Stochastic and delayed stochastic models of gene expression and regulation

Andre S. Ribeiro$^{a,b}$

$^a$Computational Systems Biology Research Group, Dept. of Signal Processing, Tampere University of Technology, Finland.
$^b$Centre for Computational Physics, Univ. Coimbra, P-3004-516 Coimbra, Portugal

*Contact information:
   Mailing address
       Andre S. Ribeiro, Department of Signal Processing, Tampere University of Technology. P.O Box 553, 33101 Tampere, FINLAND
       E-mail address: andre.sanchesribeiro@tut.fi

ABSTRACT

Gene expression and gene regulatory networks dynamics are stochastic. The noise in the temporal amounts of proteins and RNA molecules in cells arises from the stochasticity of transcription initiation and elongation (e.g. due to RNA polymerase pausing), translation and post-transcriptional regulation mechanisms, such as reversible phosphorylation and splicing. This is further enhanced by the fact that most RNA molecules and proteins exist in cells in very small amounts. Recently, the time needed for transcription and translation to be completed once initiated were shown to affect the stochasticity in gene networks. This observation stressed the need of either introducing explicit delays in models of transcription and translation or to model processes such as elongation at the single nucleotide level. Here we review stochastic and delayed stochastic models of gene expression and gene regulatory networks. We first present stochastic non-delayed and delayed models of transcription, followed by models at the single nucleotide level. Next, we present models of gene regulatory networks, describe the dynamics of specific stochastic gene networks and available simulators to implement these models.
Keywords: Gene expression, Gene Regulatory Network, Stochastic, Time delay.

1. INTRODUCTION

Stochasticity in gene expression affects the functioning of cells and organisms and contributes to the phenotypic diversity in populations of genetically identical individuals [1][2][3][4]. Sources of this diversity, both genotypic and phenotypic, are of importance by creating variability on which natural selection can act upon [5][6].

Stochasticity in gene expression is used by cells to cope with fluctuating environments [7] and may drive certain types of cell differentiation [8][9]. Single cell experimental assays demonstrated the stochasticity of transcription [1][10][11][12] and translation [13][14] (reviewed in [15]). Importantly, the noise in gene expression propagates within genetic circuits implying that gene regulatory networks have intrinsic stochastic interactions and dynamics [16](see [17] regarding methods to quantify noise in proteins and RNA levels).

One of the earliest observations of phenotypic dependence on the stochasticity of a genetic circuit was reported in [18]. Escherichia coli cells lysogenic for a defective prophage λ-CI8B7 which was unable to kill the host were found to be in one of two regulatory phases: im+ and im-. The im+ phase corresponds to stable lysogeny: immunity is established and a superinfecting homoimmune phage is repressed. In the im- phase, the immunity is absent, some early genes of the prophage are unrepressed, and a superinfecting homoimmune phage is channeled towards the lytic response. Importantly, the same cellular behavior persists for many generations in either phase. The regulatory pattern of each phase is transmissible to an unlimited number of cell generations, but in a small portion of cells a spontaneous immunity phase shift occurs (from im + to im- or vice versa).
These observations are a consequence of the underlying stochastic dynamics of the genetic circuit regulating this mechanism [19]. Arkin and colleagues developed a stochastic kinetic model of this genetic circuit at the molecular level, whose dynamics was simulated by the stochastic simulation algorithm (SSA) [20]. They validated the model by confronting it with measurements in variable conditions and showed that when two independently produced regulatory proteins, acting at low cellular concentrations, competitively control a switch in a pathway, stochastic variations in their concentrations produce probabilistic pathway selection, so that an initially homogeneous cell population partitions into distinct phenotypic subpopulations [21]. They also showed how the pathogenic organism phage-λ uses this lysis-lysogeny decision circuit to randomly switch surface features to evade host responses. Finally, they showed that deterministic kinetics could not predict the statistics of the system due to its probabilistic outcomes.

Initially, stochastic models of genetic circuits assumed gene expression as an instantaneous process. The models’ dynamics was driven by the stochastic simulation algorithm (SSA) or variations (for a review see [22][23]). However, gene expression is a multi-step process that comprises thousands of consecutive chemical reactions. Several recent studies [24-32] have shown that the time that it takes for gene expression to be completed once initiated is not negligible in the dynamics of gene regulatory networks (GRN). This duration also has a random nature, varying significantly even between transcription events of the same gene [33][34] and thus, is also a source of stochasticity in genetic circuits.

Elongation is one step whose time length varies from one transcription event to the next. Events during transcriptional elongation such as RNA polymerase (RNAP) pausing, arrests, misincorporations and editing, pyrophosphorolysis, and premature
terminations, are an important source of noise in transcript levels [28][35][36][37]. RNAp pausing is particularly important by contributing with long delays whose occurrence is probabilistic [35]. Since the rate of occurrence and duration of pauses are sequence-dependent, this regulatory mechanism of transcriptional dynamics is likely evolvable. The further dependence of pause propensity on regulatory molecules makes “RNAp pausing” an adaptable mechanism to changes in the environment.

Here we present a brief review of the SSA, the delayed SSA, and delayed, non-delayed, and detailed stochastic models of gene expression and GRNs. We also present examples of how the noise in gene expression affects key cellular processes, namely cell differentiation, and how it can be regulated.

The development of these computational models is of importance since they allow a detailed study of gene expression and GRN dynamics and can be used to guide and support experimental research. They offer different perspectives from those found in experimental research. Thus, the two methodologies should be used together in investigating GRN dynamics and exploring new possibilities in fields such as synthetic biology.

2. THE STOCHASTIC SIMULATION ALGORITHM

The Stochastic Simulation Algorithm (SSA) is a Monte Carlo simulation of the chemical master equation and thus, is an exact procedure for numerically simulating the time evolution of a well-stirred reacting system [20]. Each chemical species quantity is treated as an independent variable and each reaction is executed explicitly. Time advances in discrete steps and at each step a reaction is executed and the number of molecules of each of affected species is updated according to the reaction formula. After, the algorithm advances to the next event. Since each reaction and the
time for the next reaction to occur are independent of the preceding ones, the temporal evolution of the system is a Markov process.

The algorithm is exact in that each simulation of a system of chemical reactions in the conditions required by the SSA provides an exact temporal trajectory, matching one of the system's possible trajectories in the state space. The necessary condition for the SSA to be valid is that the system is kept homogenous during the simulation, either by direct mixing or by requiring that non-reactive molecular collisions occur far more frequently than reactive molecular collisions [20]. For the collision probability of two molecules to be spatially homogenous, each time a reaction occurs due to a collision between two potentially reacting molecules, this event will be followed by many non-reactive collisions, which cause the molecules to be once again uniformly distributed in space before the next reactive event occurs.

Each reaction rate constant, \( c_\mu \), depends on the reactive radii of the molecules involved in the reaction and their relative velocities. The velocities depend on the temperature and molecular masses. After setting the initial species populations \( X_i \) and reactions rate constants \( c_\mu \), the SSA calculates the propensity \( a_\mu = c_\mu h_\mu \), for all possible reactions, where \( h_\mu \) is the number of distinct molecular reactants combinations available at a given moment. The SSA then generates two random numbers, \( r_1 \) and \( r_2 \), used to compute \( \tau \), the time interval until the next reaction occurs, and \( \mu \), that determines which reaction occurs. Finally, the system time \( t \) is increased by \( \tau \) and the \( X_i \) quantities are adjusted to account for the occurrence of reaction \( \mu \), assuming that it occurred instantaneously. This process is repeated until no more reactions can occur or for a defined time interval. The probabilities for events to occur are converted into the expected time it takes to actually occur, allowing computation of the system state evolution. The SSA runs as follows:
Step 0 (Initialization). Input the desired values for the $M$ reaction rate constants $c_1, \ldots, c_M$ and the $N$ initial molecular population numbers $X_1, \ldots, X_N$. Set the time variable $t$ and the reaction counter $n$ both to zero. Initialize the unit-interval uniform random number generator (URN).

Step 1. Calculate and store the $M$ quantities, $a_1 = c_1 h_1, \ldots, a_M = c_M h_M$ for the current molecular population numbers, where $h_\mu$, is the number of distinct molecular reactant combinations available, given the system current state $(X_1, \ldots, X_N)(\mu = 1, \ldots, M)$. Calculate and store as $a_0$ the sum of the $M$ $a_\mu$ values.

Step 2. Generate two random numbers $r_1$ and $r_2$ from a unitary uniform distribution and calculate $\tau$ and $\mu$ according to: $\tau = (1/a_0) \ln(1/r_1)$, and $\mu$ is an integer such that:

$$\sum_{\nu=1}^{\mu-1} a_\nu < r_2 a_0 < \sum_{\nu=1}^{\mu} a_\nu.$$

Step 3. Using the $\tau$ and $\mu$ values obtained in step 2, increase $t$ by $\tau$, and adjust the molecular population levels to reflect the occurrence of the reaction chosen to occur. Then increase the reaction counter $n$ by 1 and return to step 1.

Several algorithms have since been proposed that allow simulating faster versions of the SSA, provided some restrictions and approximations (e.g. the Gibson-Bruck algorithm [38]). In [2] it was first proposed to simulate gene expression and GRNs with the SSA. When simulating gene expression, it was observed that proteins were produced in short bursts of variable quantities that occur at random time intervals. These findings explain numerous observations of phenotypic variability in isogenic populations of prokaryotic and eukaryotic cells, since the results imply that there can be large differences in the time between successive events in regulatory cascades across a cell population and in probabilistic outcomes in switching mechanisms that select between alternative regulatory paths.
3. DELAYED STOCHASTIC MODELS OF GENE EXPRESSION

While the first stochastic models assumed transcription to be instantaneous, it takes a considerable time for an RNAP to create an RNA molecule. This interval depends on the gene length, thus, varies from gene to gene. Therefore, recent models introduced delays in the appearance of products of gene expression such as proteins and, in some cases, RNA [26][28][29]. While non-delayed models can mimic gene expression fluctuations [39] under restricted conditions, models of complex GRN (e.g. with feedback loops) require delayed reactions to reproduce the dynamics [26][40].

Several steps in gene expression, such as transcripts assembly, mRNA translation, post-translation modifications and folding, are time consuming [41]. Transcription elongation is the process by which the RNAP slides along the template strand and adds bases to the transcript according to the DNA template. Its duration depends on the gene length and transcription speed. This varies between different events, potentially for the same gene, since the underlying reactions are stochastic. Measurements of elongation times showed that the velocities of different transcription events followed a wide normal distribution [42].

Several studies [26][27] proposed delayed SSA algorithms that allow explicit delays in protein production. Bratsun and colleagues [26] explored the combined effects of time delay and intrinsic noise on gene regulation. Their results indicate that delays in gene expression can cause a GRN to be oscillatory even when its deterministic counterpart is not. Further, they showed how delay-induced instabilities can reduce the effects of noise in negative feedback loops. Barrio et al showed that the delays in transcription and translation allow matching observed sustained
oscillations in expression levels of hes1 mRNA and Hes1 protein in mouse better than continuous deterministic models [27]. A subsequent work [84] proposed another delayed SSA, for reactions with a single delayed event. This algorithm considered two types of reactions with a delayed event. In “nonconsuming” reactions, the reactants of an unfinished reaction cannot participate in new reactions, while in “consuming” ones, the reactants change immediately. An important result was reported, namely, this delayed SSA, able to model reactions with one delayed event, is as exact as the original SSA [20]. However, it is noted that transcription was assumed to be a non-consuming reaction [84], which, in a strict sense, it is not true, due to events in elongation, such as RNAp “fall-off” and collisions between RNAs [28][36]. When using these single-delayed SSA’s one can account for the time taken by the promoter open complex formation by modeling transcription as a multi-step process, the last reaction being delayed, modeling elongation and termination.

The algorithm proposed in [28] differs from the previous in that it can handle multiple delayed events in one reacting event. Given that multiple delays, such as the promoter release delay and the elongation delay, have been shown to be crucial in GRN dynamics, we opted for presenting in detail the delayed SSA [28], which allows implementing the modeling strategy for GRNs proposed in [29]. As shown ahead, each delay in transcription and translation has a unique effect on the GRN dynamics.

The ‘delayed SSA’ [28] uses a waiting list to store delayed output events. The waiting list is a list of elements (e.g., proteins being produced and occupied promoter regions), each to be released after a certain time interval (also stored in the waitlist). The algorithm proceeds as follows:

1) Set $t = 0$, $t_{\text{stop}} = \text{stop time}$, set initial number of molecules and reactions, and create empty waitlist $L$. Go to step 2.
2) Generate an SSA step for reacting events to get the next reacting event $R_1$ and the corresponding occurrence time $t + t_1$. Go to step 3.

3) Compare $t_1$ with the least time in $L$, $t_{\text{min}}$. If $t_1 < t_{\text{min}}$ or $L$ is empty, set: $t = t + t_1$.

Update the number of molecules by performing $R_1$, adding to $L$ both any delayed products and the time delay for which they have to stay in $L$. This time can be chosen from a defined distribution. Go to step 4.

4) If $L$ is not empty and if $t_1 \geq t_{\text{min}}$, set $t = t + t_{\text{min}}$. Update the number of molecules and $L$, by releasing the first element in $L$; otherwise go to step 5.

5) If $t < t_{\text{stop}}$, go to step 2; otherwise stop.

Note that after step 4, the algorithm goes to step 5 instead of performing the event $R_1$ previously generated in step 3. In step 5, it then jumps to step 2 if the simulation time has not elapsed. When it does so, it generates a new reaction event rather than performing the previous reaction event that was not executed because the delayed event occurred first. This sequence of steps is necessary since delayed events add new substances into the system and so the propensity of reactions might change. Generating a new reaction event while discarding the one previously generated due to the occurrence of the delayed event, does not introduce errors as the process is memoryless.

4. MULTI-DELAYED STOCHASTIC MODELS OF TRANSCRIPTION AND TRANSLATION.

Recently, the real-time production of individual protein molecules by a repressed lac promoter in individual $E. coli$ cells was monitored by epifluorescence microscopy [14]. In a constructed $E. coli$ strain, a single copy of the chimeric gene tsr-venus was incorporated into the chromosome, replacing the native lacZ gene and leaving and the endogenous lac promoter intact. The lac promoter was maintained at a highly
repressed state. Since the endogenous tsr gene is expressed in high quantities, the addition of the extra tsr gene (albeit repressed) did not affect the cell’s normal functioning. When the infrequent spontaneous dissociation of the repressor from the operator occurs, transcription can begin. A single mRNA is typically generated in this instance. When mRNA is produced, ribosomes bind to it and proteins are produced. Protein production is detected after completion of their assembly, including folding, insertion into the inner cell membrane, and maturation of the Venus fluorophore [14]. By detecting radiation emission from the fluorophore, it was found that mRNA are translated in bursts, with the distribution of the bursts per cell cycle fitting a Poisson distribution, and that the number of proteins produced per burst follows a geometric distribution [14]. The set of chemical reactions of the delayed stochastic model of gene expression and repression are:

\[
\begin{align*}
&\text{Pro + RNAp } \xrightarrow{k_{1}} \text{Pro}(\tau_1) + \text{RBS}(\tau_1) + \text{RNAp}(\tau_2) + \text{R}(\tau_2) \tag{4.1} \\
&\text{Rib + RBS } \xrightarrow{k_{3}} \text{RBS}(\tau_3) + \text{Rib}(\tau_4) + \text{P}(\tau_5) \tag{4.2} \\
&\text{RBS } \xrightarrow{d_{\text{rate}}} \emptyset \tag{4.3} \\
&\text{Pro + Rep } \xrightarrow{k_{\text{rate}}} \text{ProRep} \tag{4.4} \\
&\text{ProRep } \xrightarrow{k_{\text{unrep}}} \text{Pro} + \text{Rep} \tag{4.5}
\end{align*}
\]

This model, first proposed in [30], is based on a previous model proposed in [44]. Reactions (4.1) and (4.2) model, respectively, prokaryotic transcription and translation. Pro represents the promoter region of the gene, RNAp is an RNA polymerase, Rib is ribosome, and R is a transcribed RNA molecule. The RBS (ribosome binding site) is the initial sequence of the RNA to which the ribosomes bind to and initiate translation. In prokaryotes this can occur when the RBS is produced (\(\tau_1\) seconds after transcription starts). When a product X has a delay \(\tau\),
represented by $X(\tau)$, it implies that when the reaction occurs, it takes $\tau$ seconds after that for $X$ to appear in the cell. Note that $\tau$ can be a random variable following a predefined distribution, thus vary from one reaction event to the next. The RBS degrades via reaction (4.3).

Reaction (4.4) models transcription repression by a repressor molecule (Rep) while (4.5) models the unbinding of the repressor from the promoter. Only when the promoter is free can transcription occur and, since this reaction rate constant is small, this occurs at sparse intervals. The expected fraction of time that the promoter is available for reactions is given by:

$$\frac{1}{1 + \left(\frac{1}{k_{unrep}} + \tau\right) k_{rep} \text{Rep}}.$$ 

The rates set for reactions (4.1-5) are: $k_t = 0.01 \text{ s}^{-1}$, $k_{tr} = 0.00042 \text{ s}^{-1}$, $d_{\text{RBS}} = 0.01 \text{ s}^{-1}$, $k_{\text{rep}} = 1 \text{ s}^{-1}$, and $k_{\text{unrep}} = 0.1 \text{ s}^{-1}$. Initially $\text{RNAP} = 40$, Pro = 1, R = 0, Rib = 100, RBS = 0, P = 0, ProRep = 0, and Rep = 100.

Values for the delays were extracted from measurements. $\tau_1$ is set to 40 s, within the range from 10 s to several minutes [33]. The length of the gene tsr-venus driven by a Lac promoter is 2500 nucleotides [14]. Since the average rate of transcriptional elongation in $E. coli$ is $\sim$50 nt/s, it follows that $\tau_2$ is, on average, 90 s ($\tau_2 = \tau_1 + 2500/50$). The post-translational protein assembly process was observed to take $420\pm140$ s [14], thus $\tau_5$ is randomly generated in accordance, each time the reaction occurs. The time of the RBS clearance, $\tau_3$, is 2 s [45]. The average translation rate is 15 amino acids/s, thus $\tau_4 = \tau_3 + 2500 \text{ nt} / (45\text{nt/s}) = 58$ s. Finally, transcription initiation includes the transition from the closed to the open promoter complex (accounted by $\tau_1$) [33]. All delays, except $\tau_5$, were held constant [30] since measurements suggest that they are not highly variable between transcription events of a gene [35][33][14][30] (although vary widely from gene to gene [46]).
example, measurements on an unrepessed Lac promoter [34] suggest that it follows a Gaussian distribution with a mean of 40 s and standard deviation of 4 s. This model was implemented in “SGNSim” [47] since this simulator allows multi-delayed reactions. Another simulator able to implement the SSA with delayed reactions (but only one delayed event per reaction) is “Dizzy” [48].

Fig. 1 shows the number of proteins produced during a time interval of approximately 3 cell life cycles. The proteins are produced in bursts as reported [14]. Each time a repressor unbinds the promoter, an RNAp can bind to the promoter producing one RNA molecule that is translated into several proteins before decaying. In Fig. 2.A is shown the distribution of the number of transcription initiations from 1000 simulations which fit a Poisson distribution. Fig. 2.B shows the number of translation reactions which fit an exponential distribution [14].

FIGURE 1.
FIGURE 2.

Subsequent studies [31] focused on the dynamical properties of this model and studied the noise in the proteins level. Several factors were examined, such as gene length, the transcription initiation rate, the translation initiation rate, the decay of RNA and proteins, the promoter delay, and the RBS delay. The results showed that there is a linear scaling behavior between the protein variance and the mean. Also, the most dominant noise sources are promoter fluctuations and the half-time of mRNA. At the translational level, either increasing the translation initiation frequency or decreasing the protein decay raises the noise level. As the period of promoter activity determined the interval between mRNA bursts and the RBS life period determined the time range of the protein bursts, it was inferred that the frequency and duration of promoter activation are the main factors determining transcriptional noise.
A detailed analysis of noise sources in gene expression can be found in [17]. This work also discusses in detail various noise sources and methodologies to measure noise. As well, it provides a detailed review on studies focusing on noise in protein levels. Pioneering works in this field [49][50] predicted mRNA and protein fluctuations in cells. Importantly, they predicted that mRNAs produce geometrically distributed protein bursts if they are either translated or degraded with constant probabilities. This was later on experimentally confirmed [11][14]. The stationary distribution for geometrically distributed translation bursts was then deduced [51]. An identical model was independently analyzed in [52] and a more detailed model also accounting for replication, cell growth and division was proposed in [53].

Recent studies focused on the effects that delays have on transcriptional and translational noise [30][31][36]. Among their findings is the observation that each delay (e.g. on the promoter, on the RBS, and on the RNAp in (reaction 4.1) affects gene expression dynamics, and consequently the RNA noise level. For example, the promoter delay, $\tau_1$, prevents two transcription events from initiating within a smaller interval than $\tau_1$ [33]. $\tau_2$ affects the number of RNAp available for transcription, which affects the propensity of this reaction at run-time. Finally, the time needed to produce an RBS ($\tau_1$), causes a temporal translation in the amount of RBS in the cell, in comparison to a non-delayed model. Since the production of RBS allows initiating translation in prokaryotes, this delay is also relevant in the context of genetic circuits. In [30] it was observed that if $(|RNAp| \times k_t)^{-1} + \tau_1 >> \tau_2$, then $\tau_2$ does not affect the dynamics significantly. If $\tau_1 << (|RNAp| \times k_t)^{-1}$ then the promoter delay is not significantly relevant ($|.|$ denotes the amount of a substance in a cell).

Most importantly, this analysis suggests that delays can actually act as to decrease transcriptional noise level. That is, given a delayed and non-delayed model of
transcription (reaction 4.1, with all delays set to zero in the non-delayed case) with identical mean transcriptional activity for the same time length, the existence of the delay on the promoter complex release implies that the transcription initiation rate in the delayed model needs to be higher and thus, the noise level will be lower. This is because, while on the waitlist due to the delay, the promoter is unavailable for initiating new transcription events, thus, in comparison to the non-delayed case, the rate of transcription initiation must be higher so has to compensate the lesser time that the promoter is available for transcription.

In figure 3 are shown the expected distributions between transcription events in the delayed and the non-delayed case. The fact that the non-delayed distribution is broader implies that the RNA bursts will be higher than in the delayed case. This conclusion depends on the variability of the time delay, namely, it depends on the variability of the promoter delay to be much smaller than its mean value as in the case of the lac promoter [34].

FIGURE 3

The toggling behavior of the model of toggle switch here shown is caused by the fluctuations in the proteins’ levels and, as shown in other studies [43], the toggling frequency is affected by the time delays in gene expression, especially the promoter delay. The effect of noise and delays in a genetic circuit are not always evident. E.g., in some cases noise enhances periodic behaviors. One example of this phenomenon of noise-induced (or noise-amplified) oscillations is presented in [85], on a model of circadian rhythms in Drosophila. Using a reduced model of the circadian oscillator in Drosophila simulated according to the SSA, and with delayed negative and positive feedback loops, it was shown that the internal noise sustains oscillations, which otherwise would not exist. The fact that there is an optimal noise intensity that
maximizes periodicity robustness indicates the occurrence of intrinsic coherence resonance. In addition, a delay in the positive feedback was found to affect the robustness of the noise-sustained oscillations. Similarly, a stochastic analysis of the repressilator circuit [86], which consists of three transcription factors that negatively regulate each other in a cyclic manner [87], showed that the fluctuations modify the range of conditions in which oscillations appear as well as their amplitude and period, compared to the deterministic equations.

5. ENSEMBLE APPROACH FOR DELAYED STOCHASTIC GENE REGULATORY NETWORKS

Since the genes of the human genome (as well as others) have been identified, one next step is to understand the behavior of GRN. For example, the human genome has ~30,000 genes which are regulated by a network of their own products, among other factors such as environmental. The genome can be seen as a parallel processing nonlinear dynamical system.

The large scale topologies of real GRNs, at a detailed level, are still unknown. Only partial information exists on the expected number of connections between genes, number of genes, etc. One example study reporting measurements aiming to determine the large scale topology of a GRN is [88], which used microarrays to detect the effect of genes’ deletion and overexpression in Yeast, from which one can then infer the topology of the GRN. However, these experiments and the algorithms used for inferring the structure from expression profiles are still quite inaccurate. Thus, in order to study the dynamics of large GRNs comprising thousands of genes, modeling strategies have been developed from which one can generate ensembles of GRNs to study general dynamical properties of these systems. These modeling strategies aim at
generating GRNs given mean topological features, e.g., a given mean connectivity. In this section we reference a few modeling strategies that allow using the ensemble approach [58] to study general dynamical properties of large scale GRNs, and then describe in detail one of these, that allows implementing delayed stochastic models of GRNs using the ensemble approach.

Models of generic large scale GRNs, whose general properties can be studied using the ensemble approach, have been developed using several modeling strategies, such as random Boolean networks [54][55], differential equations [56], piecewise linear differential equations [57] and stochastic equations [2][48].

While it is known that GRNs are stochastic, the use of large scale stochastic models is not very common so far, due to the necessary computational complexity of simulating its dynamics. One proposal has been made of a modeling strategy that allows applying an ensemble approach to investigate the general dynamic properties of large scale delayed stochastic models of GRN [29]. The procedure to design GRNs is a generalization of the one to generate random Boolean networks (RBN) [54]. In RBNs, each gene is assigned a random Boolean function, creating a combinatorial logic. This can be attained by defining reactions between gene expression products and assigning the resulting multimers as activators or repressors of genes. For that one defines multiple operator sites for genes and randomly assigns all possible binding combinations as activations or inhibitions, and by defining how gene expression products bind competitively to binding sites, each with a different effect on the gene’s transcription rate. These features allow Boolean functions and topologies to be represented in a delayed stochastic GRN. However, one RBN could be mapped to an infinite number of stochastic GRNs since many parameters, such as rate constants, are not defined in RBNs.
The regulation of a gene’s expression occurs via transcription factors, or indirectly by protein degradation. Proteins can form oligomeric structures which can feedback into the GRN. These interactions define the GRN topology. Since genes can have multiple operator sites, the following notation is used: Pro_{i,(op)}, where i is the gene identification index, and (op) is an array of all operator sites whose values represent the state of the gene's operator sites, i.e., what transcription factor is bound to the site, if any.

For each combination of input states (promoter state), a regulating function is assigned that determines the gene’s expression rate. Depending on its promoter occupancy state, the gene is either repressed or activated at a certain level. A fraction of genes can be assigned to have basal level of expression.

This model of GRNs [29] can be implemented in SGNSim [47]. GRNs are generated from the following reactions defined below (5.1 to 5.8). For gene i = 1,…,N there is basal transcription reaction of promoter Pro_{i} by one RNAp (reaction 5.1). The model includes transcription reactions for promoters with specific sets of transcription factors bound (reaction 5.2) and translation of RNA by ribosomes (Rib) into proteins (reaction 5.3). These are all time-delayed reactions. Delays differ between products of each reaction, and between similar reactions for different genes.

The binding/unbinding of a transcription factor from operator site j of a gene i are represented in reaction 5.4. If the complex is not able to transcribe, the two reactions represent repression/depression reactions. Note that reaction 5.4 is bidirectional, corresponding to the binding of the repressor, and its spontaneous unbinding. Reaction 5.6 represents the derepression due to an element that removes the repressor from the operator site. Decay of RNA, as represented by its ribosome binding site RBS, and proteins, occur via reaction 5.7. Decay of a protein while bound to a
promoter occurs via reaction 5.5. Finally, protein polymerization (here, limited to dimers for simplicity) and protein dissociation occur via the bidirectional reaction (5.8). Proteins and RBS degrade at constant rates via the uni-molecular reaction 5.7:

\[
\text{Pro}_i + \text{RNAp} \xrightarrow{k_{\text{max}}} \text{Pro}_i(\tau_i) + RBS_i(\tau_i) + \text{RNAp}(\tau_i) + R_i(\tau_i) \quad (5.1)
\]

\[
\text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} + \text{RNAp} \xrightarrow{k_{1,\text{opt}}} \text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}}(\tau_i^1) + RBS_i(\tau_i^1) + \text{RNAp}(\tau_i^2) + R_i(\tau_i) \quad (5.2)
\]

\[
\text{Rib} + RBS_i \xrightarrow{k_{\text{arg}}} \text{Rib}(\tau_3) + RBS_i(\tau_3) + p_i(\tau_4) \quad (5.3)
\]

\[
\text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} + p_w \xleftarrow{k_{w,ij}} \text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} \quad (5.4)
\]

\[
\text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} \xrightarrow{k_w} \text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} \quad (5.5)
\]

\[
\text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} + p_w \xrightarrow{k_{\text{inc}}} \text{Pro}_{i,\{\cdot\}_{\cdot,\cdot}} + p_z + p_w \quad (5.6)
\]

\[
\text{RBS}_i \xrightarrow{k_{\text{inc}}} \emptyset, \quad p_i \xrightarrow{k_{\text{inc}}} \emptyset, \quad (5.7)
\]

\[
p_i \xrightarrow{k_{\text{and}}} p_i \quad (5.8)
\]

Ensembles of GRN [58] can be generated by choosing random integers for all indexes in the reactions modeling interactions between genes i, j, z and w. The choice of which dimers can form can also be random. Each different set of choices corresponds to a unique GRN topology. Since the effect of transcription factors in gene expression levels can be randomly chosen, the stochastic version of any Boolean or more complex transfer function can be implemented.

Other works have proposed stochastic and delayed stochastic models of genetic circuits which could be used to develop general modeling strategies of ensembles of GRNs. For example, in [85] was proposed a model of the circadian oscillator in Drosophila, in [19] was proposed a model of the genetic circuit responsible for the pathway bifurcation in phage λ, in [89] was modeled the circuit responsible for the
Yeast cell cycle, in [90] was proposed a model of the Notch signaling pathway, in [91] a model was proposed for a network within the *Salmonella* GRN, and in [92] a stochastic kinetic model was built of the intracellular growth of vesicular stomatitis virus (VSV), comprising five genes, using delayed reactions modeling transcription and genome replication. In several of these and other models, hybrid approaches are used to model reactions that occur at a very fast pace (for a review on approximate and hybrid approaches see e.g. [93]).

6. MODELS OF TRANSCRIPTION AT THE SINGLE NUCLEOTIDE LEVEL

During elongation the RNAp is in constant kinetic competition with other regulatory pathways [59] and regulatory mechanisms can act at this stage in both prokaryotes and eukaryotes [42]. Noise during elongation also affects transcriptional dynamics [35]. To account for this noise one needs models of transcription at the single nucleotide level.

One of the first detailed stochastic models of transcription at the single nucleotide level was proposed in [28]. Using that model they characterized the distributions of transcriptional delays and elongation rates. The model considers a DNA template with one promoter site and \( n \) nucleotide sites, and five basic types of reaction processes. Importantly the model predicts that different RNAp molecules move at varying rates along the template strand in agreement with measurements [33][60][35].

In a similar molecular multistep model of transcription elongation [37], it was observed that the transcription times were, in general, non-Poisson-distributed. This model introduced transcriptional pauses due to “backtracking” of the RNA polymerase as a first passage process. Due to such pauses, they obtained a broad, heavy-tailed distribution of transcription elongation times, which can be significantly
longer than if no pauses occurred. When pauses result in long transcription times, it leads to bursts of mRNA production and non-Poisson statistics of mRNA levels.

More recently, a new model based on the previous ones was proposed [36] that includes additional events occurring during elongation, such as RNA polymerase halting [59] and promoter complex formation [33]. Such alternative pathways to stepwise elongation play a role in transcription regulation [59][61]. For example, their occurrence can amplify collisions between preceding RNAP molecules. The delayed stochastic model of transcription at the single nucleotide level proposed in [36] incorporates the promoter occupancy time, pausing, arrest, misincorporation and editing, pyrophosphorolysis, premature termination, and accounts for the range occupied by an RNAP when bound to the template [59][62]. Since most measurements of transcriptional dynamics are from *E. coli*, all parameters values in the model were taken from measurements in *E. coli*.

Besides matching the measurements of gene expression at the single protein level [14] of a strongly repressed gene, the dynamics did not fully match the dynamics of single-step delayed models, indicating that the noise at elongation cannot be neglected. Importantly, this difference in delayed and detailed model was not solely due to the emergence of traffic. Events such as pauses and arrests have a non-negligible role even when not sufficiently frequent or long to cause collisions between RNA polymerases. This result is of importance since single-step multi-delayed models of transcription were assumed accurate as long as there were no collisions between RNA polymerases in the strand [28].

The model matched, at the single RNA level, the “bursty” dynamics of transcription of an induced gene [11] i.e., the distribution of intervals between RNA completions. The time length for the promoter complex formation [33] was also
relevant since it affected the number of collisions between RNAp in the strain. One important feature observed in a modeled induced gene is the RNA pulsing over time (RNA completions separated by a time interval smaller than the minimum interval between transcription initiations). Also, it was shown that the events during elongation affect the switching frequency of a toggle switch, even though the frequency of toggling events is much smaller than the frequency of fluctuations in RNA levels due to noise at elongation. This result implies that events during elongation can affect the dynamics of genetic circuits.

In [63] an analytically solvable model of molecular network burst and general measures for characterization of bursts (burst size, significance, and duration) were proposed. They report that the simulations of their model suggest that collisions between neighboring RNAp molecules can attenuate bursts originating at the initiation site. Motor protein pausing can give rise to bursts. These results are in agreement with the previous studies described using the stochastic models at the single nucleotide level, and with experimental evidence [1]. In this work, the biochemical contribution to phenotypic noise of molecular fluctuations at the single cell level was isolated. By independent variation of the transcription and translation rates of a fluorescent reporter gene in the chromosome of *Bacillus subtilis*, and measuring the changes in the phenotypic noise, they found that increased translational efficiency is the main source of increased phenotypic noise. This effect is consistent with a stochastic model of gene expression in which proteins are produced in random and sharp bursts [30]. Their results provided the first direct experimental evidence of the biochemical origin of phenotypic noise, demonstrating that the phenotypic variation in isogenic populations can be regulated by genetic parameters [1].
Simulations of the model proposed in [36] showed that, e.g., the delay accounting for the process of elongation cannot be constant, and more importantly, the distribution from which these delays should be drawn is sequence dependent. E.g., a sequence with more pause-prone sites is likely to cause a broader distribution of elongation delays than otherwise.

As a side note on the importance of modeling explicitly the process of elongation, some delayed O.D.E. models of GRNs were recently shown to exhibit “artificial” oscillations. Namely, using more realistic models of the same GRNs where processes are modeled by explicit reactions rather than delays, some of the periodic attractors (oscillations) cease to exist. When using such models, is it usually worthwhile to confer with an explicit model if oscillations cease to exist and were, thereby, artificially created by the modeling strategy rather than existing in the real system. For an interesting discussion regarding this problem see [94].

7. THE TOGGLE SWITCH

The recent engineering of a toggle switch, a bistable GRN in E. coli of two mutually repressing genes [64], has spawned many studies that model this genetic circuit in an effort to further understand its dynamics. Using a synthetic toggle switch, the conditions necessary for bistability were studied [64]. The switch was constructed from two repressible promoters arranged in a mutually inhibitory network. It could be flipped between stable states using transient chemical or thermal induction, and it exhibited a well defined switching threshold. The importance of this work is that the toggle switch can be used as a cellular memory unit, thus having applications in biotechnology, biocomputing and gene therapy [64], and as a source of phenotypic
diversity [19][64][65][66]. The toggle switch might also be used by cells as decision circuits of differentiation pathways, and for preventing reversibility [67][69].

It has long been hypothesized that some types of cell differentiation are based on bistable genetic sub-circuits controlling many downstream genes [67]. In this process, a stem cell takes on a ‘stable’ phenotype. The genetic decision circuit of differentiation must be (at least) bistable, to allow branching into distinct cell types, and reliable, to prevent reversibility [64]. Recent evidence suggests that the neutrophil cell lineage has such bistable switch-like behavior [68]. Single cell analysis of the expression kinetics of the differentiation marker CD11b (Mac-1) revealed an all-or none switch-like behavior in HL60 promyelocytic precursor cells that transit to the neutrophil cell lineage. The progression from the precursor to the differentiated state is a discrete transition between low-CD11b and high-CD11b expressing subpopulations which are distinguishable in a bimodal distribution [68].

The bistability of the toggle switch, i.e., having two noisy attractors (with both high and low gene expression levels), depends on its internal noise level [43][70]. Its stability can be enhanced, such as by overlapping upstream regulatory domains [71].

Conditions for bistability of stochastic models of toggle switches were investigated, showing that, for a range of biologically realistic conditions, a suitable combination of network structure and stochastic effects gives rise to bistability even without cooperative binding [75]. A subsequent study showed that with realistic time delays in transcription and translation, self-activation reactions are unnecessary to attain bistability and switching [43]. Also, given sufficient noise, the two genes can express simultaneously for long periods of time. Thus, the toggle switch can be tri-stable [72]. Importantly, this ‘unstable attractor’ can be the region of the state space where the toggle switch spends most time [72].
A recent work explored how the stochasticity of the toggle switch can regulate patterns of cell differentiation [73]. Modeling single cells, each with a toggle switch whose dynamics are driven by the delayed SSA [28] they measured the distribution of differentiation pathway choices of an initially undifferentiated cell population. Assuming that in each cell the protein levels (high vs. low) of the toggle switch determine which one of four possible differentiation pathways the cell chooses, they varied protein degradation rates, as these vary widely between proteins [74] and transcription initiation rates, which are sequence dependent [35], and observed the patterns of cell differentiation. The delayed stochastic model of toggle switch consists of reactions (7.1) to (7.6) [72] with $i = 1, 2$ (when only the index $i$ is present) or $i, j = 1, 2$ with $i \neq j$ (when both indices are present):

\begin{align*}
\text{Pr} o_i + Rp & \xrightarrow{k_i} \text{Pr} o_i (\tau_i) + Rp(\tau_2) + R_i (\tau_i) \quad (7.1) \\
R_i + \text{Rib} & \xrightarrow{k_{ri}} R_i (\tau_i) + \text{Rib}(\tau_4) + P_i (\tau_5, \tau_{sad}) \quad (7.2) \\
R_i & \xrightarrow{d_{ib}} \emptyset \quad (7.3) \\
P_i & \xrightarrow{k_{2i}} \emptyset \quad (7.4) \\
\text{Pr} o_i + P_j & \xleftarrow{k_{rep}} \xrightarrow{k_{unrep}} \text{Pr} o_i P_j \quad (7.5) \\
\text{Pr} o_i P_j & \xrightarrow{k_{2}} \text{Pr} o_i \quad (7.6)
\end{align*}

Gene expression is modeled by multi-delayed reactions for transcription (7.1) and translation (7.2), where Pro$_i$ is the promoter of gene i, Rp is an RNA polymerase, Rib is a ribosome, and R$_i$ is the ribosome binding site of each RNA. The delays ($\tau_1$ to $\tau_5$) account for the duration of transcription and translation. Reaction (7.2) for translation accounts for the variability of the time needed to generate a functional protein (translation, folding, activation, etc.) given that the delay of P$_i$ follows a normal
distribution, with a mean of $\tau_5$ and a standard deviation of $\tau_{5,\text{std}}$ [30]. Each protein $P_i$ represses the other gene’s promoter. Reactions (7.5) model binding and dissociation of the repressor from the promoter, which defines the toggle switch. Reactions (7.4) and (7.6) model degradation of proteins and (7.3) models degradation of RNA.

The rates (in s$^{-1}$) of reactions (7.1) to (7.6) were $k_t = 0.005$, $k_tr = 0.00042$, $d_{bs} = 0.01$, $k_{rep} = 1$, $k_{unrep} = 0.1$, and $k_d = 0.0012$. Time delays (in seconds) are $\tau_1 = 40$, $\tau_2 = 90$, $\tau_3 = 2$, $\tau_4 = 58$, $\tau_5 = 420$, and $\tau_{5,\text{std}} = 140$. Each ‘cell’ was initialized with $P_1 = 0$, $P_2 = 0$, $R_1 = 0$, and $R_2 = 0$, and with one promoter of each gene ($\text{Pro}_1 = 1$, $\text{Pro}_2 = 1$), 40 RNA polymerases ($R_p = 40$), and 100 ribosomes ($\text{Rib} = 100$).

This study found that by varying the transcription rate $k_t$ and the protein decay rate $k_d$, cells could tune their pluripotency and distribution of lineage choice in a population, suggesting that the stochastic switch has high plasticity regarding differentiation pathway choice regulation. One example of this is the variation of the switch noise level, by varying simultaneously $k_t$ and $k_d$ so as to maintain the mean protein level constant. As the noise level changed, they observed significant changes in the pattern of cell differentiation (Fig. 4). In Fig. 4, cell type 1 corresponds to the differentiation pathway chosen by a cell when only $P_1$ is present, cell type 2 when only $P_2$ is present, cell type 3 when both proteins are present and, cell type 4 when both proteins are absent.

FIGURE 4

8. NOISY ATTRACTORS AND ERGODIC SETS. WHAT IS A CELL TYPE?

At the molecular dynamical level, it is not well defined what a cell type is. Even in the simplest case of a Boolean network with 30,000 genes, its state space has $2^{30,000}$ possible states. Since it takes seconds to minutes for genes to turn on and off,
whatever a cell type may be, it must be a very restricted subset of the possible states of the GRN, otherwise it would be impossible for a cell to be in a reliable state during its lifetime, i.e., the phenotypic diversity of cells of the same type would be immense, making cells’ performance unreliable. In the deterministic framework it is almost an inevitable hypothesis that cell types correspond to attractors of the network dynamics [54] as these, in the absence of noise, are the asymptotic behaviors of the network. To be biologically plausible, attractors need to be constrained to very small regions of the state space of the system (i.e., the number of states of an attractor must be immensely small in comparison to the size of the state space of a “binary” GRN). In this scenario, if a cell type is an attractor, then a pathway of differentiation is either triggered by a perturbation from one attractor into a new basin of attraction or triggered by a change in the GRN internal dynamics due to noise [54]. Cell types would correspond to the attractors of the dynamics, and differentiation would occur due to bifurcations in the dynamics as variables change, e.g., morphogenesis. These two possibilities do not exclude one another.

An important criticism [76] to this hypothesis is that noise may render such attractors as a poor model of cell types, since closure of an attractor (a state cycle) in the discrete dynamics is delicate. This is an important criticism that leads to questions on the effect of noise on the “attractor as cell type” hypothesis.

Some attempts were made to address this problem. Klemm and Bornholdt [77] tested the stability of attractors with respect to infinitesimal deviations from synchronous updates and found that most attractors are artifacts arising from synchronous clocking. The remaining attractors are stable against fluctuating delays, and its average number grows with the number of nodes, within the numerically tractable range. A similar scaling law was observed using a different approach [78].
These works confirm that the models predict multiple attractors assuming asynchronous updating. Yet, note that these models do not assume any probability of genes “misbehaving”, i.e., acting contrary to what input states and Boolean transfer function determine. Only the moment at which nodes update is assumed stochastic.

In [70] the concept of “ergodic sets” for Boolean networks was introduced as a set of states from which the system, once entering, does not leave when subject to internal noise using the Boolean network model. It was shown that if all nodes of states on attractors are subject to internal state change with a probability p due to noise, multiple ergodic sets are very unlikely. However, if a fraction of those nodes are “locked” (not subject to state fluctuations caused by internal noise e.g., due to epigenetic regulation), multiple ergodic sets emerge. Additionally, simulations of a delayed stochastic model of two coupled bistable switches showed that this GRN has two distinct ergodic sets which are stable within a wide range of parameters variations and, to some extent, to external perturbations. A subsequent study [79] proposed a gamma-bernoulli mixture model clustering algorithm to determine the noisy attractors of stochastic GRN, using multiple data sources, namely, protein and RNA levels, and promoter occupancy state. The method was also applied to model of the MeKS module of Bacillus subtilis, and the results matched the experimental data, in particular the expected number of cells competent for DNA uptake at any given moment. The clustering algorithm in [79] was also used to confirm the correlation between gene expression profiles and the resulting patterns of cell differentiation driven by a model genetic switch [73].

Several other recent studies also focused on genetic circuits responsible for regulating differentiation pathway choices [9][68]. For example, in [97] the authors proposed a generic minimal model of a GRN controlling cell fate determination, i.e.,
able to regulate hierarchical branching of developmental paths, which exhibited five elementary characteristics of cell differentiation: stability, directionality, branching, exclusivity, and promiscuous expression. The GRN model was implemented using both O.D.E.’s and stochastic differential equations.

In [95] was studied the dynamics of two-gene switches responsible for deciding eukaryotic cell differentiation pathways, and the results analyzed in terms of their response specificity, their ability to store short-term memory of signaling events, and stability of noisy attractors. Their results showed that genetic switches are capable of responding differently to stimulus differing in strength and duration.

An interesting model of the formation of somites in vertebrate embryos, based on a delayed SSA, was recently studied [96]. This model accounts for delays in gene expression and for the influence on neighboring cells in the dynamics of each cell’s GRN. Namely, the model had a line of neighbor cells driving the oscillation patterns responsible for the differentiation patterns, each cell having two neighbor cells. The study focused on how the interactions between cells affect the oscillatory dynamics of the GRN in each cell. The results exemplify the importance of accounting for stochasticity and time delays in gene expression when studying processes arising from cell-to-cell interactions, e.g., synchronization, and cell differentiation in tissues, usually driven by a very small number of molecules.

9. APPLICABILITY OF STOCHASTIC MODELS, STUDY OF LARGE SCALE MODELS OF GRN, AND SOFTWARE

Currently, it is prohibitively costly and time consuming, if not impossible, to experimentally measure ranges of behaviors of GRNs for varying conditions, including extreme ones [23], and varying parameters values, making the use of models necessary. We reviewed stochastic models of gene expression at the single
molecule level, namely, non-delayed models [19], with delayed protein release [38], multi-delayed [29] and detailed models, i.e. with explicit stepwise elongation, [28][36][37]. All were proven useful in analyzing specific aspects of GRNs dynamics. Non-delayed models capture the stochasticity of transcription initiation, multi-delayed models additionally capture the time needed for transcription and translation to be completed once initiated, and detailed models are required to, besides the two previous features, account also for events such as pauses during elongation.

The choice of modeling strategy depends on the aspect of the GRN dynamics studied and what features affect it significantly. In general, non-delayed models are valid to determine equilibrium states of small GRNs [19][98][99], where delayed and non-delayed versions differ only in transient time to reach equilibrium.

The multi-delayed model is required if the dynamics is affected by delays in transcription and translation. One such case is when negative feed-back loops exist [100]. Also, if noise plays an important role in the dynamics, delays might have to be explicitly modeled as these were shown to affect gene expression noise [31].

Multi-delayed models are appropriate if the GRN dynamics’ noise is mostly dependent on the stochasticity of transcription initiation and fluctuations in transcription factors amounts, but less appropriate when non-linear events in elongation play a significant role. One such case is when there are pause-prone sequences (see [101] for a review on elongation), as these pauses, being probabilistic events and lasting seconds to minutes [68][102] can lead to collisions between RNAs in the strain, and thus transcriptional bursting, that can only be captured by models with stepwise elongation [28][36][37].

29
In some cases it is not evident \emph{a priori} how the GRN dynamics, even with few genes, is affected by not accounting for a given feature. If so, one needs to compare models with different levels of detail to confer. Importantly, these effects are likely to depend on the dynamical regime, e.g., varying with protein levels’ amplitude. The difficulty of assessing the necessary level of detail increases with the GRN size, as the behavior becomes less predictable. Meanwhile, increased computational complexity with the increase of number of reactions demands simplifications. Further, many parameters values will be unknown. Thus, how should one study large scale stochastic models of GRNs?

In principle, the questions one whishes to address in large scale GRNs with thousands of genes and TFs differ from those for single genes or small GRNs, thus the methods to address them differ as well. Also, given that large scale GRN are likely to have a wider range of behaviors, the number of independent simulations needed to capture the distribution of behaviors is much larger. More importantly, the dependency on initial conditions is stronger, as these, e.g., determine to some extent which of the multiple noisy attractors are more likely to be reached [70], thus it is necessary, for large scale GRNs, to impose a wide range of initial conditions.

The topology of interactions and the distribution of transfer functions are likely to be key factors in large scale GRN dynamics, by determining how the various signals and noise propagate throughout the network [103], long term behaviors and transients. For these reasons, the characterization of the behavior of large scale GRN can only be done by distributions of behaviors rather than mean values alone.

Such distributions of behaviors are strongly dependent on the many parameters values, thus these needed to be varied to assess their role and importance in the dynamics. Imagine one is studying the effects on the network dynamics of the mean
propensity with which TFs bind to their TFBS. To do so, in practice, one cannot analytically predict the range of behaviors for all possible values of the parameter. One common procedure is thereby to first inspect the network behavior for a “high” and a “low” value of the parameter [31] (e.g., connectivity) by running several simulations with different initial conditions (such as protein and RNA levels) for each value. Depending on the results one might afterwards want to inspect the dynamics for a set of values of the parameter within a given interval. Also important is to additionally change another parameter, e.g., average connectivity, and again test if the differences in the dynamics when having a high or low affinity of TF to the TFBS are the same. In some cases, it might happen that the effect is the opposite. Few such studies have so far been made using purely stochastic models of large scale GRN (see, e.g., [47][104]). Usually, other approaches such as Boolean networks are preferred given their simplicity.

Recently, several simulation tools have been proposed that can execute systems of chemical reactions following deterministic and stochastic dynamics, e.g., DBsolve [108], GEPASI [109], KINSIM [110], and Cellware [111]. The most popular is probably “Dizzy” [103], a chemical kinetics simulation tool that can simulate systems of chemical reactions according to the SSA [20], the Gibson-Bruck algorithm [38] (which allows one delayed event per reaction), the \(\tau\)-leap method [105], and ODEs for any specific initial conditions. It is not yet possible to implement the multi-delayed SSA in Dizzy, for which one should rather use SGNSim [47].

Briefly, the \(\tau\)-leap method and its later versions [106][107][112][113] were developed so as to allow faster simulations than the original SSA. The method recently proposed in [113], named “binomial \(\tau\)-Delayed SSA” extends previous ones
in that it allows a delayed event per reaction, very important to allow modeling complex chemical processes such as transcription.

All methods above are based on, or, are derivations of the SSA, thus, they assume the compartments where reactions take place (e.g., the cell cytoplasm) to be well-stirred environments. Recent results suggest that spatial distribution of molecules in cells might affect gene expression dynamics considerably (see, e.g., [114]).

To model environments with heterogeneous environments one has to use simulators where space is explicitly modeled [115]. One such simulator, able to model chemical reactive systems in a spatial 3-dimensional environment, is MCell. This software tool simulates diffusion by 3D random walk movements for individual molecules. The outcome of encounters between molecules is decided by comparing the value of a random number to the probability of a reactive event per encounter. Such outcomes depend on the properties of the surface of the molecules. Random numbers are also used to decide between all other possible reaction mechanism transitions that might occur during a time step such as decays. This and other spatial simulators of chemical reactive systems [116][117][118] do not yet allow implementing delayed events.

10. FUTURE DIRECTIONS AND CLOSING REMARKS

Delayed stochastic models of gene expression and GRN have rich dynamics and a multitude of regulatory mechanisms. The ability of delayed stochastic models to accurately match dynamic measurements of gene expression through mean expression levels, noise and dynamic patterns of small GRNs suggests that they are suitable to explore the dynamics of GRNs in combination with experimental studies.
Regarding future directions, it should be noted that many regulatory mechanisms in GRN are yet to be incorporated in stochastic models. One such case is phosphorylation which has been established as a crucial regulatory mechanism by which the activity of many eukaryotic transcription factors is both positively and negatively controlled [80]. Multisite phosphorylation allows a sophisticated regulation of transcription factors activity and is now being established as a dynamic mechanism that allows fine-tuning of transcription factor activity. There are various levels of regulation by phosphorylation [81][82] that can directly modulate the activity of transcription factors such as affecting their stability and cellular localization, or by facilitating protein–protein interaction and oligomerization, as in the case for CREB–CBP association and STAT dimerization [83].

The fast pace at which this area of research is advancing, both at the computational and the experimental levels, guarantees that this and other candidates for regulatory mechanisms of GRN will be suggested in the near future and their role investigated with further developed stochastic models of GRN.

ACKNOWLEDGMENTS
This work was supported by the Academy of Finland project no. 126803, and the FiDiPro program of the Finnish Funding Agency for Technology and Innovation (40284/08). I thank Richard Bertram for inviting me to do this review, the reviewers for very useful comments, and S. Healy for insightful comments and suggestions.

REFERENCES


[34] R. Lutz, T. Lozinski, T. Ellinger, H. Bujard, Dissecting the functional program of *Escherichia coli*


[93] Pahle, J., Biochemical simulations: stochastic, approximate stochastic and hybrid approaches


FIGURES LEGENDS

Figure 1. Time series of proteins produced in ~3 cell cycles.
Figure 2. (A) Number of transcription events for 4 cell cycles in 1000 simulations. (B) Number of translation events for each RNA transcribed in 1000 simulations. The smaller figures show the experimental data [14] for direct comparison.

Figure 3. Distribution of time intervals between transcription events in delayed and in non-delayed stochastic model of transcription.

Figure 4. Distribution of the amount of cells (y-axis) that are differentiated into each of four cell types (x-axis) of the 9 cell populations (1000 cells per cell population), differing in noise levels while maintaining the same protein mean level (noise increases from population 1 to 9).