

Computing intrinsic noise of the genetic regulation modeled by Hill functions

Riccardo Ziraldo and Lan Ma*

Department of Bioengineering, The University of Texas at Dallas, Richardson, USA

*Corresponding author: Lan Ma, Department of Bioengineering, The University of Texas at Dallas, Richardson, TX, USA, Tel: 9144005317, E-mail: lan.ma@utdallas.edu

Citation: Riccardo Ziraldo, Lan Ma (2018) Computing intrinsic noise of the genetic regulation modeled by Hill functions. *J Comp Sys Bio* 2(1): 102

Abstract

Intrinsic noise embedded in stochastic gene regulation due to low copy number of species has been studied using the approach of theoretical modeling and computational simulation, including the standard methods of stochastic simulation algorithm (SSA) and linear noise approximation (LNA). At average cell level, Hill functions are widely used as a compact format to represent gene regulation involving multi-transcription-factor binding and cooperativity. Heuristic SSA and LNA methods (hSSA and hLNA) have been applied to study stochastic models employing Hill functions. It is however unclear which modeling and simulation method is suitable to characterize intrinsic noise of Hill-type gene regulation with sufficient accuracy and computational efficiency. In this work, we perform noise analysis of two gene regulatory models represented by second-order activating and inhibitory Hill functions, seeking to evaluate the performance of five existing noise modeling methods. Specifically, SSA and LNA are applied to the full models that are expanded from the Hill functions containing only elementary reactions, while hSSA and hLNA are applied to reduced models where the Hill function is heuristically used. In addition, we characterize intrinsic noise using the slow-scale LNA (ssLNA) method that is recently proposed to deal with models with both fast and slow time scales. Using SSA as ground truth, we find that hSSA and hLNA underestimate the level of intrinsic noise in the Hill-type models, despite of high computational efficiency. The ssLNA approach calculates noise with comparable accuracy as SSA and LNA, while requesting much less computational resources. In addition, the chemical Langevin equation (CLE) under the same slow-scale framework simulates single-cell stochastic trajectories as accurately as SSA yet with significantly lower computational demands. This study shows that ssLNA complemented by slow-scale CLE offers a computational platform that out-performs the other four methods in characterizing intrinsic stochasticity of the Hill-type genetic models.

Keywords: Intrinsic Noise; Gene Regulation; Hill Functions; Linear Noise Approximation

Introduction

Gene regulation is among the most important yet intricate biological processes. Mathematical modeling of gene regulatory networks has provided helpful insights in understanding natural genetic networks and designing synthetic genetic circuits. One type of widely accepted approach for modeling gene regulation is through deterministic rate equations (i.e. ordinary differential equations) [1,2]. In the formulation of complex transcriptional events regulated by transcription factors, Hill-type rate law has been shown to successfully approximate the multibinding-site mechanism and cooperativity [3-5]. In particular, the application of Hill functions to represent activating or inhibitory regulation by transcription factors greatly reduces the complexity of deterministic modeling of gene regulation [6,7]. Nevertheless, it is noteworthy that the deterministic models using ordinary differential equations only reflect the macroscopic behavior of genetic networks averaged over many cells, while neglecting the intrinsically random nature of gene regulation in a single cell, namely the stochasticity inherent in gene expression [8,9].

It has been shown that the dynamics of gene regulation fluctuate stochastically [10]. One source of stochasticity in gene expression is due to a series of events that involve a small number of molecular species such as mRNAs and proteins [11]. Such random fluctuations in the molecular abundance due to the low copy number of species is called intrinsic noise by biologists [8]. Considerable efforts have been made to develop theoretical and computational methods to model the intrinsic stochasticity in biochemical reactions, including gene regulation [11-18]. To this end, the chemical master equation (CME) offers an accurate stochastic description of general chemical reaction systems on the mesoscopic scale, assuming that biochemical kinetics are Markovian [19]. CME can be

applied to describe intrinsic noise by constructing the probability evolution equation (aka master equation) from rate equations, but solving the CME is difficult in general. Gillespie developed a stochastic simulation algorithm (SSA), based on a Monte Carlo numerical procedure, that can exactly simulate stochastic trajectories characterized by the CME [20]. Using SSA, intrinsic noise of a biochemical network at single-cell level can be calculated accurately. The SSA approach implements the Monte Carlo simulation of each reaction at each time step for an individual cell, and such procedure needs to be repeated N times if one wishes to obtain the probability distribution of the intrinsic noise among N number of cells. This leads to high computational demands associated with SSA. As a practical alternative, the linear noise approximation (LNA) of CME is analytically tractable and computationally efficient. Specifically, LNA uses van Kampen's size-expansion of the CME to separate macroscopic concentration (i.e. mean) from the fluctuations around it [21]. As a result, the dynamics of mean, variances and covariances of molecular species can be described by a set of ordinary differential equations. It is shown that LNA accurately approximates intrinsic stochasticity when the molecular numbers are relatively large [22,23].

Notably, the standard methods of SSA and LNA require the theoretical model to be composed of elementary chemical reactions. Hill-type rate law, however, is not elementary in its apparent form. How the existing methods can be used to accurately and efficiently calculate the intrinsic noise of the gene expression described by Hill functions, a well-accepted representation that considerably simplifies genetic models, remains unclear. Efforts have been made to heuristically apply SSA and LNA to the Hill-type dynamics (termed as hSSA and hLNA respectively), whereby the Hill rate is simply treated as a single reaction despite of the non-elementary nature [24-26]. The heuristic approach successfully captures the time trajectory of the mean value of the concentration, but it seems to neglect higher order statistics in concentration fluctuations [27]. Similarly, Thomas et al. showed that, when the heuristic representation of the system with Hill propensity functions is used in the hLNA, the contribution to the overall noise levels from the fast reacting species is underestimated [28]. Indeed, a series of analysis of stochastic simulation methods suggest that further computational modification beyond heuristic approach is needed for Hill-type dynamics to improve the estimation of system noise [15,27,29].

The slow-scale LNA (ssLNA) has recently been introduced as a compromise between the need for network simplification and the precise representation of the system's noise levels. The ssLNA approach reduces the network size when calculating the concentration means, while still accounting for the intrinsic noise generated by the fast reacting species that is neglected by hLNA [28].

In this study we investigate the intrinsic stochasticity of the genetic models corresponding to second-order activating and inhibitory Hill functions, using five different computational methods. First, we apply SSA and LNA to the full-scale elementary-reaction models that are expanded from the two Hill functions. In comparison, the heuristic methods hSSA and hLNA are applied directly to the (non-elementary) Hill functions. Furthermore, the ssLNA is also implemented for Hill genetic activation and inhibition. While all three LNA methods (LNA, hLNA, and ssLNA) show much faster computational speed when compared to the SSA counterparts, only the full-scale LNA and the ssLNA compute the correct levels of intrinsic noise when compared to the result of the full-scale SSA method, which is considered exact. In addition, we simulate the chemical Langevin equations for the Hill functions developed under the ssLNA framework and show that they can produce stochastic trajectories with the similar noise accuracy as SSA, yet requiring much less computational expenses. The investigation indicates that ssLNA together with the associated chemical Langevin equations may provide a relatively accurate as well as efficient computational platform to delineate intrinsic stochasticity of the Hill-type genetic models.

Results and Discussions

In the present study, we focus on exploring the intrinsic noise of two theoretically basic and useful genetic models, namely the second-order Hill-type gene activation and gene inhibition networks. For both cases, the genetic reactions include the cooperative binding of two transcription factor molecules (denoted as T) to the gene promoter (denoted as G), where the T molecules form a transcription dimer (denoted as T₂) first, and subsequently the binding of dimer T₂ to the promoter G respectively initiates or inhibits the mRNA transcription. The mRNA molecule is then translated into protein product.

We seek to understand the performance, including computational accuracy and efficiency, of the five existing stochastic modeling methods in calculating intrinsic noise of the Hill-type gene regulations. To this end, we apply two Gillespie algorithm-based methods, namely SSA and heuristic SSA (hSSA), as well as three LNA-based methods, namely LNA, heuristic LNA (hLNA) and slow-scale LNA (ssLNA), to the aforementioned second-order Hill-type gene activation and gene inhibition networks, via calculating the stochastic molecular abundances of mRNA and protein. Due to its exact implementation of CME, the results of the SSA method are used as the gold standard. The average-cell level behavior, represented by mean mRNA and protein molecule count, and the single-cell level intrinsic noise, represented by coefficient of variation (CV), defined as the ratio between standard deviation and mean of mRNA and protein molecules [30-32], are obtained using each method and compared with the results of the SSA method.

The Gillespie algorithm-based approach differs from the LNA-based approach in that the former needs to simulate multiple molecule-number trajectories for noise analysis across a population of single cells. To generate meaningful statistical data for the characterization of model stochasticity (mean and CV) using the SSA and the hSSA methods, we compute 10^2 - 10^3 single-cell trajectory samples, the number used in previous computational as well as experimental noise studies [30,33,34].

The LNA framework uses the van Kampen's Ω -expansion method to approximate the Chemical Master Equation, where Ω represents the system volume. Specifically, the CME is Taylor-expanded about the macroscopic dynamics in powers of $1/\sqrt{\Omega}$. The first-order terms give the rate equations governing macroscopic dynamics (mean abundance), and the next-order terms give the Linear Noise Approximation (LNA) for dynamics of fluctuations [21]. The intrinsic fluctuation approximated by the solution of LNA has been shown to be multivariate Gaussian, whereby the ordinary differential equation (ODE) of its correlation matrix can be also obtained [21,35]. By solving these ODEs for the means and the variances of the stochastic variables (i.e. molecular numbers of mRNA and protein), LNA allows for direct calculation of the intrinsic noise (i.e. CV) of the models of Hill-type genetic activation and inhibition.

The standard SSA and LNA methods are rigorously developed for biochemical models consisting of elementary reactions. For the models that incorporate nonlinear non-elementary (loosely called 'phenomenological') terms such as Michaelis-Menten kinetics and Hill functions, the SSA and the LNA method have been applied heuristically, by assuming quasi-steady-state of the fast-reacting species and using the nonlinear terms as approximate propensity functions [24,28,36]. To compare with the standard approach of SSA and LNA, we also explore the performance of hSSA and hLNA in quantifying the intrinsic stochasticity of Hill-type gene regulations, where the fast binding/unbinding processes are assumed to reside at quasi-steady states.

The slow-scale LNA is a new general method derived based on standard LNA, which promises to rigorously quantify the statistics of intrinsic noise of the reduced models under quasi-steady-state approximation (QSSA), without being expanded into their elementary-reaction full models [28]. A major goal of this study is to evaluate the performance of ssLNA in describing the noise statistics of the Hill-type gene regulations. We seek to explore how QSSA together with the associated time-scale separation can be applied within the framework of LNA to improve the computational efficiency of noise analysis of the class of compact models containing 'phenomenological' terms while maintaining accuracy.

The detailed procedures of LNA, ssLNA, as well as hLNA and hSSA methods are described in the Methods section and Supplementary Materials.

Intrinsic Stochasticity of Hill-type Gene Activation

We first explore the intrinsic stochasticity of a Hill-type gene regulation process represented by a transcription rate $\frac{T^2}{T^2 + k_d^2}$, where T represents the concentration of transcription factor, and k_d represents Hill constant. This second-order Hill function corresponds to 'phenomenological' approximation of a full model that includes five elementary reactions represented by the scheme in formula (1)-(3) below. Specifically, two monomers of transcription factor T form a dimer T_2 . T_2 binds to gene G and transactivates mRNA molecule M .



The remaining three elementary reactions in the gene regulation process include translation from mRNA M to protein P and degradation of M and P molecules as listed in equation (4)-(6).



Since binding and unbinding reactions in general occur at much faster rate than transcription, translation and degradation, we can apply QSSA to assume that the fast variables T_2 and GT_2 reach equilibrium for modeling simplification. As a result, the five elementary reactions in the equations (1)-(3) are reduced to a 'phenomenological' step of transcription, whose rate is proportional to the steady-state portion of the GT_2 complex among total gene (denoted as G_T) as given in equation (7) (see Supplemental Materials for the derivation).

$$\frac{[GT_2]}{G_T} = \frac{[T]^2}{[T]^2 + \frac{k_2 k_4}{k_1 k_3}} = \frac{[T]^2}{[T]^2 + k_d^2} \quad (7)$$

This yields a second-order Hill function provided we assume $k_d = \sqrt{\frac{k_2 k_4}{k_1 k_3}}$.

We use five methods to calculate the intrinsic noise of the above gene activation process. First, the standard SSA and LNA methods are applied to the full model consisting of all the 8 elementary reactions in equations (1) through (6). Second, heuristically using the ‘phenomenological’ Hill function (7) as the propensity function for the simplified transcription reaction, together with the remaining elementary reactions in equations (4)-(6), we apply hSSA and hLNA. Lastly, we apply ssLNA to the gene regulation described by equations (1)-(6), assuming that G, T, T₂ and GT₂ are fast variables reaching quasi-steady state, and M and P are slow variables [28]. Each method is used to compute the time trajectories of macroscopic mean and intrinsic noise of mRNA (M) and protein (P) in the gene regulation. The intrinsic noise is quantified by the coefficient of variation (CV), defined as standard deviation divided by mean. For the Gillespie-based methods, SSA and hSSA, we compute >10³ number of stochastic samples to obtain means and standard deviations that are statistically meaningful.

We seek to evaluate the accuracy in computing the trajectories of mean and intrinsic noise, as well as the computational efficiency of computing the noise. To this end, the standard SSA of the full gene activation model is used as the gold standard for accuracy evaluation. In Figures 1-3, trajectories of the means and CVs for the activating Hill regulation obtained by five methods are compared with the results of the standard SSA method. Each of the figures corresponds to the same model of gene activation, with parameters calibrated to evaluate performance under different ranges of steady-state copy numbers. Specifically, we simulate three scenarios where the mean molecular number of proteins at steady state is respectively in the range of ~10² (Figure 1), ~10³ (Figure 2), and ~10⁴ (Figure 3).

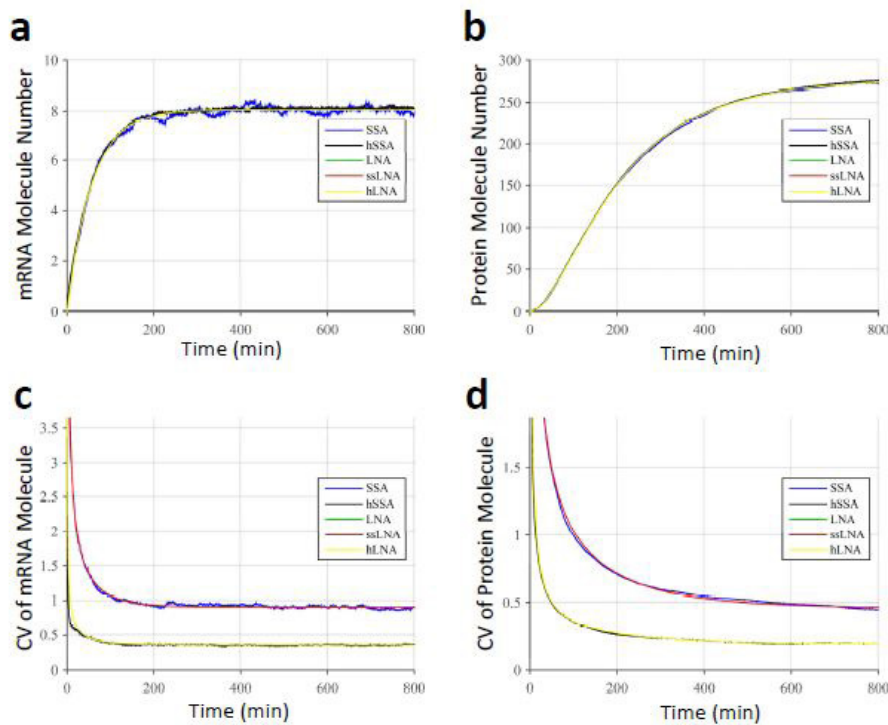


Figure 1: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order activating Hill-type gene regulation, calculated using SSA, hSSA, LNA, hLNA and ssLNA methods. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Model parameters are tuned for ~10² number of protein molecules at steady state

As shown in the panels (a) and (b) of Figures 1-3, all the trajectories of the macroscopic means of M and P converge toward the results of standard SSA (note that the LNA and ssLNA trajectories are hidden beneath the hLNA trajectory), indicating that all the five methods give rise to similar degree of accuracy in describing the average-cell behavior of the activating gene regulation. Notably, using Hill functions directly as propensity functions in the heuristic methods, hSSA and hLNA, do not seem to affect the accuracy of computing the mean behavior, which is consistent with previous reports [24,25]. Such agreement persists across all the three ranges of molecule number, albeit the higher number of molecules, the less degree of stochasticity in the mean trajectories computed by the Gillespie-based SSA and hSSA methods, which is expected for stochastic simulations.

With regard to the precision of noise calculation, however, we find some discrepancies between the results of the five methods. Figure 1-3 (c) and (d) show that, for all the three ranges of molecule numbers, the CV trajectories computed by SSA, LNA and

ssLNA converge together (note that the LNA trajectory is hidden beneath the ssLNA trajectory). But the CV trajectories computed by the heuristic methods hSSA and hLNA converge to another steady-state value smaller than that computed by the standard SSA, the ground truth. For instance, when the number of protein molecules is in the hundreds at steady state, the CV of mRNA and protein is respectively underestimated by $\sim 60\%$ and $\sim 30\%$ at steady state. The degree of underestimation is alleviated as the copy number increases. In particular, when the number of protein molecules is in the 10^4 at steady state, the CV of mRNA and protein is respectively underestimated by $\sim 10\%$ and $\sim 4\%$ at steady state. On the other hand, the results of the standard LNA and ssLNA agree well with the ground truth under all the three copy-number scenarios. That is, the ssLNA method based on reduced model of slow variables is able to achieve the same precision of noise calculation as that of the standard SSA and standard LNA methods, which are based on the full model of all elementary reactions.

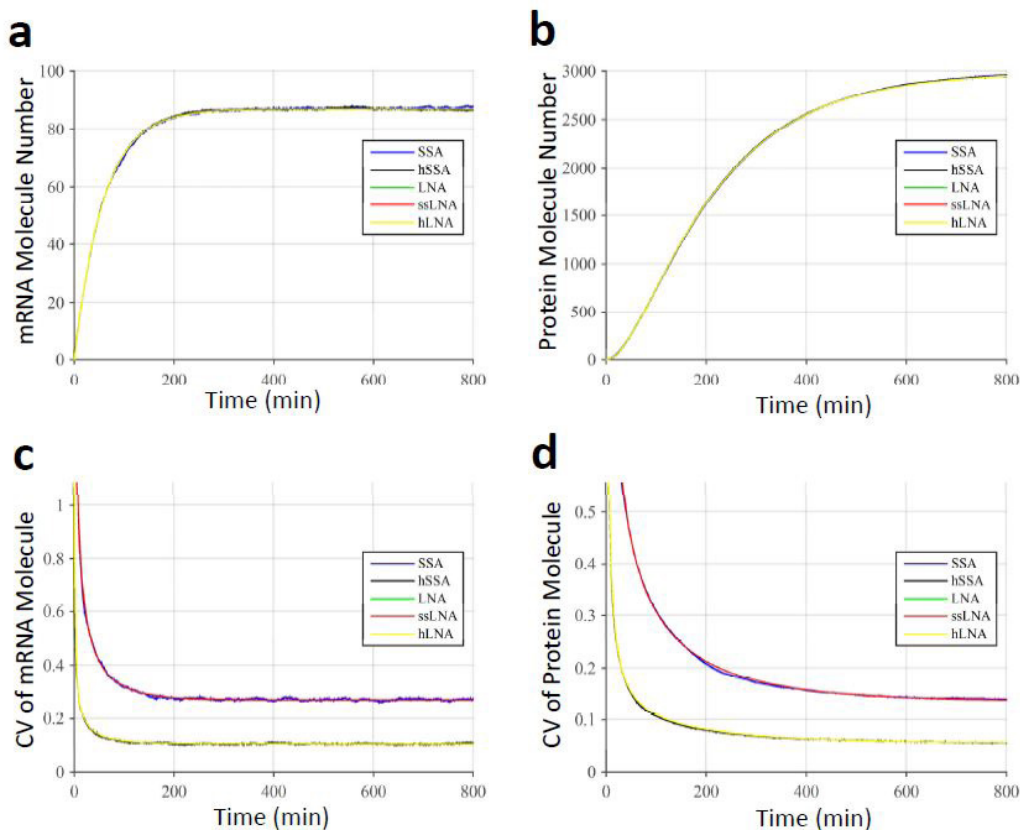


Figure 2: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order activating Hill-type gene regulation, calculated using SSA, hSSA, LNA, hLNA and ssLNA methods. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Model parameters are tuned for ~ 103 number of protein molecules at steady state

CPU Time(sec) Method	Activation			Inhibition		
	$\sim 10^2$ Protein Molecules	$\sim 10^3$ Protein Molecules	$\sim 10^4$ Protein Molecules	$\sim 10^2$ Protein Molecules	$\sim 10^3$ Protein Molecules	$\sim 10^4$ Protein Molecules
SSA (1000 runs)	43.751	698.438	1252.501	12382.813	3553.125	1754.692
hSSA (1000 runs)	1.719	39.063	60.002	14.063	13.751	140.625
LNA	13.484	14.25	13.953	9.719	7.406	7.344
ssLNA	0.047	0.031	0.047	0.031	0.047	0.047
ssCLE (1000 runs)	3.125	4.688	3.438	96.875	91.254	81.253
hLNA	0.016	0.016	0.016	0.016	0.031	0.016

Table 1: CPU time comparison for six methods (SSA, hSSA, LNA, hLNA, ssLNA, SSA, hSSA and ssCLE) under three copy-number scenarios. The SSA and hSSA simulations as well as the numerical solution of the slow-scale chemical Langevin equation (ssCLE) require the averaging of multiple runs to estimate the intrinsic noise metrics. Specifically, we compare their CPU time of using 1000 runs for CV computation.

Although the SSA method provides the highest accuracy in quantifying intrinsic noise, it has significantly high demand for computation times, evaluated by CPU times (see the Methods section for details), especially when the number of molecules increase (Table 1). For instance, the CPU time used for SSA to obtain the CVs of 1,000 samples increases from 4.375 seconds for $\sim 10^2$ protein molecules to 125.25 seconds for $\sim 10^4$ protein molecules, a 28 fold increment. Note that the CVs reported above are calculated using 2000-3000 runs to produce relatively smooth mean trajectory, indicating that high computational demand is

required to obtain reliable statistical estimation of CV using SSA. The LNA approach yields considerably greater computational efficiency, especially for high-molecule case, as it directly computes population-level variances without sampling individual stochastic trajectories. Specifically, each of the three copy-number cases takes only ~ 14 seconds. As expected, the computation time of the heuristic methods hSSA and hLAN is greatly reduced in comparison with their full-model counterpart by about 20 and 300 fold respectively. The computation efficiency of ssLNA is also very good whereby the time is reduced by ~ 300 -400 fold comparing to the standard LNA. The above results show that, among the five methods, the ssLNA method achieves not only high accuracy but also fast computation in quantifying the intrinsic noise of the model of Hill-type gene activation.

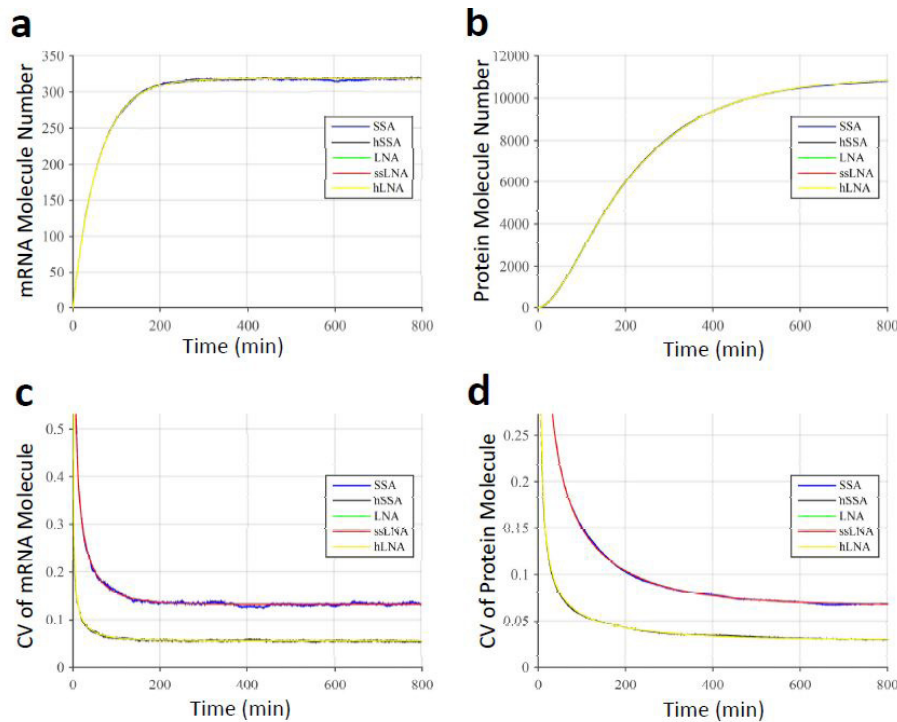


Figure 3: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order activating Hill-type gene regulation, calculated using SSA, hSSA, LNA, hLNA and ssLNA methods. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Model parameters are tuned for $\sim 10^4$ number of protein molecules at steady state

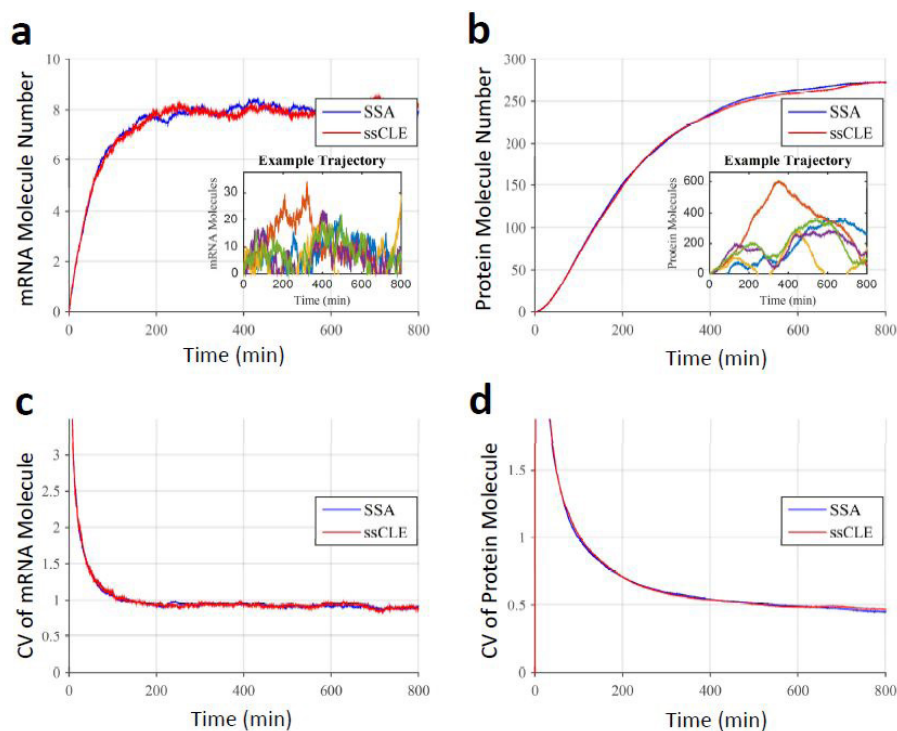


Figure 4: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order activating Hill-type gene regulation, calculated using SSA as well as the chemical Langevin equation (CLE) under the ssLNA framework. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Shown in the insets of a) and b) are five CLE sample paths. Model parameters are tuned for $\sim 10^2$ number of protein molecules at steady state

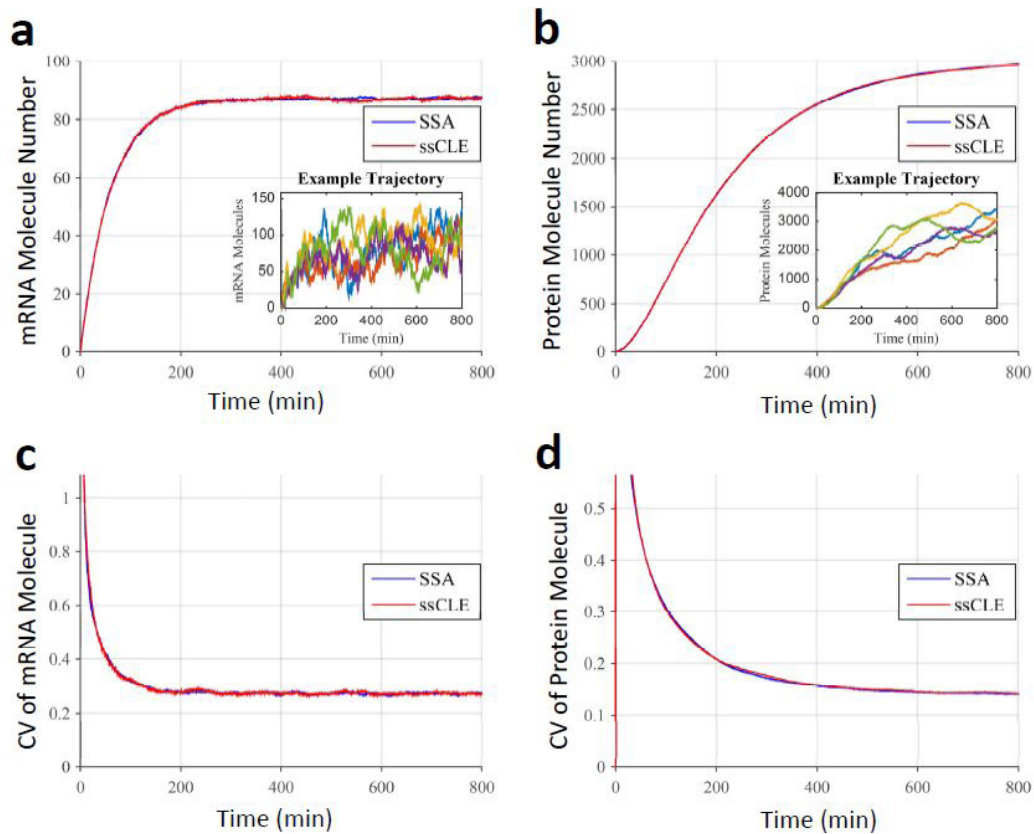


Figure 5: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order activating Hill-type gene regulation, calculated using SSA as well as the chemical Langevin equation (CLE) under the ssLNA framework. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Shown in the insets of a) and b) are five CLE sample paths. Model parameters are tuned for $\sim 10^3$ number of protein molecules at steady state

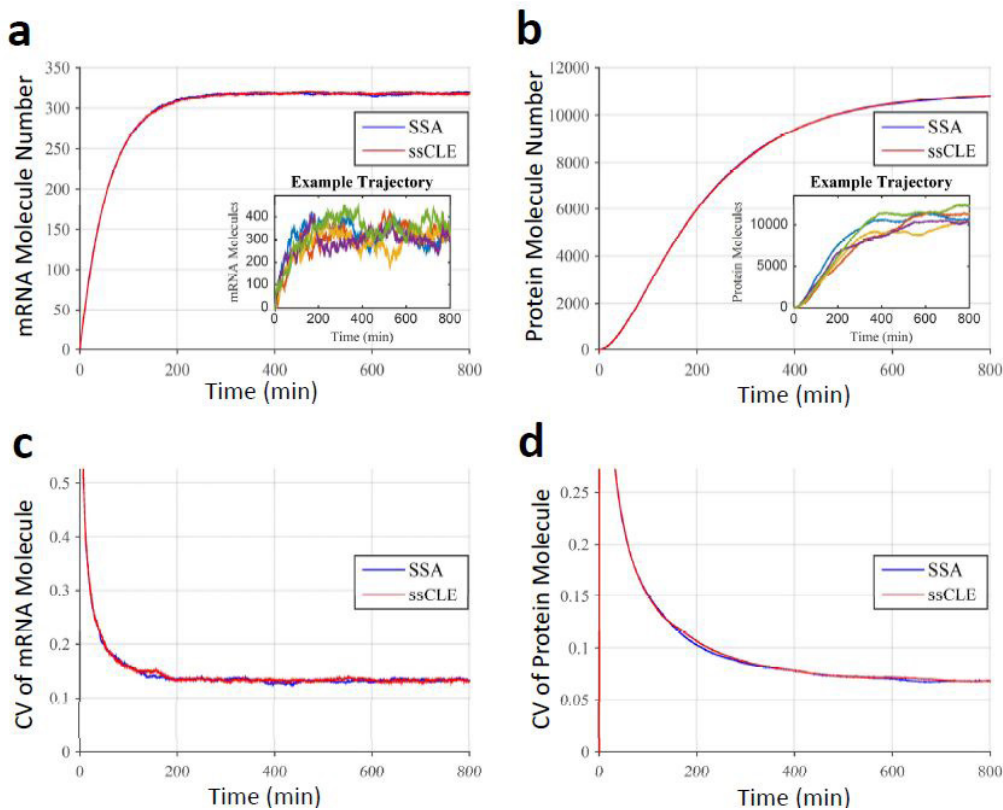


Figure 6: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order activating Hill-type gene regulation, calculated using SSA as well as the chemical Langevin equation (CLE) under the ssLNA framework. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Shown in the insets of a) and b) are five CLE sample paths. Model parameters are tuned for $\sim 10^4$ number of protein molecules at steady state

While the ssLNA method calculates variances efficiently without running stochastic simulations, the chemical Langevin equation (CLE) provides a complementary solution to approximate stochastic individual trajectories. The CLE formulates as a stochastic differential equation (Methods), and thus is much faster to solve than SSA, which exactly simulates the trajectories of the CME. In particular, we apply the slow-scale CLE method that has been developed under the timescale-separation framework of ssLNA [28]. The computation time of stochastic trajectories using CLE indeed reduces considerably (Table 1), especially for the $\sim 10^4$ copy-number case (36-fold reduction compared with SSA). To validate accuracy of the stochastic simulation by CLE under the ssLNA framework, we further compute the macroscopic mean and CVs of multiple CLE simulation samples. For the three different molecular copy numbers, they all agree well with the results of the standard SSA method (Figures 4-7). These results demonstrate that, for the noise analysis employing QSSA-based reduced model, CLE could serve as a well-behaved single-cell simulation tool in complement to ssLNA that only generates macroscopic mean trajectory rather than individual sample paths.

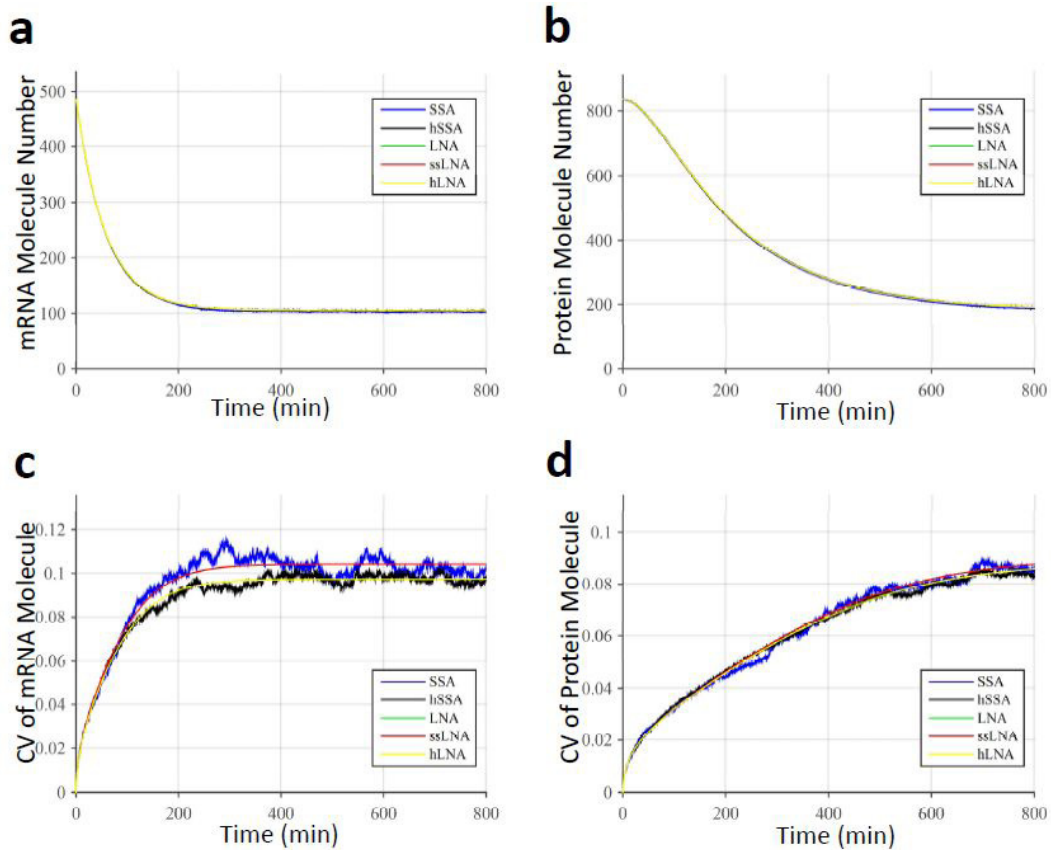


Figure 7: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order inhibitory Hill-type gene regulation, calculated using SSA, hSSA, LNA, hLNA and ssLNA methods. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Model parameters are tuned for $\sim 10^2$ number of protein molecules at steady state

Intrinsic Stochasticity of Hill-type Gene Inhibition

We further study the intrinsic noise of an inhibitory gene regulation process that is classically lumped as a second-order Hill-type transcription rate $\frac{k_d^2}{T^2 + k_d^2}$. The corresponding eight elementary reactions are very similar to the gene activation case. Specifically, the transcription-factor dimer formation and its binding to gene are exactly the same as (1)-(2), where the same QSSA can be applied to derive a reduced model. The inhibitory transcription, however, assumes that mRNA transcription is activated only if gene (G) is not bound by T, which is represented by regulation (8):



As a result, under the QSSA condition, the transcription rate is proportional to the fraction of free DNA (G) to total gene at steady state (equation (9)).

$$\frac{[G]}{G_T} = 1 - \frac{[GT_2]}{G_T} = \frac{k_d^2}{[T]^2 + k_d^2} \quad (9)$$

For the calculation of intrinsic noise among a cell population, standard SSA and LNA are applied to the full model accounting for the elementary reactions (1)-(2), (8) and (4)-(6), while hSSA and hLNA are applied to the reduced model of the reactions (4)-(6) associated with the non-elementary transcription rate (9). In addition, intrinsic noise is also computed by ssLNA, and single-cell stochastic trajectories are obtained using the slow-scale CLE method.

The noise level computed by SSA of the full model is again used as the gold standard for accuracy evaluation. We simulate with steady-state protein copy-number in three ranges (again $\sim 10^2$, $\sim 10^3$, and $\sim 10^4$) by varying the amount of transcription factor. Figure 7-9 (a) and (b) show that the mean dynamics of mRNA and protein computed by hSSA, LNA, hLNA and ssLNA agree well with the result by SSA (again the LNA and ssLNA trajectories are hidden beneath the hLNA trajectory). The cellular noise in Figure 7-9 (c) and (d), however, show that the population CV trajectory computed by the heuristic methods (hSSA and hLNA) is slightly smaller compared to that computed by the full-model methods (SSA and LNA) as well as ssLNA (again the LNA trajectory is hidden beneath the ssLNA trajectory). That is, the intrinsic noise of inhibitory gene model is still underestimated by the heuristic methods, although the degree of underestimation is much less (up to $\sim 1\%$) compared to that for the activation gene model (up to 60%). Note that the overall noise levels are relatively small in the gene inhibition model, because large numbers of initial molecules have to be assumed for the model to decay toward the desired steady-state copy number levels, which are in the ranges consistent with the activating gene regulation case. This could be the main reason why the degree of underestimation by the heuristic methods is much lower for the inhibitory Hill model. On the other hand, the LNA and ssLNA methods yield population CV paths comparable to those produced by SSA for all the three copy-number cases, proving again their reliable accuracy of noise computation. In terms of simulating individual stochastic trajectory, the slow-scale CLE method again results in considerable accuracy in the mean and CV paths, as demonstrated by their good agreement with the results by the full-model SSA method (Figure 10-12).

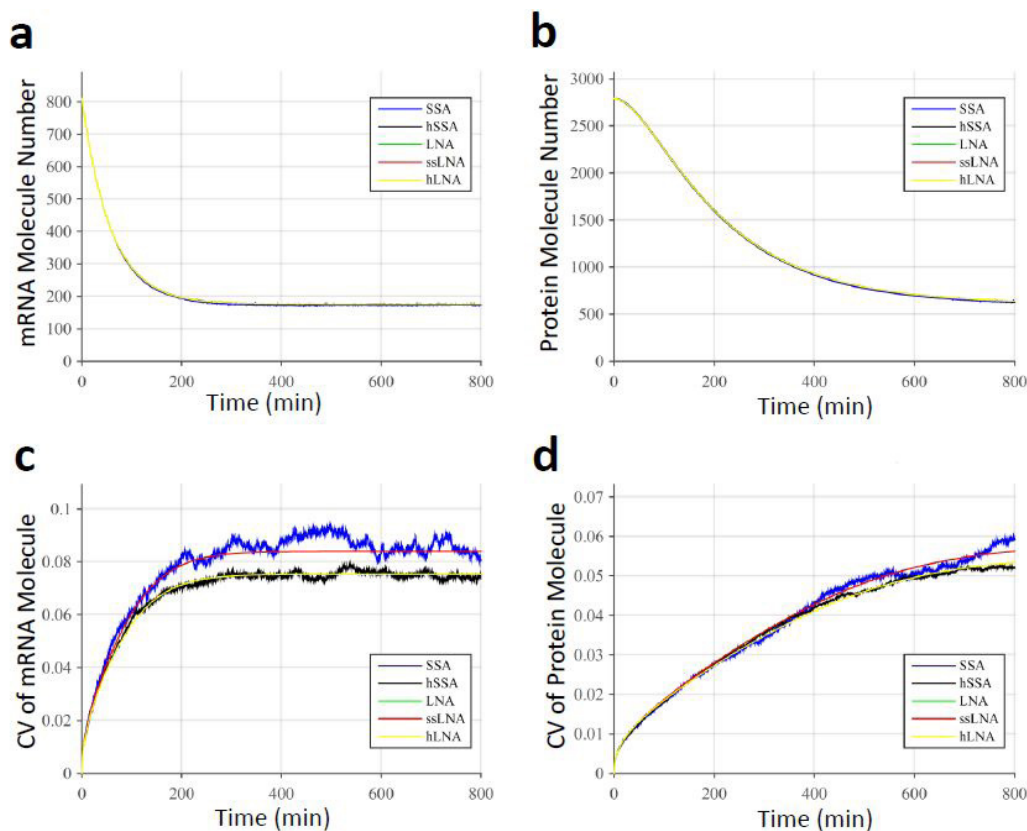


Figure 8: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order inhibitory Hill-type gene regulation, calculated using SSA, hSSA, LNA, hLNA and ssLNA methods. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Model parameters are tuned for $\sim 10^3$ number of protein molecules at steady state

With regard to efficiency of noise calculation, the full-model SSA method poses the highest computational demand, while the LNA-based methods are significantly faster (Table 1). In particular, the ssLNA method reduces the computation time by at least 10^3 fold comparing to SSA and by at least 150 fold comparing to LNA. Model reduction in the heuristic methods (hSSA and hLNA) leads to a computational time ~ 1 -3 orders of magnitude smaller than their full-model counterpart at the price of noise underestimation. For simulating individual stochastic path, the slow-scale CLE method reduces the computation time of the standard SSA method by at least 20 times, showing its computational efficiency in addition to accuracy.

In summary, we evaluate the performance of SSA-based methods (SSA, hSSA) and LNA-based methods (LNA, hLNA, ssLNA) in computing intrinsic noise of a gene inhibition model. The results are similar to those of the gene activation model in that

heuristically integrating the reduced model (i.e. Hill function) to SSA or LNA impairs the accuracy of noise computation. In addition, using the slow-scale framework, we obtain considerable accuracy as well as computational efficiency in quantifying intrinsic noise via ssLNA, as well as in delineating individual paths via CLE.

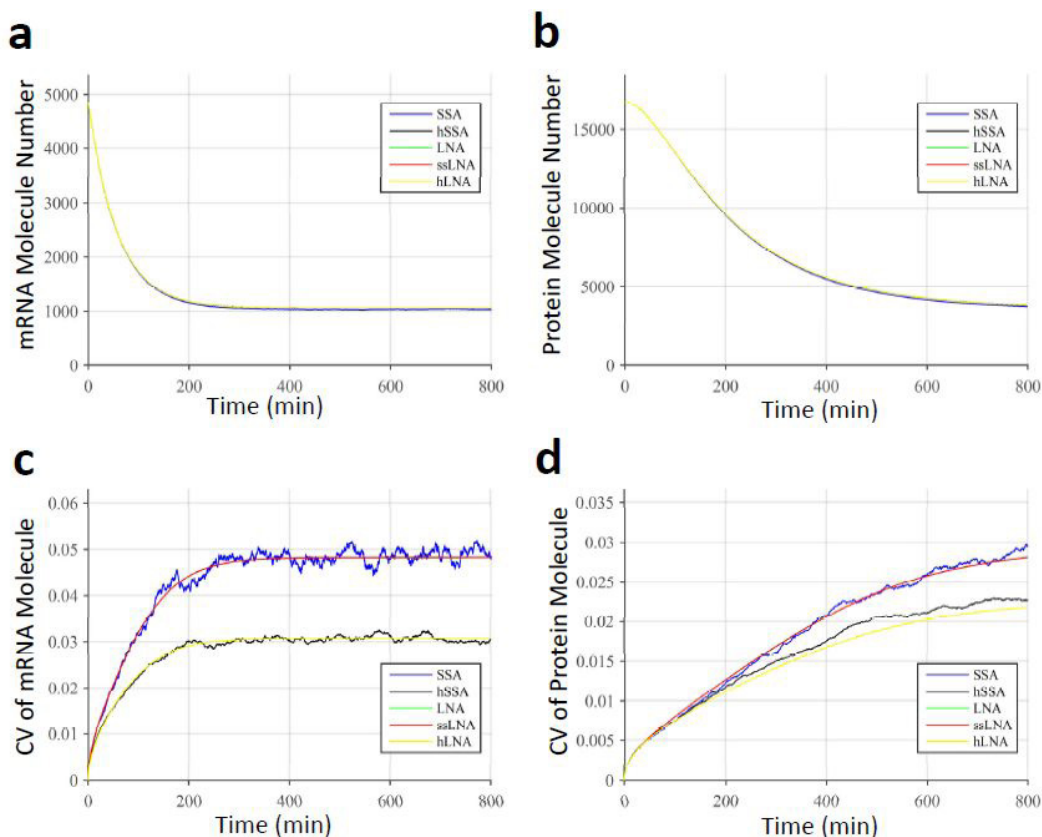


Figure 9: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order inhibitory Hill-type gene regulation, calculated using SSA, hSSA, LNA, hLNA and ssLNA methods. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Model parameters are tuned for ~10⁴ number of protein molecules at steady state

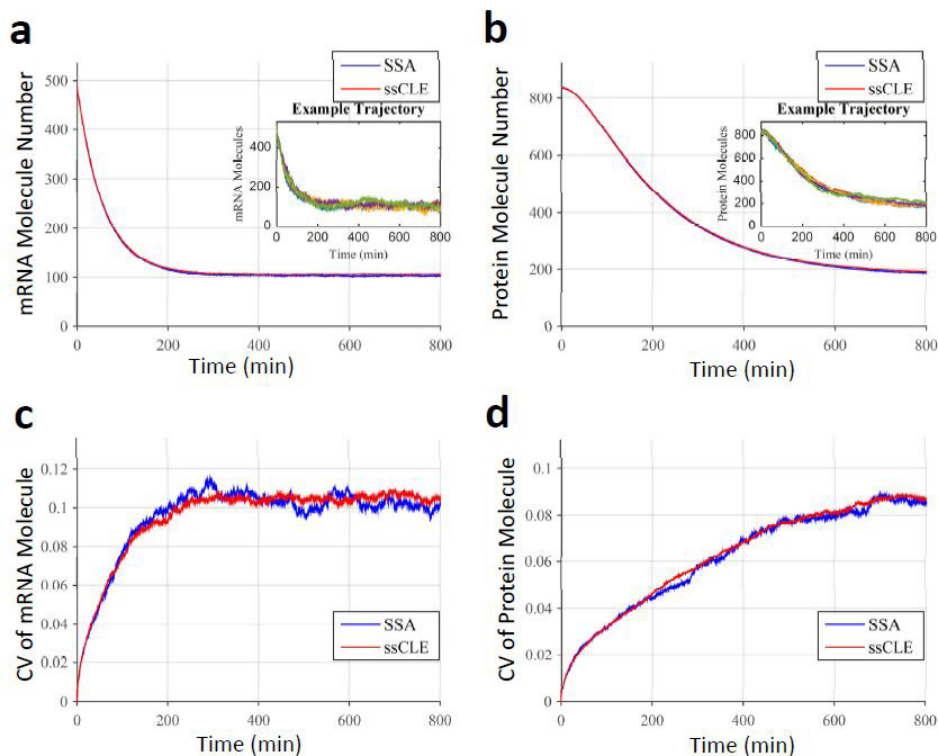


Figure 10: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order inhibitory Hill-type gene regulation, calculated using SSA as well as the chemical Langevin equation (CLE) under the ssLNA framework. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Shown in the insets of a) and b) are five CLE sample paths. Model parameters are tuned for ~10² number of protein molecules at steady state

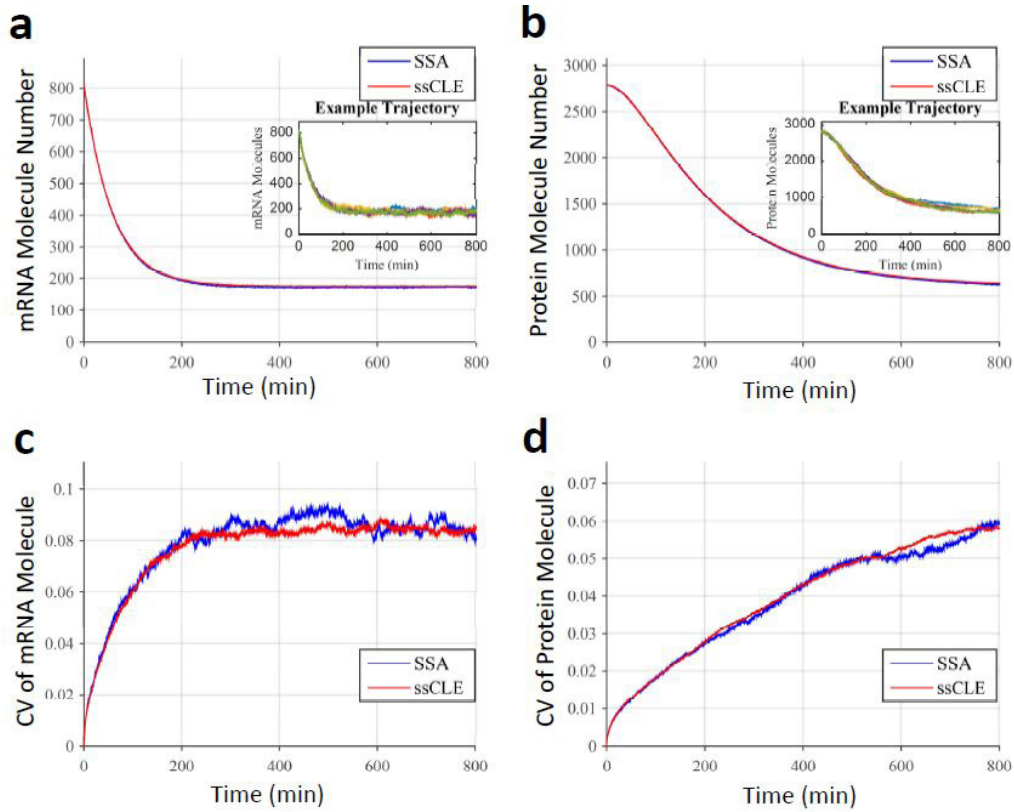


Figure 11: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order inhibitory Hill-type gene regulation, calculated using SSA as well as the chemical Langevin equation (CLE) under the ssLNA framework. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Shown in the insets of a) and b) are five CLE sample paths. Model parameters are tuned for $\sim 10^3$ number of protein molecules at steady state

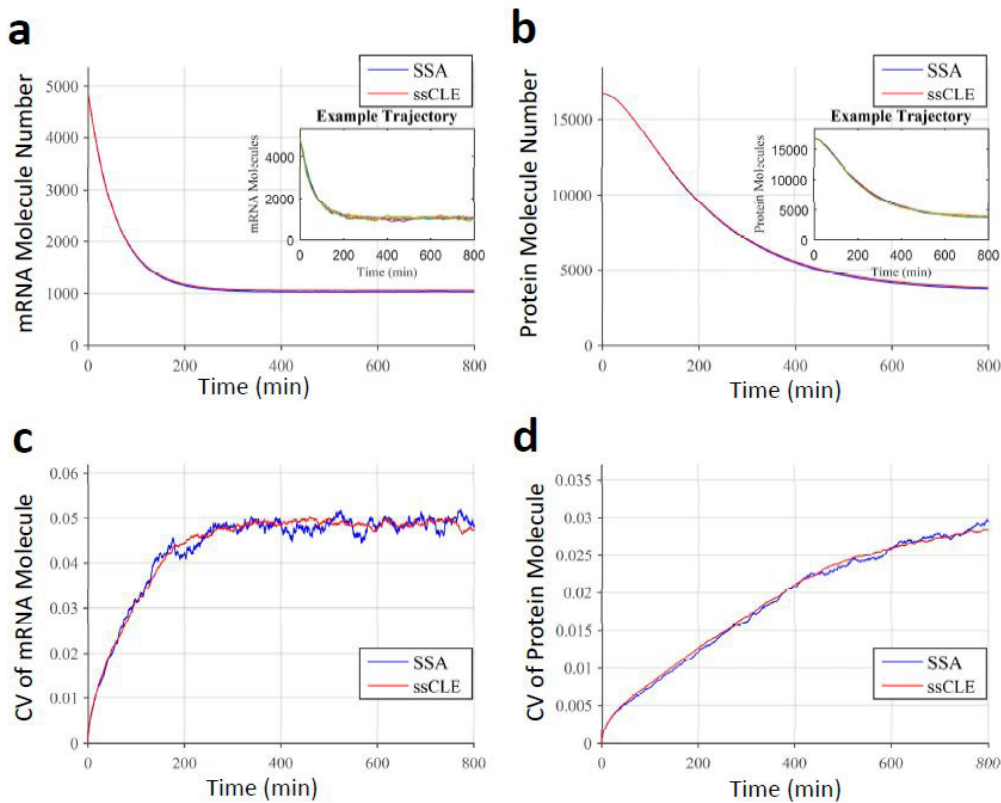


Figure 12: Mean (a, b) and CV (c, d) trajectories of mRNA and protein molecules for second-order inhibitory Hill-type gene regulation, calculated using SSA as well as the chemical Langevin equation (CLE) under the ssLNA framework. a) Mean of mRNA. b) Mean of protein. c) CV of mRNA. d) CV of protein. Shown in the insets of a) and b) are five CLE sample paths. Model parameters are tuned for $\sim 10^4$ number of protein molecules at steady state

Methods

Standard linear noise approximation (LNA) method

Suppose the state vector X denotes the macroscopic concentrations of a model containing elementary reactions (considered as full-scale model). With respect to cellular dynamics, the macroscopic concentrations represents mean molecular behavior averaged over a cell population. Due to intrinsic stochastic fluctuations, the microscopic state of each individual cell is random variable: $x=X+\varphi$, where φ represents the linear-term fluctuations and is Gaussian centered around X within the LNA framework. If the matrix S denotes the stoichiometry matrix, where each column represents the variation in molecule counts of the species for a certain elementary reaction, and f is reaction propensity vector, then the governing deterministic ODE of the means X is written as: $\dot{X} = S\vec{f}$ (10).

The ODE of the correlation matrix $H_F = \left\{ \text{cov}(\varphi_i(t), \varphi_j(t)) \right\}$, containing the variances and covariances for the fluctuations of all the components in the system, is given by $\frac{dH_F}{dt} = J_F H_F + H_F J_F^T + D_F$ (11), where $J_F = \frac{\partial(S\vec{f})}{\partial x}$ is the Jacobian of the rate

equation (10), and $D_F = S \text{diag}(\vec{f}) S^T$ is called the diffusion matrix [21]. Using the covariance trajectories computed from equation (11) and the means modeled by (10), we can compute the time courses of the coefficient of variation (CV = standard deviation divided by mean), which is commonly used to characterize the intrinsic noise across a cell population.

Slow-scale LNA (ssLNA) method and slow-scale chemical Langevin equation

Suppose the species in a model can be time separated into fast and slow species, where the fast species are those assumed to reach pseudo equilibrium under the quasi-steady state approximation (QSSA). We partition the state vector as $X = (X_s, X_f)$, where X_s and X_f denote the vector of fast and slow variables, respectively. Under QSSA of fast species, the deterministic model can be reduced to a slow-scale rate equation for X_s , $\dot{X}_s = S_h \vec{f}_h$ (12), which contains 'phenomenological' terms such as Michaelis-Menten and Hill functions.

The ssLNA method can be implemented in such a slow-scale framework by first partitioning the J_F and S matrices of the full model as follows:

$$J_F = \frac{\partial(S\vec{f})}{\partial x} = \begin{bmatrix} J_s & J_{sf} \\ J_{fs} & J_f \end{bmatrix}, \quad S = \begin{bmatrix} S_s \\ S_f \end{bmatrix}, \quad \begin{array}{l} s, \text{ slow species} \\ f, \text{ fast species} \end{array} \quad (13)$$

The variances of the fluctuations around the means of X_s can be solved using the ODE of the slow-scale correlation matrix H_{ss} : $\frac{dH_{ss}}{dt} = J_s H_{ss} + H_{ss} J_s^T + D_{ss}$ (14), where J_s is shown to be the same as the Jacobian matrix J_h of the reduced rate equation (12) of slow species [28], and the diffusion matrix is given as:

$$D_{ss} = \left(S_s - J_s J_f^{-1} S_f \right) \text{diag}(\vec{f}) \left(S_s - J_s J_f^{-1} S_f \right)^T \quad (15)$$

The time courses of the coefficient of variation (CV) are obtained using the covariance trajectories computed from equation (14) and the means modeled by the slow-scale rate equations (12).

In the framework of LNA, chemical Langevin equation (CLE) has been proposed to effectively approximate single-cell stochastic time trajectory of the species in CME. In this study, we use the CLE developed under ssLNA, $\frac{d\varphi_s}{dt} = J_{ss}\varphi_s + (S_s - J_s J_f^{-1} S_f) \sqrt{S_h \text{diag}(\vec{f}_h) S_h^T} \vec{\Gamma}(t)$ (16), where each element of the $\vec{\Gamma}(t)$ vector is a temporally uncorrelated, independent Gaussian white noise [21, 28]. Euler-Maruyama method is used to numerically solve the stochastic differential equation (16). Same as the SSA method, one run of simulation of CLE represents a single-cell path sample. We use similar sample size as SSA (10^2 - 10^3) to obtain the statistics of cell-population level noise (i.e. standard deviation and mean) for the ssLNA-based CLE.

Metrics for stochastic computation

To compute the means and intrinsic noise of cell-population dynamics, ~2000-3000 runs of SSA, hSSA, and CLE trajectories are used to obtain a converging mean trajectory. The CV is used to measure the population-level intrinsic noise. For the comparison of computational efficiency, CPU time is used to evaluate each method's performance. The CPU time, obtained using the `cputime` function in MATLAB, evaluates the time used for each method. The time is intended as used by a function for the computation, excluding the possible setup time. LNA, hLNA and ssLNA times are averaged over a set of ~100 runs. SSA, hSSA, and CLE times

are also averaged over ~100 runs, but the time taken for 1000 runs is reported in Table 1 for comparison, since several runs are required to approximate the solution of the CME. All computations are performed using Dell Precision T3500 with Intel® Xeon® W3530 (8MB Cache, 2.80 GHz, 4.80 GT/s Intel® QPI) and 6 GB Memory.

Conclusions

Gene regulation is a molecular process that could present high intrinsic noise. For the gene activation and the gene repression systems studied here, we apply five different stochastic analysis methods to provide a comparison on the quantitative measure of the intrinsic noise. The comparisons of the performance of the five methods are summarized in Figure 13 (normalized CV) and Figure 14 (normalized CPU time). In both model cases, while the SSA can provide the most accurate noise representation, it requires large computational resources (Figure 14), which could make the realization of large models infeasible. The heuristically reduced SSA (hSSA) provides a good characterization of the mean molecule counts, but fails to faithfully reproduce fluctuations around the mean (Figure 13). The linear noise approximation (LNA) approach significantly reduces computational times (Figure 14) while preserving noise accuracy (Figure 13, overlapped with ssLNA). The heuristic reduction of LNA models (hLNA), however, presents the same degree of inaccuracies as the hSSA in noise calculation (Figure 13). The slow-scale approach (ssLNA) not only performs

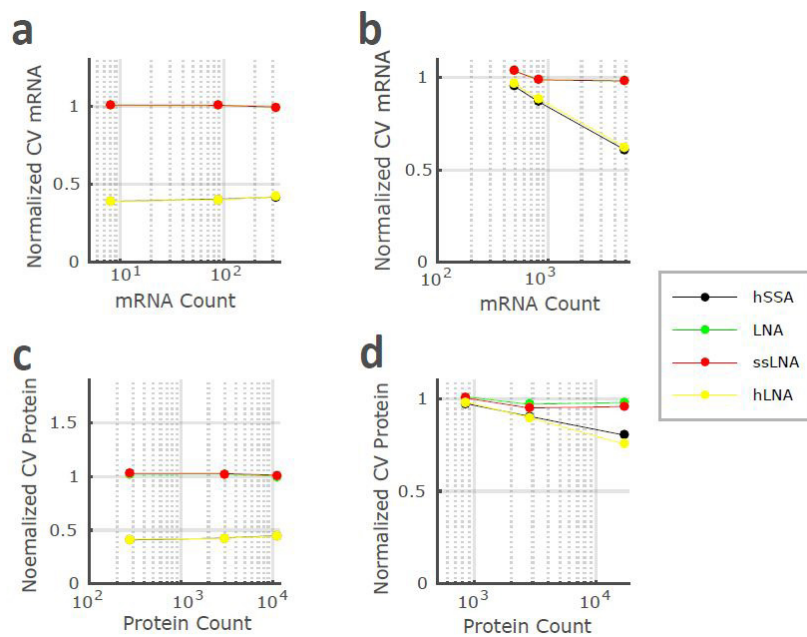


Figure 13: CVs via the hSSA, LNA, ssLNA and hLNA methods normalized by CV via the SSA method. Normalized mRNA and protein CVs of the second-order activating Hill-type gene regulation model in (a, c) show that noise given by ssLNA (red) overlaps with those by standard SSA (normalized value 1) and LNA (green) methods. In (a, c) the noise by hSSA (black) and hLNA (yellow) overlap, both of which are underestimated comparing to the noise given by standard SSA. Normalized mRNA and protein CVs of the second-order inhibitory Hill-type gene regulation model in (b, d) show the similar trend in CV comparison of the five methods, except that when species molecule number is low, the underestimation is not as significant.

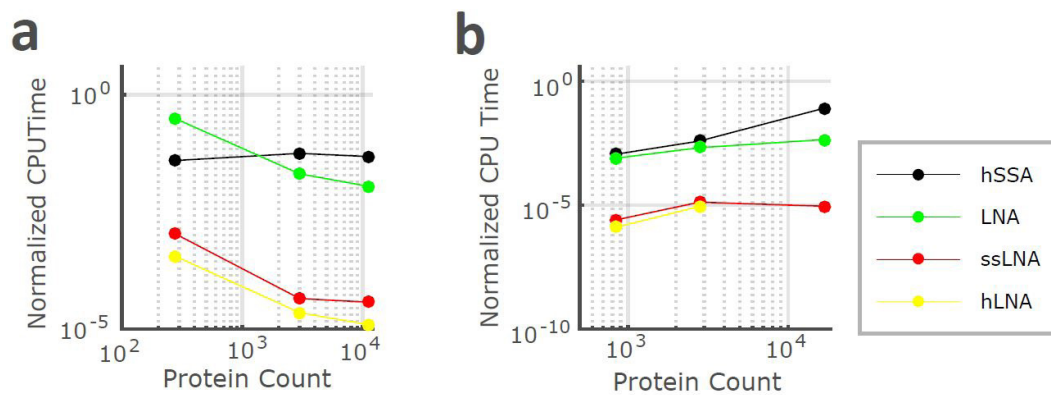


Figure 14: CPU times of the hSSA, LNA, ssLNA and hLNA methods normalized by that of the SSA method for the second-order activating Hill-type gene regulation model (a) and the second-order inhibitory Hill-type gene regulation model (b). For hSSA and SSA, CPU times of 1000 runs of simulation for CV calculation are used in the plots. Comparing to the CPU time of SSA (normalized value 1), all other methods use less CPU time for both models. In particular, comparing to SSA, the LNA (green) and hSSA (black) methods reduce the CPU time by similar degree, while ssLNA (red) and hLNA (yellow) further reduce the CPU time by another similar degree.

as accurately as the full-scale LNA models in terms of noise quantification (Figure 13), but also improves considerably on the computational demand compared to the SSA and LNA methods (Figure 14). Furthermore, the slow-scale framework allows us to precisely and efficiently simulate individual stochastic paths via CLE. Therefore, we find that to characterize intrinsic stochasticity of the gene regulation represented by Hill functions the best choice is the ssLNA modeling framework, which provides the best compromise between theoretical accuracy using partially reduced model instead of full model and computational efficiency using ODE solution instead of stochastic simulations.

Acknowledgements

The work was supported by a NSF/NIGMS grant DMS1361355, which is gratefully acknowledged.

Supplementary Info

References

- de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9, 67-103.
- Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* 9, 770-780.
- Santillan, M. (2008). On the Use of the Hill Functions in Mathematical Models of Gene Regulatory Networks. *Math. Model. Nat. Phenom.* 3, 85-97.
- Meister, A., Du, C., Li, Y.H., and Wong, W.H. (2014). Modeling stochastic noise in gene regulatory systems. *Quant Biol* 2, 1-29.
- Alon, U. (2006). *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Boca Raton, FL: Chapman & Hall/CRC Press).
- Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339-342.
- Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335-338.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183-1186.
- Blake, W.J., M, K.A., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature* 422, 633-637.
- Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* 99, 12795-12800.
- Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* 98, 8614-8619.
- Cook, D.L., Gerber, A.N., and Tapscott, S.J. (1998). Modeling stochastic gene expression: implications for haploinsufficiency. *Proc Natl Acad Sci U S A* 95, 15641-15646.
- Cai, X., and Wang, X. (2007). Stochastic Modeling and Simulation of Gene Networks. *IEEE Signal Processing Magazine* 24, 27-36.
- McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* 94, 814-819.
- Kim, J.K., Josic, K., and Bennett, M.R. (2014). The validity of quasi-steady-state approximations in discrete stochastic simulations. *Biophys J* 107, 783-793.
- Komorowski, M., Miekisz, J., and Kierzek, A.M. (2009). Translational repression contributes greater noise to gene expression than transcriptional repression. *Biophys J* 96, 372-384.
- Sanft, K.R., Gillespie, D.T., and Petzold, L.R. (2011). Legitimacy of the stochastic Michaelis-Menten approximation. *IET Syst Biol* 5, 58.
- Tomioka, R., Kimura, H., T, J.K., and Aihara, K. (2004). Multivariate analysis of noise in genetic regulatory networks. *J Theor Biol* 229, 501-521.
- Gillespie, D.T. (1992). A rigorous derivation of the chemical master equation. *Physica A* 188, 404-425.
- Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *J Phys Chem* 81, 2340-2361.
- van Kampen, N.G. (2007). *Stochastic Processes in Physics and Chemistry, Third Edition* (Amsterdam: North Holland).
- Wallace, E.W., Gillespie, D.T., Sanft, K.R., and Petzold, L.R. (2012). Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET Syst Biol* 6, 102-115.
- Elf, J., and Ehrenberg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res* 13, 2475-2484.
- Rao, C.V., and Arkin, A.P. (2003). Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.* 118.
- Gonze, D., and Goldbeter, A. (2006). Circadian rhythms and molecular noise. *Chaos* 16, 026110.
- Gonze, D., Halloy, J., Leloup, J.C., and Goldbeter, A. (2003). Stochastic models for circadian rhythms: effect of molecular noise on periodic and chaotic behaviour. *C R Biol* 326, 189-203.
- Smadbeck, P., and Kaznessis, Y. (2012). Stochastic model reduction using a modified Hill-type kinetic rate law. *J Chem Phys* 137, 234109.
- Thomas, P., Straube, A.V., and Grima, R. (2012). The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Syst Biol* 6, 39.
- Wu, J., Vidakovic, B., and Voit, E.O. (2011). Constructing stochastic models from deterministic process equations by propensity adjustment. *BMC Syst Biol* 5, 187.
- Ooi, H.K., and Ma, L. (2012). Modeling heterogeneous responsiveness of intrinsic apoptosis pathway. *BMC Syst Biol* 7, 65.
- Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459, 428-432.
- Ooi, H.K., and Ma, L. (2012). Modeling cell-to-cell stochastic variability in intrinsic apoptosis pathway. *Conf Proc IEEE Eng Med Biol Soc* 2012, 5498-5501.
- Munsky, B., Trinh, B., and Khammash, M. (2009). Listening to the noise: random fluctuations reveal gene network parameters. *Mol Syst Biol* 5, 318.
- Spencer, S.L., and Sorger, P.K. (2011). Measuring and modeling apoptosis in single cells. *Cell* 144, 926-939.
- Toni, T., and Tidor, B. (2013). Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology. *PLoS Comput Biol* 9, e1002960.
- Kim, H., and Gelenbe, E. (2012). Stochastic gene expression modeling with Hill function for switch-like gene responses. *IEEE/ACM Trans Comput Biol Bioinform* 9, 973-979.

Submit your next manuscript to Annex Publishers and benefit from:

- ▶ Easy online submission process
- ▶ Rapid peer review process
- ▶ Online article availability soon after acceptance for Publication
- ▶ Open access: articles available free online
- ▶ More accessibility of the articles to the readers/researchers within the field
- ▶ Better discount on subsequent article submission

Submit your manuscript at
<http://www.annepublishers.com/paper-submission.php>