Stochasticity in multi-phosphorylation and quasi steady state approximation in stochastic simulation

S. Das and D. Barik*

Abstract

Quantitative and qualitative nature of chemical noise propagation in a network of chemical reactions depend crucially on the topology of reaction networks. Multisite reversible phosphorylation-dephosphorylation of target proteins is one such recurrently found topology in various cellular networks regulating key functions in living cells. Here we analytically calculated the stochasticity in multistep reversible chemical reactions by determining variance of phosphorylated species at the steady state using linear noise approximation. We investigated the dependence of variance on the rate parameters in the reaction chain and the number of phosphorylation sites on the species. Assuming a quasi steady state approximation on the multistep reactions, originating from the disparity in time scales in the network, we propose a simulation scheme for coupled chemical reactions to improve the computational efficiency of stochastic simulation of the network. We performed case studies on signal transduction cascade and positive feedback loop with bistability to show the accuracy and efficiency of the method.

Introduction

An isogenic population of cells, both from single and multi-cellular organisms, in an identical environmental condition show remarkable cell-to-cell variability in protein abundance, cell cycle properties, cell size and shape, timescales of key signaling events resulting in population heterogeneity (1-7). The population heterogeneity is due to the inevitable intrinsic and extrinsic noise, collectively known as chemical noise, affecting chemical reactions inside living cells(8). While chemical noise may act as nuisance, it has potential beneficial roles in cellular physiology and functions (9-11). While intrinsic noise originates from the low molecular abundances of chemical species and produces stochastic trajectories of products over time, extrinsic noise originates from global factors such as variation in cell volume or size, cell cycle phases, abundances of regulatory molecules such as transcription factors affecting all chemical reactions in a given cell equally (12). Using probabilistic description of chemical reactions, theoretical models of gene expression noise were able to quantitatively explain many experimental observations on protein noise (13-15). However cellular functions are regulated by chemical reaction networks involving many proteins and these networks often consists of small regulatory network motifs with distinct properties (16). Therefore further investigations were carried out to understand the effects of network topology such as signaling cascades, feedback loops and feed-forward loops on the propagation of chemical noise (17-22). These protein interaction networks often consist of multisite reversible covalent modifications of target proteins such as phosphorylation, acetylation, methylation, ubiquitylation. Such multisite modifications lead to change in catalytic activity, binding capability, transport or degradation of target proteins. Multisite phosphorylation is one such reversible covalent modification where a target protein is phosphorylated multiple times either by distributive or processive manner (23). When the same enzyme causes multiple phosphorylation in a distributive manner, it generates ultrasensitive signal response – a requirement for generating multistability and oscillations in signaling networks (23-25). Many crucial events in eukaryotic cell cycle are regulated by multisite phosphorylations of proteins Whi5, Cdh1, Sic1, Cdc25 by cyclin dependent kinase. Therefore, multisite phosphorylation mechanism has been widely used in mathematical and computational modeling of bistability and oscillations in cell cycle (26-30) – a fundamental property of any living organism. Therefore it is highly relevant to study the characteristics of chemical noise in a multisite phosphorylation chain. In particular study on the dependence of steady state variability of phosphorylated species on the various rate parameters of phosphorylation chain or the equilibrium constant of phosphorylation-dephosphorylation reactions.

Further, stochastic simulation algorithm (SSA) has been an obvious choice for accurate estimation of noise in chemical reaction networks of various cellular events (31). However due to the disparity in time scales of chemical reactions, particularly reactions with faster times scales tend to slow down the simulation as these reactions ‘fire’ recurrently in a given time interval. As phosphorylation/dephosphorylation of proteins usually occurs in faster time scales as compared to synthesis or degradation of proteins in living cells (32), computational cost of SSA are usually quite high in networks with multisite phosphorylation reactions. In order to improve the efficiency of SSA various methods have been developed exploiting the firing rate or the timescale separation of chemical reactions. Methods such as tau leaping, slow-scale SSA, stochastic quasi steady state approximation on enzyme catalyzed reactions are useful in this context with their own advantages and disadvantages (33-43) .

Keeping in view of the important role of multisite phosphorylation, in this paper we investigated two aspects of stochasticity in multisite phosphorylation chain. In order to investigate the quantitative and qualitative nature of variability in phosphorylated species, we used van Kampen’s system size expansion method to quantify noise at steady state. We examined the dependence of variability on the equilibrium constant of the reactions and on the total number of phosphorylation sites on the species. On account of timescale separation we applied quasi steady state approximation on the phospho chain and proposed a method to improve the computational efficiency of SSA. We applied the method on signal transduction cascade and positive feedback loop to validate the quantitative accuracy of the method.

Methods

We studied noise propagation in ordered distributive multisite phosphorylation where each enzyme-substrate encounter leads to single phosphorylation or dephosphorylation of target protein and enzymatic events happen in a specific order(26). We present a three-component reversible distributive multisite phosphorylation scheme (Figure 1a). We considered that a single enzyme catalyzes all the phosphorylation events and similarly another enzyme catalyzes all the dephosphorylation events. In addition we also assumed that the enzymes are catalyze the forward and backward reactions are in much excess concentrations compared to their substrates such that the reactions can be represented by pseudo-first order mass action rate laws. The chemical master equation that three component reaction follows can be represented as:

$$\begin{aligned} \frac{dP(n_0, n_1, n_2; t)}{dt} = & k_1(n_0 + 1)P(n_0 + 1, n_1 - 1, n_2; t) + k_2(n_1 + 1)P(n_0, n_1 + 1, n_2 - 1; t) \\ & + k_3(n_1 + 1)P(n_0 - 1, n_1 + 1, n_2; t) + k_4(n_2 + 1)P(n_0, n_1 - 1, n_2 + 1; t) \\ & - (k_1n_0 + k_2n_1 + k_3n_1 + k_4n_2)P(n_0, n_1, n_2; t) \end{aligned} \quad 1$$

$P(n_0, n_1, n_2; t)$ represents the joint probability that at time t there will be n_0 , n_1 and n_2 number of molecules for the species X_0 , X_1 and X_2 respectively. The total number of molecules of X , $n_T (= n_0 + n_1 + n_2)$, is fixed and holds a mass conservation. Due to the mass conservation, the chemical master equation can also be written as a function of any two variables of the phospho chain (44).

In order to calculate the steady state fluctuations in phospho species we used linear noise approximation on master equation also known as van Kampen's system size expansion or Ω expansion of master equation (45). Identifying Ω as the volume of the reaction chamber of the homogeneous reaction mixture, the main ansatz in the system size expansion is that the size of the fluctuations from the macroscopic average varies inversely with $\sqrt{\Omega}$. Systematic expansion of master equation to the first order in $1/\sqrt{\Omega}$ leads to macroscopic average equation. Whereas the expansion to second order in $1/\sqrt{\Omega}$ generates a linear Fokker-Planck equation of the random variable. Since in the process of generating Fokker-Planck equation, higher order terms above $1/\Omega^0$ are excluded it is called linear noise approximation. This method has been used extensively to calculate gene expression noise and noise in biochemical reaction networks (44, 46, 47).

Figure 1: Schematic representation of three (a) and four component (b) phosphorylation chains. The rate constants of individual reactions are mentioned on top of arrows indicating chemical reactions. The subscript in the name of chemical species represents the number of phosphorylation.

At the steady state the drift and diffusion matrices in the Fokker-Plank equation are connected by a fluctuation-dissipation like relation given by

$$\mathbf{A}\boldsymbol{\sigma} + \boldsymbol{\sigma}\mathbf{A}^T + \mathbf{B} = \mathbf{0} \quad 2$$

where $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\sigma}$ are the drift or Jacobian matrix, diffusion matrix and covariance matrix respectively. The covariance matrix holds information about the variance and covariance of all the molecular species in the network. The elements in the drift matrix $\mathbf{A}$ are given by

$$A_{ij} = \frac{\partial}{\partial \bar{n}_j} \frac{\partial \bar{n}_i}{\partial t} \quad 3$$

Where $\bar{n}_i$ is the average number of molecules for the i -th chemical species and can be obtained from the macroscopic rate equations (the derivative with respect to time on the right hand side). The elements in the diffusion matrix $\mathbf{B}$ are given by

$$B_{ij} = \sum_k v_{jk} v_{ik} V_k \quad 4$$

Where v_{ik} and V_k are the stoichiometric coefficient of the i -th species in the k -th reaction and the rate of the k -th reaction respectively. For the three component reaction scheme the drift and diffusion matrices are given by

$$\mathbf{A} = \begin{bmatrix} -(k_1 + k_2 + k_{-1}) & (k_{-2} - k_1) \\ k_2 & -k_{-2} \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 2(k_1 + k_{-1})\bar{n}_1 & -2k_2\bar{n}_1 \\ -2k_{-2}\bar{n}_2 & 2k_{-2}\bar{n}_2 \end{bmatrix} \quad 5$$

Using the above expressions of $\mathbf{A}$ and $\mathbf{B}$ in the fluctuation-dissipation relation (Eq.2) and applying the symmetry in covariance ($\sigma_{ij} = \sigma_{ji}$) we obtained a matrix equation for the steady state variances of stochastic variables as

$$\begin{bmatrix} p_1 & -q_1 & 0 \\ k_2 & -s_1 & q_1 \\ 0 & k_2 & -k_4 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} r_1\bar{n}_1 \\ 2k_2\bar{n}_1 \\ -k_{-1}\bar{n}_2 \end{bmatrix} \quad 6$$

Where $p_1 = k_1 + k_2 + k_{-1}$, $q_1 = k_{-2} - k_1$, $r_1 = k_1 + k_{-1}$ and $s_1 = k_1 + k_2 + k_{-1} + k_{-2}$. Note that as we have applied the mass conservation relation, the Eq.6 does not contain the variability term for the unphosphorylated species X_0 . Solution of the above system of linear equations lead to the variances of species X_1 and X_2 as

$$\sigma_{11} = \frac{k_2 q_1 (q_1 - r_1 - 2k_{-2}) \bar{n}_1 + k_{-2} r_1 s_1 \bar{n}_1}{s_1 (p_1 k_{-2} - k_2 q_1)} \quad 7$$

$$\sigma_{22} = \frac{k_{-2} (p_1 s_1 - k_2 q_1) \bar{n}_2 + k_2 k_{-2} (r_1 - 2p_1) \bar{n}_2}{s_1 (p_1 k_{-2} - k_2 q_1)} \quad 8$$

In the special case where all the forward and backward rate constants are equal i.e. $k_1 = k_2 = k_f$ and $k_{-1} = k_{-2} = k_b$, the steady state variances can be represented as a function of equilibrium constant ($K = k_f/k_b$) and they take simple form as

$$\sigma_{11} = \frac{(1+K^2) \bar{n}_1}{(1+K)^2 - K} \text{ and } \sigma_{22} = \frac{(1+K) \bar{n}_2}{(1+K)^2 - K} \quad 9$$

We proceed further to calculate steady state variances in a four-component equilibrium (Figure 1b). We followed the similar procedure as described for the three-component system. The Jacobian matrix is given by

$$\mathbf{A} = \begin{bmatrix} -(k_1 + k_2 + k_{-1}) & (k_{-2} - k_1) & -k_1 \\ k_2 & -(k_3 + k_{-2}) & k_{-3} \\ 0 & k_3 & -k_{-3} \end{bmatrix} \quad 10$$

and the diffusion matrix is given by

$$\mathbf{B} = \begin{bmatrix} 2(k_2 + k_{-1}) \bar{n}_1 & -(k_2 \bar{n}_1 + k_{-2} \bar{n}_2) & 0 \\ -(k_2 \bar{n}_1 + k_{-2} \bar{n}_2) & 2(k_3 + k_{-2}) \bar{n}_2 & -2k_3 \bar{n}_2 \\ 0 & -2k_{-3} \bar{n}_3 & 2k_{-3} \bar{n}_3 \end{bmatrix} \quad 11$$

Making use of the fluctuation-dissipation relation (Eq. 2) we again obtained a matrix equation for the covariance as

$$\begin{bmatrix} -p_2 & q_2 & -k_1 & 0 & 0 & 0 \\ k_2 & -(p_2 + r_2) & k_{-3} & q_2 & -k_1 & 0 \\ 0 & k_3 & -(p_2 + k_{-3}) & 0 & q_2 & -k_1 \\ 0 & k_2 & 0 & -r_2 & k_{-3} & 0 \\ 0 & 0 & k_2 & k_3 & -(r_2 + k_{-3}) & k_{-3} \\ 0 & 0 & 0 & 0 & k_3 & -k_{-3} \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{21} \\ \sigma_{22} \\ \sigma_{33} \end{bmatrix} = \begin{bmatrix} -(k_2 + k_{-1})\bar{n}_1 \\ (k_2\bar{n}_1 + k_{-2}\bar{n}_2) \\ 0 \\ -(k_3 + k_{-2})\bar{n}_2 \\ 2k_3\bar{n}_2 \\ -k_{-3}\bar{n}_3 \end{bmatrix} \quad 12$$

Solution of these linear equations provide the steady state variance of phospho species X_1 - X_3 . We have analytically solved these simultaneous linear equations to get the steady state variances. However due to complex nature of the solutions we prefer not to present them for the general case. However, in the special case where all the forward and backward rate constants are equal ($k_1 = k_2 = k_3 = k_f$ and $k_{-1} = k_{-2} = k_{-3} = k_b$), the variances of X_1 , X_2 and X_3 can be expressed as a function of K as

$$\sigma_{11} = \frac{[Q_3Q_2 + K\{2(K-1)(5K+3) - K(K-1)(K^2-1) - K^2(K+1)(K^2+1) + 3K^2(K+1)(2K+1) + (K-1)(K^2-1)(K^2+4)\}]\bar{n}_1}{Q_1Q_2 + K\{K(2K+1)(K^2+4) + 3(2K+1)(K^2-2) - K^2(K^2-2)\}} \quad 13$$

$$\sigma_{22} = \frac{[Q_1Q_2 + K\{K^2(K^2+1) - K(K^2+4) - 6(2K+1)\}]\bar{n}_2}{Q_1Q_2 + K\{K(2K+1)(K^2+4) + 3(2K+1)(K^2-2) - K^2(K^2-2)\}} \quad 14$$

$$\sigma_{33} = \frac{[Q_1(K^2 + K + 4) + K\{2(2K-3)(2K+1) + K^2(K-1)\}]\bar{n}_3}{Q_1Q_2 + K\{K(2K+1)(K^2+4) + 3(2K+1)(K^2-2) - K^2(K^2-2)\}} \quad 15$$

where $Q_1 = 2(K+1)(2K+1) + 3K^2$, $Q_2 = K^3 + K^2 + K + 4$ and $Q_3 = (K+1)(3K+2) + 2K(K-1)$.

Although we have calculated only the variances using Eq. 12 one can also calculate covariances. Further the macroscopic averages can be obtained from the deterministic dynamical equations. The general expression of the average for i -th phospho state is given by (26)

$$\bar{n}_i = \frac{(k_f/k_b)^i}{\sum_{i=0}^N (k_f/k_b)^i} n_T \quad 16$$

N is the total number of phosphorylation site in the chemical species.

Results

In Figure 2 we present the dependence of steady state variance of phosphorylated species with the variation of rate constant of phosphorylation (left panel), dephosphorylation (middle) and equilibrium constant of phosphorylation-dephosphorylation reactions (right panel) for three (top row) and four (bottom row) component phospho chains. In case of three- and four-component chain, we used Eq. 9 and Eq. 13-15, respectively to get the equilibrium constant dependence on variance. To obtain the individual rate constant dependence for three-component chain we used Eq. 7-8 and for four-component chain we solved the matrix equation (Eq. 12) using Matlab. The variance increases sharply with increase of equilibrium constant and passing through a maximum it decreases with a long tail. The value of K at which the variance is maximum increases with the number of phosphorylation. Further comparisons with the numerical simulations of the chain using SSA show complete agreement of analytical and numerical calculations.

Figure 2: Dependence of steady state variance on the rate constants of chemical equilibria in phosphorylation chain. Solid lines and open circles represent analytical and numerical results respectively for three- (upper row) and four-component (bottom row) chains. Numerical results were from stochastic simulations of chain using SSA. Other numerical parameters were: for k_1 variation

$k_2 = k_3 = 1$ and $k_{-1} = k_{-2} = k_{-3} = 0.5$; for k_{-1} variation $k_1 = k_2 = k_3 = 1$ and $k_{-2} = k_{-3} = 0.5$ and $n_T = 300$.

Next we extended linear noise approximation calculation to five- and six-component phospho chains. Following the similar methodology as described for the previous cases we generated matrix equations for the covariance vector at the steady state. Applying the symmetry condition $\sigma_{ij} = \sigma_{ji}$ the sizes of σ vectors for five- and six-component chains were 10 and 15, respectively. Therefore the resulting matrix equations were quite big and hard to solve analytically (see Supplementary text). To get the values of σ_{ii} we solved these matrix equations using Matlab. We present the comparisons of analytical and numerical results of steady state variance in Figure 3. We further calculated coefficient of variation ($CV = \sqrt{\sigma_{ii}} / \bar{n}_i$) for all the phospho species and compared with numerical results. CV provides information about the ‘noisiness’ exhibited by the chemical species in a network. Our calculations indicate that both for five- and six-component chains, below a critical value of equilibrium constant ($K = 1$) the noise increases with the number of phosphorylations and above the critical value the trend is reverse. Similar behavior was also obtained for three- and four-component chains (not shown). Therefore the crossover nature of noise with the variation of K is independent of the total number of phospho species in the chain. The crossover was due to the shift in distribution of abundance of phospho species with variation of equilibrium constant (26).

Figure 3: Comparison of steady state variance and CV obtained from analytical calculations (solid line) and numerical simulations (circle) for five- (upper row) and six-component (bottom row) phosphorylation chains. The total population (n_T) of X was 300 in both the cases.

Many proteins in living cells are phosphorylated multiple times to generate threshold and ultrasensitive signal response in protein activity (48-50). Furthermore in a chain increased number of phospho states are known to cause sharper signal response curve with higher Hill coefficient values (23, 26). Keeping in mind this role of multi-phosphorylation, next we estimated the effect of increased phospho states on the noise of chemical species in a chain. In this regard we compared the analytically calculated CV of a given phospho-state (X_i) from chains with different total number of phosphorylation sites or in other words the 'length' of the chain (Figure 4) keeping n_T fixed. We found that for every phospho species the variability increased with increase of total number of phosphorylation sites. Therefore although higher number of phospho states are known to increase the ultrasensitive nature of signal response curve however our calculations indicated that it may lead to increase of variability in response. As we kept the total population (n_T) same for all chains, with the increase of the number of states the average abundance of every phospho state decreased resulting in increased noise. Therefore a chain with more number of states may be noisier than a chain with less number of states. To determine the overall variability of the phospho states, we next calculated the total variance of all the phospho species by

summing up their variances and found that the total variance increased with increase in the number of components in the chain (Figure 5).

Figure 4: Comparison of noise in phosphorylated species from chains of different total number of phosphorylation states with $n_T = 300$ for every chain. Different colors or line styles represent phospho chains with 3, 4, 5 and 6 number of components. The phospho states are indicated inside each plot.

Figure 5: Variation of the total variance ($=\sum \sigma_{ii}$) with the equilibrium constant. Each color or line style represents chain with different total number of phosphorylation sites.

Application: quasi steady state approximation

Population abundance heterogeneity and disparity in timescales in coupled chemical reactions often pose a significant challenge towards computational efficiency of SSA (51). For a coupled chemical reaction network involving multi-phosphorylation chain, we applied quasi steady state approximation on the phospho chain under the condition of time scale separation between the reactions in the chain and the rest of the network. Specifically when the reactions in phospho chain occur in much faster timescale as compared to other reactions in the network, it allows the conditional probability distribution of fast variables to reach a quasi steady state in the timescale of other reactions. This is stochastic analogue of deterministic quasi steady state approximation (52). Under this condition the conditional probability distribution reaches a stochastic quasi steady state (36, 39, 40)

$$\frac{dP(\mathbf{n}_f|\mathbf{n}_s)}{dt} \approx 0$$

Where $\mathbf{n}_f$ and $\mathbf{n}_s$ are the population of fast and slow variables, respectively. The QSSA assumption on the probability leads to reduction in dimension of the reacting system (40, 53). Thus the reactions evolving in slower time scales are simulated using SSA where the propensities of slow reactions depend parametrically on the steady state average of the fast variables obtained from the steady state solution of their macroscopic dynamical equations. However as the values of the fast variables are taken from the deterministic equations, they lack the effect of intrinsic noise of chemical reactions. Therefore in order to include the effect of intrinsic noise in the fast variables, at each instant to the deterministic average of fast variables we added Gaussian white noise terms. The added noise was sampled from a Gaussian distribution with zero mean and the variance was taken from the analytical solution provided by the linear noise approximation of the chain. For example in a network with four-component phospho chain, for the species X_3 we added a zero mean Gaussian noise having variance given by σ_{33} in Eq. 15. The result of linear noise approximation is a Fokker-Planck equation having a solution of Gaussian form for the probability distribution of random variables. Therefore adding a Gaussian noise with the variance given by analytical solution of fluctuation-dissipation relation is justified. The steps of modified QSSA were as follows:

1. Initialized of all chemical species.

2. Calculated of steady state deterministic average ($\bar{n}_i$) and variance (σ_{ii}) of phospho species using the analytical expressions given by linear noise approximation.
3. Added Gaussian noise to the deterministic average: $\bar{n}_i + \eta_i(t) (= n'_i)$ where $\eta(t)$ is a Gaussian white noise having statistical properties: $\langle \eta_i(t) \rangle = 0$ and $\langle \eta_i(t)\eta_i(t') \rangle = \sqrt{\sigma_{ii}}\delta(t - t')$.
4. Calculated the propensities of all chemical reactions in the reduced system i.e. excluding the reactions of the chain; calculated next reaction time interval following SSA method and updated the time.
5. Determined the next reaction and updated the species abundance as per the stoichiometry of the chemical reaction occurred.
6. Repeated the process from step 2.

Using this recipe we applied our proposed method in linear cascade signal transduction network and positive feedback loop – two commonly found reaction motifs in living cells.

We first applied QSSA method to a signal transduction cascade network where information flows only in one direction without any feedback loop. The network diagram is given in Figure 6a. A four-component phospho chain was embedded in the network where the species Y acted as the enzyme catalyzing phosphorylation reactions in the chain and only the terminal phosphorylated species, X_3 , was assumed to catalyze synthesis of the species Z. The full network consisted of 10 chemical reactions and upon QSSA approximation the reduced system consisted only 4 chemical reactions associated with the synthesis and degradation of Y and Z (Table 1). At each time for a given value of Y, average value of X_3 was calculated using the relation given by Eq. 16. A Gaussian random number with zero mean was added to the average of X_3 where the variance of the random noise was given by the Eq. 15. Further, to investigate the effect of different total number of X on the variability, the zeroth and second order rate constants were multiplied and divided by n_T , respectively leading to the appropriate scaling of the system. Note that in the average and variance equation (Eq. 15 and Eq. 16) k_f was replaced by $k_f n_Y / n_T$ due to the second order nature of phosphorylation reaction. In Figure 6b we present a comparison between steady state CV of X_3 and Y calculated using full SSA and QSSA methods for different values of synthesis rate of Y (k_s). We further extended the comparison for different values of the equilibrium constant of the chain (Figure 6b-d) by varying k_b keeping k_f constant. We found that that in all cases the QSSA accurately recaptures the noise in all the regulators in the network. The accuracy of QSSA method was limited if the Gaussian noise was not added to the average of X_3 (Supplementary Figure S1). Further in order to quantitatively estimate the accuracy of the QSSA method, we estimated a deviation

error by calculating the sum-squared-deviation ($=\sum_i (x_{i,QSSA} - x_{i,SSA})^2$). The deviation error in CV was quite less ($<10^{-3}$) both for X_3 and Y (Figure 6e).

Figure 6: a) Schematic diagram of signal transduction cascade model where Y catalyzes phosphorylation of X and X_3 catalyzes production of Z . Chemical reaction and catalytic activity to a reaction are represented by solid and dashed arrows, respectively. (b-d) Comparison of CV from full SSA (solid line) and QSSA (circle or square) simulations for X_3 (circle) and Y (square) for indicated values of K . (e) Error in CV, defined as sum-squared-deviation QSSA from full SSA, for different values of K . The parameter values used were: $\gamma = 0.1$, $k_Z = 0.5$, $k_f = 10$ and $n_T = 50$. To vary the equilibrium constant different values of $k_b (= 10, 5, 1)$ were chosen.

Table 1: List of chemical reactions in the reduced system of signal transduction cascade network. $\eta_3(t)$ is a Gaussian random number following $\langle \eta_3(t) \rangle = 0$ and $\langle \eta_3(t)\eta_3(t') \rangle = \sqrt{\sigma_{33}}\delta(t - t')$.

Reaction	Propensity
$\rightarrow Y$	$k_s n_T$
$Y \rightarrow$	γn_Y
$\rightarrow Z$	$k_Z n'_3$ where $n'_3 = \bar{n}_3 + \eta_3(t)$

$$Z \rightarrow$$

$$\gamma n_Z$$

We next applied the QSSA method to a positive feedback loop network that generate bistable signal response. In living cells bistability is quite a common feature of chemical reaction network with positive feedback loop. All-or-none type of response generated by a bistable switch helps cell in making various decision making processes in proliferation, death and differentiation (54). Previously multisite phosphorylation in a positive feedback loop has been predicted to generate bistability even with mass action reaction rates in the chain (26). Here, as in previous network Y acted as kinase to X and the terminally phosphorylated X_3 , assumed to be active form of X, catalyzed synthesis of Y thereby creating a positive feedback loop between X and Y (Figure 7a). On application of QSSA on the chain, the reduced system consisted only two reactions associated with the synthesis and degradation of Y with propensities $k_s n_T + k_Y n'_3$ and γn_Y , respectively. We made a comparison of steady state population of Y generated from full SSA and QSSA of the model while the bifurcation parameter, the unregulated synthesis rate constant of Y (k_s), was varied (Figure 7b-c). Here the steady state distribution of Y is the stochastic analogue of the deterministic one-parameter bifurcation diagram that was overlaid on top of the population distributions. We made the comparison initializing the simulations at the lower (Figure 7b) and the upper steady states (Figure 7c) of the deterministic bifurcation diagram and in both cases results from QSSA simulations were almost identical to the full SSA. A full comparison of steady state distribution of Y across the bifurcation diagram can be found in the Supplementary Figure S2. The distributions obtained from QSSA simulations were quite similar to the distributions from SSA simulations. For quantitative comparison of noise, next we calculated average and CV of Y for three different values of n_T that scales the abundances of all the species without shifting the bifurcation diagram in x-axis (Figure 8). Again we found that QSSA method was able to capture noise accurately in all three cases. The comparison at different values of n_T highlights the important fact that even with low population abundance ($n_T = 50$), the QSSA method was able to capture the noise accurately. Further the accuracy of QSSA does not depend on the initial conditions used in the simulations of bistable switch.

Figure 7: Schematic diagram of a positive feedback loop involving 4-component multi-phosphorylation chain (a). Comparison of steady state populations of Y from full SSA and QSSA simulations initialized from lower (b) and upper (c) steady states of deterministic bifurcation diagram. The sloid line represents one-parameter bifurcation diagram of the model. The stable and unstable steady states are represented by black and red lines, respectively. The parameter values used were: $\gamma = 0.1$, $k_Y = 0.7$, $k_f = 1$, $k_b = 2$ and $n_T = 50$.

So far we assumed that the terminally phosphorylated species X_3 was the only active form X. However in many protein regulatory networks two or more phospho species of same protein may act as active form of the protein. Therefore in order to investigate the effect of other phospho species on the accuracy of QSSA, next we assumed that both X_3 and X_2 are capable of catalyzing the production of Y. Now in this case the propensity of production of Y was given by $k_s n_T + k_Y n'_3 + k'_Y n'_2$. A comparison of full SSA and QSSA showed that in this case also QSSA accurately recaptures the noise in Y (Figure 9). Note that the region of bistability in this case was a bit smaller (Supplementary Figure S3) as compared to the previous case due to decreased ultrasensitivity in the chain (26). Further to investigate the effect of time scale separation, we performed full SSA simulation of the model with reduced values of both k_f and k_b keeping the equilibrium constant of the chain same. We found that by slowing down the reactions in the chain, the noise in Y changes significantly both for simulations initialized at the upper and lower (Supplementary Figure S4) steady states. Therefore the accuracy of QSSA would critically depend on the timescale separation which requires faster k_f and k_b in the chain of phosphorylation-dephosphorylation reactions.

Figure 8: Comparison of steady state average and CV of Y for the indicated values of n_T . Upper and lower panels are results from simulations initialized at the upper and lower steady states of the bifurcation diagram, respectively. The parameter values are same as Figure 7.

Figure 9: Upper row: Comparison of CV from full SSA and QSSA for the positive feedback network where both X_3 and X_2 catalyzes production of Y . Parameter values were: $k_Y = 0.5$, $k'_Y = 0.25$ and others were same as Figure 7.

Conclusion

Chemical reactions in living cells experience unavoidable chemical noise resulting in cell-to-cell variability of cellular functions or characteristics. The origin of cell-to-cell variability was found to be the gene expression noise whose propagation critically depends on the topology of network. Multisite phosphorylation of protein is often responsible for activation, inactivation, recognition, binding, degradation of target proteins. Further it is also responsible for ultrasensitive signal response in signaling network. Therefore understanding of chemical noise propagation in a chain of multisite phosphorylation reactions is important due to its vast role in cellular signaling. In this paper we analytically investigated chemical noise propagation in a multisite phosphorylation chain using van Kampen's system size expansion method. We calculated steady state variance of phosphorylated species while we varied the equilibrium constant of phosphorylation-dephosphorylation reaction and the total number of phosphorylated states in the chain or the chain length. The variability or noise in phospho species, measured by the coefficient of variation, showed a crossover at a critical value of equilibrium constant ($K = 1$). Below the critical value the variability increases with increasing number of phosphorylations whereas above the critical value the variability decreases with increasing number of phosphorylations. This behavior is consistent for chains of different lengths. The noise at a given phosphorylated state depends on the total number of phosphorylated states in the chain – noise increases with increasing number of states in the chain. Further we found that the total variance of all the phosphorylated species increases with increasing number of states in the chain. Previous studies have shown that more phosphorylated states (or 'longer' chain) lead to increased ultrasensitive nature of signal response curve(25, 26), however our calculations indicate that it may result in increase of variability in the response.

Next we made use of the steady state variability given by linear noise approximation to propose a recipe for improving computational efficiency of stochastic simulation algorithm applying a quasi steady state approximation on the phosphorylation chain assuming a timescale separation between the phospho chain and the rest of the network. The key ingredient in the method is that a zero-mean Gaussian white

noise was added to the steady state average of phospho species in order to include the effect of inherent stochasticity of phosphorylation-dephosphorylation chain. The variance of Gaussian noise was given by the analytical solution of fluctuation-dissipation relation from linear noise approximation. We applied the recipe to a signaling cascade and to a positive feedback loop network motifs. We found that under the condition of timescale separation our recipe estimated the average and the noise accurately in these systems. As our analytical method correctly estimated the noise in the phospho species, the calculated variability in the downstream signaling was accurate in QSSA method. As application of QSSA lead to dimensional reduction of the system, the number of reactions fired in a given time interval reduced significantly. For example in the positive feedback loop network with a total of 50 molecules in the chain ($n_T = 50$) the average number of reactions fired were $\sim 3.6 \times 10^5$ and $\sim 2.2 \times 10^6$ for QSSA and full SSA methods respectively till a simulation time of 10000 arbitrary unit in the bistable region ($k_s = 0.035$). In the same situation with 100 total molecules in the chain, the number of fired reactions were $\sim 8.3 \times 10^5$ and $\sim 4.6 \times 10^6$ for QSSA and full SSA methods, respectively. Due to the less number of reactions the QSSA method was computationally inexpensive as compared to the full SSA method. In the mentioned cases QSSA method took about 3 and 7 minutes for 50 and 100 molecules in the chain, respectively. Whereas full SSA method spent about 9 and 17 minutes, respectively. Therefore the proposed recipe lowers the computational cost significantly without compromising the accuracy of calculation. Particularly the method will be quite useful for networks where the simulation spends most of the time in simulating fast reactions.

Acknowledgments

The work was supported by funding from the Science and Engineering Research Board, Department of Science and Technology (India), grant no. EMR/2015/001899, to DB. SD acknowledges fellowship from INSPIRE program of Department of Science and Technology, India.

References

1. 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.
2. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic Gene Expression in a Single Cell. *Science* 297:1183.
3. Nelson, D. E., A. E. C. Ihekweba, M. Elliott, J. R. Johnson, C. A. Gibney, B. E. Foreman, G. Nelson, V. See, C. A. Horton, D. G. Spiller, S. W. Edwards, H. P. McDowell, J. F. Unitt, E. Sullivan, R. Grimley, N. Benson, D. Broomhead, D. B. Kell, and M. R. H. White. 2004. Oscillations in NF- κ B Signaling Control the Dynamics of Gene Expression. *Science* 306:704-708.
4. Geva-Zatorsky, N., N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon. 2006. Oscillations and variability in the p53 system. *Molecular Systems Biology* 2.
5. Di Talia, S., J. M. Skotheim, J. M. Bean, E. D. Siggia, and F. R. Cross. 2007. The effect of molecular noise and size control on the variability in the budding yeast cell cycle. *Nature* 448:947-951.
6. Feinerman, O., J. Veiga, J. R. Dorfman, R. N. Germain, and G. Altan-Bonnet. 2008. Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science* 321:1081.
7. Cerulus, B., Aaron M. New, K. Pougach, and Kevin J. Verstrepen. 2016. Noise and Epigenetic Inheritance of Single-Cell Division Times Influence Population Fitness. *Current Biology* 26:1138-1147.
8. Raj, A., and A. van Oudenaarden. 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* 135:216-226.
9. Eldar, A., and M. B. Elowitz. 2010. Functional roles for noise in genetic circuits. *Nature* 467:167-173.
10. Ackermann, M. 2015. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology* 13:497.
11. Jolly, M. K., P. Kulkarni, K. Weninger, J. Orban, and H. Levine. 2018. Phenotypic Plasticity, Bet-Hedging, and Androgen Independence in Prostate Cancer: Role of Non-Genetic Heterogeneity. *Frontiers in Oncology* 8.
12. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* 99:12795-12800.
13. Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* 98:8614-8619.
14. Pedraza, J. M., and J. Paulsson. 2008. Effects of molecular memory and bursting on fluctuations in gene expression. *Science* 319:339-343.
15. Jia, T., and R. V. Kulkarni. 2011. Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting. *Physical Review Letters* 106:058102.
16. Tyson, J. J., K. C. Chen, and B. Novak. 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology* 15:221-231.
17. Dey, A., and D. Barik. 2017. Parallel arrangements of positive feedback loops limit cell-to-cell variability in differentiation. *PLOS ONE* 12:e0188623.
18. Barik, D., D. A. Ball, J. Peccoud, and J. J. Tyson. 2016. A Stochastic Model of the Yeast Cell Cycle Reveals Roles for Feedback Regulation in Limiting Cellular Variability. *PLOS Computational Biology* 12:e1005230.

19. Pedraza, J. M., and A. van Oudenaarden. 2005. Noise propagation in gene networks. *Science* 307:1965–1969.
20. Brandman, O. 2005. Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions. *Science* 310:496-498.
21. Bruggeman, F. J., N. Blüthgen, and H. V. Westerhoff. 2009. Noise Management by Molecular Networks. *PLOS Computational Biology* 5:e1000506.
22. Chepyala, S. R., Y.-c. Chen, C.-c. S. Yan, and C.-y. D. Lu. 2016. Noise propagation with interlinked feed-forward pathways. *Nature Publishing Group*:1-15.
23. Gunawardena, J. 2005. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proceedings of the National Academy of Sciences of the United States of America* 102:14617-14622.
24. Markevich, N. I., J. B. Hoek, and B. N. Kholodenko. 2004. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of Cell Biology* 164:353-359.
25. Salazar, C., and T. Höfer. 2009. Multisite protein phosphorylation: From molecular mechanisms to kinetic models. *FEBS Journal* 276:3177-3198.
26. Kapuy, O., D. Barik, M. R. Domingo Sananes, J. J. Tyson, and B. Novak. 2009. Bistability by multiple phosphorylation of regulatory proteins. *Prog. Biophys. Mol. Biol.* 100:47-56.
27. Barik, D., W. T. Baumann, M. R. Paul, B. Novak, and J. J. Tyson. 2010. A model of yeast cell cycle regulation based on multisite phosphorylation. *Molecular Systems Biology* 6.
28. Yao, G., T. J. Lee, S. Mori, J. R. Nevins, and L. You. 2008. A bistable Rb-E2F switch underlies the restriction point. *Nat Cell Biol* 10:476-482.
29. Ferrell, J. E., Jr, and S. H. Ha. 2014. Ultrasensitivity part II: multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends in Biochemical Sciences* 39:556-569.
30. Yang, L., W. R. MacLellan, Z. Han, J. N. Weiss, and Z. Qu. 2004. Multisite Phosphorylation and Network Dynamics of Cyclin-Dependent Kinase Signaling in the Eukaryotic Cell Cycle. *Biophysical Journal* 86:3432-3443.
31. Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22:403-434.
32. Kholodenko, B. N. 2006. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology* 7:165.
33. Cao, Y., D. T. Gillespie, and L. R. Petzold. 2005. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics* 122:014116.
34. Cao, Y., D. T. Gillespie, and L. R. Petzold. 2006. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics* 124:044109.
35. Auger, A., P. Chatelain, and P. Koumoutsakos. 2006. R-leaping: Accelerating the stochastic simulation algorithm by reaction leaps. *The Journal of Chemical Physics* 125:084103.
36. Wu, S., J. Fu, and L. R. Petzold. 2015. Adaptive deployment of model reductions for tau-leaping simulation. *The Journal of Chemical Physics* 142:204108.
37. Pu, Y., L. T. Watson, and Y. Cao. 2011. Stiffness detection and reduction in discrete stochastic simulation of biochemical systems. *The Journal of Chemical Physics* 134:054105.
38. Goutsias, J. 2005. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *The Journal of Chemical Physics* 122:184102.
39. Rao, C. V., and A. P. Arkin. 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *The Journal of Chemical Physics* 118:4999-5010.
40. Barik, D., M. R. Paul, W. T. Baumann, Y. Cao, and J. J. Tyson. 2008. Stochastic Simulation of Enzyme-Catalyzed Reactions with Disparate Timescales. *Biophysical Journal* 95:3563-3574.

41. Kim, Jae K., K. Josić, and Matthew R. Bennett. 2014. The Validity of Quasi-Steady-State Approximations in Discrete Stochastic Simulations. *Biophysical Journal* 107:783-793.
42. Kim, J. K., and E. D. Sontag. 2017. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. *PLOS Computational Biology* 13:e1005571.
43. Herath, N., and D. D. Vecchio. 2018. Reduced linear noise approximation for biochemical reaction networks with time-scale separation: The stochastic tQSSA+. *The Journal of Chemical Physics* 148:094108.
44. Paulsson, J. 2005. Models of stochastic gene expression. *Physics of Life Reviews* 2:157-175.
45. Van Kampen, N. G. 2007. Chapter X - THE EXPANSION OF THE MASTER EQUATION. In *Stochastic Processes in Physics and Chemistry (Third Edition)*. Elsevier, Amsterdam. 244-272.
46. Elf, J., and M. Ehrenberg. 2003. Fast Evaluation of Fluctuations in Biochemical Networks With the Linear Noise Approximation. *Genome Research* 13:2475-2484.
47. Hayot, F., and C. Jayaprakash. 2004. The linear noise approximation for molecular fluctuations within cells. *Physical Biology* 1:205.
48. Ferrell, J. E. 1996. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends in Biochemical Sciences* 21:460-466.
49. Deshaies, R. J., and J. E. Ferrell. 2001. Multisite Phosphorylation and the Countdown to S Phase. *Cell* 107:819-822.
50. Nash, P., X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, and M. Tyers. 2001. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 414:514.
51. Gillespie, D. T. 2007. Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry* 58:35-55.
52. Segel, L. A., and M. Slemrod. 1989. The Quasi-Steady-State Assumption: A Case Study in Perturbation. *SIAM Review* 31:446-477.
53. MacNamara, S., A. M. Bersani, K. Burrage, and R. B. Sidje. 2008. Stochastic chemical kinetics and the total quasi-steady-state assumption: Application to the stochastic simulation algorithm and chemical master equation. *The Journal of Chemical Physics* 129:095105.
54. Tyson, J. J., and B. Novák. 2010. Functional Motifs in Biochemical Reaction Networks. *Annual Review of Physical Chemistry* 61:219-240.