

Evolution of complex probability distributions in enzyme cascades

Yueheng Lan, Garegin A. Papoian*

Department of Chemistry, University of North Carolina, Chapel Hill, USA

Received 9 March 2007; received in revised form 29 May 2007; accepted 6 June 2007

Available online 12 June 2007

Abstract

Unusual probability distribution profiles, including transient multi-peak distributions, have been observed in computer simulations of cell signaling dynamics. The emergence of these complex distributions cannot be explained using either deterministic chemical kinetics or simple Gaussian noise approximation. To develop physical insights into the origin of complex distributions in stochastic cell signaling, we compared our approximate analytical solutions of signaling dynamics with the exact numerical simulations. Our results are based on studying signaling in 2-step and 3-step enzyme amplification cascades that are among the most common building blocks of cellular protein signaling networks. We have found that while the multi-peak distributions are typically transient, and eventually evolve into single peak distributions, in certain cases these distributions may be stable in the limit of long times. We also have shown that introducing positive feedback loops results in diminution of the probability distribution complexity.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Nonlinear chemical kinetics; Signal transduction; Discrete noise; Strong fluctuations; Master equation

1. Introduction

Cells are highly nonequilibrium systems that require constant energy flux to maintain their heterogeneous structures and power their life activities. The gene transcription, metabolism and cell motion are realized and regulated by interconnected networks of complicated biochemical reactions, including protein signal transduction processes (Gomperts et al., 2002; Bray, 1995). In cells, these biochemical reactions are grouped into different pathways that may crosstalk to each other and are usually localized in space and coordinated in time. In recent years, with large amount of data coming from high throughput experiments, much effort has been devoted to model and understand in detail how these signaling cascades work (Heinrich et al., 2002; Chaves et al., 2004; Hansel and Mato, 2001; Markevich et al., 2004; Huang et al., 2003; Cao et al., 2004; Meng et al., 2004; Turner et al., 2004; Bratsun et al., 2005). For example, one would like to predict the cellular response to a specific external cue, for example, a chemoattractant or a repellent. To answer these

types of questions, it is necessary to develop qualitative and quantitative understanding of the signaling network dynamical behaviors that arise from both steady state and transient responses (Kærn et al., 2005; Yildirim et al., 2004; Paulsson et al., 2000; Berg et al., 2000; Kepler and Elston, 2001; Metzler and Wolynes, 2002; Blake et al., 2003).

Cells constantly interact with their environment and each other, inducing regulated response reactions in the cell. These responses are often initiated at random points on the cell surface and transduced into the cytosol by discrete chemical reactions. If the number of protein copies is large for each species in a cascade, the law of large numbers guarantees the validity of a deterministic description based on the chemical kinetics equations (Heinrich et al., 2002; Chaves et al., 2004; Hansel and Mato, 2001; Markevich et al., 2004; Huang et al., 2003). If the protein number of one species in a cascade is small, stochastic description of signaling dynamics becomes necessary. It is well established now that stochasticity plays a central role in various signal transduction contexts (Cao et al., 2004; Lubchenko and Silbey, 2004; Meng et al., 2004; Turner et al., 2004; Bratsun et al., 2005; Stukalin and Kolomeisky, 2006; Cai et al., 2006; Thattai and van Oudenaarden, 2001;

*Corresponding author. Tel.: +19199628037; fax: +19199622388.

E-mail address: gpapoian@unc.edu (G.A. Papoian).

Süel et al., 2006; Grossman, 1995; Sasai and Wolynes, 2003). Furthermore, in certain cases, cells may control and even utilize the stochasticity intrinsic to signaling reactions (Berg et al., 2000; Vilar and Rubí, 2001; Arias and Hayward, 2006).

Signaling network stochastic dynamics in the limit of large protein numbers is well approximated by the chemical Langevin equation and the Fokker–Planck equation. The validity of these approximations often extend to systems containing a moderate number of protein molecules (Weinberger et al., 2005; Thattai and van Oudenaarden, 2002; Swain, 2004; Shibata and Fujimoto, 2005). These equations invariably contain a skeleton deterministic part and a diffusive fluctuating part around the deterministic orbit. They can be derived from the more precise microscopic description, the master equation, by linear noise approximation in which the random part is almost always assumed to be modeled by Gaussian white noise (Gillespie, 2000; van Kampen, 1992; Kærn et al., 2005; Thattai and van Oudenaarden, 2002). When the protein number of even one species is very low, however, the approximation is likely to break down and one needs to solve the original master equation (Lan and Papoian, 2006). In this paper, we will show how complex, non-Gaussian probability distribution functions (PDFs) arise in simple reaction cascades when protein copy numbers are small.

For the examples studied in this work, the deterministic evolution is a monotonic relaxation to a unique stationary state. When only the Langevin noise is considered, the obtained PDFs often turn out nearly Gaussian, having only a single peak (van Kampen, 1992; Kærn et al., 2005; Lan et al., 2006) in the distribution profile. However, when the exact master equation is solved, in some cases distributions with multiple peaks arise. Consequently, in these parameter regimes, the simple Langevin description of noise is inadequate. This, in turn, suggests a strong connection between the discreteness of the protein numbers and the correlation among the chemical reaction events. It was suggested that each peak of the PDF profile corresponds to a stable steady state of the noisy system (Garca-Ojalvo and Sancho, 1999). Thus, for the multi-peak distributions discussed below, multiple stable states are a common occurrence. These states may be classified as either transient or permanent. In a recent paper by Arkin et al. (Samoilov and Arkin, 2006), these kinds of phenomena were termed “deviant effect”, in analogy to quantum deviation from classical mechanics in physics.

In the next section, we discuss a master equation which describes stochastic signaling a specific 2-step cascade most commonly encountered in cell enzymatic cascades. For the parameters regime where multiply peaked distributions arise, we provide approximate analytical solution to the master equation (Lan and Papoian, 2006). In Section 3, we compare our approximate results with exact Gillespie simulation results, elucidating how the complexity of the PDF profile emerges for this 2-step cascade. To show the

ubiquity of this phenomenon, we extend the computations to a 3-step cascade. When a positive feedback loop is introduced in the 3-step cascade, the distribution complexity becomes diminished, which is discussed next. Finally, we speculate on the possible biological significance of the complex PDF profiles.

2. Analytical approach to stochastic dynamics in signal amplification cascades

The potential complexities of the protein number PDF profiles may be well demonstrated using the example of a 2-step signal amplification cascade. This seemingly simple cascade shows nontrivial dynamical behavior, allowing us to gain useful physical insights into cellular stochastic signaling (Lan and Papoian, 2006; Lan et al., 2006). Interestingly, the PDF evolution in the 2-step cascade can be approximated analytically in the parameter regime where multi-peak distributions are produced (Lan and Papoian, 2006). This, in turn, allows us to gain additional insights into the physical reasons behind the emergence of complex protein number PDF profiles in stochastic signaling.

The 2-step amplification cascade without feedback is shown in the dashed box in Fig. 1. It is described by a simple reaction scheme, where, R represents an inactive receptor, which becomes activated into R^* upon binding of an external ligand (stimulus) with a rate g . When the receptor is activated, it acts as an enzyme, catalyzing the phosphorylation of the next kinase downstream ($A + R^* \rightarrow A^* + R^*$) with a rate μ . A^* spontaneously decays to A with a rate λ and R^* to R with a rate k . Although the R^* reaction is unary and independent of the A reaction, the latter one is binary, making the system *nonlinear*, thus, different from commonly modeled linear reactions (Thattai and van Oudenaarden, 2001; Kierzek et al., 2001; Swain et al., 2002; Ozbudak et al., 2002). Additional downstream reactions are common in specific biological pathways, where the way signals are transduced is tuned by the network architecture. For example, a subsequent downstream reaction is introduced in Fig. 1, $B + A^* \rightarrow B^* + A^*$, where the activation of B up-regulates

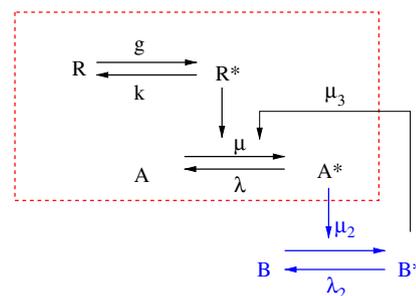


Fig. 1. A 3-step cascade with one positive feedback is shown. The smaller network in the dashed box is a 2-step amplification cascade, which has been the main focus of our recent works (Lan and Papoian, 2006; Lan et al., 2006).

A activation, which is an example of a positive feedback loop. We studied the 2-step cascade using both analytical and numerical computations, while the 3-step cascade was investigated only numerically.

Within a volume with linear dimensions of the Kuramoto length, ζ , diffusion mixes the reagents in a nearly uniform manner (Kuramoto, 1974; van Kampen, 1992; Markevich et al., 2004). If the reactions are considered in the Kuramoto volume, as done in the current work, it is reasonable to neglect the spatial heterogeneity. The typical values of ζ are estimated in the range from 0.3 to 5 μm for the MAPK cascade (Schoeberl et al., 2002; Thorne and Hrabětová, 2004; Elowitz et al., 1999). In our computation, $\zeta \sim 1 \mu\text{m}$ was assumed. If the linear dimensions of the cell are significantly larger than 1 μm , then multiple Kuramoto compartments may need to be considered to study spatial patterns of signal activation on the scale of the whole cell.

The reaction rate parameters vary widely among various signal transduction pathways. In the visual signal transduction cascade, for example, rhodopsin (Rh) activation is controlled by the incident photons which may arrive, one by one, within 1 s or even longer time interval in the single photon experiment (Schneeweis and Schnapf, 1995) (thus, $g = 1 \text{ s}^{-1}$). The deactivation rate from Rh^* to Rh is approximately $k = 0.5 \text{ s}^{-1}$. The activation rate of the transducin is approximately $\mu = 120 \text{ s}^{-1}$ and the deactivation rate is $\lambda = 100 \text{ s}^{-1}$ (Burns and Baylor, 2001). In a different example, the rate constants may be around $\mu = 0.0167 \text{ s}^{-1}$ in a MAPK cascade (Markevich et al., 2004). Furthermore, even for similar pathways, the reaction rates may significantly vary depending on the cell type and the stage of development. In the computations carried out in this work, some biologically reasonable parameter values were chosen to illustrate the phenomenon of multi-peak distributions. Since the compartment size for all our computations is fixed to ζ , all the reaction rates are given in units of s^{-1} . Correspondingly, the time unit of seconds is used to report the results of our numerical calculations.

If $P(m, n)$ is used to denote the probability of having m R^* 's and n A^* 's, then, the master equation for the 2-step cascade in Fig. 1 can be written as

$$\begin{aligned} \frac{dP}{dt}(m, n) = & \mu[-m(N - n)P(m, n) \\ & + m(N - n + 1)P(m, n - 1)] \\ & + \lambda[-nP(m, n) + (n + 1)P(m, n + 1)] \\ & + g[-P(m, n) + P(m - 1, n)] \\ & + k[-mP(m, n) + (m + 1)P(m + 1, n)], \end{aligned} \quad (1)$$

where N , a constant of motion, is the total number of A and A^* proteins. In Eq. (1), the first two terms describe the $A - A^*$ reaction and the rest the $R - R^*$ reaction. The nonlinear terms describe catalytic action of A^* on R . If a large number of inactivated receptors R are present, the rate of conversion is almost independent of m and the $R \rightarrow R^*$ reaction becomes Poissonian. For simplicity, we assume that this is the case in all the following calculations. If the

$R \rightarrow R^*$ reaction is formulated as the usual birth–death problem, our computation still applies with only minor changes, with the same qualitative trends, as discussed below.

Eq. (1) can be efficiently solved using numerical simulations of Monte Carlo type, called the Gillespie technique in the context of stochastic chemical kinetics. In addition, in the parameter regime we are interested in, a good analytical approximation of the solution can be derived based on the generating function formalism (Lan and Papoian, 2006). We analyze below the structure of this solution and explain how it may lead to complex PDF profiles.

For the 2-step cascade, the generating function $\Psi(x, y) = \sum_{m,n} P(m, n)x^m y^n$ satisfies

$$\begin{aligned} \frac{\partial \Psi}{\partial t} = & (1 - y) \left(\mu xy \frac{\partial^2}{\partial y \partial x} - \mu N x \frac{\partial}{\partial x} + \lambda \frac{\partial}{\partial y} \right) \Psi \\ & + g(x - 1)\Psi - k(x - 1) \frac{\partial \Psi}{\partial x}. \end{aligned} \quad (2)$$

In the limit of slow $R - R^*$ reaction, through time scale separation and application of the method of characteristics, an approximate solution of Eq. (2) was obtained (Lan and Papoian, 2006)

$$\begin{aligned} \Psi(x, y) = & \sum_{m=0}^{\infty} \exp\left(-\frac{g}{k}(1 - e^{-kt})\right) \\ & \times (1 - e^{-kt})^m \left(\frac{g}{k}\right)^m \phi_m(y) \frac{x^m}{m!}, \end{aligned} \quad (3)$$

where

$$\phi_m(y) = \left[1 + \frac{\mu m}{\lambda + \mu m} (1 - e^{-(\lambda + \mu m)t})(y - 1) \right]^N \quad (4)$$

is the generating function of the A^* distribution starting with zero A^* and at fixed m . From the generating function, we can immediately obtain the average $\bar{n} = \partial \Psi / \partial y|_{(x=1, y=1)}$ and the variance $\sigma^2 = (\partial^2 \Psi / \partial y^2 + \bar{n} - \bar{n}^2)|_{(x=1, y=1)}$.

Our next goal is to obtain physical insights into the solution, Eq. (3). In the limit of slow $R - R^*$ reaction, the $A - A^*$ reaction closely follows each $R - R^*$ reaction trajectory. The conditional distribution $P_m(n)$ of A^* at time t is completely determined by the protein R^* number at the same time t in the given realization of the stochastic trajectory. As can be inferred from Eq. (3), a binomial distribution of A^* corresponds to each m term in the summation, centered around the conditional average $\bar{n}_m = N f_m = \mu m N (1 - e^{-(\lambda + \mu m)t}) / (\lambda + \mu m)$. In fact, the conditional average \bar{n}_m determines the whole binomial distribution. The conditional variance $\sigma_m^2 = N f_m (1 - f_m)$ indicates that when \bar{n}_m approaches 0 or N (i.e., $f_m \rightarrow 0$ or 1), the distribution becomes narrower; while in the middle $\bar{n}_m = N/2$, the distribution achieves its maximum width. Also note that the relative width w_m of each peak

$$w_m = \sigma_m / \bar{n}_m = \sqrt{\frac{1 - f_m}{N f_m}} \quad (5)$$

decreases with increasing N , thus, the discreteness represented by these peaks becomes easier to identify for larger

N . Therefore, the total distribution of A^* may be approximately considered as a superposition of Gaussian packets. Depending on the height, width and spacing between these packets, the total composite distribution can be near Gaussian, strongly distorted Gaussian or being comprised multiple peaks. The exact features of the distribution are determined by the rate constant parameter values and the initial conditions.

Generalization of the above analytical approximation may be difficult for longer cascades. However, the physical picture, which it provides, remains valid, even if complicated by additional reactions and protein species. The multi-peak character of probability distributions can be inherited, diminished or completely suppressed in larger cascades, depending on the reaction network architecture and reaction rate parameters. The example of the 3-step cascade is elaborated in the next section. In general, as the total number of upstream proteins increases, the number of distribution peaks may decrease and their discreteness may become less clear. Since the initial discreteness originates from the small number of protein copies, individual peaks in the distribution may become smeared as the signal propagates through the pathway, possibly merging into a smooth, single-peak distribution. Next, we discuss the time evolution of the PDF profiles for a few examples.

3. Exploring dynamical behaviors using numerical simulations

3.1. Simulation details

We used the Gillespie algorithm to calculate the PDFs at different times for the 2-step and the 3-step cascades. As discussed below, the PDFs exhibit very different profiles in different parameter regimes, demonstrating the dynamical richness of the stochastic dynamics in these signaling networks. For the 2-step cascade, we compare the results from Eqs. (3) and (4) with the Gillespie computation, in order to show the validity of our analytical approximation, which, in turn, is used to uncover the origin of the PDF profile complexity. For the 3-step cascade, we show that this complexity may be reduced by introducing a positive feedback loop. All Gillespie simulations were carried out over 4×10^5 trajectories to ensure sufficient convergence. To clearly show the peak structure in the figures below, the distribution graphs were trimmed near the origin in the plot windows, when the probability of having few activated proteins is very large. Also, some parts of the distributions that were essentially zero were omitted. All computations were started with zero copies of all activated species, R^* , and, A^* , B^* .

3.2. Transient and permanent multi-peak distributions in the 2-step cascade

In Figs. 2 and 4, the A^* PDFs at two different times are shown that correspond to the transient stage and equi-

brated stage of the distribution evolution. The circles mark the results obtained from the Gillespie simulations, while the solid lines are obtained from the analytical expressions, Eqs. (3) and (4). These two different computations perfectly agree with each other, indicating the effectiveness of our approximation.

The PDF profile at the initial stage $t = 20$ in Fig. 2(a) is characterized by three peaks. The peak at the origin corresponds to the realization of zero activated proteins in the system. The other two peak correspond to the generation of one, two or more protein R^* 's. The second peak is near $N_{A^*} = 13$ which is the asymptotic average of N_{A^*} for $m = 1$. For $m = 2$, this average should be around 21 where the third peak is truly located. At this stage, the $m = 1$ peak is perceivably higher than the $m = 2$ peak. At $t = 400$, the system equilibrates. We only see one peak in Fig. 2(b) at $N_{A^*} = 57$ which is the average of N_{A^*} for $m = 10$. Therefore, in this case, the multi-peak distribution profile is only transient. It implies a maximum trajectory variability at the initial stage of the dynamics. When the system becomes mature, the PDF turns Gaussian.

The cascade plot of the actual time evolution of the system in Fig. 2 is shown in Fig. 3(a), where a snapshot of the PDF was taken at time intervals of 10. Initially, all the probability concentrates to the first peak at the origin. Then the second peak appears and grows. Subsequently, the third peak appears while the second one keeps growing. When the fourth peak appears, the second peak starts to decay. During the whole evolution, the first peak decays monotonically while other peaks emerge, grow larger and decay chronically in a time-ordered way until the asymptotic average is reached. Finally, all the peaks merge into one large Gaussian-like distribution (see Fig. 2(b)).

The above observations may be explained as follows. At the initial stage $t = 20$, the R^* distribution concentrates at the small R^* number states $m = 0, 1, 2, \dots$. The discreteness is succinctly pronounced so that the conditional averages \bar{n}_m of protein A^* are relatively well separated. Hence the individual peaks are visible even when they are juxtaposed to form the total PDF. At the final equilibrium stage $t = 400$, R^* distribution approaches its steady Poisson distribution centered around the average $\bar{n} = g/k = 10$. Compared to the $t = 20$ distribution, two things happened: the difference of \bar{n}_m 's between neighboring m 's decreases so that the neighboring binomial distributions have a larger overlap; and also, the width of each distribution increases when n_m 's move closer to $N/2$ as discussed in Section 2. Under these two effects, all the peaks merge into a single large peak, which only appears to be slightly distorted from a Gaussian distribution.

The multi-peak character of distribution profiles can be preserved even when the equilibrium is reached if the discreteness of protein numbers remains pronounced. In Fig. 4, the PDF of A^* for a different set of parameter values is shown. In this case, the asymptotic average $\bar{n} = g/k$ of R^* is decreased to 2 to emphasize the discreteness. In addition, the protein number N_{A^*} is increased from 100 to

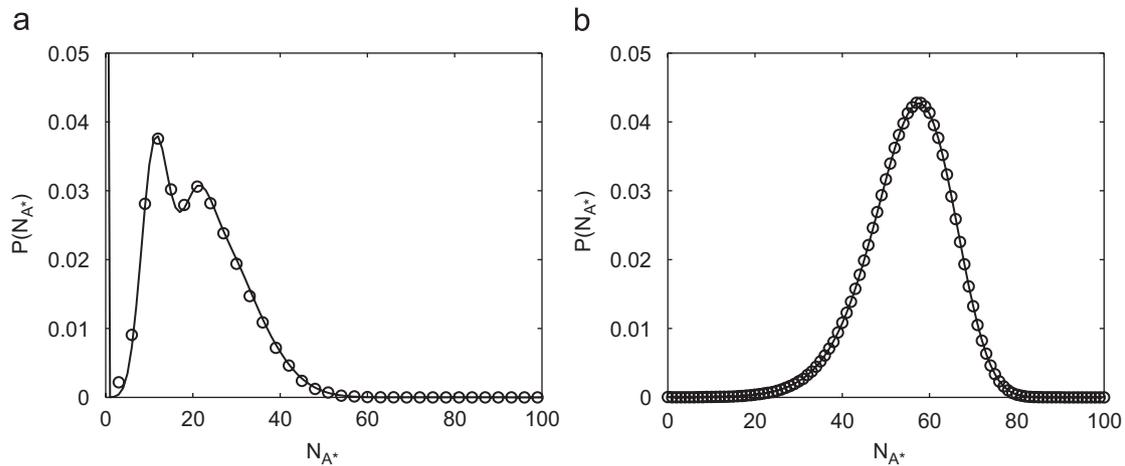


Fig. 2. The PDF of A^* at two different times. (a) At the transient phase $t = 20$, (b) at the equilibrium $t = 400$. $g = 0.1$, $k = 0.01$, $\mu = 0.2$, $\lambda = 1.5$ with initial condition $(N_R, N_{R^*}, N_A, N_{A^*}) = (100, 0, 100, 0)$.

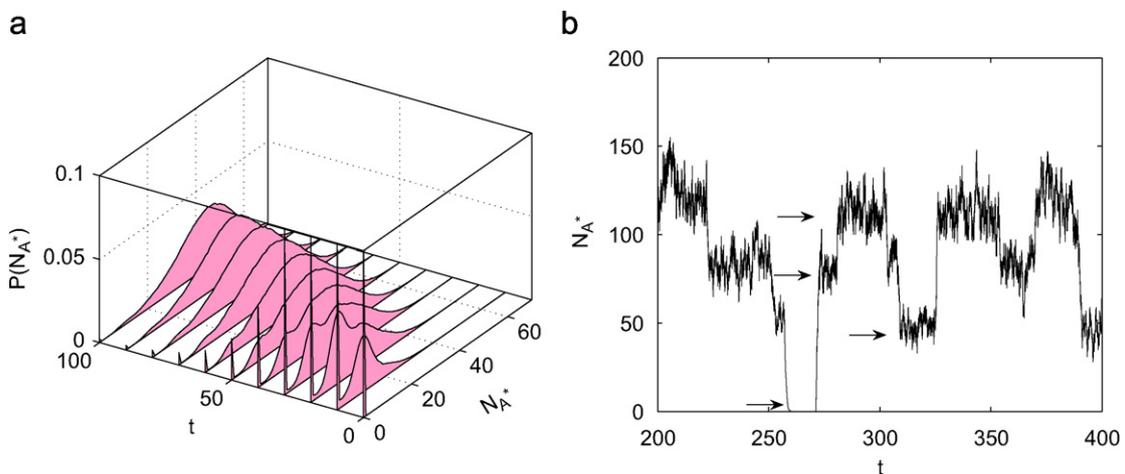


Fig. 3. (a) The time evolution of the transient multi-peak distribution for the same set of parameters as in Fig. 2. (b) One typical trajectory at the equilibrated state with the same parameters as in Fig. 4. The arrows point to different “stable states”.

400 to sharpen the peaks around each R^* trajectory by decreasing the relative width of each peak of the A^* PDF, w_m (see Eq. (5)). At the transient phase $t = 60$ in Fig. 4(b), four clear peaks of the protein A^* PDF are seen that correspond to different R^* states $m = 0, 1, 2, 3$. With increasing m values, the peak heights monotonically decrease while their widths increase. At equilibration time of $t = 400$, the peak for $m = 0$ has decreased and all other peaks have grown and become more distinct. Furthermore, even a clearly discernible fifth peak is observed.

In Fig. 3(b), a typical reaction trajectory is shown, generated after the equilibrium had been reached. From bottom up, the four arrows in the figure clearly mark four stable steady states that correspond to $m = 0, 2, 3, 4$. Hence, the multi-peak distribution profiles have clear physical manifestation that can be observed.

The examples above demonstrate that multi-peak distributions could either be transient or permanent in which multiple stable points emerge. Both phenomena are completely nonclassical that would not be reproduced

either from the deterministic description or within the Gaussian noise approximation. Thus, when modeling stochastic signaling in protein networks, one has to consider the possibility of having these highly distorted distributions.

3.3. Signal propagation in a 3-step cascade without feedback

In this section we demonstrate that multi-peak distribution profiles may be propagated and modulated along longer cascades. We carried out Gillespie simulations for a 3-step cascade without feedback, shown in Fig. 1, to investigate the PDF time evolution. Here, the $A - A^*$ and $R - R^*$ reaction rates are similar to the ones reported in Fig. 2. A snapshot of B^* PDF at $t = 20$ is shown in Fig. 5(a), displaying two peaks. As discussed above, the first peak at the origin indicates no reaction taking place. The second peak around $N_{B^*} = 20$ is very broad and exhibits a plateau. Although there are fewer peaks than in Fig. 2(a), a clear deviation from the Gaussian

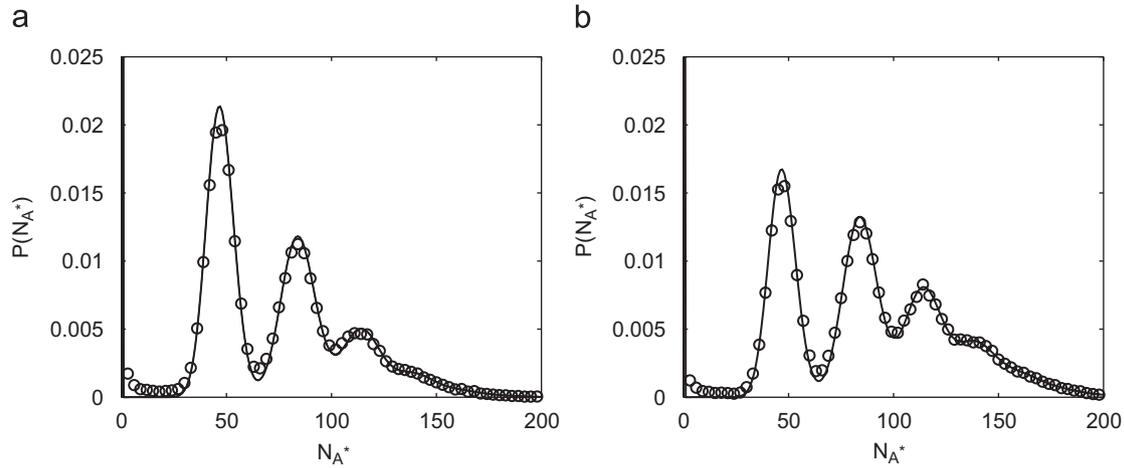


Fig. 4. The PDF of A^* at two different times. (a) At the transient phase $t = 60$, (b) At the equilibrium $t = 400$. Another set of parameters and initial conditions was chosen (cf. Fig. 2): $g = 0.04$, $k = 0.02$, $\mu = 0.2$, $\lambda = 1.5$ and $(N_R, N_{R^*}, N_A, N_{A^*}) = (100, 0, 400, 0)$.

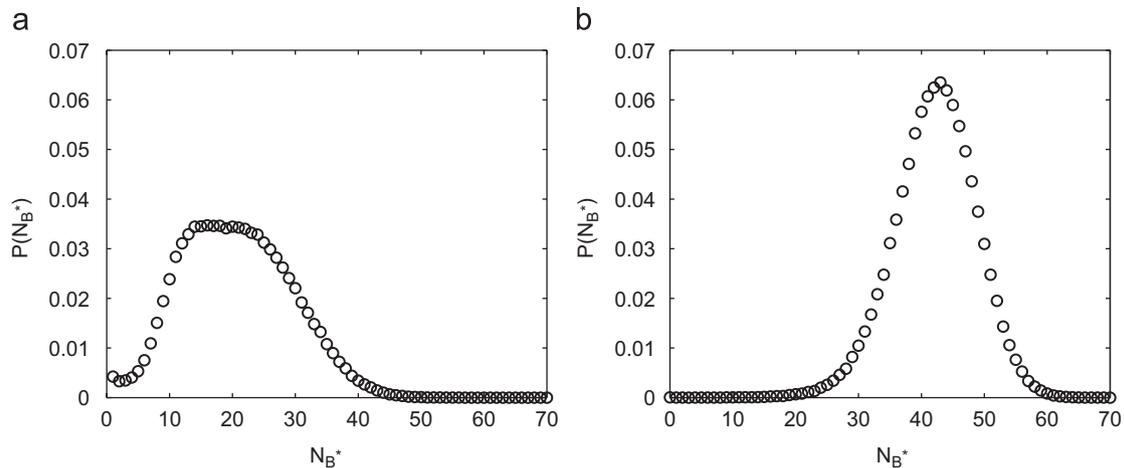


Fig. 5. The PDF of B^* at two different times. (a) At the transient phase $t = 20$, (b) at the equilibrium $t = 400$. For parameter values $g = 0.1$, $k = 0.01$, $\mu = 0.2$, $\lambda = 1.5$, $\mu_2 = 0.02$, $\lambda_2 = 1.5$ with initial conditions $(N_R, N_{R^*}, N_A, N_{A^*}, N_B, N_{B^*}) = (100, 0, 100, 0, 100, 0)$.

distribution is apparent, as demonstrated by the broad plateau consisting of several closely spaced peaks. The equilibrium PDF of B^* at $t = 400$ (see Fig. 5(b)) is nearly Gaussian, peaked at the deterministic value of $N_{B^*} = 44$ (cf. Fig. 2(b)).

More pronounced multi-peak distribution profiles are also observed in the 3-step signaling cascade. The profiles shown in Fig. 6 were generated using the same parameter values for the $R - R^*$ and the $A - A^*$ reactions as reported in Fig. 4. Besides the peak at the origin, two additional peaks are observed for both the transient case (Fig. 6(a)) and the equilibrium case (Fig. 6(b)). As the system moves towards equilibrium, the probability flows to the major peak at $N_{B^*} = 60$. Compared with the profiles shown in Fig. 4, the number of peaks decreases. However, they become much broader. This is rationalized by a much larger quantity of the upstream protein A^* , which results in the nearby peaks being superimposed to give just one broad peak. Two distinct peaks that appear in the PDF

profile carry the memory of the discreteness in the most upstream $R - R^*$ reaction. Even though protein B^* is present at large copy numbers, the fluctuations are still very large and cannot be accounted for by the usual Gaussian statistics. This, in turn, indicates that the deterministic dynamics cannot be relied upon to even qualitatively model signaling in this system.

3.4. Introducing feedback in the 3-step cascade allows to control distribution profiles

It is believed that cell reaction networks may take advantage of noise in signal transduction. The noise, either external or internal, can be modulated to suit different needs of the cell. On the other hand, the large fluctuations due to the discreteness of the initial triggering species, as discussed above, can also be further controlled by the downstream signaling networks. If a positive feedback loop is added to the 3-step cascade, as shown in Fig. 1, then the

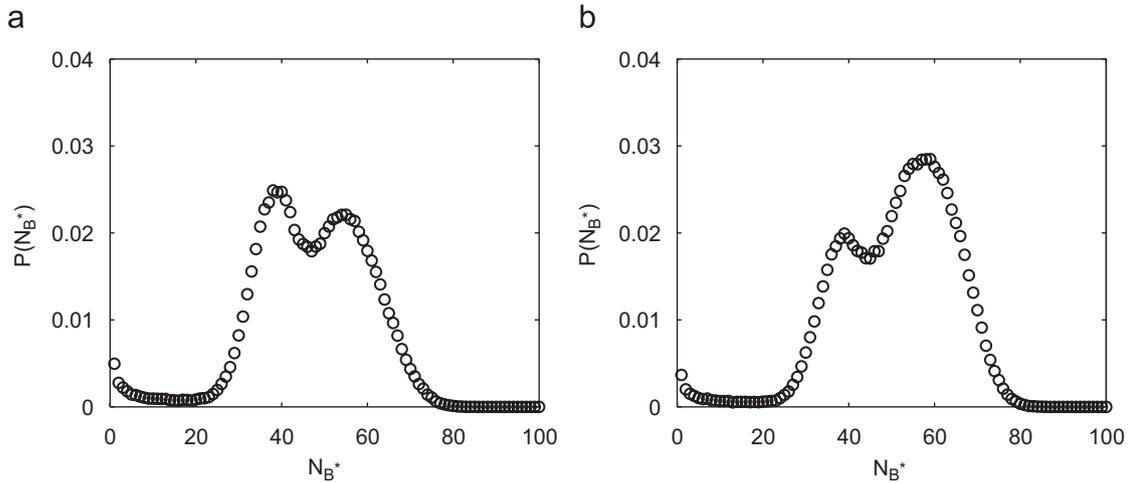


Fig. 6. The PDF of B^* at two different times. (a) At the transient phase $t = 60$, (b) at the equilibrium $t = 400$. Another set of parameters and initial conditions was chosen (cf. Figs. 4 and 5): $g = 0.04$, $k = 0.02$, $\mu = 0.2$, $\lambda = 1.5$, $\mu_2 = 0.02$, $\lambda_2 = 1.5$ and $(N_R, N_{R^*}, N_A, N_{A^*}, N_B, N_{B^*}) = (100, 0, 400, 0, 100, 0)$.

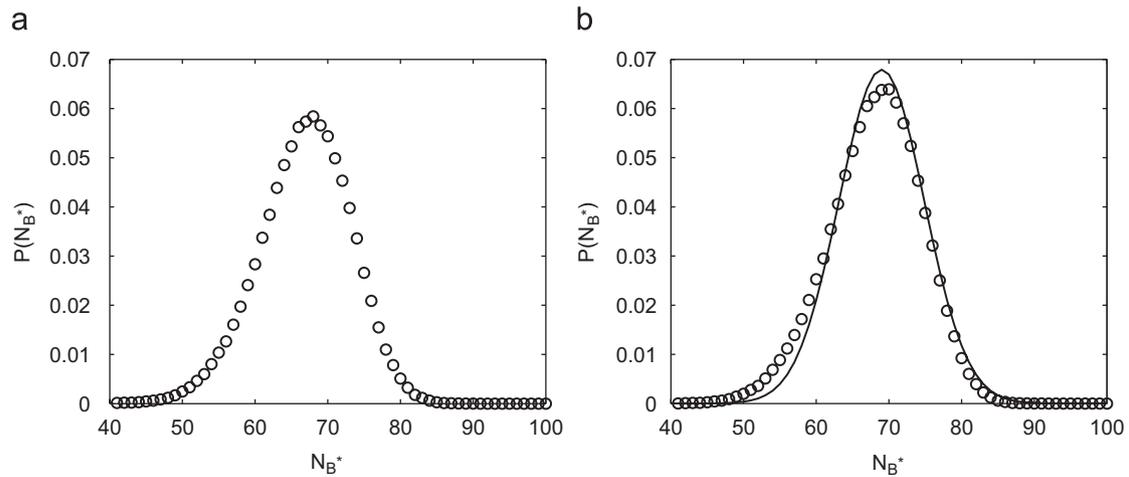


Fig. 7. The PDF of B^* at two different times. (a) At the transient phase $t = 60$, (b) at the equilibrium $t = 400$. For parameters values $g = 0.04$, $k = 0.02$, $\mu = 0.2$, $\lambda = 1.5$, $\mu_2 = 0.02$, $\lambda_2 = 1.5$, $\mu_3 = 0.01$ with initial conditions $(N_R, N_{R^*}, N_A, N_{A^*}, N_B, N_{B^*}) = (100, 0, 400, 0, 100, 0)$.

noise fluctuations can be greatly reduced. The B^* PDFs at the transient stage $t = 60$ and the equilibrated stage $t = 400$ are shown in Fig. 7. The main peak of both PDFs is nearly Gaussian. However, the profile shown in Fig. 7(a) (the PDFs at $t = 60$) indicates that there still exists a peak at the origin (outside of the window shown in Fig. 7(a)) signifying zero reaction events in the system. Therefore, the system exhibits a transient bimodal distribution. However, the second peak entrains a much larger probability, such that the majority of reaction events follow a deterministic route to the classical peak center at $N_{B^*} = 69$. A small fraction, however, defy the “classical” dynamics and remain at the origin. At equilibrium, the PDFs obtained from the Gillespie simulations are shown in Fig. 7(b). The exotic set at the origin disappears and all the probability is concentrated in the classical package at $N_{B^*} = 69$. The

solid line is a plot of the Gaussian approximation of the profile:

$$P(N_{B^*}) = \frac{1}{\sqrt{69\pi}} \exp\left(\frac{-(N_{B^*} - 69)^2}{69}\right), \tag{6}$$

which largely agrees with the true distribution. Therefore, the introduction of the positive feedback diminishes the non-Gaussian behavior and partly recovers the effectiveness of the Langevin or deterministic equation.

The effect of the positive feedback can be explained as follows. Once A^* is generated, it catalyzes the formation of B^* . Conversely, the B^* feedback provides an extra reaction channel for producing A^* , independent of the R^* catalytic channel. The B^* feedback alone can sustain the average of B^* at $N_{B^*} = 60$. Thus, when A^* and B^* are being switched

on, their averages follow a deterministic path to the final large asymptotic values. On the other hand, in a specific realization, A^* may not be activated for some time, thus, both A^* and B^* numbers would remain at the origin. This explains the transient bimodal distribution, discussed above in the context of Fig. 7(a). On the other hand, every reaction trajectory will eventually result in activation of A^* and B^* , where the protein numbers will approach and fluctuate around the classical averages. Therefore, at equilibrium, only one peak is produced, as observed in Fig. 7(b). The positive feedback here acts as a unidirectional two-state switch, manifestly reducing the fluctuations.

4. Summary

We have demonstrated the possibility of obtaining complex protein number PDF profiles in the context of stochastic dynamics of cell signaling cascades. Different PDF profiles were explored for stochastic signaling in the 2-step and the 3-step cascades by either solving the master equation analytically or using numerical simulations. For various parameter values, the PDFs are characterized by a spectrum of qualitatively distinct profiles, ranging from multi-peak profiles, to those having a plateau, and to nearly Gaussian profiles. In addition, the distribution characteristics evolve continuously with time, resulting in interesting transient dynamics.

This nonclassical behavior cannot be captured by the equations of deterministic chemical kinetics, or even the commonly used chemical Langevin equation. This, in turn, indicates that the discreteness of the chemical species in a signaling cascade may qualitatively change the system behavior, resulting in significant deviation from the linear noise approximation. For these cases, the master equation formulation becomes a necessity for obtaining reliable results.

One could hypothesize that the transient multi-peak distributions may play important biological roles. For example, in the embryonic development, the fate decision of the early cell differentiation may depend on the protein fluctuations in each individual cell (Yakoby et al., 2005). The transient multi-peak PDF profile could provide a convenient mechanism for the cell to choose its developmental path. In the later stages, when the tissues and organs are localized and start to grow, presumably, stability and uniformity of the cell within each strain is required (Arias and Hayward, 2006; von Dassow et al., 2000), thus, the fluctuations should be suppressed. Hence, only transient variability of the cell is needed. Of course, the real biological network for controlling this developmental process is extremely complicated (Freeman and Gurdon, 2002). However, the idea of nonclassical, transient complexity of the PDF profiles, put forward in this work, can be extended to study these more complex networks.

References

- Arias, A.M., Hayward, P., 2006. Filtering transcription noise during development: concepts and mechanisms. *Nature* 7, 34.
- Berg, O.G., Paulsson, J., Ehrenberg, M., 2000. Fluctuations in repressor control: thermodynamic constraints on stochastic focusing. *Biophys. J.* 79, 2944–2953.
- Blake, W.J., Kaern, M., Cantor, C.R., Collins, J.J., 2003. Noise in eukaryotic gene expression. *Nature* 422, 633.
- Bratsun, D., Volfson, D., Tsimring, L.S., Hasty, J., 2005. Delay-induced stochastic oscillations in gene regulation. *Proc. Natl Acad. Sci. USA* 102 (41), 14593–14598.
- Bray, D., 1995. Protein molecules as computational elements in living cells. *Nature* 376 (27), 307.
- Burns, M.E., Baylor, D.A., 2001. Activation, deactivation, and adaptation in vertebrate photoreceptor cells. *Annu. Rev. Neurosci.* 24, 779–805.
- Cai, L., Friedman, N., Xie, X.S., 2006. Stochastic protein expression in individual cells at the single molecular level. *Nature* 440, 358.
- Cao, Y., Li, H., Petzold, L., 2004. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys.* 121 (9), 4059.
- Chaves, M., Sontag, E.D., Dinerstein, R.J., 2004. Optimal length and signal amplification in weakly activated signal transduction cascades. *J. Phys. Chem. B* 108, 15311–15320.
- Elowitz, M.B., Surette, M., Wolf, P., Stock, J., Leibler, S., 1999. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 181, 197–203.
- Freeman, M., Gurdon, J.B., 2002. Regulatory principles of developmental signaling. *Annu. Rev. Cell Devel. Biol.* 18, 515–539.
- Garca-Ojalvo, J., Sancho, J.M., 1999. *Noise in Spatially Extended Systems*. Springer, New York.
- Gillespie, D.T., 2000. The chemical Langevin equation. *J. Chem. Phys.* 113 (1), 297.
- Gomperts, B.D., Kramer, I.M., Tatham, P.E.R., 2002. *Signal Transduction*. Academic Press, San Diego.
- Grossman, A.D., 1995. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus Subtilis*. *Annu. Rev. Genet.* 29, 477–508.
- Hansel, D., Mato, G., 2001. Existence and stability of persistent states in large neuronal networks. *Phys. Rev. Lett.* 86 (18), 4175.
- Heinrich, R., Neel, B.G., Rapoport, T.A., 2002. Mathematical models of protein kinase signal transduction. *Mol. Cell* 9, 957–970.
- Huang, K.C., Meir, Y., Wingreen, N.S., 2003. Dynamic structures in *Escherichia coli*: spontaneous formation of minE rings and minD polar zones. *Proc. Natl Acad. Sci. USA* 100, 12724–12728.
- Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6, 451.
- Kepler, T.B., Elston, T.C., 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81, 3116–3136.
- Kierzek, A.M., Zaim, J., Zielenkiewicz, P., 2001. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J. Biol. Chem.* 276, 8165–8172.
- Kuramoto, Y., 1974. Effects of diffusion on the fluctuations in open chemical systems. *Prog. Theor. Phys.* 52, 711.
- Lan, Y., Papoian, G.A., 2006. The interplay between discrete noise and nonlinear chemical kinetics in a signal amplification cascade. *J. Chem. Phys.* 125, 154901.
- Lan, Y., Wolynes, P., Papoian, G.A., 2006. A variational approach to the stochastic aspects of cellular signal transduction. *J. Chem. Phys.* 125, 124106.
- Lubchenko, V., Silbey, R.J., 2004. Control of chemical equilibrium by noise. *J. Chem. Phys. B* 108, 19852–19858.
- Markevich, N.I., Hoek, J.B., Kholodenko, B.N., 2004. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.* 164 (3), 353–359.

- Meng, T.C., Somani, S., Dhar, P., 2004. Modeling and simulation of biological systems with stochasticity. *In Silico Biol.* 4, 0024.
- Metzler, R., Wolynes, P.G., 2002. Number fluctuations and the threshold model of kinetic switches. *Chem. Phys.* 284, 469.
- Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., van Oudenaarden, A., 2002. Regulation of noise in the expression of a single gene. *Nature Genet.* 31, 69–73.
- Paulsson, J., Berg, O.G., Ehrenberg, M., 2000. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc. Natl Acad. Sci. USA* 97 (13), 7148–7153.
- Samoilov, M.S., Arkin, A.P., 2006. Deviant effects in molecular reaction pathways. *Nat. Biotech.* 24, 1235.
- Sasai, M., Wolynes, P.G., 2003. Stochastic gene expression as a many-body problem. *Proc. Natl Acad. Sci. USA* 100 (5), 2374–2379.
- Schneeweis, D.M., Schnapf, J.L., 1995. Photovoltage of rods and cones in the macaque retina. *Science* 268, 1053–1056.
- Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., Müller, G., 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* 20, 370.
- Shibata, T., Fujimoto, K., 2005. Noisy signal amplification in ultrasensitive signal transduction. *Proc. Natl Acad. Sci.* 102 (2), 331–336.
- Stukalin, E.B., Kolomeisky, A.B., 2006. Atp hydrolysis stimulates large length fluctuations in single actin filaments. *Biophys. J.* 90, 2673–2685.
- Süel, G.M., Ojalvo, J.G., Liberman, L.M., Elowitz, M.B., 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440, 545.
- Swain, P.S., 2004. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.* 344, 965–976.
- Swain, P.S., Elowitz, M.B., Siggia, E.D., 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci.* 99 (20), 12795–12800.
- Thattai, M., van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci.* 98 (15), 8614–8619.
- Thattai, M., van Oudenaarden, A., 2002. Attenuation of noise in ultrasensitive signaling cascades. *Biophys. J.* 82, 2943.
- Thorne, R.G., Hrabětová, S., 2004. Diffusion of epidermal growth factor in rat brain extracellular space measured by integrative optical imaging. *J. Neurophysiol.* 92, 3471.
- Turner, T.E., Schnell, S., Burrage, K., 2004. Stochastic approaches for modelling in vivo reactions. *Comput. Biol. Chem.* 28, 165–178.
- van Kampen, N.G., 1992. *Stochastic Processes in Physics and Chemistry*. North Holland Personal Library, Amsterdam.
- Vilar, J.M.G., Rubí, J.M., 2001. Noise suppression by noise. *Phys. Rev. Lett.* 86 (6), 950.
- von Dassow, G., Meir, E., Munro, E.M., Odell, G.M., 2000. The segment polarity network is a robust developmental module. *Nature* 406, 188.
- Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., Schaffer, D.V., 2005. Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuation drive phenotypic diversity. *Cell* 122, 169–182.
- Yakoby, N., Bristow, C.A., Gouzman, I., Rossi, M.P., Gogotsi, Y., Schüpbach, T., Shvartsman, S.Y., 2005. Systems-level questions in *Drosophila oogenesis*. *IEE Proc.—Syst. Biol.* 152, 276.
- Yildirim, N., Hao, N., Dohlman, H.G., Elston, T.C., 2004. Mathematical modeling of RGS and G-protein regulation in yeast. *Methods Enzymol.* 389, 384.