A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics

ANDRE RIBEIRO,¹,³ RUI ZHU,² and STUART A. KAUFFMAN¹

ABSTRACT

A stochastic genetic toggle switch model that consists of two identical, mutually repressive genes is built using the Gillespie algorithm with time delays as an example of a simple stochastic gene regulatory network. The stochastic kinetics of this model is investigated, and it is found that the delays for the protein productions can highly weaken the global fluctuations for the expressions of the two genes, making the two mutually repressive genes coexist for a long time. Starting from this model, we propose a practical modeling strategy for more complex gene regulatory networks. Unlike previous applications of the Gillespie algorithm to simulate specific genetic networks dynamics, this modeling strategy is proposed for an ensemble approach to study the dynamical properties of these networks. The model allows any combination of gene expression products, forming complex multimers, and each one of the multimers is assigned to a randomly chosen gene promoter site as an activator or inhibitor. In addition, each gene, although it has only one promoter site, can have multiple regulatory sites and distinct rates of translation and transcription. Also, different genes have different time delays for transcription and translation and all reaction constant rates are initially randomly chosen from a range of values. Therefore, the general strategy here proposed may be used to simulate real genetic networks.

Key words: gene regulatory networks (GRNs), genetic toggle switch, Gillespie algorithm, non-Markov Processes, random Boolean networks, stochastic dynamics

1. INTRODUCTION

Since the genes of the human genome (as well as others) have now been identified, one of the next steps is to understand the behavior of the genetic regulatory networks. As an example, the human genome has between 30,000 and 45,000 genes whose activities are regulated by a network of their own products. The genome can be seen as a parallel processing nonlinear dynamical system. This system has been modeled by several approaches, such as random Boolean networks (Kauffman, 1969; Aldana,
2003), differential equations (Mestl et al., 1995), piecewise linear differential equations (Glass, 1975), and stochastic equations (McAdams and Arkin, 1997; Ramsey et al., 2005). The first approach to model Gene Regulatory Networks (GRNs) was made by Stuart Kauffman, who introduced the Boolean network model. In such models, gene states are represented by binary variables with two possible values, 1, when a gene is being expressed, and 0 if not. In such models, all genes states are synchronously updated. The state of a gene is regulated by genes directly connected to it, to which a random Boolean function is assigned, determining its state in the next time step from the previous inputs’ states. Although its dynamics is very rich and allows applying the ensemble approach, useful at determining general dynamical behavior, real genetic networks are not synchronous and genes’ level of expression are not binary quantities. Also, the Boolean model does not allow a correct simulation of “stochasticity” or molecular noise because it mimics noise by randomly bit flipping genes states (Kauffman, 1993). On the other hand, when the system has many molecules, its chemical dynamics—that is, the variations of the concentrations of the chemical species present—can be computed approximately using continuous differential equations. However, since in the GRN model here proposed, genes are treated as chemical species, and, in the real systems, exist only in very small quantities (usually one or two copies of the same gene), a “mean field” approach is not an accurate approach. For example, until now, attempts to simulate noise using Langevin equations were not very successful (Toulouse et al., 2005). The favored approach here is the Gillespie algorithm (Gillespie, 1976, 1977), recently used to model genetic networks (Ramsey et al., 2005), and gene expression (Kierzek et al., 2001). Using the Gillespie algorithm one attains temporal stochastic dynamics by calculating the probability of each possible chemical reaction event and the resulting changes in the number of each molecular species. The Gillespie algorithm has proved reasonable to model transcription and translation as “single step” discrete molecular events. To account for the fact that these are not single step reactions, an improvement of these models consists in introducing time delays in the translation and transcription reactions. Accordingly, we use a delay stochastic simulation algorithm, where translation and transcription are modeled as time delayed reactions. We use this algorithm to model simple genetic networks, and then increasingly complex networks. Our main goal is to create a realistic model of GRNs, which allows an ensemble approach (Kauffman, 1993) to study general properties of these networks, such as dependence on the topology (the chemical reactions possible) and initial state (initial concentrations). On the other hand, if there are sets of states to which the system is driven, we can simulate GRNs for the purpose of inferring the structure from the functioning and characterize its dynamics (i.e., if they are ordered, critical, or chaotic). This paper extends a simple GRN modeling strategy previously proposed by the authors (Ribeiro et al., 2006) and tests this strategy by doing a variety of numerical simulations of small GRNs.

2. PROBLEM FORMULATION AND SOLUTION

The main problem that we deal with here is to how to model GRNs with stochastic dynamics taking into consideration the time duration of reactions and where genes affect one another with realistic interactions. GRNs involve very complex biological processes, such as transcription and translation. Also, the products of these processes with feedback and feed forward effects regulate the gene expression itself. Many chemical pathways are involved in these processes. It is not yet possible, to simulate the entire gene and protein network, or the whole body chemistry in a single system. Reasonable reduction methods are needed to study the dynamics of GRNs. Here, we have transcription and translation, modeled as non-Markov events. A non-Markov process allows multiple step reactions to be modeled as a single step (Gibson and Bruck, 2000). To do so, it models each single step as a reaction with time delays in the appearance of products. This can reduce dramatically the number of reactions and steps, and thereby diminish the complexity of the model. As mentioned, the other problem in modeling GRNs is that there are only a small number of interactive molecules, namely the number of copies of a gene, each with only one promoter site. This gives rise to large fluctuations in behavior due to high molecular noise. To take into consideration this key dynamical feature, we use the Gillespie algorithm. Accordingly, our method for simulating biochemical processes involved in the GRN is to use the Gillespie algorithm to drive the system stochastically, with time delayed reactions, to skip the multiple steps in transcription and translation (using a single delayed reaction for both these processes). Also, we skip the whole cascade of events such as proteins networks and pathways,
and suppose that gene expression products feedback into the GRN. To build the relationships among genes (i.e., the network topology), we assign each gene expression product to another gene’s promoter, acting as its activator or inhibitor. For these connections, just like Boolean networks, a regulating function is assigned, which will determine, depending on the inputs state, the genes’ next states. The advantage of Boolean networks is that they allow assigning random Boolean functions, thereby creating very diverse dynamics, depending on initial conditions. This kind of advantage and level of complexity can be attained here in three ways: (i) either by allowing some reactions between the genes expression products, creating complex molecules, and then assign such molecules to a randomly chosen gene promoter site, choosing what effect that complex has on that gene expression level when connected to the promoter site; or, (ii) by allowing a gene to have more than one binding site and randomly assigning the effects of all the possible binding combinations as activators or inhibitors, or, finally, (iii) by allowing different genes expression products to bind to a single binding site, each with its own effect. These features allow that all Boolean functions and all topologies of interactions have equivalents in this model. For simplicity, in the present work, we restrict ourselves to monomers and dimers, and each gene has only one activator and inhibitor. However, the approach is easily extendable to more complex networks. We now present the general formulation of GRNs.

3. RESULTS AND DISCUSSION

We begin by building a toggle switch model using this approach, and then generalize it to networks, where topology and initial concentrations are initially chosen to allow these systems to be studied using the ensemble approach.

3.1. Stochastic simulation of gene expression

The gene expression process contains transcription and translation. These are very complex biological processes that contain a series of elementary chemical reactions. Series of elementary chemical reactions can be, in certain cases, modeled by a single time delayed reaction (Gibson and Bruck, 2000). We use an algorithm developed by Roussel and Zhu that models the whole gene expression process in a single delayed chemical reaction. A similar approach, for the same purpose, was independently developed in (Brastsun et al., 2005). To represent the production of the protein resulting from gene expression, we use the following Equation (1) where Pro\(_i\)\((t)\) is the promoter site, RNAP is the RNA Polymerase, and \(r_i\) is the resulting protein created from the translation of the RNA formed in transcription. The probability rate constant for the RNAP to bind to the DNA is represented by \(k_i\). The values \(\tau\) are the times taken for each of the products of the reaction to become available in the system. The \(n_i\) constant is an integer associated with the rate of translation of each distinct mRNA and the mRNA rate of transcription. The higher its value, the more times will a single mRNA be translated.

\[
\text{RNAP}(t) + \text{Pro}_i(t) \xrightarrow{k_i} \text{Pro}_i(t + \tau_1) + \text{RNAP}(t + \tau_2) + n_i \times r_i(t + \tau_3)
\]

For simplicity, in the examples provided in the present article, \(n_i\) is equal to one. Yet these values will vary from system to system, and even vary over time, depending on the number of copies of a specific mRNA, and the number of ribosomes binding to the mRNA and will influence the dynamics. Using this equation to represent transcription and translation, we built a stochastic model of the toggle switch, presented below.

3.2. The toggle switch model with stochastic dynamics

The gene toggle switch system (Gardner and Collins, 2000) consists of two repressor molecules and two promoter sites. The repressor that is transcribed by the other gene inhibits each gene promoter site. The inducers have the ability to reactivate the inhibited promoter. With the following chemical reactions, we simulate a toggle model with two identical genes using the Gillespie algorithm:
Equations (2) and (3) represent the chemical processes of gene expression. Reactions (4) and (5) are repressing processes of the promoters by forming Pro$_1$r$_2$ and Pro$_2$r$_1$, and (6) and (7) reactivate the promoters’ expression ability with inducers, Ind$_1$ and Ind$_2$. The last two reactions, (8) and (9) are the decay processes of the gene expression products. As for $n_1$ and $n_2$, they are associated with the rates of translation, here equal in both genes; thus, we set them to 1 for simplicity. In our simulations, the stochastic rate constants of all reactions are equal to 1 s$^{-1}$, except the decay reactions, with a stochastic rate constant of 0.001 s$^{-1}$.

The delay times are random variables following some distribution. For simplicity, we use constant delays, and set delay $\tau_1$ to 1 s, $\tau_2$ to 20 s and $\tau_3$ to 10 s. The initial numbers of the reactants are: RNAP = 50, Pro$_1$ = 1, Pro$_2$ = 1 and Ind$_1$ = 1. All other elements are not present initially.

From these initial conditions, we attain (after some transients) a stable state where gene 2 is off and 1 is on (Fig. 1). By removing inducer 1 and introducing inducer 2, we toggle to the other possible stable state (gene 2 on, gene 1 off). The variation of $r_1$ and $r_2$ quantities in Figure 1 confirms the robustness of the toggle switch control mechanism. Initially, from 0 to 10 s, after the necessary transient, both $r_1$ and $r_2$ are expressed, due to reactions (2) and (3). After this transient, since only inducer 1 exists, which releases Pro$_1$ by reaction (6), only $r_1$ is now produced. At about time 100 s, due to the toggle of the inducers, $r_1$ production stops, and $r_2$ begins, for the same reasons. If both inducers are absent, $r_1$ and $r_2$ are not produced after an initial transient, corresponding to the state where both genes expression are repressed. Thus, it seems that deterministic attractors can be stable in the face of stochastic behavior. With both inducers present, two stable states are possible: one where gene 1 is on and gene 2 is off, and the other where gene 2 is on and gene 1 is off. In this case, if the decay rates are too small, the system settles in

![Toggle switch model](image-url)
Having both inducers present, 1 of each, the two genes switch on and off, depending on the concentrations, making the system state switch frequently. If there is intermediate decay, then the system will switch from one attractor to the other (Fig. 2). We find that the decay equations play a very important role in the states switching dynamics when both inducers exist (the decay rate constant used here is $0.001 \, s^{-1}$ for $r_1$ and $r_2$). The higher are the probability rate constants for decay, the more likely it is for the genes to toggle. This is because according to the Gillespie algorithm, increasing the probability rate constant of a reaction it becomes more likely that the reaction is chosen. Thus, as one gene represses the other, the corresponding repressor will start decaying faster with the higher probability rate constant, indicating more probability for the gene to be toggled by the other. Since another study in real gene networks shows that repression can be made by homo-dimers (dimmers of two equal molecules) of the protein products of translation (Burz et al., 1994), we now suppose that the proteins $r_1$ and $r_2$ form homo-dimers and that these act as repressors, instead of the monomers ($r_1$ and $r_2$ alone). Thus, we introduce the following reactions:

\[
\begin{align*}
    r_1(t) + r_1(t) & \xrightarrow{k_9} r_{11}(t) \\
    r_{11}(t) & \xrightarrow{k_{10}} r_1(t) + r_1(t) \\
    r_2(t) + r_2(t) & \xrightarrow{k_{11}} r_{22}(t) \\
    r_{22}(t) & \xrightarrow{k_{12}} r_2(t) + r_2(t)
\end{align*}
\]

Also, we replace in equations (4), (5), (6), and (7), $r_1$ and $r_2$, respectively, by $r_{11}$ and $r_{22}$. As for the probability rate constants, we set the dimers production at $1 \, s^{-1}$, and their dissociation reactions at

![Toggle switch model—both inducers present](image1)

**FIG. 2.** Having both inducers present, 1 of each, the two genes switch on and off, depending on the concentrations, making the system state switch frequently.

![Toggle switch model with dimmers](image2)

**FIG. 3.** Same model as in Figure 1, with both using dimers ($r_{11}$ and $r_{22}$) as repressors. Dimers formation and dissociation are given in equations (10), (11), (12), and (13).
FIG. 4. Using dimers as repressors with the two inducers initially present. As before, the system state switches frequently.

Using dimers as repressors with the two inducers initially present. As before, the system state switches frequently. 

To be compared with the monomer case shown in Figure 1, the corresponding dimer case is given in Figure 3. Notice that we choose to follow the quantity of the dimers in this case, not the monomers, since these are now the ones responsible for repression, and the genes states will depend directly on them, and not on the monomers population (e.g., if the probability rate constant of the dimmers binding to the promoter sites was null, no repression would occur independently of the quantity of monomers). As expected, the dynamic is faster, that is, more reactions occur in the same time interval. Due to this, \( r_{11} \) quantity is able to overcome \( r_{22} \) quantity, while as seen in Figure 1, there is not enough time for \( r_1 \) quantity to become larger than \( r_2 \) quantity, having both experiments the same total running time. One of the reasons for the stronger fluctuations is that the total number of dimers is smaller than that of the monomers (of the first case). The other reason for these stronger fluctuations is the higher level of “decay” of the dimers into monomers, than the decay rate of monomers in the first experiment. For these reasons \( r_{11} \) and \( r_{22} \) absolute fluctuations (Fig. 3) are higher than \( r_1 \) and \( r_2 \) in Figure 1. For comparison, we also tested the dimers system with similar conditions of those in Figure 2 (Fig. 4). Again, the system dynamics is very similar, characterized only by faster toggling and stronger fluctuations, for the reasons mentioned above. The toggle becomes less observable due to these higher fluctuations.

In order to explore the effects of time delays, we set to null the three delays for each gene in the dimer case as in Figure 4. A representative simulation is shown in Figure 5. Different from Figure 4 with delays, the non-delay case in Figure 5 shows clearly that one gene represses the other. A plot of the standard deviations of \( r_{11} \)’s population versus time for the delay and non-delay cases is given in Figure 6. Since the two genes are identical, the standard deviation evolutions of \( r_{11} \) and \( r_{22} \) are almost the same for each case, thus we plot only \( r_{11} \)’s standard deviation for the delay and non-delay cases. The standard deviations for

FIG. 5. Same as in Figure 4, except that all three delays are set to zero.
the non-delay toggle switch are much higher (due to stronger fluctuations) than those for the delay toggle switch, indicating that the delays highly weaken the global fluctuations in the toggle switch model. Note that the term of global fluctuations here means differences in state (numbers of $r_{11}$ and $r_{22}$ molecules) between different simulations at the instants when measures were taken. The fluctuations mentioned earlier are local fluctuations, the result of differences in state during a short period of time for a single simulation. We tested the three types of delays separately in the model. We found that $\tau_3$'s, the delays for protein productions, are most responsible for the global fluctuations variation. This is due to the fact that neither of the two genes with long delays for protein productions can be turned off immediately by repressors' binding to their promoters since several “baby” proteins are stored (protected) in the waiting list (a list of delayed events). It could be a very important effect, in gene regulations, for delayed protein productions to make repressive genes coexpress for a long time. However, the delays cannot change the fates of the two genes: one becomes statically on, and the other statically off, because there are only two stable states in a long term for this system. Notice that, in the deterministic, continuous description of this system, only one state is observable for these initial conditions, representing the average of the two stable states that in the stochastic dynamics toggle constantly until, after a long transient, settling on one of the stable states. We have confirmed this by doing long simulations for the delay toggle model. Accordingly, from this point of view, the delays can adjust the ability for genes to survive in gene regulations: long delays may imply long-term existing (that is, ability to coexpress during a more or less long transient). Based on these illustrative examples we now move on to designing a general framework for the simulation of GRNs.

3.3. A general set of reactions

To create a model simulator able to generate stochastic GRNs, we consider, for the sake of simplicity, that each gene has a single promoter site and can either be directly activated (15) or repressed (16), by one activator or repressor, and possessing, independently of these two states, a basic level of expression (14). Reaction (16) of repression is reversible, through reaction (17). Also, gene expression products ($r_i$ for gene $i, i = 1, \ldots, N$) decay through equation (18). Each gene, although it has only one “promoter” site, can have multiple internal regulatory sites. We represent this by allowing more than one molecule to bind to the promoter site in its control region at the same time. Different combinations of bound molecules will have different effects in the gene transcription function. The number and identity of molecules that can bind and have an effect on the gene transcription function can be either randomly chosen or specifically assigned to mimic a known network. Also, gene expression products can combine forming multimers (dimers in our case, for the sake of simplicity) through reaction (19), with the possibility of reverse reaction through reaction (20). Activation is possible through reaction (21). Also, indirect activation and repression are
possible due to these dimers creation, which will be part of the set of molecules that can act as activators and repressors. Finally, we allow that a gene may have no direct activator or repressor. Some parameters have their values chosen randomly at each independent simulation, although it’s possible to set them to desired values. Their values are chosen from a specified range of values, set at the beginning. Each reaction rate constant, time delays, rate of production of proteins [variable $n$ in Equations (14) and (15)], initial quantity of RNAP, monomers, dimers and the initial state of the promoter sites are set at the beginning of each simulation within certain intervals of values. From this, we developed a set of equations for each gene, which, from gene to gene, vary only in which molecules are repressors ($r_j$) and which molecules are activators ($r_w$). Notice that there can be more than one equation of the type of (15), (16) and (17) for a single gene, allowing more than one activator or repressor.

\begin{align*}
RNAP(t) + Pro_i(t) & \xrightarrow{k_{1,1}} Pro_i(t + \tau_{1,1}) + RNAP(t + \tau_{1,2}) + n_i \times r_i(t + \tau_{1,3}) \hspace{1cm} (14) \\
RNAP(t) + Pro_i r_w(t) & \xrightarrow{k_{1,2}} Pro_i r_w(t + \tau_{1,1}) + RNAP(t + \tau_{1,2}) + n_i \times r_i(t + \tau_{1,3}) \hspace{1cm} (15) \\
Pro_i(t) + r_j(t) & \xrightarrow{k_{1,3}} Pro_i r_j(t) \hspace{1cm} (16) \\
Pro_i r_j(t) & \xrightarrow{k_{1,4}} Pro_i(t) + r_j(t) \hspace{1cm} (17) \\
r_i(t) & \xrightarrow{k_{1,5}} \hspace{1cm} (18) \\
r_i(t) + r_z(t) & \xrightarrow{k_{1,6}} r_i z(t) \hspace{1cm} (19) \\
r_i z(t) & \xrightarrow{k_{1,7}} r_i(t) + r_z(t) \hspace{1cm} (20) \\
Pro_i(t) r_j(t) + r_w(t) & \xrightarrow{k_{1,8}} Pro_i(t) + r_j(t) + r_w(t) \hspace{1cm} (21)
\end{align*}

In this model of general stochastic networks, the ability to generate an ensemble of networks comes from the fact that $i$ and $z$ are randomly chosen integer numbers (from 1 to $N$) different from one another and from the ability to form randomly chosen dimers. Thus, when the network of interactions between the genes is being created, since $j$ and $w$ are randomly chosen from all existing monomers and dimers, a different wiring diagram of influences is generated at the beginning of each independent simulation. Inserting equations of reactions between the $r_i$’s, allows the creation of both homogenous and heterogeneous dimers, trimers, and so on, which would act as activators and inhibitors. In general, we can allow the creation of polymers from any combination of $r_i$’s. The random combination between the $r_i$’s in polymers and assignment as activators and inhibitors enables the same level of complexity as the one using random transfer Boolean functions. Also, the random assignment of these polymers to genes (and allowing more than one activator and inhibitor to each gene) corresponds to a random topology between genes. From this, it is now possible to do the ensemble approach (Kauffman, 2004), as in random Boolean networks models.

4. CONCLUSION

We built a stochastic genetic toggle switch model, the dynamics of which are driven by the Gillespie algorithm, as an example of a simple stochastic GRN. From this, we proposed a general strategy to model GRNs. The mechanisms illustrated here can be modified to adapt to a specific network. While the results presented in the examples might not represent real gene expression profiles, to do so would only require changing the mechanism but not the method, which is general. This flexible method can be used to mimic specific known gene regulatory sub-networks by providing the specific mechanism (topology and transfer functions), or to generate random artificial stochastic networks for an ensemble study of these networks’ general properties. A recent experimental paper (Gu et al., 2002) reported a library of gene networks with
varying connectivity between three genes and five promoters in *E. coli*. One of the interesting results is that the behavior of simple networks built out of a few, well-characterized components cannot always be inferred from connectivity diagrams alone. This may not be well explained by the Boolean-type models. However, it seems promising to see the insight of this experimental result by using the model proposed here because of the fact that each gene in our gene network can be unique with different probability rate constants, delay times and various possibilities of being directly or indirectly expressed or repressed. In addition, the proposed modeling strategy allows mimicking the key features of the Random Boolean networks approach, such as its richness in the wiring diagram and transfer functions, while also having a stochastic dynamics driven by the Gillespie algorithm, a good approximation of the chemical master equation. The mechanism can be very simple or very complex, depending on the system size, the complexity of the transfer functions that activate or inhibit genes, and the level of reactions between the expression products. Also, other features, such as protein networks, can be easily introduced. We believe that the method here proposed for building general stochastic GRNs allows realistic simulation of a genetic network. We also believe that this method will be useful for developing better inference algorithms of the structure and logic of gene networks by allowing the construction of realistic artificial networks. We are currently performing such analysis in small numbers of genes networks, yet we show that simulations up to a few thousands genes are not computationally very expensive. Finally, in the future, we wish to continue to tune the model and check its validity by studying real genetic networks, the dynamics of which are known.

**ACKNOWLEDGMENTS**

A.S.R. is currently funded by FCT, Portuguese Foundation for Science and Technology, with a Post Doctoral Fellowship (SFRH/BPD/14565/2003). R.Z. is supported by a grant from the Natural Sciences and Engineering Research Council of Canada 10174. We all thank iCORE, a funding agency of the Alberta government, and the Department of Physics and Astronomy of the University of Calgary.

**REFERENCES**


Address reprint requests to:
Dr. Andre Ribeiro
Institute for Biocomplexity and Informatics
Department of Physics and Astronomy
University of Calgary
2500 University Dr. NW
Calgary, Alberta, Canada T2L 1N4

E-mail: ARibeiro@ucalgary.ca