frac{\alpha_1}{\delta}\frac{z}{\beta+z}. \tag{23}$$

At steady state, activation, and dimerization in Fig. 1 are at equilibrium. So we obtain

$$\theta_a = \frac{[SdiA][AI2]}{[SdiAact]}, \tag{24}$$

$$\theta_d = \frac{[SdiAact]^2}{[Dimer]}. \tag{25}$$

Multiply Eq. 25 by the square of Eq. 24, and use Eq. 19 to obtain

$$[AI2] = (\theta z)^{1/2}[SdiA]^{-1} = (\theta z)^{1/2}\frac{\delta(\beta+z)}{\alpha_0\beta+\alpha_1 z}, \tag{26}$$

where $\theta = \theta_a^2\,\theta_d$.

From Eq. 25, the number of activated protein complex is expressed as

$$[SdiAact] = (\theta_d z)^{1/2}. \tag{27}$$

From Eqs. 21 and 22, the number of DNA binding dimer is

$$[DNA1] = \frac{z}{\beta+z}. \tag{28}$$

When we consider the total signal molecule number, we must count free signal molecules, signal molecules binding to the activated protein complex, and those binding to dimers (including the dimer binding DNA). Therefore the total number of signal molecules is

$$\begin{aligned}[AI2total] &= [AI2] + [SdiAact] + 2[Dimer] + 2[DNA1] \\ &= (\theta z)^{1/2}\frac{\delta(\beta+z)}{\alpha_0\beta+\alpha_1 z} + (\theta_d z)^{1/2} + 2z + \frac{2z}{\beta+z}.\end{aligned} \tag{29}$$

Equation 29 is used to obtain a plot of [AI2total] as a function of z, shown in Fig. 10. The full model assumes that the total number of signal molecules is a constant. It can be found from Fig. 10 that when this constant is within a certain range, the curve has more than one intersection point with the

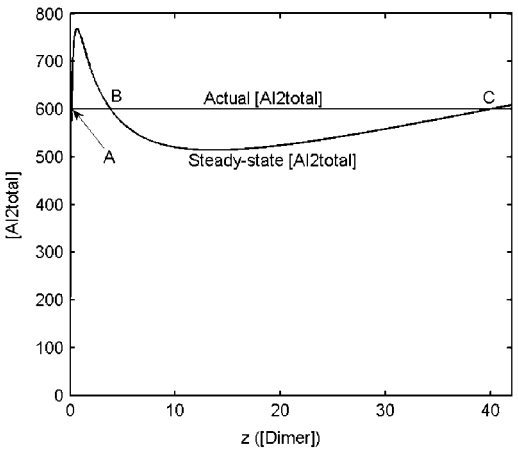

FIGURE 10 Total signal molecule number ([*AI2total*]) versus z([*Dimer*]).

horizontal line, which means that the deterministic model has multiple steady states. In these cases, pathway bifurcation is possible and the stochastic distribution may also be bistable. Therefore the curve for [AI2total] versus z can indicate the existence of bistability.

Let us suppose that the total AI-2 number is equal to 600, as it was in the stochastic simulation. Then the points A–C in Fig. 10 represent three steady states. The protein monomer number at each steady state is calculated from Eq. 23 and shown in Table 4. Consider a point between A and B. In that case the steady-state total signal number is higher than the actual total signal number. This means there are not enough signal molecules, so the dimers and activated complexes will disassociate and their number will decrease. The system goes to point A. As for a state between B and C, the actual signal number is higher than the steady-state total signal number. This means there is a surplus of signal molecules, so more activated complexes and dimers will form and the state goes to point C. Therefore A and C are stable states for the deterministic model and B is a switch point. The protein monomer numbers for A and C are close to the peaks in Fig. 9.

DISCUSSION

We present two gene-regulation network models, both of which contain positive autoregulation. In the full model, shown in Fig. 1, the positive autoregulation is mediated through signal molecules. The gene product protein binds signal molecules first, then forms dimers which subsequently bind the gene and improve its expression. In the simplified model, shown in Fig. 2, no signal molecule is involved. The gene product protein forms a dimer directly, then the dimers upregulate the gene expression. The simplified model can be obtained from the full model with two simplifications.

Both models can exhibit bistable distributions of protein numbers as shown in Figs. 3 and 9, which means that an initially uniform cell population can evolve into two subpopulations. Cells in one population have low protein numbers, while cells in the other have high protein numbers. The bistability of the two models is the result of the positive autoregulation mechanism and their stochastic nature. However, obtaining the bistable protein number distributions for the two models involves different levels of difficulty. For comparison purposes, on a standard single processor, our SPN simulation takes >10 h for the full model, when one thousand batches (single path simulations) are used to obtain the distribution. For the simplified model, the same type of simulation takes <1 h. The calculation of the analytical steady-state distribution for the simplified model is essentially instantaneous, but it requires the additional small-noise approximation, which requires large protein number, and the fast-transition approximation, which needs fast but finite transition between the gene states.

TABLE 4 Steady states when [*AI2total*] = 600

Steady states	A	B	C
[*Dimer*]	0.1394	3.8215	39.9287
[*SdiA*]	9.9422	55.0765	238.4189

Let us further discuss the meaning and effect of the two simplifications we introduced. The first is that the transcription and translation steps are merged into one single step of gene expression, which assumes the mRNA number is always at steady state and can be calculated from Eq. 6. This assumption has two effects. One is that the dynamic simulation is faster because the time it takes for mRNA number to reach steady state is ignored. This effect is not important though, when considering steady state or long times in simulation. The second effect is caused by ignoring the fluctuation of mRNA number, because with this assumption, when the full model is at steady state and the state of the gene does not change, the mRNA number is fixed. The mRNA fluctuation does not play an important role in the presence or not of bistability. In the simplified model, the fluctuation in the gene states can induce bistability, while the fluctuation of monomer number can reduce bistability (11), suggesting that the fluctuation in gene states is the main source of bistability. The results in Fig. 3–5 have confirmed this point. The fact that the distributions in Fig. 3 and Fig. 9 are both bistable and have similar shapes, indicates that mRNA fluctuation is mostly unrelated to bistability and does not have a significant influence. The peaks in Fig. 3 are a little thinner and higher than in Fig. 9, indicating that ignoring the mRNA fluctuation only causes a slight reduction of the protein number fluctuation. Therefore, the second effect is also not important and the assumption behind this simplification is good.

The second simplification is that the activation and dimerization of protein monomers are combined into one step. Here it is assumed that the number of unbound signal molecules is a constant. This assumption can influence molecule numbers at steady state. This is why the position of the peak for the higher protein number in Fig. 3 has an observable difference from the position of the corresponding peak in Fig. 9. To avoid unacceptable differences, the steady-state free signal number for the full model cannot be too far away from the assigned constant value in the simplified model. For this simplification, there is an additional assumption that activation is much faster than dimerization so that it can be considered at quasi-equilibrium. However, usually even dimerization is fast enough to be thought at equilibrium (11). So this assumption is expected to be reasonably satisfied.

The discussion above indicates that the number of signal molecules present can be a determining factor on the presence or not of stochastic bistability. The difficulty involved in the direct mathematical analysis of the full stochastic model led us to consider the use of the corresponding deterministic model. We saw that such an analysis can be a very useful indicator of stochastic bistability. Furthermore, we were able to directly focus on the effect of the number of signal mol-

ecules as shown in Fig. 10. This supports that further work is needed in connecting the signal-mediated positive autoregulation model with uptake models for the signal molecules. We have such work underway for the AI-2 uptake in *E. coli*.

Finally, we note that the transient simulation results in Fig. 7 indicate that stochastic simulation is essential for a bistable system. For a monostable system, the mean for the stochastic simulation usually matches the deterministic path and the standard deviations are relatively small. Therefore, deterministic simulation is usually sufficient and the time-consuming stochastic simulation may not be necessary. However, for a bistable system, the mean path for the stochastic simulation lies between the two deterministic stable steady states, while the deterministic simulation only goes to one of the two stable steady states depending on the initial condition. As for the standard deviations, they are very large as a result of the development of two distinct cell subpopulations.

Verification of the possible presence of bistability in SdiA regulation might be based on the involvement of the protein in cell division. A bifurcation in a division-promoting protein number distribution may lead to a cell age distribution with distinct subpopulations that exhibit differences in characteristics that may be experimentally observable.

We are grateful to the Performability Engineering Research Group (PERFORM) at the University of Illinois at Urbana-Champaign for the use of the software package Mobius (http://www.mobius.uiuc.edu).

This work was supported by National Science Foundation grant No. BES-0222687.

REFERENCES

1. Raser, J., and E. O'Shea. 2005. Noise in gene expression: origins, consequences, and control. *Science*. 309:2010–2013.
2. Rao, C., D. Wolf, and A. Arkin. 2002. Control, exploitation and tolerance of intracellular noise. *Nature*. 420:231–237.
3. McAdams, H., and A. Arkin. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* 15:65–69.
4. McAdams, H., and A. Arkin. 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*. 94:814–819.
5. Zhou, B., D. Beckwith, L. Jarboe, and J. Liao. 2005. Markov chain modeling of pyelonephritis-associated pili expression in uropathogenic *Escherichia coli*. *Biophys. J.* 88:2541–2553.
6. Li, J., L. Wang, Y. Hashimoto, C.-Y. Tsao, T. Wood, J. Valdes, E. Zafiriou, and W. Bentley. 2006. A stochastic model of *Escherichia coli* AI-2 quorum signal circuit reveals alternative synthesis pathways. *Mol. Sys. Biol.* DOI:10.1038/msb4100107.
7. Arkin, A., J. Ross, and H. McAdams. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics*. 149:1633–1648.
8. Isaacs, F., J. Hasty, C. Cantor, and J. Collins. 2003. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*. 100:7714–7719.
9. Becskei, A., B. Seraphin, and L. Serrano. 2001. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* 20:2528–2535.
10. Ferrell, J. 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Chem. Biol.* 6:140–148.
11. Kepler, T., and T. Elston. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.
12. Shadel, G., and T. Baldwin. 1991. The *Vibrio fischeri* LUXR protein is capable of bidirectional stimulation of transcription and both positive and negative regulation of the *luxr* gene. *J. Bacteriol.* 173:568–574.
13. DeLisa, M., C.-F. Wu, L. Wang, J. Valdes, and W. Bentley. 2001. DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*. *J. Bacteriol.* 183:5239–5247.
14. Fuqua, W., S. Wintans, and E. Greenberg. 1994. Quorum sensing in bacteria: the Luxr-Luxi family of cell density-responsive transcriptional regulators. *J. Bacteriol.* 176:269–275.
15. Ahmer, B. 2004. Cell-to-cell signaling in *Escherichia coli* and *Salmonella enterica*. *Mol. Microbiol.* 82:933–945.
16. Kanamaru, K., I. Tatsuno, T. Tobe, and C. Sasakawa. 2000. SdiA, an *Escherichia coli* homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohemorrhagic *Escherichia* coli o157:h7. *Mol. Microbiol.* 38:805–816.
17. DeLisa, M., J. Valdes, and W. Bentley. 2001. Mapping stress-induced changes in autoinducer AI-2 production in chemostat-cultivated *Escherichia coli* k-12. *J. Bacteriol.* 183:2918–2928.
18. Daga, A. 2001. Quorum sensing in silico: kinetic modeling in *Escherichia coli* through stochastic Petri nets. Master of Science thesis. Department of Chemical Engineering, University of Maryland, College Park, MD.
19. Ptashne, M. 1992. A Genetic Switch: Phage λ and Higher Organisms, 2nd Ed. Cell Press and Blackwell Scientific Publications, Cambridge, MA.
20. Record, M. T., Jr., W. Reznikoff, M. Craig, K. McQuade, and P. Schlax. 1996. *Escherichia coli* RNA polymerase ($e\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. *In Escherichia coli* and *Salmonella*, Vol. 2, 2nd Ed. F. Neidhart, R. Curtiss, J. Ingraham, E. Lin, K. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. Umbarger, editors. ASM Press, Washington, DC.
21. Pedersen, S., and S. Reeh. 1978. Functional mRNA half lives in *Escherichia coli*. *Mol. Gen. Genet.* 166:329–336.
22. Varshavsky, A. 1996. The *N*-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA*. 93:12142–12149.
23. Chen, W., J. Bailey, and S. Lee. 1991. Molecular design of expression systems: comparison of different repressor control configurations using molecular mechanism models. *Biotechnol. Bioeng.* 38:679–687.
24. Goss, P., and J. Peccoud. 1998. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA*. 95:6750–6755.
25. Sanders, W. 2004. Mobius User Manual, Ver. 1.5.0. PERFORM Performability Engineering Research Group, University of Illinois at Urbana-Champaign, Illinois.