

Classical versus Stochastic Kinetics Modeling of Biochemical Reaction Systems

John Goutsias

Whitaker Biomedical Engineering Institute, The Johns Hopkins University, Baltimore, Maryland

ABSTRACT We study fundamental relationships between classical and stochastic chemical kinetics for general biochemical systems with elementary reactions. Analytical and numerical investigations show that intrinsic fluctuations may qualitatively and quantitatively affect both transient and stationary system behavior. Thus, we provide a theoretical understanding of the role that intrinsic fluctuations may play in inducing biochemical function. The mean concentration dynamics are governed by differential equations that are similar to the ones of classical chemical kinetics, expressed in terms of the stoichiometry matrix and time-dependent fluxes. However, each flux is decomposed into a macroscopic term, which accounts for the effect of mean reactant concentrations on the rate of product synthesis, and a mesoscopic term, which accounts for the effect of statistical correlations among interacting reactions. We demonstrate that the ability of a model to account for phenomena induced by intrinsic fluctuations may be seriously compromised if we do not include the mesoscopic fluxes. Unfortunately, computation of fluxes and mean concentration dynamics requires intensive Monte Carlo simulation. To circumvent the computational expense, we employ a moment closure scheme, which leads to differential equations that can be solved by standard numerical techniques to obtain more accurate approximations of fluxes and mean concentration dynamics than the ones obtained with the classical approach.

INTRODUCTION

The design of predictive models of cellular regulation is an important problem in computational systems biology. The majority of models published in the literature assume that cells are well-stirred, homogeneous biochemical reaction systems at thermal equilibrium, an assumption that we also follow in this article. A widely used approach to modeling cellular regulation characterizes the dynamic evolutions of molecular concentrations by deterministic first-order ordinary differential equations, known as chemical kinetics equations (CKEs) (1). However, to take into account that reactions in cells occur by random collisions of reactant molecules, we must employ a stochastic approach to modeling cellular regulation. A popular approach characterizes the dynamic evolution of the joint probability mass function of the state of cellular regulation by a first-order partial differential equation known as the chemical master equation (CME) (2–4). This leads to a modeling methodology that has been employed in several biological settings with remarkable success (5–9).

It has been increasingly recognized that cellular regulation should be studied at the level of single cells. Despite a growing effort to develop experimental methods for observing biochemical activities in single cells (10–12), these methods can only be used to simultaneously observe a limited number of molecular dynamics. Most experimental techniques used today estimate molecular concentrations in tissues containing a large number of cells (13,14). As a consequence, appreciable research activity is focused on studying the aggregate behavior of cellular regulation in a large population of cells.

For the purpose of this work, we may assume that a tissue is composed of K genetically identical cells that express the same set of genes independently from each other. This is a convenient albeit reasonable approximation, since it frees us from modeling tissue inhomogeneities and biological effects due to complex interactions among cells. We may model cellular activities in each cell by a stochastic biochemical reaction system that consists of N molecular species and use the random variable $X_{nk}(t)$ to denote the number of molecules of the n^{th} species present in the k^{th} cell at time t . Since cellular regulation is observed by pooling together molecules extracted from all cells in the tissue, we may characterize its state at time t by the molecular concentrations $Y_n(t) = (X_{n1}(t) + X_{n2}(t) + \dots + X_{nK}(t))/KAV$, where V is the cellular volume and $A = 6.0221415 \times 10^{23} \text{ mol}^{-1}$ is the Avogadro constant. The mean value and variance of $Y_n(t)$ are given by $E(Y_n(t)) = \mu_{X,n}(t)/AV$ and $\text{Var}(Y_n(t)) = v_{X,nn}(t)/KA^2V^2$, where $\mu_{X,n}(t)$ and $v_{X,nn}(t)$ are the mean and variance of $X_{nk}(t)$, respectively. This implies that the mean value of the molecular concentration $Y_n(t)$ of the n^{th} species is independent of the number of cells in the tissue, whereas its variance tends to zero as the number of cells grows to infinity (provided that the variance $v_{X,nn}(t)$ is finite). As a consequence, we may approximately characterize cellular regulation in a large population of cells by the mean concentration vector

$$\mathbf{u}(t) = \frac{\boldsymbol{\mu}_X(t)}{AV}, \quad (1)$$

where $\boldsymbol{\mu}_X(t)$ denotes the $N \times 1$ mean vector with elements $\mu_{X,n}(t)$, $n = 1, 2, \dots, N$.

An important question that arises here is whether the molecular concentration dynamics predicted by the CKEs coincide with the mean concentration dynamics predicted by

Submitted July 21, 2006, and accepted for publication December 19, 2006.

Address reprint requests to J. Goutsias, Tel.: 410-516-7871; E-mail: goutsias@jhu.edu.

© 2007 by the Biophysical Society

0006-3495/07/04/2350/16 \$2.00

doi: 10.1529/biophysj.106.093781

the underlying CME and Eq. 1. It turns out that, given the CME, we can uniquely construct the corresponding CKEs and vice versa. Therefore, we may expect that the two approaches lead to the same dynamics. However, by analyzing a number of simple chemical reactions such as $2A \rightarrow B$, $A + B \rightarrow C$, $A \rightarrow B$, $A \rightarrow B \rightarrow C$, $2A \rightarrow B \rightarrow C$, and $2A \rightleftharpoons B$, it was previously shown in the literature (15–17) that this may not be true in general. A notable exception occurs at the thermodynamic limit, in which the number of molecules and cellular volume tend to infinity while the molecular concentrations remain finite, or when all reaction mechanisms are linear. However, both of these cases are clearly not realistic.

In Zheng and Ross (18), they extended the previous work by focusing on the autocatalytic cubic Schlögl model $A + 2B \rightleftharpoons 3B$, $B \rightleftharpoons C$. These investigators noted that differences between classical and stochastic chemical kinetics are due to a coupling of correlation effects with system nonlinearities. By focusing on parameter values that lead to the same stationary states (concentrations of the molecular intermediate B) for both models, they showed that the deterministic model may result in quantitatively different transient behavior for the mean concentration of B than the corresponding stochastic model, with the maximum deviation between the concentration trajectories decreasing as the model parameters are modified toward a linear kinetic mechanism.

In view of the fact that cellular regulation is controlled by a complex network of biochemical reactions, it is necessary to investigate the relationship between classical and stochastic chemical kinetics in a more general setting than the one considered in the literature (15–18). In this article, we derive fundamental relationships between the two approaches for a biochemical reaction system that consists of elementary (monomolecular or bimolecular) irreversible reactions. We can use this system to model any set of biochemical reactions, since we can decompose any reaction that involves more than two molecules (a rare possibility in practice) into a cascade of bimolecular reactions and split a reversible reaction into two separate irreversible reactions (19). We have chosen to illustrate our results by employing two reaction mechanisms: a unidirectional dimerization and a quadratic autocatalator with positive feedback. These mechanisms allow us to clearly demonstrate that intrinsic stochastic fluctuations may appreciably influence the qualitative and quantitative behavior of cellular regulation and to analytically investigate the origins of such influence. Note, however, that our approach is very general and can be applied to more complex regulatory mechanisms as well.

In this article, we show that the mean concentration dynamics predicted by the CME are governed by first-order ordinary differential equations similar to the ones obtained by classical chemical kinetics, expressed in terms of the stoichiometry coefficients and the time-dependent fluxes of the underlying reactions. However, the flux is now decomposed into a macroscopic and a mesoscopic term. The macroscopic

term is analytically identical to the classical flux and accounts for the effect of mean reactant concentrations on the rate of product synthesis, whereas the mesoscopic term accounts for the effect of statistical correlations among interacting reactions. When all mesoscopic fluxes are zero, a situation that occurs when the biochemical reaction system consists of only monomolecular reactions (which leads to linear reaction mechanisms), the concentration dynamics predicted by the CKEs will be identical to the mean concentration dynamics predicted by the CME. However, and by using the two aforementioned examples, our analytical and numerical investigations show that nonzero mesoscopic fluxes may induce appreciable qualitative and quantitative differences in transient and stationary system behavior from that predicted by classical chemical kinetics. In addition to the conclusions reached by Zheng and Ross (18), we show that the mean concentration dynamics predicted by the CME may converge to different stationary values than those predicted by the CKEs, thus supporting the fact that intrinsic stochastic fluctuations may play an important role in determining a cell's phenotype by quantitatively influencing cell regulation at steady state. Moreover, we analytically and numerically demonstrate that intrinsic stochastic fluctuations may also affect the epigenetic properties of cell regulation in a qualitative manner by introducing novel modes of stationary behavior not accounted for by the CKEs. Hence, a sufficiently accurate model of mean concentration dynamics must necessarily include all mesoscopic fluxes in its formulation. These developments provide a theoretical understanding of the role that intrinsic stochastic fluctuations may play in inducing biochemical function.

The mesoscopic fluxes cannot be evaluated analytically. We can estimate them by Monte Carlo simulation, but the resulting method is computationally intensive in most cases of interest. To circumvent the computational expense, we employ a moment closure scheme that allows us to approximate the underlying covariances (and thus the mesoscopic fluxes) by first-order ordinary differential equations that are similar to the CKEs and can be solved by standard numerical techniques. We show that, at least for the quadratic autocatalator, this approximation leads to more accurate predictions of fluxes and mean concentration dynamics than the CKEs.

We should mention that, in a recent work (20), Samoilov et al. have used a simple example (an enzymatic futile cycle) to analytically and numerically demonstrate that extrinsic stochastic fluctuations in biochemical reaction systems may also produce dynamic behavior not accounted for by classical chemical kinetics. Our work is complementary to theirs, since it focuses on the effects of intrinsic stochastic fluctuations on system behavior. Moreover, it supports, both analytically and computationally, the general belief that stochastic fluctuations may play an important role in determining biological function in cells and, therefore, must be accounted for by computational models of cellular regulation.

BIOCHEMICAL REACTION SYSTEMS

Deterministic description

In this article, we consider a well-stirred biochemical reaction system at thermal equilibrium that consists of M elementary (monomolecular or bimolecular) irreversible reaction channels. By assuming that the system contains N molecular species, we may characterize its state at time $t \geq 0$ by an $N \times 1$ deterministic vector $\mathbf{q}(t)$ whose dynamic evolution is governed by the following CKEs (1):

$$\frac{d\mathbf{q}(t)}{dt} = \mathbb{S}\boldsymbol{\rho}(t), \quad t \geq 0. \quad (2)$$

In Eq. 2, $\mathbb{S} = (s_{nm})$ is the $N \times M$ stoichiometry matrix of the underlying biochemical reactions. Moreover, $\boldsymbol{\rho}(t)$ is an $M \times 1$ (time-dependent) vector with elements $\rho_m(t) = (1/V) d\xi_m(t)/dt$, where $\xi_m(t)$ is the extent of the m^{th} reaction, defined as the amount (in moles) of a species produced or consumed by the m^{th} reaction during the time interval $[0, t)$, divided by the corresponding stoichiometric coefficient. Note that $\rho_m(t)$ is the rate of change in the extent of reaction per unit volume at time t and, hence, it quantifies the reaction rate of the m^{th} reaction. These parameters are frequently referred to as time-dependent fluxes (velocities of molecular flow) and play a fundamental role in the analysis of biochemical reaction systems (1). The mass action rate law implies that the m^{th} element of $\boldsymbol{\rho}(t)$ is given by

$$\rho_m(t) = \kappa_m \psi_m(\mathbf{q}(t)), \quad (3)$$

where κ_m is the reaction rate constant of the m^{th} reaction and $\psi_m(\mathbf{q})$ is a product of the reactant concentrations of the m^{th} reaction, given by

$$\psi_m(\mathbf{q}) = \begin{cases} q_n, & \text{for monomolecular reactions} \\ q_n(q_n - 1/AV), & \text{for bimolecular reactions} \\ & \text{with identical reactants} \\ q_n q_{n'}, & \text{for bimolecular reactions} \\ & \text{with different reactants} \end{cases}. \quad (4)$$

Note that in classical chemical kinetics it is assumed that molecular concentrations are appreciably larger than $1/AV$, in which case $\psi_m(\mathbf{q}) = q_n^2$ for bimolecular reactions with identical reactants. To account for the possibility that some molecular concentrations may be comparable to $1/AV$, we set in this article $\psi_m(\mathbf{q}) = q_n(q_n - 1/AV)$.

Although it is commonly believed that $\mathbf{q}(t)$ provides a sufficient approximation to the mean concentration vector $\mathbf{u}(t)$, given by Eq. 1, because of stochastic fluctuations in biochemical activity, this may not be true. Moreover, the derivation of Eq. 2 requires that the numbers of molecules in the system are very large compared to 1. Otherwise, $\mathbf{q}(t)$ will not be a continuous function of t and differentiation of $\mathbf{q}(t)$ will not be possible. Therefore, we may not be able to justify the CKEs when modeling biochemical reaction systems with

appreciable stochastic fluctuations and small numbers of reactant molecules. In view of the fact that reactions occur by random collisions of reactant molecules, it is intuitive to believe that it will be more appropriate if we employ a stochastic approach.

Stochastic description

If we use a stochastic biochemical reaction system to model cellular regulation in single cells, then the mean molecular concentrations $\mathbf{u}(t)$, given by Eq. 1, will satisfy the system of first-order differential equations

$$\frac{d\mathbf{u}(t)}{dt} = \mathbb{S}\boldsymbol{\nu}(t), \quad t \geq 0, \quad (5)$$

where $\boldsymbol{\nu}(t)$ is an $M \times 1$ (time-dependent) flux vector with elements $\nu_m(t) = (1/V) d\chi_m(t)/dt$, and $\chi_m(t)$ is the extent of the m^{th} reaction at time t , which is now defined as the mean degree of advancement (DA) of the m^{th} reaction at time t divided by the Avogadro number (see Appendix A for a brief review of stochastic chemical kinetics). Notably, the flux $\nu_m(t)$ of the m^{th} reaction is the rate of change in its mean DA per unit volume divided by the Avogadro number.

Equation 5 is similar to the CKEs with one important difference. The m^{th} element $\nu_m(t)$ of the flux vector $\boldsymbol{\nu}(t)$ is now given by

$$\nu_m(t) = \rho_m(t) + \theta_m(t), \quad (6)$$

where

$$\rho_m(t) = \kappa_m \psi_m(\mathbf{u}(t)), \quad (7)$$

as in the deterministic case, and (see Appendix A)

$$\theta_m(t) = \frac{1}{2AV} \sum_{k=1}^M \sum_{l=1}^M \eta_{m,kl} \nu_{z,kl}(t). \quad (8)$$

In Eq. 8, $\eta_{m,kl}$ is the second-order partial derivative of the propensity function of the m^{th} reaction with respect to the DAs of the k^{th} and l^{th} reactions, whereas, $\nu_{z,kl}(t)$ is the covariance between the DAs of those reactions. We refer to $\rho_m(t)$ as the macroscopic flux and to $\theta_m(t)$ as the mesoscopic flux of the m^{th} reaction. Clearly, the macroscopic flux accounts for the effect of mean reactant concentrations on the rate of product synthesis, whereas the mesoscopic flux accounts for the effect of statistical correlations among interacting reactions on that rate. Because the propensity functions are at most quadratic functions of the DAs (see Appendix A), the mesoscopic flux of a given reaction depends only on the (nonzero) Hessian elements of the propensity function of that reaction and on the corresponding DA covariances. As a consequence of Eq. 8, the mesoscopic flux of a monomolecular reaction will be zero, since the propensity function of such reaction will be linear and the corresponding Hessian matrix will be zero (i.e., $\eta_{m,kl} = 0$, for every k, l). However, this may not be true for the mesoscopic flux of a bimolecular

reaction, whose value will depend on the DA covariances between reactions that affect the molecular population of one reactant species and reactions that affect the population of the other reactant species.

If all reactions are monomolecular, the biochemical reaction system will be linear (i.e., all propensity functions will be linear). In this case, the second-order partial derivatives of the propensity functions with respect to the DAs will be zero, which, together with Eq. 8, implies that $\theta_m(t) = 0$, for every m (since $\eta_{m,kl} = 0$, for every m, k, l). Hence, when all reactions are monomolecular, Eq. 5 is identical to Eq. 2 of classical chemical kinetics. This result was derived in Gillespie (21).

Equations 5–8 extend the classical CKEs 2 and 3 to account for intrinsic stochastic fluctuations in biochemical activity and are an exact consequence of the conservation of mass and the CME underlying the biochemical reaction system (Eqs. 21 and 27). Note that

$$\frac{d\mathbf{u}(t)}{dt} = \mathbb{S}\boldsymbol{\rho}(t) + \boldsymbol{\varepsilon}(t), \quad t \geq 0, \quad (9)$$

where

$$\boldsymbol{\varepsilon}(t) = \mathbb{S}\boldsymbol{\theta}(t) = \sum_{m=1}^E s_{nm}\theta_m(t). \quad (10)$$

We refer to these equations (and Eqs. 5–8) as statistical chemical kinetics equations (SCKEs), where we use the term “statistical” to emphasize that the equations account for correlations among reactions. Like the CKEs, the SCKEs provide a macroscopic description of a biochemical reaction system. However, this description is now controlled by the mesoscopic behavior of the system through the forcing term $\boldsymbol{\varepsilon}(t)$. If $\boldsymbol{\varepsilon}(t)$ were known for every $t \geq 0$, then we could evaluate $\mathbf{u}(t)$ by integrating Eq. 9 using standard numerical techniques. However, this is not true and $\mathbf{u}(t)$ must be estimated by computationally intensive Monte Carlo simulation using the Gillespie algorithm (22). In an effort to circumvent this computational expense, we later discuss a method that allows us to approximately evaluate $\mathbf{u}(t)$ by numerically integrating an appropriately derived system of first-order ordinary differential equations.

Deterministic versus stochastic description

Equation 9 reveals that intrinsic stochastic fluctuations may influence the mean concentration dynamics $\mathbf{u}(t)$ through the mesoscopic forcing term $\boldsymbol{\varepsilon}(t)$. This fact is not considered by the CKEs and its importance should not be underestimated. If, for some species n , $\varepsilon_n(t) \neq 0$, for every $t \geq 0$, then at least one function ψ_m will be quadratic, where m is a reaction that consumes or produces the n^{th} molecular species, and the n^{th} SCKE will therefore be nonlinear. Indeed, if ψ_m is linear, for any reaction m that consumes or produces the n^{th} molecular species, then its second-order partial derivative $\eta_{m,kl}$ will be

zero, for every k, l , which implies that $\theta_m(t) = 0$, by virtue of Eq. 8. Since the mesoscopic forcing term $\varepsilon_n(t)$ is given by Eq. 10, this implies that $\varepsilon_n(t) = 0$ for every $t \geq 0$, which contradicts our assumption that $\varepsilon_n(t) \neq 0$ for every $t \geq 0$. By combining this observation with the fact that a nonzero forcing term may substantially affect the solution of a nonlinear differential equation, we may expect that nonzero mesoscopic forcing terms could appreciably affect the mean behavior of a nonlinear biochemical reaction system.

It is clear from our previous discussion that we may use the CKEs to characterize the mean concentration dynamics if and only if, at any time $t \geq 0$, the mesoscopic flux vector $\boldsymbol{\theta}(t)$ is in the null space of the stoichiometry matrix \mathbb{S} . In this case, $\boldsymbol{\varepsilon}(t) = \mathbb{S}\boldsymbol{\theta}(t) = 0$, for every $t \geq 0$, and the SCKEs will be reduced to the CKEs. Eq. 8 suggests that this will happen if all underlying reactions are monomolecular (see also (18,21)). We may also use the CKEs when all covariances are zero, a condition that will be satisfied in the thermodynamic limit. However, most biochemical reaction systems of interest are nonlinear and there is no way to know a priori whether the covariances are zero or, more generally, whether $\boldsymbol{\theta}(t)$ is in the null space of \mathbb{S} . Therefore, to account for the influence of nonzero mesoscopic fluxes on system dynamics, we must include them in the formulation.

Numerical example

To provide a simple illustration of our discussion so far, we consider the following unidirectional dimerization reaction,



with specific probability rate constant c_1 , initialized with S molecules P and S molecules Q (see Appendix B for details). In Fig. 1, we depict the normalized (with respect to the steady-state dimer concentration $s = S/AV$) dimer concentration and flux dynamics, predicted by the underlying SCKE (solid lines) versus the ones predicted by the corresponding CKE (dotted lines), for $S = 1$, in Fig. 1 A, and $S = 10$, in Fig. 1 B. We also depict the dynamics of the normalized mesoscopic forcing term. We have estimated the concentrations, fluxes, and forcing terms by Monte Carlo simulation using the Gillespie algorithm (22,23), and calculated the CKE concentrations and fluxes analytically (see Eq. 38). It turns out that, as $t \rightarrow \infty$, both models converge to the same steady-state concentration s .

When the initial number of reactant molecules is very small (e.g., when $S = 1$ in Fig. 1 A), the CKE concentration dynamics do not match the SCKE dynamics obtained by Monte Carlo simulation. According to the results depicted in Fig. 1 A, small differences in flux dynamics may lead to substantial differences in concentration dynamics. However, a sufficient increase in the initial number of reactant molecules (e.g., by tenfold in Fig. 1 B) may drastically alleviate this difference. For sufficiently large S ($S \geq 100$) the flux

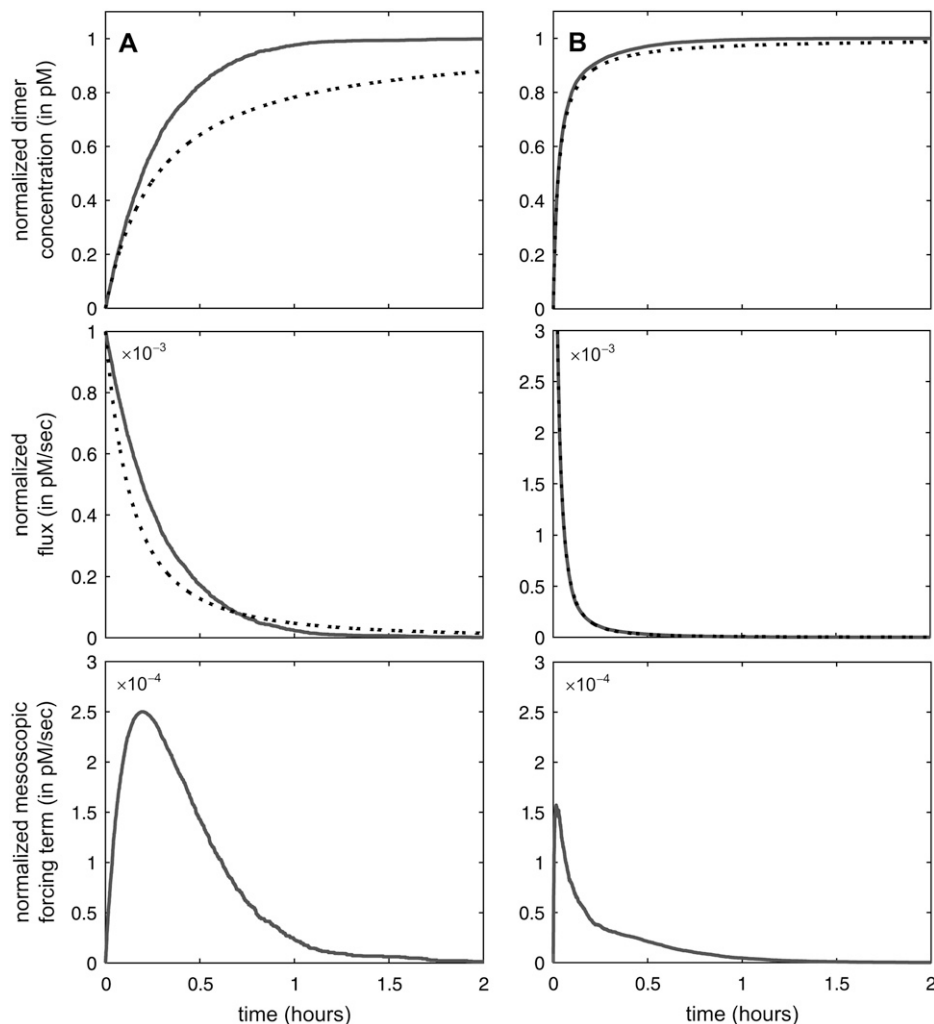


FIGURE 1 Normalized dimer accumulation in the unidirectional dimerization reaction, given by Eq. 11, predicted by the SCKEs (solid lines) and CKEs (dotted lines). The dynamics obtained by the SCKEs have been computed by Monte Carlo simulation using the Gillespie algorithm, whereas the dynamics obtained by the CKEs have been computed analytically from Eq. 38. The system is initialized with (A) one molecule P and one molecule Q, (B) 10 molecules P, and 10 molecules Q. The associated normalized flux and mesoscopic forcing term dynamics are depicted as well. Parameters used are $c_1 = 10^{-3} \text{ s}^{-1}$, $V = 2 \text{ pL}$, and $K = 6000 \text{ cells}$.

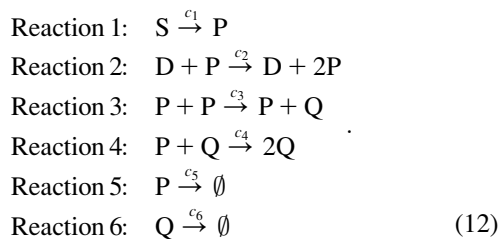
and concentration dynamics are virtually identical (data not shown).

The observed differences are due to the mesoscopic forcing term, which coincides in this case with the mesoscopic flux of the reaction. Since all reactants are eventually transformed into dimers, the mesoscopic forcing term tends to zero as $t \rightarrow \infty$. Its magnitude and rate of convergence to zero affect the SCKE concentration dynamics and the time it takes for the system to reach steady state. Because the flux is given by Eq. 37, larger values of the mesoscopic forcing term will promote faster reaction rates and thus faster relaxation to steady state (for this example, the mesoscopic forcing term is nonnegative). Fig. 1 shows that, when $S = 1$, the mesoscopic forcing term converges to zero slower than when $S = 10$. Our simulations show that the response predicted by the SCKE reaches steady state at $\sim 2 \text{ h}$, whereas the response predicted by the CKE requires substantially more time ($\sim 24 \text{ h}$) to reach steady state. This example provides an analytical justification, by means of the mesoscopic forcing term, of a previously recognized fact that intrinsic stochastic fluctuations in biochemical activity may produce

quantitative differences between the transient concentration dynamics predicted by classical and stochastic chemical kinetics (15,18).

A QUADRATIC AUTOCATALATOR

Although the dimer concentration dynamics predicted by the CKE of the reaction equation of the previous example (Eq. 11) may follow a different trajectory than the dynamics predicted by the corresponding SCKE, eventually the two trajectories reach the same steady state. We will now show that this may not be necessarily true. To do so, we turn to a more complex example, governed by the following six reactions:



These reactions convert substrate molecules S into proteins Q . An intermediate protein P is first produced by Reaction 1 and, subsequently, by Reaction 2 via transcription and translation in which P acts as a transcription factor to promote its own synthesis from a DNA template D . P is then transformed into Q via the intermolecular reactions 3 and 4, with P and Q , respectively. Finally, Reactions 5 and 6 model degradation of P and Q . Due to Reactions 2 and 3, we refer to this system as quadratic autocatalator with positive feedback, since Reaction 3 is autocatalytic with quadratic concentration dependence (see also (24)) and Reaction 2 applies (positive) feedback on the synthesis of P . The resulting system is similar to a reaction cascade considered in Kaufman et al. (25), which involves the autophosphorylation of protein tyrosine kinase activity in T cell stimulation, obtained by ignoring all dephosphorylation reactions. This simplification leads to a biologically relevant example, which allows us to analytically demonstrate that intrinsic stochastic fluctuations may appreciably affect, both qualitatively and quantitatively, the stationary behavior of a biochemical reaction system. We could use more complicated reaction schemes (e.g., schemes that involve phosphorylation/dephosphorylation, transcription, translation, etc.), but it would not be possible to proceed analytically.

Quantitative stationary behavior

The steady-state concentrations \bar{u}_1 of P and \bar{u}_2 of Q , predicted by the SCKEs associated with the quadratic autocatalator with feedback are given by (see Appendix C for details)

$$\bar{u}_1(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s - \bar{\epsilon}(s)}{\alpha}} \quad (13)$$

$$\bar{u}_2(s) = \frac{\kappa_1 s + (\kappa_2 d - \kappa_5)\bar{u}_1(s)}{\kappa_6}, \quad (14)$$

where

$$\alpha = \kappa_3 + \frac{\kappa_4}{\kappa_6}(\kappa_2 d - \kappa_5) \quad \text{and} \quad \beta(s) = \frac{\kappa_1 \kappa_4}{\kappa_6} s - \frac{\kappa_3}{AV} + \kappa_5 - \kappa_2 d, \quad (15)$$

provided that $\alpha > 0$ and $\kappa_1 s > \bar{\epsilon}(s)$ (see also the first row of Table 1). In these equations, $s = S/AV$ and $d = D/AV$, where

S, D are the numbers of S and D molecules, respectively. Moreover, $\bar{\epsilon}(s)$ is the sum of the mesoscopic fluxes of the third and fourth reactions at steady state, which depends on s , and the κ -values are the reaction rate constants, given by Eq. 42. We use the notation $\bar{u}_1(s), \bar{u}_2(s)$, and $\bar{\epsilon}(s)$ to explicitly denote that the stationary quantities \bar{u}_1, \bar{u}_2 , and $\bar{\epsilon}$ depend on the input substrate concentration s . By setting $\bar{\epsilon}(s) = 0$ in Eq. 13 (i.e., by ignoring the mesoscopic fluxes), we obtain the steady-state concentrations predicted by the CKEs (see Table 2).

According to Eq. 14, when $\kappa_5 = \kappa_2 d$, the steady-state concentration \bar{u}_2 of Q predicted by the SCKEs will be identical to the one predicted by the CKEs; this concentration is given by $\kappa_1 s / \kappa_6$. However, this may not be true for the steady-state concentration \bar{u}_1 of P , since this concentration depends on $\bar{\epsilon}(s)$, according to Eq. 13. If $\lim_{s \rightarrow \infty} \bar{\epsilon}(s)/s = 0$, then Eqs. 13 and 14 imply that $\lim_{s \rightarrow \infty} \bar{u}_1(s) = \kappa_6 / \kappa_4$ (this is also true when $\kappa_5 \neq \kappa_2 d$; see Appendix C), in which case, both models will asymptotically (as $s \rightarrow \infty$) reach the same steady-state concentration for P as well. However, for finite input substrate concentrations, we may not be able to ignore the steady-state mesoscopic forcing term $\bar{\epsilon}(s)$, in which case the steady-state concentration of P predicted by the SCKEs will be different than the one predicted by the CKEs, with the difference being controlled by the sign and magnitude of $\bar{\epsilon}(s)$. This is illustrated in Fig. 2, which depicts the concentration dynamics of P and Q and the flux dynamics of the third and fourth reactions predicted by the SCKEs (solid lines), estimated by Monte Carlo simulation using the Gillespie algorithm, and the CKEs (dotted lines), obtained numerically. In this case, the steady-state P concentration predicted by the CKEs is larger than the one predicted by the SCKEs, since $\bar{\epsilon} = 3.65 \times 10^{-3} \text{ pM/s} > 0$. Note the quantitative differences between the transient mean concentration dynamics and fluxes. As a matter of fact, the CKEs wrongly predict that the mean concentration of Q will be zero during the first minute, whereas the SCKEs predict a gradual increase in the mean concentration of Q from 0 pM to ~ 0.033 pM.

Similar remarks apply when $\kappa_5 \neq \kappa_2 d$. However, the steady-state concentrations of both P and Q depend now on $\bar{\epsilon}$ (the concentration of Q depends on $\bar{\epsilon}$ through the concentration of P ; recall Eq. 14) and the CKEs may not provide good approximations at finite input substrate concentrations. We illustrate this case in Fig. 3.

TABLE 1 Stationary concentrations of P, predicted by the SCKEs, in the quadratic autocatalator with feedback

| $\alpha > 0$ | $\beta(s) \geq 0$ | $\beta(s) < 0$ |
|----------------------------------|---|---|
| $\kappa_1 s > \bar{\epsilon}(s)$ | $\bar{u}_1(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s - \bar{\epsilon}(s)}{\alpha}}$ | $\bar{u}_1(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s - \bar{\epsilon}(s)}{\alpha}}$ |
| $\kappa_1 s = \bar{\epsilon}(s)$ | $\bar{u}_1(s) = 0$ | $\bar{u}_1(s) = 0$ or $\bar{u}_1(s) = -\frac{\beta(s)}{\alpha}$ |
| $\kappa_1 s < \bar{\epsilon}(s)$ | No steady-state concentration | $\bar{u}_1(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s - \bar{\epsilon}(s)}{\alpha}}$ |

TABLE 2 Stationary concentrations of P, predicted by the CKEs, in the quadratic autocatalator with feedback

| $\alpha > 0$ | $\beta(s) \geq 0$ | $\beta(s) < 0$ |
|--------------|---|---|
| $s = 0$ | $\bar{q}_1(0) = 0$ | $\bar{q}_1(0) = 0$ or $\bar{q}_1(0) = -\frac{\beta(0)}{\alpha}$ |
| $s > 0$ | $\bar{q}_1(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s}{\alpha}}$ | $\bar{q}_1(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s}{\alpha}}$ |

Our previous investigation shows that intrinsic stochastic fluctuations may produce appreciable quantitative differences between the stationary behavior of a biochemical reaction system predicted by classical chemical kinetics and the stationary behavior predicted by stochastic chemical kinetics. These differences are caused by nonzero mesoscopic forcing terms at steady state, which may influence stationary molecular concentrations and appreciably affect their values.

The stationary behavior of biochemical activity may affect cells in a biologically significant way. For example, it has been suggested that concentrations of regulatory proteins synthesized at steady state may be responsible for a cell's unique characteristics (phenotype) (26). As a consequence, the previous analytical and numerical investigations show that intrinsic stochastic fluctuations may quantitatively affect the epigenetic properties of cell regulation in a manner not accounted for by classical chemical kinetics. In addition, we show in the following that nonzero stationary mesoscopic forcing terms may influence the steady-state properties of a biochemical reaction system in a qualitative way, thus demonstrating the fact that intrinsic stochastic fluctuations may play a significant role in influencing cellular function.

Qualitative stationary behavior

In the quadratic autocatalator with positive feedback, the stationary concentration of P predicted by the SCKEs depends on the signs of parameters α and $\beta(s)$, given by Eq. 15, and the value of the input flux $\kappa_1 s$ as compared to the value of the steady-state mesoscopic forcing term $\bar{\epsilon}(s)$. The resulting concentrations are summarized in Table 1 for $\alpha > 0$ (similar results hold for $\alpha < 0$). In particular, if $\kappa_1 s > \bar{\epsilon}(s)$, the system has a unique stable stationary P concentration $\bar{u}_1(s)$, given by Eq. 13, regardless of the value of $\beta(s)$. The situation, however, is different when $\kappa_1 s \leq \bar{\epsilon}(s)$. If $\kappa_1 s = \bar{\epsilon}(s)$ and $\beta(s) \geq 0$, the system relaxes to a zero P concentration at steady state, whereas, if $\kappa_1 s < \bar{\epsilon}(s)$ and $\beta(s) \geq 0$, the system has no stationary P concentration. However, if $\kappa_1 s \leq \bar{\epsilon}(s)$ and $\beta(s) < 0$, the system has two stationary P concentrations $\bar{u}_1^*(s)$ and $\bar{u}_1^{**}(s)$, given by

$$\bar{u}_1^*(s) = -\frac{\beta(s)}{2\alpha} - \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s - \bar{\epsilon}(s)}{\alpha}},$$

and

$$\bar{u}_1^{**}(s) = -\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha}\right]^2 + \frac{\kappa_1 s - \bar{\epsilon}(s)}{\alpha}}.$$

Note that $\bar{u}_1^*(s) \leq \bar{u}_1^{**}(s)$, with the two concentrations being equal when the input substrate concentration is set to s^* , where s^* satisfies $\kappa_1 s^* = \bar{\epsilon}(s^*) - \beta^2(s^*)/4\alpha$.

On the other hand, although the stationary concentration of P predicted by the CKEs still depends on the signs of

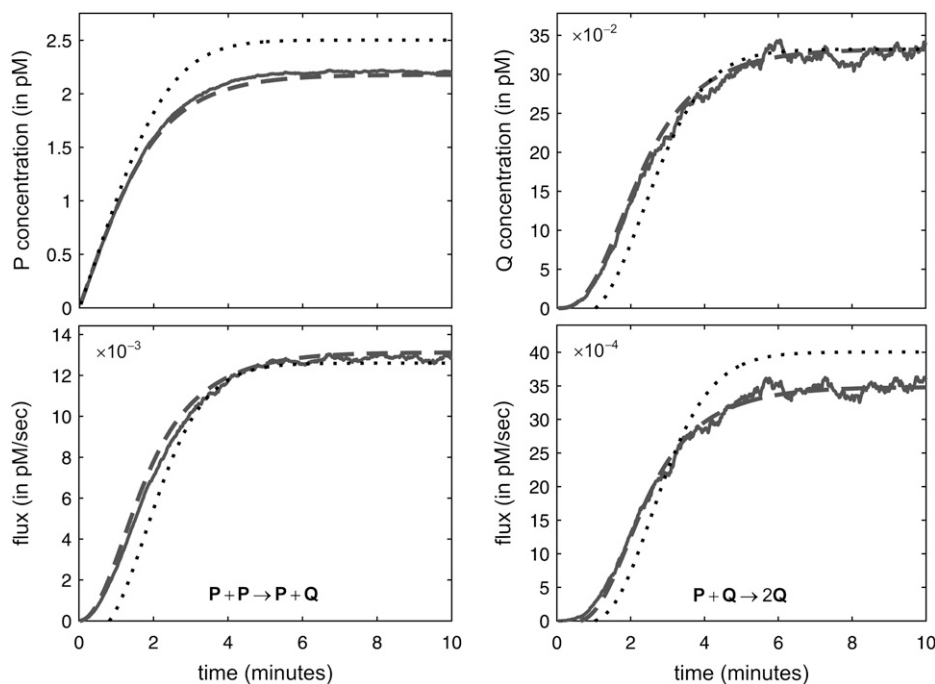


FIGURE 2 Protein accumulation in the quadratic autocatalator, given by Eq. 12, for the case when $\kappa_5 = \kappa_2 d$, initialized with 10 molecules S (concentration of 8.30 pM), two molecules D (the number of DNA copies of a particular gene per eukaryotic cell), and zero molecules P and Q, predicted by the SCKEs (solid lines) and CKEs (dotted lines). The dynamics obtained by the SCKEs have been computed by Monte Carlo simulation using the Gillespie algorithm, whereas the dynamics obtained by the CKEs have been computed numerically. The flux dynamics of the third and fourth reactions are depicted as well. Parameters used are $c_1 = 0.002 \text{ s}^{-1}$, $c_2 = 0.001 \text{ s}^{-1}$, $c_3 = 0.005 \text{ s}^{-1}$, $c_4 = 0.004 \text{ s}^{-1}$, $c_5 = 0.002 \text{ s}^{-1}$, $c_6 = 0.05 \text{ s}^{-1}$, $V = 2 \text{ pL}$, and $K = 10,000$ cells. Although the steady-state concentration of Q predicted by the CKEs is theoretically identical to the one predicted by the SCKEs, this is not true for the concentration of P. The dashed lines indicate the mean concentration and flux dynamics predicted by the second-order SCKEs discussed in this article.

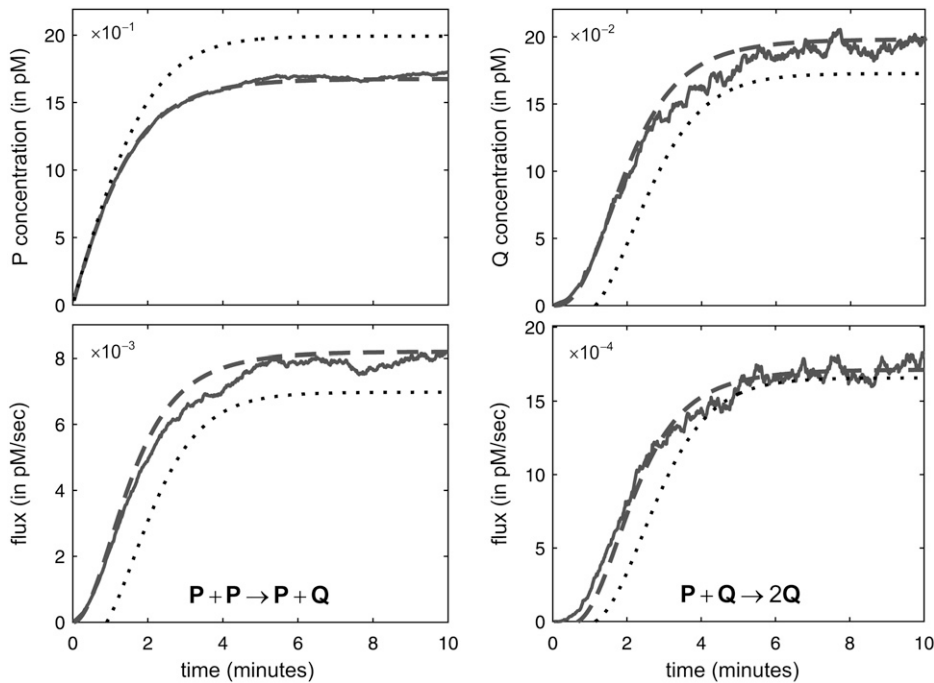


FIGURE 3 Protein accumulation and flux dynamics in the quadratic autocatalator, given by Eq. 12, for the case when $\kappa_5 > \kappa_2 d$. The parameters used are the same as in Fig. 2, but now $\kappa_5 = 0.006 \text{ s}^{-1}$. In this case, the steady-state concentrations of P and Q predicted by the CKEs (dotted lines) are both different than the actual steady-state concentrations predicted by the SCKEs (solid lines). The dashed lines indicate the mean concentration and flux dynamics predicted by the second-order SCKEs discussed in this article.

parameters α and $\beta(s)$, this concentration does not depend on the input flux values $\kappa_1 s$ but only on whether or not the input substrate concentration is zero. The resulting concentrations are summarized in Table 2, for $\alpha > 0$. When $s = 0$ and $\beta(0) \geq 0$, the CKEs predict zero stationary P concentration, whereas, when $s = 0$ and $\beta(0) < 0$, the CKEs predict two stationary P concentrations, $\bar{q}_1(0) = 0$ and $\bar{q}_1(0) = -\beta(0)/\alpha$, with the former being stable and the latter unstable. However, when $s > 0$, the CKEs predict a unique steady-state P concentration, regardless of the value of s and $\beta(s)$, which is the same as the concentration $\bar{u}_1(s)$ predicted by the SCKEs, given by Eq. 13 with $\bar{\epsilon}(s) = 0$.

To illustrate the previous analytical results and demonstrate their biological significance, we depict in Fig. 4 the stationary concentration of P as a function of the input substrate concentration s , predicted by the SCKEs (solid lines), estimated by Monte Carlo simulation using the Gillespie algorithm, and by the CKEs (dotted lines), obtained analytically. In Fig. 4 A, $\kappa_1 s > \bar{\epsilon}(s)$, for every $s > 0$, in which case, the steady-state response curve predicted by the SCKEs will be given by $\bar{u}_1(s)$. The response curve $\bar{q}_1(s)$, predicted by the CKEs, is similar to the one predicted by the SCKEs, with $\lim_{s \rightarrow \infty} \bar{q}_1(s) = \lim_{s \rightarrow \infty} \bar{u}_1(s) = \kappa_6/\kappa_4 = 10.38 \text{ pM}$, since $\lim_{s \rightarrow \infty} \bar{\epsilon}(s)/s = 0$. The situation, however, is very different in Fig. 4 B, in which $\kappa_1 s < \bar{\epsilon}(s)$, for $0 < s < s^{**} \approx 47.5 \text{ pM}$, and $\kappa_1 s > \bar{\epsilon}(s)$, for $s > s^{**}$, where s^{**} is the input substrate concentration that satisfies $\kappa_1 s^{**} = \bar{\epsilon}(s^{**})$. In this case, the steady-state response curve predicted by the SCKEs is obtained by stitching together three stable stationary P concentrations, namely, $\bar{u}_1^*(s)$, for $0 \leq s \leq s^* \approx 22.5 \text{ pM}$, $\bar{u}_1^{**}(s)$, for $s^* \leq s \leq s^{**}$, and $\bar{u}_1(s)$, for $s \geq s^{**}$. Note

that the slope of \bar{u}_1^* is larger than the slope of \bar{u}_1^{**} , whereas, the slope of \bar{u}_1 tends to zero as $s \rightarrow \infty$. As a consequence, and similarly to the behavior shown in Fig. 4 A, the system experiences appreciable protein amplification at low input substrate concentrations (i.e., for $s \leq s^*$; see the *open region* in Fig. 4 B), a moderate amplification at intermediate input concentrations (i.e., for $s^* \leq s \leq s^{**}$ —see the *light shaded region* in Fig. 4 B), and diminishing amplification at high input concentrations (i.e., for $s \geq s^{**}$; see the *dark shaded region* in Fig. 4 B). This behavior is essential to guarantee that, besides its normal operational range, the system responds quickly to low input substrate concentrations (high amplification) but very slowly to high concentrations (saturation). Note that the steady-state response curve predicted by the CKEs increases abruptly from 0 pM to $\sim 0.95 \text{ pM}$ at $s = 0$, thus failing to capture the previous “multistage” amplification property. However, since $\lim_{s \rightarrow \infty} \bar{\epsilon}(s)/s = 0$, we have that $\lim_{s \rightarrow \infty} \bar{q}_1(s) = \lim_{s \rightarrow \infty} \bar{u}_1(s) = \kappa_6/\kappa_4 = 1.04 \text{ pM}$, and the two response curves predicted by the CKEs and the SCKEs will eventually coincide for a sufficiently large input substrate concentration.

As a consequence of the previous investigations, intrinsic stochastic fluctuations may appreciably affect the qualitative properties of a biochemical reaction system at steady state. This can be analytically explained by the presence of non-zero mesoscopic forcing terms, which are responsible for introducing novel modes of behavior not accounted for by classical chemical kinetics. The significance of these modes should not be underestimated, since they may introduce behavior at low molecular concentrations that is essential for proper biological function.

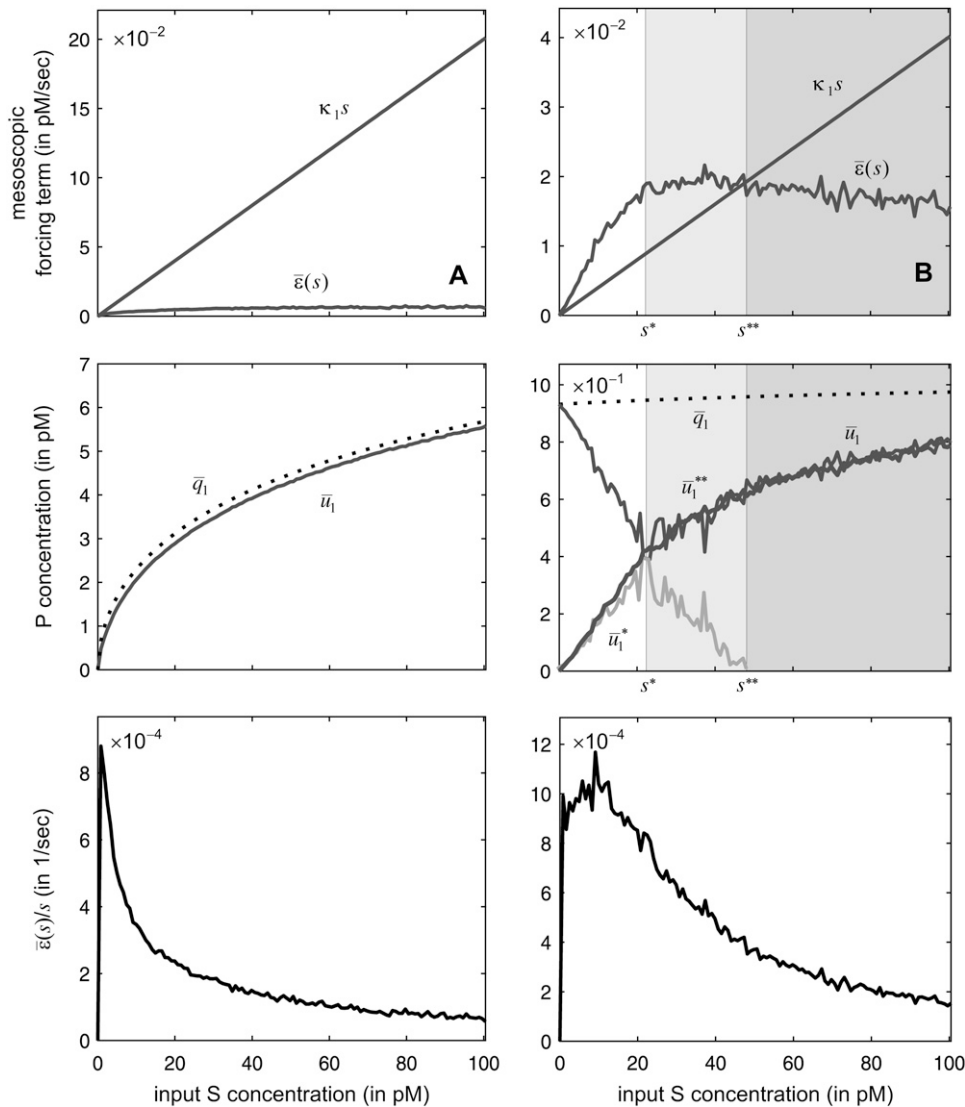


FIGURE 4 The input flux k_1s versus the stationary mesoscopic forcing term $\bar{\varepsilon}(s)$, the stationary concentration of P as a function of s , predicted by the SCKEs (solid lines) and CKEs (dotted lines), and the ratio $\bar{\varepsilon}(s)/s$ associated with the quadratic autocatalator, given by Eq. 12. The steady-state values obtained by the SCKEs have been computed by Monte Carlo simulation using the Gillespie algorithm, whereas the values obtained by the CKEs have been computed analytically. The system is initialized with two molecules D (the number of DNA copies of a particular gene per eukaryotic cell), and zero molecules P and Q. Parameters used are (A) $c_1 = 0.002 \text{ s}^{-1}$, $c_2 = 0.0005 \text{ s}^{-1}$, $c_3 = 0.005 \text{ s}^{-1}$, $c_4 = 0.004 \text{ s}^{-1}$, $c_5 = 0.004 \text{ s}^{-1}$, $c_6 = 0.05 \text{ s}^{-1}$, and (B) $c_1 = 0.0004 \text{ s}^{-1}$, $c_2 = 0.02 \text{ s}^{-1}$, $c_3 = 0.05 \text{ s}^{-1}$, $c_4 = 0.04 \text{ s}^{-1}$, $c_5 = 0.01 \text{ s}^{-1}$, and $c_6 = 0.05 \text{ s}^{-1}$. Moreover, $V = 2 \text{ pL}$ and $K = 8000 \text{ cells}$. The heavy bold line in the middle figure depicts the steady-state response curve of P, calculated by Monte Carlo simulation using the Gillespie algorithm.

Second-order SCKE approximation

It is unfortunate that we cannot compute the mesoscopic forcing term $\varepsilon(t)$ analytically. As a consequence, we cannot use standard numerical techniques to solve the SCKEs 9 (or Eqs. 5–8). Instead, we resort to Monte Carlo simulations using the Gillespie algorithm. However, to obtain accurate Monte Carlo estimates of mean concentration dynamics and fluxes, we need to uniformly sample the system state for a large number of DA trajectories. (Note that the variance of a Monte Carlo estimator with uniform sampling is $\sim 1/K$, where K is the number of samples used.) This approach is computationally intensive and especially burdensome when the biochemical reaction system is large and highly reactive.

To circumvent the computational expense of Monte Carlo simulation, we can approximate the SCKEs 5–8 by a system of first-order ordinary differential equations, which we can

solve efficiently by the same numerical techniques we use to solve the CKEs. As a matter of fact, we can approximate the mean molecular concentrations $\mathbf{u}(t)$ by concentrations $\hat{\mathbf{u}}(t)$ that satisfy the system of differential equations (see Appendix D) as

$$\frac{d\hat{\mathbf{u}}(t)}{dt} = \mathbb{S}\hat{\boldsymbol{\rho}}(t) + \hat{\boldsymbol{\varepsilon}}(t), \quad t \geq 0 \quad (16)$$

with

$$\hat{\boldsymbol{\varepsilon}}(t) = \mathbb{S}\hat{\boldsymbol{\theta}}(t), \quad (17)$$

where the m^{th} elements of $\hat{\boldsymbol{\rho}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ are given by

$$\hat{\rho}_m(t) = \kappa_m \psi_m(\hat{\mathbf{u}}(t)) \quad (18)$$

$$\hat{\theta}_m(t) = \frac{1}{2AV} \sum_{k=1}^M \sum_{l=1}^M \eta_{m,kl} \hat{v}_{z,kl}(t). \quad (19)$$

In Eq. 19, the terms $\hat{v}_{z,mm'}(t)$ approximate the DA covariances $v_{z,mm'}(t)$ and satisfy the system of first-order ordinary differential equations,

$$\begin{aligned} \frac{d\hat{v}_{z,mm'}(t)}{dt} = & AV \left[\kappa_m \psi_m(\hat{\mathbf{u}}(t)) + \frac{1}{2AV} \sum_{k=1}^M \sum_{l=1}^M \eta_{m,kl} \hat{v}_{z,kl}(t) \right] \delta_{mm'} \\ & + \sum_{k=1}^M \zeta_{m,k}(\hat{\mathbf{u}}(t)) \hat{v}_{z,m'k}(t) + \zeta_{m',k}(\hat{\mathbf{u}}(t)) \hat{v}_{z,mk}(t), \\ & t \geq 0, \quad m, m' = 1, 2, \dots, M, \end{aligned} \quad (20)$$

where $\delta_{mm'}$ is the Kronecker delta given by Eq. 31, and $\zeta_{m,k}$ is the first-order partial derivative of the propensity function of the m^{th} reaction with respect to z_k , given by

$$\zeta_{m,k}(\mathbf{u}) = \kappa_m \sum_{n=1}^N s_{nk} \frac{\partial \psi_m(\mathbf{u})}{\partial u_n}, \quad k = 1, 2, \dots, M.$$

For reasons explained in Appendix D, we refer to Eqs. 16–20 as second-order SCKEs.

We illustrate the quality of approximation obtained by the second-order SCKEs in Figs. 2 and 3, for the case of the quadratic autocatalator with feedback. For this example, the second-order SCKEs provide excellent approximations (*dashed lines*) of the mean concentration dynamics predicted by the CME (*solid lines*), which are clearly better than the approximations obtained with the CKEs (*dotted lines*). We also show in Fig. 5 that, by using the second-order SCKEs, we can obtain very good approximations of the dynamic evolutions of the coefficients of variation (CVs) associated with the intrinsic stochastic fluctuations in P and Q concentrations. (The CVs provide a measure of the relative dispersion, i.e., size, of stochastic fluctuations in the concentration

of a molecular species from the mean value; see Appendix D.) It is clear that calculation of CVs is not possible with the CKEs. The reader may also refer to Fig. 2 and Fig. 6 in Goutsias (23) for results obtained with a more complex biological system, which includes transcription, translation, protein dimerization, and molecular degradation. Therefore, in addition to satisfactorily approximating the mean concentration dynamics, we may use the second-order SCKEs to characterize intrinsic fluctuations in a stochastic biochemical reaction system by approximating CV dynamics.

Extensive simulations reveal that the quadratic autocatalator can be approximated very well by the second-order SCKEs for a wide range of parameter values and molecular concentrations (data not shown). One notable exception is at very low substrate concentrations s , in which case the biochemical reaction system will contain a very small number of molecules. We illustrate this in Fig. 6, which depicts the absolute relative errors in the steady-state mean concentrations of P and Q predicted by the second-order SCKEs with respect to the exact SCKEs (*solid lines*), by the CKEs with respect to the exact SCKEs (*dotted lines*), and by the CKEs with respect to second-order SCKEs (*dashed lines*), for the case when $\kappa_5 = \kappa_2 d$, in Fig. 6 A, and $\kappa_5 > \kappa_2 d$, in Fig. 6 B. Clearly, the second-order SCKEs provide consistently good and better approximations than the CKEs. As expected, when $\kappa_5 = \kappa_2 d$, the errors in approximating the steady-state Q concentration are zero. Moreover, the errors in the concentrations of P and Q predicted by the two models gradually diminish for large input substrate concentrations. Note however that the accuracy of the second-order SCKEs decreases at very small input substrate concentrations.

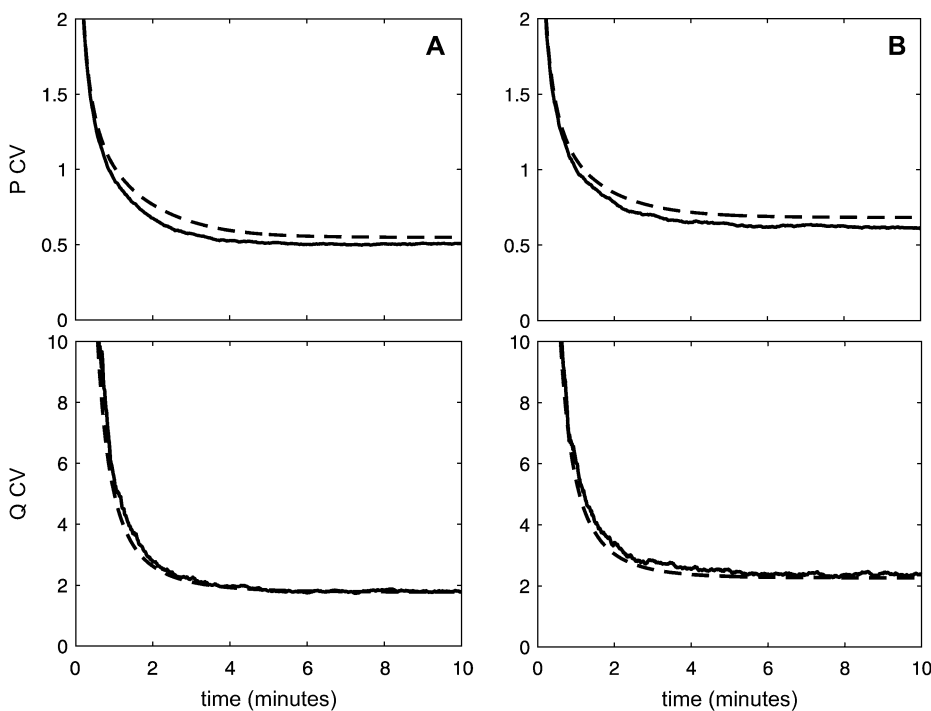


FIGURE 5 CV dynamics in the quadratic autocatalator, given by Eq. 12, associated with intrinsic stochastic fluctuations in the concentrations of P and Q, for the case when $\kappa_5 = \kappa_2 d$, in panel A, and $\kappa_5 > \kappa_2 d$, in panel B, predicted by the exact SCKEs (*solid lines*) and second-order SCKEs (*dashed lines*). The dynamics obtained by the exact SCKEs have been computed by Monte Carlo simulation using the Gillespie algorithm, whereas the dynamics obtained by the second-order SCKEs have been computed numerically. The parameters used are the same as in Figs. 2 and 3.

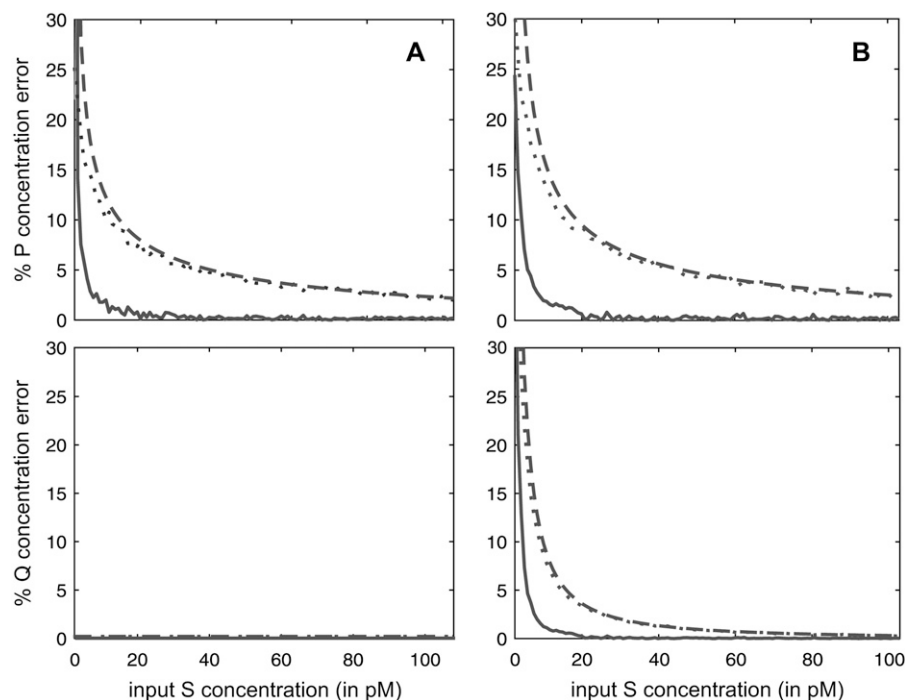


FIGURE 6 Absolute relative errors in the steady-state concentrations of P and Q associated with the quadratic autocatalator, given by Eq. 12, as a function of the input substrate concentration. (Solid lines) Second-order SCKEs with respect to the exact SCKEs; (dotted lines) CKEs with respect to the exact SCKEs; and (dashed lines) CKEs with respect to the second-order SCKEs. The steady-state values obtained by the exact and second-order SCKEs have been respectively computed by Monte Carlo simulation using the Gillespie algorithm and numerically, whereas the values obtained by the CKEs have been computed analytically from Eq. 44. The system is initialized with two molecules D (the number of DNA copies of a particular gene per eukaryotic cell), and zero molecules P and Q. Parameters used are $c_1 = 0.002 \text{ s}^{-1}$, $c_2 = 0.001 \text{ s}^{-1}$, $c_3 = 0.005 \text{ s}^{-1}$, $c_4 = 0.004 \text{ s}^{-1}$, $c_6 = 0.05 \text{ s}^{-1}$, $V = 2 \text{ pL}$, and $K = 8000 \text{ cells}$. Moreover, $c_5 = 0.002 \text{ s}^{-1}$ in panel A, and $c_5 = 0.006 \text{ s}^{-1}$ in panel B.

A good match between the predictions obtained by the second-order SCKEs and the ones obtained by the exact SCKEs indicates that the mean and covariances provide a sufficient description of intrinsic stochastic fluctuations, in which case, the molecular distributions will approximately follow a Gaussian distribution. However, the observation that, at very small input substrate concentrations, the second-order SCKEs may not sufficiently approximate the dynamics obtained by the exact SCKEs strongly suggests that higher-order (≥ 3) central moments may play a significant role in determining these dynamics. In this case, the underlying reactions will be subject to appreciable higher-order statistical interactions and the underlying probability distributions will not be Gaussian. Recent findings suggest that molecular distributions are often non-Gaussian and that such distributions may play an important role in cellular regulation (27,28). In those cases, it will be necessary to derive higher-order SCKE approximations, by including higher-order moments in the formulation (see our discussion in Appendix D).

CONCLUSIONS

In this article, by adopting a general framework for modeling the macroscopic behavior of a biochemical reaction system consisting of elementary irreversible reactions, we have shown that a classical chemical kinetics approach to modeling biochemical reaction systems may not be appropriate. The flux of each reaction is decomposed into two terms, a macroscopic term that accounts for the effects of mean molecular concentrations on the macroscopic behavior of the system and a mesoscopic term that accounts for the effects of pairwise

correlations among reactions. Based on this decomposition, we may characterize a biochemical reaction system by a system of exact first-order ordinary differential equations, the SCKEs, which provide a straightforward extension to the classical CKEs. The SCKEs require use of mesoscopic forcing terms, obtained by linearly transforming the mesoscopic fluxes through the stoichiometry matrix, whose calculation requires computationally expensive Monte Carlo simulations or evaluation of correlation dynamics among pairs, triplets, quadruplets, and larger groups of biochemical reactions.

To avoid such calculations, we have focused on a second-order approximation to the SCKEs, which includes only first- and second-order reaction statistics (i.e., means and pairwise correlations). These equations can be solved by standard numerical procedures and may lead to versatile tools for the analysis of biochemical reaction systems, similar to the ones used in classical chemical kinetics. Notably, a first-order approximation to the SCKEs produces the equations of classical chemical kinetics.

Our analysis indicates that pairwise correlation effects may lead, through mesoscopic forcing terms, to a dynamic behavior not accounted for by classical chemical kinetics. Numerical analysis of a quadratic autocatalator with positive feedback shows that the proposed second-order approximation faithfully reproduces system behavior, for a wide range of molecular concentrations and kinetic parameters. The success of this approximation demonstrates that the second-order SCKEs may provide substantial simplification in describing and analyzing stochastic biochemical reaction systems. Moreover, it suggests that pairwise statistical interactions among reactions may be sufficient for determining biological

function and supports the use of multivariate Gaussian distributions for modeling biochemical reactions. However, this may not be true at very low molecular concentrations in which case higher-order approximations may be necessary. The need to include higher-order moments in the approximation highlights the importance of higher-order interactions among biochemical reactions and the inappropriateness of Gaussian modeling at very low molecular concentrations.

APPENDIX A: STOCHASTIC CHEMICAL KINETICS

Since biochemical reactions occur by random collisions of reactant molecules, the number of molecules of a particular species present in the system at time t may fluctuate randomly. It is therefore appropriate to characterize the state of a biochemical reaction system at time t by an $N \times 1$ random vector $\mathbf{X}(t)$ whose n^{th} element $X_n(t)$ is the number of molecules of the n^{th} species present in the system at time t . In addition, we may use the degree of advancement (DA) $Z_m(t)$ to describe the (random) progress of the m^{th} reaction during the time interval $[0, t)$, where $Z_m(t) = z \geq 0$ means that the m^{th} reaction has occurred z times during the time interval $[0, t)$ (29). Note that, due to conservation of mass, we can uniquely determine $\mathbf{X}(t)$ from the $M \times 1$ random vector $\mathbf{Z}(t)$ with elements $Z_m(t)$, $m = 1, 2, \dots, M$, since

$$\mathbf{X}(t) = \mathbf{x}(0) + \mathbb{S}\mathbf{Z}(t), \quad t \geq 0. \quad (21)$$

Recall that our objective is to model the dynamic evolutions of molecular concentrations in a tissue containing a large population of cells by the $N \times 1$ vector $\mathbf{u}(t)$ given by Eq. 1. By taking expectations on both sides of Eq. 21 and by dividing with AV , we obtain

$$\mathbf{u}(t) = \mathbf{u}(0) + \frac{1}{AV}\mathbb{S}\boldsymbol{\mu}_{\mathbf{Z}}(t), \quad t \geq 0, \quad (22)$$

where $\mathbf{u}(0) = \mathbf{x}(0)/AV$ and $\boldsymbol{\mu}_{\mathbf{Z}}(t) = E[\mathbf{Z}(t)]$. If we assume that the mean DA $\boldsymbol{\mu}_{\mathbf{Z}}(t)$ is differentiable with respect to t (see below why this is true), then, by differentiating both sides of Eq. 22, we obtain Eq. 5. The average reaction rate (flux) $\nu_m(t)$ of the m^{th} reaction is given by

$$\nu_m(t) = \frac{1}{V} \frac{d\chi_m(t)}{dt}, \quad (23)$$

where $\chi_m(t)$ is the average extent of the m^{th} reaction defined as the mean DA of the reaction divided by the Avogadro number; i.e.,

$$\chi_m(t) = \frac{\boldsymbol{\mu}_{\mathbf{Z},m}(t)}{A}. \quad (24)$$

Computation of $\nu_m(t)$ requires calculation of the derivative of the mean DA $\boldsymbol{\mu}_{\mathbf{Z},m}(t)$ with respect to t . We show how to calculate this derivative in the following.

We denote by $\varphi_m(\mathbf{x})$ the number of all possible distinct combinations of the reactant molecules of the m^{th} reaction channel when the system is at state \mathbf{x} , given by (compare with Eq. 4)

$$\varphi_m(\mathbf{x}) = \begin{cases} x_n, & \text{for monomolecular reactions} \\ x_n(x_n - 1)/2, & \text{for bimolecular reactions with} \\ & \text{identical reactants} \\ x_n x_{n'}, & \text{for bimolecular reactions with} \\ & \text{different reactants} \end{cases}. \quad (25)$$

We also denote by c_m the specific probability rate constant of the m^{th} reaction (i.e., the probability per unit time that a randomly chosen combination of reactant molecules will react through the m^{th} reaction channel). Then, given that the biochemical reaction system is at state $\mathbf{X}(t) = \mathbf{x}$ at time t , the probability that one m^{th} reaction will occur during the time interval $[t, t + dt)$ is $\pi_m(\mathbf{x})dt + o(dt)$, for a sufficiently small dt , where $o(dt)$ is defined so that $o(dt)/dt \rightarrow 0$, as $dt \rightarrow 0$, and

$$\pi_m(\mathbf{x}) = c_m \varphi_m(\mathbf{x}) \quad (26)$$

is the propensity function of the m^{th} reaction channel (30,31). Moreover, the probability that more than one reaction will occur during $[t, t + dt)$ is $o(dt)$.

If $P_{\mathbf{Z}}(\mathbf{z}; t) = Pr(\mathbf{Z}(t) = \mathbf{z})$ is the probability that the DA vector $\mathbf{Z}(t)$ takes value \mathbf{z} at time t , then (23,32,33)

$$\frac{\partial P_{\mathbf{Z}}(\mathbf{z}; t)}{\partial t} = \sum_{m=1}^M \alpha_m(\mathbf{z} - \mathbf{e}_m) P_{\mathbf{Z}}(\mathbf{z} - \mathbf{e}_m; t) - \alpha_m(\mathbf{z}) P_{\mathbf{Z}}(\mathbf{z}; t), \quad t \geq 0, \quad (27)$$

where \mathbf{e}_m is the m^{th} column of the $M \times M$ identity matrix, and

$$\alpha_m(\mathbf{z}) = \pi_m(\mathbf{x}(0) + \mathbb{S}\mathbf{z}). \quad (28)$$

This chemical master equation (CME) describes the dynamic evolution of the joint probability mass function of the DA process $\mathbf{Z}(t)$. As a consequence, we can show that the means $\boldsymbol{\mu}_{\mathbf{Z},m}(t)$ and covariances $\nu_{\mathbf{Z},mm'}(t)$ of the DA process $\mathbf{Z}(t)$ satisfy the system of first-order ordinary differential equations (23),

$$\frac{d\boldsymbol{\mu}_{\mathbf{Z},m}(t)}{dt} = E[\alpha_m(\mathbf{Z}(t))], \quad m = 1, 2, \dots, M \quad (29)$$

$$\begin{aligned} \frac{d\nu_{\mathbf{Z},mm'}(t)}{dt} &= E[\alpha_m(\mathbf{Z}(t))\delta_{mm'}] \\ &+ E[Z_m(t)\alpha_{m'}(\mathbf{Z}(t))] - \boldsymbol{\mu}_{\mathbf{Z},m}(t)E[\alpha_{m'}(\mathbf{Z}(t))] \\ &+ E[Z_{m'}(t)\alpha_m(\mathbf{Z}(t))] - \boldsymbol{\mu}_{\mathbf{Z},m'}(t)E[\alpha_m(\mathbf{Z}(t))], \\ m, m' &= 1, 2, \dots, M, \end{aligned} \quad (30)$$

for $t \geq 0$, where $\delta_{mm'}$ is the Kronecker delta, given by

$$\delta_{mm'} = \begin{cases} 1, & \text{if } m = m' \\ 0, & \text{otherwise} \end{cases}. \quad (31)$$

Note that the derivatives $d\boldsymbol{\mu}_{\mathbf{Z},m}(t)/dt$ and $d\nu_{\mathbf{Z},mm'}(t)/dt$ always exist at finite times, regardless of the number of molecules present in the system, since the CME (Eq. 27) is valid only when the joint probability mass function $P_{\mathbf{Z}}(\mathbf{z}; t)$ is a continuous function of t , which in turn implies that the means $\boldsymbol{\mu}_{\mathbf{Z},m}(t)$ and covariances $\nu_{\mathbf{Z},mm'}(t)$ are continuous functions of t as well.

If we expand the propensity function $\alpha_m(\mathbf{z})$ by a Taylor series about the mean vector $\boldsymbol{\mu}_{\mathbf{Z}}(t)$, we have

$$\begin{aligned} \alpha_m(\mathbf{Z}) &= \alpha_m(\boldsymbol{\mu}_{\mathbf{Z}}(t)) + [\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}}(t)]^T \boldsymbol{\zeta}_m(\boldsymbol{\mu}_{\mathbf{Z}}(t)) \\ &+ \frac{1}{2} [\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}}(t)]^T \mathbb{H}_m [\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}}(t)], \end{aligned} \quad (32)$$

where $\boldsymbol{\zeta}_m$ and \mathbb{H}_m denote the gradient vector (of the first-order partial derivatives with respect to \mathbf{z}) and the Hessian matrix (of the second-order partial derivatives with respect to \mathbf{z}) of $\alpha_m(\mathbf{z})$, respectively, and T denotes vector (matrix) transposition. From Eqs. 25, 26, and 28, note that the propensity function α_m is at most a quadratic function of the DAs. Therefore, its derivatives of order > 2 are zero and Eq. 32 is exact. Moreover, the Hessian \mathbb{H}_m does not depend on \mathbf{z} . By taking expectations on both sides of Eq. 32 and by using Eq. 29, we obtain

$$\frac{d\boldsymbol{\mu}_{\mathbf{z},m}(t)}{dt} = \alpha_m(\boldsymbol{\mu}_{\mathbf{z}}(t)) + \frac{1}{2} \sum_{k=1}^M \sum_{l=1}^M \eta_{m,kl} \nu_{\mathbf{z},kl}(t),$$

$$t \geq 0, m = 1, 2, \dots, M. \quad (33)$$

In Eq. 33, $\eta_{m,kl}$ is the (k, l) element of the Hessian matrix \mathbb{H}_m (i.e., the second-order partial derivative of the propensity function $\pi_m(\mathbf{x})$ with respect to the DAs of the k^{th} and l^{th} reactions), given by

$$\eta_{m,kl} = \frac{\kappa_m}{AV} \sum_{n=1}^N \sum_{n'=1}^N s_{nk} s_{n'l} \frac{\partial^2 \psi_m(\mathbf{u})}{\partial u_n \partial u_{n'}}, \quad k, l = 1, 2, \dots, M,$$

where ψ_m is given by Eq. 4 and

$$\kappa_m = \begin{cases} c_m, & \text{for monomolecular reactions} \\ AVc_m/2, & \text{for bimolecular reactions with} \\ & \text{identical reactants} \\ AVc_m, & \text{for bimolecular reactions with} \\ & \text{different reactants.} \end{cases} \quad (34)$$

Equation 34 relates the specific probability rate constants c with the reaction rate constants κ . Equations 6–8 are now obtained from Eqs. 4, 23–26, 28, 33, and 34.

APPENDIX B: UNIDIRECTIONAL DIMERIZATION

Let us consider the unidirectional dimerization reaction in Eq. 11. We denote its DA by $Z_1(t)$ and use random variables $X_n(t)$, $n = 1, 2, 3$, to characterize its state at time $t \geq 0$, where each variable denotes the number of molecules of a reactant or product species, as identified by the following assignment:

$$X_n \leftrightarrow \begin{cases} \text{P}, & \text{for } n = 1 \\ \text{Q}, & \text{for } n = 2. \\ \text{P} \cdot \text{Q}, & \text{for } n = 3 \end{cases}$$

Note that the stoichiometry matrix is given by

$$\mathbb{S} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad (35)$$

whereas, the propensity function is given by $\pi_1(x_1, x_2, x_3) = c_1 x_1 x_2$. If we initialize the reaction with $x_1(0) = S$ molecules P, $x_2(0) = S$ molecules Q, and $x_3(0) = 0$ molecules P·Q, then, from Eq. 21 and Eq. 35, we have that

$$\begin{aligned} X_1(t) &= S - Z_1(t) \\ X_2(t) &= S - Z_1(t) \\ X_3(t) &= Z_1(t). \end{aligned} \quad (36)$$

Moreover, Eq. 28 implies that $\alpha_1(z_1) = c_1(S - z_1)^2$.

Since the reaction is bimolecular with different reactants, $\psi_1(u_1, u_2, u_3) = u_1 u_2$ (see Eq. 4). Moreover, from Eqs. 1 and 36, we have

$$u_1(t) = s - u_3(t) \quad \text{and} \quad u_2(t) = s - u_3(t),$$

where $s = S/AV$ is the concentration of initial P or Q molecules. In this case, Eq. 9 implies the following SCKE for the dimer concentration $u_3(t)$,

$$\frac{du_3(t)}{dt} = \kappa_1 [s - u_3(t)]^2 + \varepsilon_3(t), \quad t \geq 0,$$

where the mesoscopic forcing term $\varepsilon_3(t)$ is given by

$$\varepsilon_3(t) = \theta_1(t) = \frac{\kappa_1}{A^2 V^2} \nu_{\mathbf{z},11}(t),$$

and $\kappa_1 = AVc_1$, with initial conditions $u_3(0) = \varepsilon_3(0) = 0$. Moreover, Eq. 6 implies the following expression for the flux:

$$\nu_1(t) = \kappa_1 [s - u_3(t)]^2 + \varepsilon_3(t), \quad t \geq 0. \quad (37)$$

Finally, we can verify that the concentration $q_3(t)$ and flux $\rho_1(t)$ predicted by the CKE (obtained by setting $\varepsilon_3(t) = 0$, for every $t \geq 0$), are given by

$$q_3(t) = \frac{\kappa_1 s^2 t}{1 + \kappa_1 s t} \quad \text{and} \quad \rho_1(t) = \frac{\kappa_1 s^2}{(1 + \kappa_1 s t)^2}, \quad t \geq 0. \quad (38)$$

APPENDIX C: QUADRATIC AUTOCATALATOR

Let us now consider the six reactions in Eq. 12. We use variables $X_1(t)$ and $X_2(t)$ to characterize the molecular state of the system at time $t \geq 0$, where each variable is identified by the following assignment:

$$X_n \leftrightarrow \begin{cases} \text{P}, & \text{for } n = 1 \\ \text{Q}, & \text{for } n = 2. \end{cases}$$

In this case, the stoichiometry matrix is given by

$$\mathbb{S} = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & -1 \end{bmatrix}, \quad (39)$$

whereas, the propensity functions are given by

$$\begin{aligned} \pi_1(x_1, x_2) &= c_1 S \\ \pi_2(x_1, x_2) &= c_2 D x_1 \\ \pi_3(x_1, x_2) &= c_3 x_1 (x_1 - 1)/2 \\ \pi_4(x_1, x_2) &= c_4 x_1 x_2 \\ \pi_5(x_1, x_2) &= c_5 x_1 \\ \pi_6(x_1, x_2) &= c_6 x_2, \end{aligned}$$

with S and D being the number of S and D molecules, respectively, which we assume to be fixed. Note also that

$$\begin{aligned} \psi_1(u_1, u_2) &= s \\ \psi_2(u_1, u_2) &= d u_1 \\ \psi_3(u_1, u_2) &= u_1 (u_1 - 1/AV) \\ \psi_4(u_1, u_2) &= u_1 u_2 \\ \psi_5(u_1, u_2) &= u_1 \\ \psi_6(u_1, u_2) &= u_2. \end{aligned} \quad (40)$$

Here, $s = S/AV$ and $d = D/AV$. We initialize the reactions by setting $X_1(0) = X_2(0) = 0$. Then, from Eqs. 21 and 39, we have

$$\begin{aligned} X_1(t) &= Z_1(t) + Z_2(t) - Z_3(t) - Z_4(t) - Z_5(t) \\ X_2(t) &= Z_3(t) + Z_4(t) - Z_6(t). \end{aligned}$$

Equations 9, 39, and 40 lead to the following SCKEs for characterizing the quadratic autocatalator with feedback,

$$\begin{aligned} \frac{du_1(t)}{dt} &= \kappa_1 s + \kappa_2 d u_1(t) - \kappa_3 u_1(t) [u_1(t) - 1/AV] \\ &\quad - \kappa_4 u_1(t) u_2(t) - \kappa_5 u_1(t) + \varepsilon_1(t) \\ \frac{du_2(t)}{dt} &= \kappa_3 u_1(t) [u_1(t) - 1/AV] + \kappa_4 u_1(t) u_2(t) \\ &\quad - \kappa_6 u_2(t) + \varepsilon_2(t), \end{aligned} \quad (41)$$

for $t \geq 0$, where

$$\varepsilon_1(t) = -\varepsilon_2(t) = -\theta_3(t) - \theta_4(t),$$

and (recall Eq. 34)

$$\kappa_1 = c_1, \kappa_2 = AVc_2, \kappa_3 = \frac{AV}{2}c_3, \kappa_4 = AVc_4, \kappa_5 = c_5, \kappa_6 = c_6. \quad (42)$$

Note that $\theta_1(t) = \theta_2(t) = \theta_5(t) = \theta_6(t) = 0$, since the corresponding propensity functions are linear in \mathbf{z} (reactions 1, 2, 5, and 6 are monomolecular) and their Hessian matrices are thus zero. The corresponding CKEs are obtained from Eq. 41 by setting $\varepsilon_1(t) = \varepsilon_2(t) = 0$, for every $t \geq 0$.

Finally, if we make the reasonable assumption that, for every input substrate concentration, the steady-state mesoscopic forcing term $\bar{\varepsilon}(s)$ is bounded, then

$$\lim_{s \rightarrow \infty} \frac{\bar{\varepsilon}(s)}{s} = 0, \quad (45)$$

which implies that

$$\lim_{s \rightarrow \infty} \bar{u}_1(s) = \frac{\kappa_6}{\kappa_4}.$$

Indeed, from Eq. 45 and for a sufficiently large s , we have that $\kappa_1 s > \bar{\varepsilon}(s)$. Then,

$$\begin{aligned} \lim_{s \rightarrow \infty} \bar{u}_1(s) &= \lim_{s \rightarrow \infty} \left(-\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha} \right]^2 + \frac{\kappa_1 s - \bar{\varepsilon}(s)}{\alpha}} \right) \\ &= \lim_{s \rightarrow \infty} \frac{\left(-\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha} \right]^2 + \frac{\kappa_1 s - \bar{\varepsilon}(s)}{\alpha}} \right) \left(\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha} \right]^2 + \frac{\kappa_1 s - \bar{\varepsilon}(s)}{\alpha}} \right)}{\frac{\beta(s)}{2\alpha} + \sqrt{\left[\frac{\beta(s)}{2\alpha} \right]^2 + \frac{\kappa_1 s - \bar{\varepsilon}(s)}{\alpha}}} \\ &= \lim_{s \rightarrow \infty} \frac{\frac{1}{\alpha} \left[\kappa_1 - \frac{\bar{\varepsilon}(s)}{s} \right]}{\frac{\beta(s)}{2\alpha s} + \sqrt{\left[\frac{\beta(s)}{2\alpha s} \right]^2 + \frac{\kappa_1}{\alpha s} - \frac{\bar{\varepsilon}(s)}{\alpha s^2}}} = \lim_{s \rightarrow \infty} \frac{\kappa_1}{\beta(s)/s} = \frac{\kappa_1}{\kappa_1 \kappa_4 / \kappa_6} = \frac{\kappa_6}{\kappa_4}, \end{aligned}$$

If we set the right-hand sides of Eq. 41 (nullclines) equal to zero, we obtain the stationary solutions,

$$\begin{aligned} \bar{u}_1(s) &= -\frac{\beta(s)}{2\alpha} \pm \sqrt{\left[\frac{\beta(s)}{2\alpha} \right]^2 + \frac{\kappa_1 s - \bar{\varepsilon}(s)}{\alpha}} \quad \text{and} \\ \bar{u}_2(s) &= \frac{\kappa_1 s + (\kappa_2 d - \kappa_5) \bar{u}_1(s)}{\kappa_6}, \end{aligned} \quad (43)$$

with α and β given by Eq. 15, where $\bar{u}_1 = \lim_{t \rightarrow \infty} u_1(t)$, $\bar{u}_2 = \lim_{t \rightarrow \infty} u_2(t)$, and $\bar{\varepsilon} = -\lim_{t \rightarrow \infty} \varepsilon_1(t)$. Since only nonnegative solutions are relevant, the system will relax at steady state to molecular concentrations that depend on the signs of α , β , and on the value of the input flux $\kappa_1 s$ as compared to the value of the steady-state mesoscopic forcing term $\bar{\varepsilon}(s)$. We show this in Table 1, where we summarize the steady-state concentrations of P predicted by the SCKEs, which we obtain from Eq. 43 by assuming that $\alpha > 0$ (we can obtain similar results for the case when $\alpha < 0$). The corresponding steady-state concentrations of Q are obtained from the second Eq. 43, provided that $\kappa_1 s + (\kappa_2 d - \kappa_5) \bar{u}_1(s) \geq 0$. Note that, by setting $\bar{\varepsilon}(s) = 0$ in Eq. 43, we obtain the following stationary solutions of the CKEs:

$$\begin{aligned} \bar{q}_1(s) &= -\frac{\beta(s)}{2\alpha} \pm \sqrt{\left[\frac{\beta(s)}{2\alpha} \right]^2 + \frac{\kappa_1 s}{\alpha}} \quad \text{and} \\ \bar{q}_2(s) &= \frac{\kappa_1 s + (\kappa_2 d - \kappa_5) \bar{q}_1(s)}{\kappa_6}. \end{aligned} \quad (44)$$

For $\alpha > 0$, the steady-state P concentrations predicted by the CKEs are summarized in Table 2.

since, from Eq. 15, $\lim_{s \rightarrow \infty} \beta(s)/s = \kappa_1 \kappa_4 / \kappa_6$. This implies that, in the limit of $s \rightarrow \infty$, the steady-state P concentration will not depend on $\bar{\varepsilon}$. Therefore, for a sufficiently large input substrate concentration, the steady-state P and Q concentrations predicted by the CKEs will be approximately equal to the ones predicted by the SCKEs.

APPENDIX D: SCKE APPROXIMATIONS

To approximate the SCKEs 5–8, we should note that the behavior of a biochemical reaction system depends on the third-order central moments of the DA process $\mathbf{Z}(t)$. This dependence comes from the second and third terms on the right-hand side of Eq. 30, due to the second-order term in the Taylor series expansion of the propensity function, given by Eq. 32. Including these moments in the formulation requires an additional set of differential equations, which depend on fourth-order central moments, and so on. These nested dependencies rule out the possibility of determining the exact covariance values by solving the system of Eqs. 29 and 30.

To address this problem, we may employ a method of moment closure. For instance, it might be possible to find an expression for a higher-order moment in terms of lower-order moments, which would then make the system exactly solvable. However, we adopt a much simpler approach here, by setting the third-order central moments equal to zero (see also ((16,34)). These moments represent higher-order statistical dependencies among reaction channels due to bimolecular reactions. Note that higher-order statistical dependencies might become unimportant at some level. Since we do not know a priori when this might happen, we may assume that the third-order central moments have negligible effect on the DA means and covariances and subsequently check the resulting covariance approximations against the ones obtained by Monte Carlo estimation. If the results are not satisfactory,

then we have to include higher-order (≥ 3) moments in the formulation. Note however that, although the task of including higher-order moments is straightforward, it will increase the number of differential equations in the approximation.

To derive our approximation, we expand each propensity function by a Taylor series about the mean DA vector $\mu_z(t)$ (see Eq. 32), we use this expansion in Eqs. 29 and 30, and set the third-order central moments of the DA process $\mathbf{Z}(t)$ equal to zero. As a consequence of these steps, we obtain the system of differential equations, Eqs. 16–20, which allow us to approximate the mean molecular concentrations $\mathbf{u}(t)$ by concentrations $\hat{\mathbf{u}}(t)$. Equations 16–20 provide a second-order approximation to the SCKEs, since we derive them by including only the first- and second-order moments of the DA process (i.e., the DA means and covariances). In this sense, the classical CKEs provide a first-order approximation to the SCKEs, since we can derive them by following the same steps but by including only the first-order DA moments (i.e., the DA means). In most cases of interest, we expect that the second-order SCKEs will provide sufficiently good approximations of the mean concentration dynamics, which will be more accurate than the approximations obtained by the CKEs.

In sharp contrast to the CKEs, we can use the second-order SCKEs to approximate the CVs associated with a biochemical reaction system. (The CV associated with the n^{th} molecular species at time t is the ratio of the square-root of the n^{th} diagonal element of the covariance matrix $\mathbf{C}_x(t)$ of the molecular population process $\mathbf{X}(t)$, divided by the corresponding mean.) Moreover, we can use the second-order SCKEs to approximate the correlation dynamics between the concentrations of two molecular species, quantified by the correlation coefficient. To do so, we can use the fact that the covariance matrix $\mathbf{C}_x(t)$ is related to the covariance matrix $\mathbf{C}_z(t)$ of the DA process $\mathbf{Z}(t)$ by means of $\mathbf{C}_x(t) = \mathbb{S}\mathbf{C}_z(t)\mathbb{S}^T$, since $\mathbf{X}(t) = \mathbf{x}(0) + \mathbb{S}\mathbf{Z}(t)$ (recall Eq. 21), and approximate $\mathbf{C}_x(t)$ by $\hat{\mathbf{C}}_x(t) = \mathbb{S}\hat{\mathbf{C}}_z(t)\mathbb{S}^T$, where $\hat{\mathbf{C}}_z(t)$ is a matrix with elements $\hat{v}_{z,mm'}(t)$.

An alternative way to approximate the concentration dynamics is to use Eqs. 16–18 and recognize (by virtue of Eq. 19 and the fact that $\hat{\mathbf{C}}_x(t) = \mathbb{S}\hat{\mathbf{C}}_z(t)\mathbb{S}^T$) that

$$\hat{\theta}_m(t) = \frac{1}{2AV} \sum_{k=1}^N \sum_{l=1}^N h_{m,kl} \hat{v}_{x,kl}(t), \quad (46)$$

where $h_{m,kl}$ is the second-order partial derivative of the propensity function of the m^{th} reaction with respect to x_k and x_l . The terms $\hat{v}_{x,mm'}(t)$ in Eq. 46 approximate the covariances of the molecular population process $\mathbf{X}(t)$ and can be shown to satisfy the system of first-order ordinary differential equations

$$\begin{aligned} \frac{d\hat{v}_{x,mm'}(t)}{dt} = & AV \sum_{n=1}^M \left\{ s_{nm} s_{n'm} \left[\kappa_n \psi_m(\hat{\mathbf{u}}(t)) \right. \right. \\ & \left. \left. + \frac{1}{2AV} \sum_{k=1}^N \sum_{l=1}^N h_{m,kl} \hat{v}_{x,kl}(t) \right] \right. \\ & \left. + \sum_{k=1}^N \frac{\gamma_{m,k}(\hat{\mathbf{u}}(t))}{AV} [s_{nm} \hat{v}_{x,n'k}(t) + s_{n'm} \hat{v}_{x,nk}(t)] \right\}, \\ & t \geq 0, \quad n, n' = 1, 2, \dots, N, \end{aligned}$$

where $\gamma_{m,k}$ is the first-order partial derivative of the propensity function of the m^{th} reaction with respect to x_k , given by

$$\gamma_{m,k}(\mathbf{u}) = \kappa_m \frac{\partial \psi_m(\mathbf{u})}{\partial u_n}, \quad k = 1, 2, \dots, N.$$

When the number N of molecular species is smaller than the number M of reactions, this approach will be computationally more advantageous than calculating the DA covariances, since the number of population covariances, which is given by $N(N+1)/2$, will be smaller than the number of DA

covariances, which is given by $M(M+1)/2$. However, we have briefly discussed in Goutsias (23) that characterizing a stochastic biochemical reaction system by means of the DA process $\mathbf{Z}(t)$ may be more advantageous in certain circumstances than characterizing the system by means of the molecular population process $\mathbf{X}(t)$. Because our developments in this article are based on the DA process $\mathbf{Z}(t)$, we use Eqs. 16–20 in our numerical investigations.

We can also derive the second-order SCKEs 16–20 by assuming that the most important influence on the firing rate of a given reaction in a stochastic biochemical reaction system is exerted by the mean propensity function of that reaction through a Poisson process and use an appropriately chosen zero mean additive correction term to compensate for statistical variations not accounted for by the Poisson process. This leads to a mean-field approximation of the system whose state $\mathbf{Z}(t)$ approximately follows a normal Gibbs distribution $\hat{P}_Z(\mathbf{z}; t)$ at temperature $2/k_B$, with energy function $[\mathbf{z} - \hat{\boldsymbol{\mu}}_Z(t)]^T [\hat{\mathbf{C}}_Z(t)]^{-1} [\mathbf{z} - \hat{\boldsymbol{\mu}}_Z(t)]$, where k_B is the Boltzmann constant and the elements of $\hat{\boldsymbol{\mu}}_Z(t)$ satisfy Eq. 33 with $v_{z,kl}(t)$ being replaced by $\hat{v}_{z,kl}(t)$. The reader is referred to Goutsias (23) for details.

The SCKE approximation method employed in this article is one of several alternative strategies for approximating stochastic biochemical reaction systems (e.g., see (23)). A frequently used technique is the linear noise approximation method (29,35). This method is obtained from a Langevin approximation of the stochastic biochemical reaction system by 1), linearizing the propensity functions about the mean DA values, and 2), taking the limit of the resulting linear Fokker-Planck equation as the system volume tends to infinity (e.g., see (23)). Linearization of the propensity functions implies that their second-order partial derivatives with respect to the DAs will be zero. In turn, this implies that the mesoscopic fluxes and, therefore, the mesoscopic forcing terms will also be zero. Hence, the linear noise approximation method leads to the same system of differential equations for the mean concentration dynamics as the one obtained by classical chemical kinetics. As a consequence, the linear noise approximation method suffers from the same drawbacks as classical chemical kinetics and should be used with caution when investigating the effects of intrinsic stochastic fluctuations on biological function at low molecular concentrations.

The author thanks William Dempsey and Prof. K. Konstantopoulos for their helpful comments and suggestions.

REFERENCES

1. Heinrich, R., and S. Schuster. 1996. *The Regulation of Cellular Systems*. Chapman and Hall, New York.
2. McAdams, H. H., and A. Arkin. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* 15:65–69.
3. Rao, C. V., D. M. Wolf, and A. P. Arkin. 2002. Control, exploitation and tolerance of intracellular noise. *Nature.* 420:231–237.
4. Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6:451–464.
5. McAdams, H. H., and A. Arkin. 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA.* 94:814–819.
6. Arkin, A., J. Ross, and H. H. McAdams. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics.* 149:1633–1648.
7. Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. *Nat. Genet.* 31:69–73.
8. Blake, W. J., M. Kaern, C. R. Cantor, and J. J. Collins. 2003. Noise in eukaryotic gene expression. *Nature.* 422:633–637.
9. Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature.* 427:737–740.

10. Peixoto, A., M. Monteiro, B. Rocha, and H. Veiga-Fernandes. 2004. Quantification of multiple gene expression in individual cells. *Genome Res.* 14:1938–1947.
11. Shav-Tal, Y., R. H. Singer, and X. Darzacq. 2004. Imaging gene expression in single living cells. *Nat. Rev. Mol. Cell Biol.* 5:856–862.
12. Capodici, P., M. Donovan, K. Buchinsky, Y. Jeffers, C. Cordon-Cardo, W. Gerald, J. Edelson, S. M. Shenoy, and R. H. Singer. 2005. Gene expression profiling in single cells within tissue. *Nat. Methods.* 2:663–665.
13. Bowtell, D. D. L. 1999. Options available—from start to finish—for obtaining expression data by microarray. *Nat. Genet.* 21:25–32.
14. Lockhart, D. J., and E. A. Winzler. 2000. Genomics, gene expression and DNA arrays. *Nature.* 405:827–836.
15. McQuarrie, D. A., C. J. Jachimowski, and M. E. Russell. 1964. Kinetics of small systems. II. *J. Chem. Phys.* 40:2914–2921.
16. Thakur, A. K., A. Rescigno, and C. DeLisi. 1978. Stochastic theory of second-order chemical reactions. *J. Phys. Chem.* 82:552–558.
17. Leonard, D., and L. E. Reichl. 1990. Stochastic analysis of a driven chemical reaction. *J. Chem. Phys.* 92:6004–6010.
18. Zheng, Q., and J. Ross. 1991. Comparison of deterministic and stochastic kinetics for nonlinear systems. *J. Chem. Phys.* 94:3644–3648.
19. Savageau, M. A. 1976. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology.* Addison-Wesley, Reading, MA.
20. Samoilov, M., S. Plyasunov, and A. P. Arkin. 2005. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proc. Natl. Acad. Sci. USA.* 102:2310–2315.
21. Darvey, I. G., and P. J. Staff. 1966. Stochastic approach to first-order chemical reaction kinetics. *J. Chem. Phys.* 44:990–997.
22. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.
23. Goutsias, J. 2006. A hidden Markov model for transcriptional regulation in single cells. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 3:57–71.
24. Scott, S. K. 1991. *Chemical Chaos.* Oxford University Press, Oxford, UK.
25. Kaufman, M., F. Andris, and O. Leo. 1996. A model for antigen-induced T cell unresponsiveness based on autophosphorylative protein tyrosine kinase activity. *Int. Immunol.* 8:613–624.
26. Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2002. *Molecular Biology of the Cell*, 4th Ed. Garland Publishing, New York.
27. Banerjee, B., S. Balasubramanian, G. Ananthakrishna, T. V. Ramakrishnan, and G. V. Shivashankar. 2004. Tracking operator state fluctuations in gene expression in single cells. *Biophys. J.* 86:3052–3059.
28. Krishna, S., B. Banerjee, T. V. Ramakrishnan, and G. V. Shivashankar. 2005. Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. *Proc. Natl. Acad. Sci. USA.* 102:4771–4776.
29. van Kampen, N. G. 1992. *Stochastic Processes in Physics and Chemistry.* Elsevier, Amsterdam, The Netherlands.
30. Gillespie, D. T. 1992. A rigorous derivation of the chemical master equation. *Physica A.* 188:404–425.
31. Gillespie, D. T. 2000. The chemical Langevin equation. *J. Chem. Phys.* 113:297–306.
32. Haseltine, E. L., and J. B. Rawlings. 2002. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* 117:6959–6969.
33. Goutsias, J. 2005. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.* 122:184102.
34. Pirone, J. R., and T. C. Elston. 2004. Fluctuations in transcription factor binding can explain the graded and binary responses observed in inducible gene expression. *J. Theor. Biol.* 226:111–121.
35. Elf, J., and M. Ehrenberg. 2003. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.* 13:2475–2484.