

RESEARCH

Open Access

Efficient calculation of steady state probability distribution for stochastic biochemical reaction network

Shahriar Karim^{1,2}, Gregory T Buzzard³, David M Umulis^{1*}

From IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS) 2011 San Antonio, TX, USA. 4-6 December 2011

Abstract

The Steady State (SS) probability distribution is an important quantity needed to characterize the steady state behavior of many stochastic biochemical networks. In this paper, we propose an efficient and accurate approach to calculating an approximate SS probability distribution from solution of the Chemical Master Equation (CME) under the assumption of the existence of a unique deterministic SS of the system. To find the approximate solution to the CME, a truncated state-space representation is used to reduce the state-space of the system and translate it to a finite dimension. The subsequent ill-posed eigenvalue problem of a linear system for the finite state-space can be converted to a well-posed system of linear equations and solved. The proposed strategy yields efficient and accurate estimation of noise in stochastic biochemical systems. To demonstrate the approach, we applied the method to characterize the noise behavior of a set of biochemical networks of ligand-receptor interactions for Bone Morphogenetic Protein (BMP) signaling. We found that recruitment of type II receptors during the receptor oligomerization by itself doesn't tend to lower noise in receptor signaling, but regulation by a secreted co-factor may provide a substantial improvement in signaling relative to noise. The steady state probability approximation method shortened the time necessary to calculate the probability distributions compared to earlier approaches, such as Gillespie's Stochastic Simulation Algorithm (SSA) while maintaining high accuracy.

Introduction

Many biological networks exhibit stochasticity due to a combinatorial effect of low molecular concentrations and slow system dynamics. One important biological context where stochastic events likely have a large impact is the Bone Morphogenetic Protein (BMP) signaling pathway. BMPs make up the largest subfamily of the Transforming Growth Factor- β superfamily and are involved in numerous processes including growth, differentiation and diseases [1]. Due to their potency at driving development, they are also of great value for stem-cell differentiations in cell culture. BMPs activate near maximal signaling at $1nM$ concentration, have very slow binding kinetics and require oligomerization

between multiple receptor subunits [1]. These properties naturally lead to conditions for significant and long-duration stochastic fluctuations in cellular signaling. Interestingly, variability of BMP signaling appears to be very low *in vivo*, while it is very high in stem cell culture studies [2]. To understand the differences between *in vivo* and *in vitro* signaling and determine how various receptor oligomerization events might alter the signal and noise, a more efficient means of solving the steady state distributions for stochastic model was needed that would allow for continuation of both parameters and levels of the BMP pathway components.

Stochastic regulation can negatively impact the robustness of the system [3,4] or instead, constructively contribute to the phenotypic variation [5-7] in a species. In stochastic reaction networks, the state of a species traverses different trajectories in a probabilistic manner and the distributions of states can be difficult to predict.

* Correspondence: dumulis@purdue.edu

¹Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, USA

Full list of author information is available at the end of the article

As more biological data is available, stochastic modeling is becoming increasingly popular to estimate properties in networks where the time evolution of the system is unpredictable and dependent on unavoidable randomness inherent to the system. The complete solution can be calculated from a Chemical Master Equation (CME) [8-10], that is based on a Markovian approach that captures the inherent randomness of biochemical systems.

The Chemical Master Equation (CME) describes the dynamics of the probability distribution of a species of chemical reactions. Precisely, the CME captures the rate of change of probability that a system will be in state \mathbf{X} at time t for all the species of the system. Solution of the CME is practically intractable due to the curse of dimensionality, as the state-space of the system becomes enormously large with increases in the species number and concentrations (number of states n^N , for $N \rightarrow$ species, $n \rightarrow$ copies of each species). Moreover, the system often involves interactions between different time-scales (slow and fast reactions, frequent and infrequent transitions between states) [11], which add further complexity. Instead, numerical approaches are commonly used [12-14] to realize the CME of the stochastic system.

In the analysis of stochastic biochemical networks, steady state probability distributions for each species in the system are determined to measure variability about the deterministic steady state. The deviation around the solution contributes to stochastic noise that can be quantified by measuring the coefficient of variation $\Lambda = \frac{\sigma}{\mu}$ (the ratio between the standard deviation σ and the mean level of species concentration μ) obtained by solving the CME [9].

Frequently, Monte-carlo based simulation approaches [9,13] are used to solve stochastic problems. But, there are drawbacks in this approach for networks where the dynamics of different intermediate states of the system are unknown and continuation of several parameters is necessary to explore the system's dependency. As a screen of parameter values becomes necessary in such a scenario, the Monte-carlo based approach doesn't prove to be efficient, as it generally takes longer time to numerically simulate the process and satisfy the imposed conditions. Moreover, simulation times increase with increases in the total number of molecules, species and the number of interactions between species.

In order to ameliorate the computational cost and complexity, we present a method here to approximate the steady state probability distribution by 1) reducing the system's state-space to a finite dimension using truncated state-space method [15] and 2) subsequently, translating an eigenvalue problem associated with a CME to a system of linear equations. We illustrate that the eigenvector (for an eigenvalue = 0) that represents

the steady state probability distribution can be solved algebraically by approximating it as a system of linear equations. Previously, the influence of stochastic fluctuations on system behavior was studied also in [16], where a moment closure approach was applied to reduce the complexity associated with the identification of distributions. In contrast to the previous studies, here we use a truncated state-space for steady state probability distribution approximation, which is arguably more general since we make no assumptions on the relationships of the moments of the distribution.

The usefulness of the proposed method is illustrated by considering the example networks of BMP signaling, as described earlier in [17]. Here we examine two potential mechanisms of BMP signaling: 1) regulation between the type I and type II receptors, and 2) regulation by secreted co-factors, such as Crossveinless-2 (Cv-2). First, we apply the approach to the recruitment of Type II receptor into a BMP-bound type I receptor complex to see if such step of receptor oligomerization contributes qualitatively to the noise profile of the system. Following this, we extend our earlier work that focused on extracellular regulation of BMP signaling by SBPs to evaluate the calculation approach and compare results to the Type I/Type II receptor system.

Unlike SBPs, which tend to improve the signal to noise ratio, we did not see a significant change in stochastic variability with inclusion of Type II receptors. This result supports a previous assumption made in [4] where the recruitment of Type II receptor was excluded to simplify the modeling steps while characterizing the noise profile of a SBP regulated BMP signaling system.

Methods

Proposed approximation method

The Chemical Master Equation (CME), which is a set of first order differential (ODE) equations, demonstrates loss and gain of probabilities of discrete states of a system [10] and is often applicable to represent the stochasticity of the system. Consider a well-stirred system at thermal equilibrium of N different species $\{S_1, S_2, \dots, S_N\}$ with $\{X_1, X_2, \dots, X_N\}$ molecules respectively, participating in a total of M biochemical reactions R_μ , where $\mu = 1, 2, \dots, M$. The state of such a system is represented by the copy number (X_n molecules) of each species (S_n) at any given time t and is represented as $\mathbf{X} = [X_1(t), X_2(t), \dots, X_N(t)]$. Unless a non-zero initial state is assigned, the default initial species concentrations are always zero ($X_n(t=0) = 0$, where $1 \leq n \leq N$).

Two other quantities are further required to construct the system: 1) a state-change vector v_μ and 2) propensity functions [8,12,14,18] for the reactions R_μ , $\mu = \{1, 2, \dots, M\}$. The state-change vector v_μ for

reactions R_μ is defined as $v_\mu = [v_{1\mu}, v_{2\mu}, \dots, v_{N\mu}]^T$, where $v_{n\mu}$ represents the change in concentration of species S_n , caused by the occurrence of R_μ reaction of the underlying biochemical system. These equations fully define the system and the time evolution of the probability function $P(\mathbf{X}, t)$ can be obtained by the solution of the Chemical Master Equation(CME) [8,14,18]:

$$\frac{\partial P(\mathbf{X}, t)}{\partial t} = \sum_{\mu=1}^M (a_\mu(\mathbf{X} - v_\mu)P(\mathbf{X} - v_\mu, t) - a_\mu(\mathbf{X})P(\mathbf{X}, t)) \mathbb{1}[(\mathbf{X} + v_\mu) \in \Omega] \quad (1)$$

Here, $\mathbb{1}[(\mathbf{X} + v_\mu) \in \Omega]$ is 1 if $\mathbf{X} + v_\mu \in \Omega$ and 0 otherwise. The CME representing the rate of change of probability $P(\mathbf{X}, t)$ in an infinitely large state-space $\mathbf{X} \in \Omega$ is given by taking Ω to be the non-truncated space: $\Omega = \mathbb{N}^N$, $\mathbb{N} = \{0, 1, 2, \dots\}$

In Eq.1, a_μ represents the propensity function to account for transition from a given state \mathbf{X} to any other state, and v_μ indicates the stoichiometry of the reaction μ that results in such a transition. Eq.1 is a linear system of differential equations and may be rewritten as follows:

$$\frac{dP}{dt} = LP \quad (2)$$

where P is the probability distribution $P(\mathbf{X}, t)$ for a vector $\mathbf{X} = [X_1, X_2, \dots, X_N]$ and L is the time independent connection operator. For the steady state (SS) distribution P_{ss} , we have:

$$i) P_{ss}^X \geq 0; \quad \forall X \in \Omega \quad ii) \sum_{X \in \Omega} P_{ss}(X) = 1 \quad (3)$$

and iii) $LP_{ss} = 0$,

We assume that the deterministic steady state (SS) is unique. The non-truncated state-space Ω can be replaced with a truncated state-space $\hat{\Omega}$ [15,19] to approximate the probability distribution $P(\mathbf{X}, t)$. We define the truncated space as:

$$\{\hat{\Omega} = \{\mathbf{X} : \alpha_i \leq X_i \leq \beta_i, \quad \forall i\} \quad (4)$$

where α_i and β_i are extendable left and right boundaries of the truncated state-space. This approach is similar to that in [20], in which it is shown that the approximation based on the truncated space converges to the true steady state distribution as the limits of the truncated state-space converge to the limits of the original space.

The truncated state-space representation implies that given some $\varepsilon > 0$, for a sufficiently large $\beta_i > 0$ and sufficiently small $\alpha_i \geq 0$, the steady state probability distribution $P_{ss}(\mathbf{X})$ is approximated to within ε :

$$\sum_{X \in \hat{\Omega}} P_{ss}(\mathbf{X}) = 1 - \varepsilon$$

For an optimal SS probability approximation, ε should be made as small as possible. In the truncated state-space, Eq.3(iii) is represented as:

$$\hat{L}\hat{P}_{ss} = 0 \quad (5)$$

where \hat{L} is a matrix of propensities in $\hat{\Omega}$. To get the entries in \hat{L} we use Eq.1 modified so that $P(\mathbf{X}, t) = 0$ if $\mathbf{X} \notin \hat{\Omega}$ and $a_\mu(\mathbf{X}) = 0$ if $\mathbf{X} + v_\mu \notin \hat{\Omega}$. In the truncated state-space $\hat{\Omega}$, Eq.5 is an eigenvalue problem for eigenvalue $\lambda = 0$ and the eigenvector \hat{P}_{ss} can be obtained algebraically, or with an iterative algorithm for a large, sparse matrix \hat{L} .

Instead of finding the eigenvector, which can be an ill-conditioned problem when there are nonzero eigenvalues close to 0, we translate the problem to a well-conditioned system of linear equations as follows.

We first evaluate the deterministic steady state (Y_0) of the system, and then select state X_0 of the discrete system closest to this deterministic steady state, where $X_0 = \text{round}(Y_0)$. Taking \hat{P}_{ss} to be the solution of Eq. 5 and using the fact that the deterministic steady state solution is unique, we observe that $\hat{P}_{ss}(X_0)$ is among the largest entries of \hat{P}_{ss} . The states in $\hat{\Omega}$ are labeled as $1, 2, \dots, K$ with state X_0 denoted by j .

Then

$$\frac{\hat{P}_{ss}}{\hat{P}_{ss}^j} = [\hat{P}_{ss}^1 \dots \hat{P}_{ss}^j \dots \hat{P}_{ss}^K]^T / \hat{P}_{ss}^j \quad (6)$$

$$= [\hat{q}_1, \dots, \hat{q}_{j-1}, 1, \hat{q}_{j+1}, \dots, \hat{q}_K]^T$$

where $\hat{q}_k = \frac{\hat{P}_{ss}^k}{\hat{P}_{ss}^j}$, $k = 1 \dots K$, $\hat{q}_j = 1$. With $\hat{q} = [\hat{q}_1, \dots, 1, \dots, \hat{q}_K]$ Eq.5 now becomes $\hat{L}\hat{q} = 0$. Let \hat{L}_k be the k^{th} column of \hat{L} . Expanding $\hat{L}\hat{q}$ by column and rearranging gives the following well-conditioned problem:

$$\sum_{k=1, k \neq j}^K \hat{L}_k \hat{q}_k = -\hat{L}_j \quad \text{or,} \quad (7)$$

$$\hat{L}'\hat{q}' = -\hat{L}_j$$

In Eq.7, \hat{L}' is the matrix \hat{L} with column j removed and \hat{q}' is \hat{q} with entry \hat{q}_j removed. The error criterion for the system is checked for the calculation of \hat{P}_{ss} until a satisfactory value is obtained (see algorithm 1 for further details).

Application

In order to demonstrate the usability of the proposed steady state probability approximation method, we present

here two example networks (example network 1 and 2) from Bone Morphogenetic Protein(BMP) mediated signaling, and characterize the stochastic behavior of the systems. In the example network 1, we consider the role of a specific extracellular protein, Crossveinless-2(Cv-2), which is part of a class of proteins known as Surface-associated BMP-binding Proteins (SBPs) [4]. Cv-2 has the ability to regulate the stochastic noise in BMP signaling, and in this example we demonstrate that the role of Cv-2 is heavily dependent on reaction kinetics of the network: for some sets of parameter values, Cv-2 increases the coefficient of variation of the steady state signaling distribution, while for other parameter values it decreases the coefficient of variation.

In the second example network, we consider a model simplification strategy as used in [4,17]. This strategy is to omit a Type II receptor recruitment step from the receptor oligomerization in a BMP patterning model, under the assumption that the simplification step does not affect the outcome of a BMP-mediated patterning model. The obtained results by the use of steady state probability approximation method provide a numerical justification for the aforementioned simplification.

Background

During embryonic development, positional information is transduced by morphogens to underlying cells that respond to the concentration gradient of morphogen and eventually differentiate into distinct cell types [21]. For example, Decapentaplegic (*Dpp*), a drosophila homologue of BMP2/4, forms a spatial profile to pattern dorsal tissues in *Drosophila* development [21]. In a canonical BMP signaling pathway, dimeric ligands bind to receptors and form a heterotetrameric complex that consists of two Type I and two Type II receptors. The heterotetrameric receptor complex then phosphorylates the intracellular signal transducer Smad and the phosphorylated Smad forms a complex with a co-Smad. Subsequently, the Smad/Co-Smad complex accumulates in the nucleus and regulates gene expressions in a concentration dependent manner [17,22].

BMP regulation occurs at many points along the pathway, and a lot of focus has been on identifying and understanding how the ligand activity is regulated in the extracellular environment by secreted binding proteins. These include molecules such as Cv-2, HSPGs, among other reviewed in [1]. A focus of this work is to gain a better understanding of how regulation in the extracellular region impacts cell signaling noise, and eventually cell-to-cell variability.

Example networks

In many biochemical networks, where dynamics of the intermediate interactions of different species (proteins)

and molecular complexes are largely unknown, screening plays a significant role in the classification of dynamics-dependent network behavior. For example:

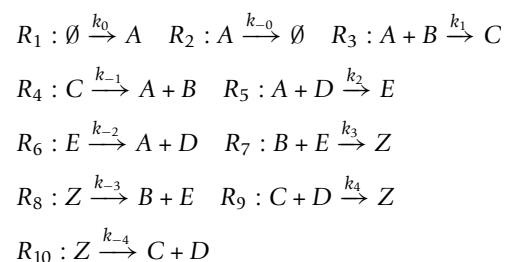
1. In a biochemical network where a class of secreted, surface-associated BMP binding proteins (SBPs) such as, Crossveinless-2 (Cv-2, node D as in Figure 1) [4] is allowed to regulate BMP signaling, the intermediate dynamics of the system that result in the formation and decoupling of a transient state BMP:Type I:Cv-2 (node M as in Figure 1) are largely unknown.
2. In the patterning modeling of BMP signaling pathways, it is often argued as a simplification strategy that omitting the step of recruitment of a Type II receptor to a bound BMP:Type I receptor complex doesn't affect the outcomes of patterning models [4,17,23,24]. While valid in the deterministic sense, it is not clear how this reduction impacts our estimates for noise in the system.

In these systems, we apply our SS probability approximation method to evaluate the probability distribution of different species and calculate the mean (μ), standard deviation (σ) and the coefficient of variation ($\Lambda = \frac{\sigma}{\mu}$; defined as the ratio between the standard deviation and the mean of any species) of the species distribution. Together with this information, we can screen the network for largely unknown dynamics of the intermediate interactions and classify solutions according to a model's ability to meet specific performance objectives.

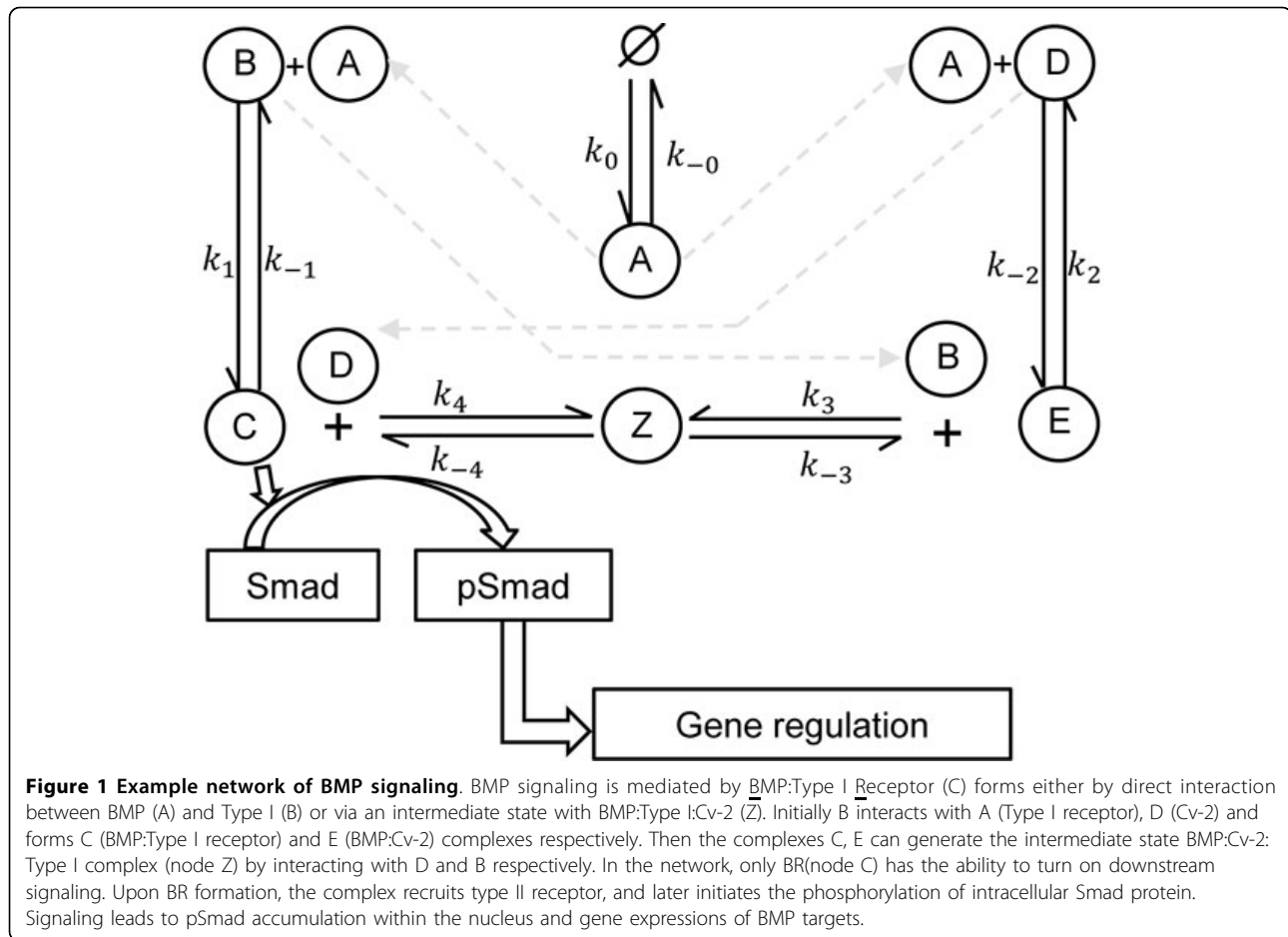
BMP-signaling regulation by SBPs

Signaling network

The single-cell local stochastic model that includes extracellular BMP(A), receptors (B), and SBPs such as, Cv-2 (D) with biochemical interactions, rate parameters, and connectivity is based on the network shown in Figure 1. Mass balance equations are listed below:



Out of all complexes ($C = \text{BMP:Type I Receptor} = BR$, $E = \text{BMP:Cv-2}$, $Z = \text{BMP:Type I receptor: Cv-2}$), available experimental evidence suggests that only ligand-bound receptors C ($\text{BMP:Type I Receptor} = BR$) initiate signaling to regulate downstream gene expression [17]. To focus on the noise compensation by regulation of receptors by



SBPs, the extracellular level of BMP (A) is treated as a parameter $A = \frac{k_0}{k_{-0}}$ and the interactions (1-10) are simplified accordingly. For example, in the simplified model reaction R_3 is represented $R'_1 : B \xrightarrow{Ak_1} C$.

The simplified model as obtained from reactions R_1 to R_{10} , has 5 species $\{S_1, S_2, \dots, S_5\}$ and is described completely by a total of 8 different chemical reactions. Time evolution of all species quantities are specified by a state vector $\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_5(t)]^T$ and state-change vector v_μ ($\mu = 1, 2, \dots, 8$), corresponding to all reactions that describe the system. For example, when $\mu = 1$, v_1 is $[-1 \ 1 \ 0 \ 0 \ 0]^T$ for reaction R'_1 of the simplified system. In this example network, techniques like those in [25] were used to verify that there is a unique steady state equilibrium and this ensures the applicability of the algorithm for this example network. Numerically, we used the polynomial root finding package hom4ps2 to ensure that there was only one equilibrium in the positive orthant [26]. It's worthwhile to mention that a similar approach is adopted in example network 2 to ensure the unique deterministic steady state. For both the networks, we numerically determined the deterministic steady state value Y_0 using Newton's method as

incorporated in SBTOOLBOX2 [27]. A generalized algorithm for simulation according to the steady state approximation as outlined in Methods section is given in algorithm 1.

Algorithm 1 Evaluate steady state (SS) distribution:
 $\hat{L}\hat{P}_{ss} = 0$

Require: Unique deterministic SS solution X_0

1. Reaction Networks with N Reaction R_1, \dots, R_N
2. Choose: $\varepsilon, \gamma_0, \gamma$
3. Solve: ODE for steady state(SS) = Y_0 and find discrete state X_0 closest to Y_0 , where $X_0 = \text{round}(Y_0)$.
4. Initiate, α_i, β_i ; where $\alpha_i = (X_0)_i - \gamma_0$, $\beta_i = (X_0)_i + \gamma_0$
5. Determine: Truncated state-space $\hat{\Omega}$ as shown in Eq.4 and \hat{L} as described after Eq.5
6. Determine: Column j of \hat{L} corresponding to X_0
7. Form \hat{L}' and \hat{L}_j as describe after Eq. 7.
8. Solve: $\hat{L}'\hat{q}' = -\hat{L}_j$
9. Find $\hat{P}_{ss} = \frac{[\hat{q}_1, \dots, \hat{q}_{j-1}, 1, \hat{q}_{j+1}, \dots, \hat{q}_K]^T}{\eta}$, where $\hat{q}' = [\hat{q}_1, \dots, \hat{q}_{j-1}, \hat{q}_{j+1}, \dots, \hat{q}_K]^T$ and $\eta > 0$ and $\eta = 1 + \sum_{l=1, l \neq j}^K \hat{q}_l$ is chosen so that $\sum_{X \in \hat{\Omega}} \hat{P}_{ss}(X) = 1$

10. Verify:

if $\sum_{X \in \hat{\Omega}, X_i = \delta_i} \hat{P}_{ss}(X) \geq \varepsilon$, for $\delta_i = \alpha_i$, or $\delta_i = \beta_i$ then
 $\alpha_i \leftarrow \alpha_i - \gamma$
 $\beta_i \leftarrow \beta_i + \gamma$
 Return to 5
 end if

In the algorithm, the values of γ_0 , γ are problem dependent based on the anticipated spread of the steady state (SS) distribution. Larger γ and γ_0 favor a larger $\hat{\Omega}$, with correspondingly better accuracy, but this comes at the expense of a larger state-space and more time required to solve the Eq.7. The parameter ε also controls the accuracy of the solution. In the truncated state-space, the tail of the distribution is essentially pushed in to the main part of the distribution and smaller ε means that less of the tail is changed.

Simulation and discussion

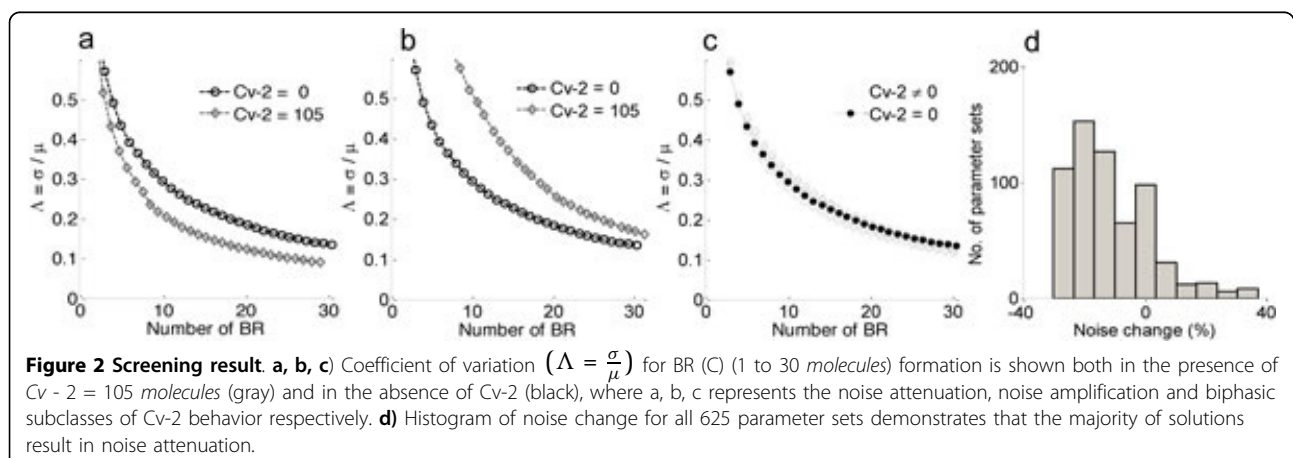
The binding kinetics between BMPs (species A, Figure 1) + receptors (species B), and BMPs + Cv-2 (species C) are largely known from the biacore analysis data [28,29]. However, the kinetic data associated with the intermediate tripartite complex BMP:Cv-2:Type I receptor (species Z, Figure 1) are currently unknown. In order to better understand the dependence of the steady state distribution on the kinetic parameters, we performed a parameter screen for the forward and reverse reaction rates ($k_{\pm s}$, $s = 3, 4$) for the formation and decoupling of species Z. For each of these four parameters, we use 5 evenly-spaced points on a logarithmic scale with the range $[10^{-1} \text{ to } 10^1] \text{ nM}^{-1}\text{s}^{-1}$ for the forward rates and $[10^{-3} \text{ to } 10^0] \text{ s}^{-1}$ for the reverse reaction rates. This produces a parameter grid that contains a total of 625 different parameter vectors.

For an appropriate comparison of the noise attenuation both in the presence and in the absence of Cv-2, species C (BMP: Type I Receptor = BR) concentration should remain the same regardless of the intermediate

dynamics. During simulation, the amount of available receptors (B) was fixed at 100, and a maximum of 30% receptor occupancy was allowed for the screening of the network. To ensure an equal amount of BR formation for each parameter vector, we modified the level of free ligand (A). For computational tractability, the screen is limited to a maximum continuation of 200 Cv-2 molecules, which allowed us to capture responses for all 625 different parameter sets.

In order to quantify noise in the system we measured the coefficient of variation ($\Lambda = \frac{\sigma}{\mu}$) that relates the standard deviation (σ) to the mean (μ) level of bound receptors. The parameter screen on the intermediate dynamics yields three primary qualitative subclasses for Cv-2 behavior in regulation of extracellular BR (C) fluctuation amplitude: i) reduced amplitude ii) increased amplitude and iii) mixed amplitude behavior [4]. Three primary types of responses of Cv-2 action on BMP fluctuations are shown in Figure 2a-c. As seen from Figure 2a, Cv-2 leads to a reduction of BR noise amplitude, and this is true for all $Cv-2 \in [0, 200]$. The value of the coefficient of variation (Λ) decreases for both increases in the level of bound receptor and the level of Cv-2. The subclass of increased amplitude demonstrates that increasing the level of Cv-2 in the system increases the value of the coefficient of variation ($\Lambda_{Cv-2 \neq 0} > \Lambda_{Cv-2 = 0}$) and is valid for the range of Cv-2 values considered in the screen (for a detailed discussion on this, interested readers can refer [4]). Lastly, mixed amplitude is classified as type iii, which demonstrates that Cv-2 can both increase and decrease the level of stochastic noise in the system (Figure 2c).

To clarify Cv-2 action further, we calculated percentage change in the amplitude using ($\% \text{ noise change} = \frac{(\Lambda_{Cv-2=105} - \Lambda_{Cv-2=0})}{\Lambda_{Cv-2=0}} \times 100$) for each parameter set output for a given amount of BR production (30 complexes) and Cv-2 value (105 molecules).



Based on the percent change of the coefficient of variation, we classify the screening outcome: negative percent change corresponds to noise attenuation whereas the positive change gives noise amplification. A histogram of all 625 parameters are shown in Figure 2d and in [4]. The implication of the screening result is that Cv-2 clearly reduces the variability of receptor activation throughout the range of Cv-2 tested. However, as demonstrated in Figure 2d, such a phenomenon as exhibited by SBPs like Cv-2 is found to be highly parameter dependent.

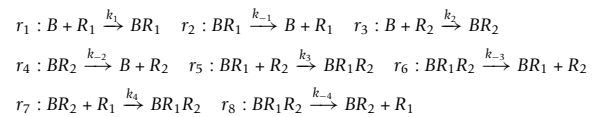
During the simulation, the kinetic rate constants for the intermediate complex BMP:receptor:Cv-2 (node Z, Figure 1) formation and decoupling were chosen from the parameter grid and representative kinetic rate constants for three different type of Cv-2 responses are enumerated in Table 1.

Analysis of Type II receptor recruitment process

In the signaling network shown in Figure 3, recruitment of Type II (= R₁) receptors can happen in two different ways: 1) BMP binds with Type I (= R₁) first and subsequently, recruits Type II receptors to form a tripartite complex BMP:Type I: Type II (BR₁R₂), and 2) BMP

directly interacts with Type I and Type II separately, and an intermediate state forms a tripartite BMP:Type I: Type II complex. In both situations, BMP:Type I:Type II complex (BR₁R₂) is the sole signaling complex responsible for the activation of downstream pathways.

All possible biochemical interactions that represent the ligand binding with Type I receptors and further recruitment of Type II receptors are:



The chemical interaction of Case II can easily be obtained from the interactions (r₁ to r₈) of Case III (Figure 3) by equaling the kinetic rate constants k_{±2} and k_{±4} of Case III to zero. For the kinetics, we relied on the published data [1]. The rate at which a Type II receptor is recruited upon formation of a BMP:Type I complex (BR₁) is comparatively faster than the rate of BMPs and Type I receptors interactions [17]. However, exact values of the rates of formulation of (BR₁R₂) complex are not readily available, and hence, parameters were screened

Table 1 Kinetic rates, Figure 2(a,b,c)

Figure	k ₃ (molecule ⁻¹ sec ⁻¹)	k ₃ (sec ⁻¹)	k ₄ (molecule ⁻¹ sec ⁻¹)	k ₄ (sec ⁻¹)
2a	1.3282	0.0100	1.3282	0.0100
2b	0.0133	1.0000	0.1328	0.0100
2c	0.1328	1.0000	0.4200	1.0000

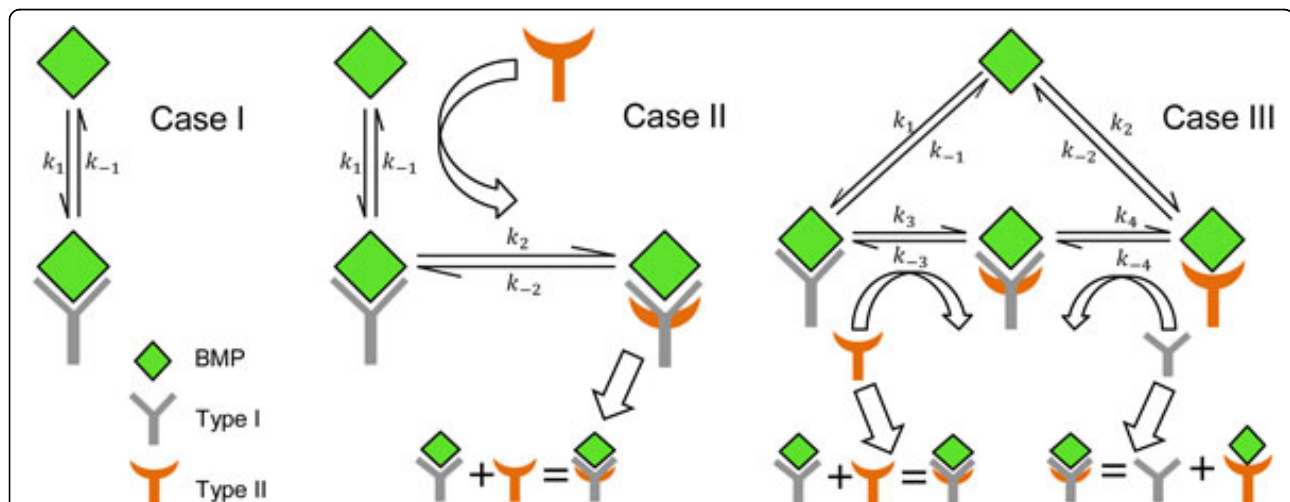


Figure 3 Network cases for Type II recruitment analysis in context of Drosophila melanogaster development. Case I) Recruitment of Type II is overlooked here and it imitates the simplified model used in previous studies. In this type of network, BMP:Type I complex (BR₁) acts as the sole signaling complex. **Case II)** Upon the formation of a BMP: Type I complex, subsequent recruitment of Type II receptor is considered here. But a direct interaction between BMP and Type II receptor doesn't happen in the network. Here, a tripartite complex BMP:Type I:Type II (BR₁R₂) activates the downstream pathways. **Case III)** Similar to Case II, but with an exception that a direct interaction between BMP and Type II receptor is allowed to form a BMP:Type II complex (BR₂) by changing BR₁ to BR₂. The kinetic equations are equivalent to the SBP system investigated in [4].

over the physiological ranges with values between $[10^{-1} \text{ to } 10^1] \text{ nM}^{-1} \text{ s}^{-1}$ for the forward rates and $[10^{-3} \text{ to } 10^0] \text{ s}^{-1}$ for the reverse reaction rates.

Simulation and discussion

To simulate the networks (as shown in Figure 3) for the calculation of the coefficient of variation Λ , we applied the truncated state-space approximation. During the simulation, a target of 1 to 30 signaling complexes (BR_1 for Case I and BR_1R_2 for Case II, Case III) in the extracellular region is considered so a direct comparison can be made for the coefficient of variation ($\Lambda = \frac{\sigma}{\mu}$) between BR_1 and BR_1R_2 .

The coefficient of variation (Λ) for BR_1R_2 remains very close to the coefficient of variation of BR_1 as shown in Figure 4a. Proximity in the coefficient of variation between BR_1 and BR_1R_2 (as shown in Figure 4a) demonstrates that the stochastic variability of the system is not affected by the recruitment of the Type II receptor. It is also found that increasing the concentration of R_2 brings the coefficient of variation of BR_1R_2 into very close agreement with the coefficient of variation of BR_1 as shown in Figure 4b. A similar outcome is obtained from the simulation of Case III of the Figure 3 and the result is shown in Figure 4c. Finally, all the outcomes are

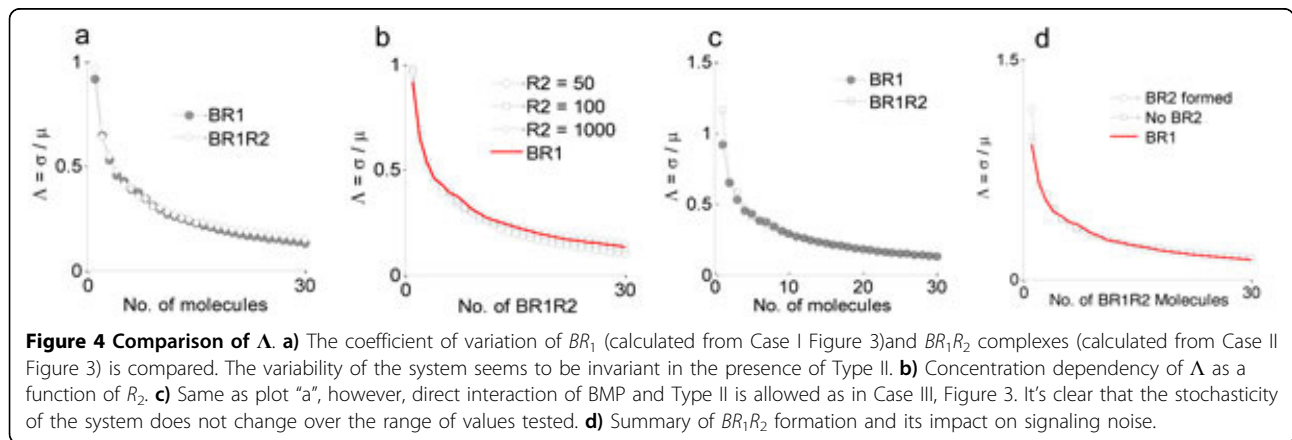


Figure 4 Comparison of Λ . **a)** The coefficient of variation of BR_1 (calculated from Case I Figure 3) and BR_1R_2 complexes (calculated from Case II Figure 3) is compared. The variability of the system seems to be invariant in the presence of Type II. **b)** Concentration dependency of Λ as a function of R_2 . **c)** Same as plot "a", however, direct interaction of BMP and Type II is allowed as in Case III, Figure 3. It's clear that the stochasticity of the system does not change over the range of values tested. **d)** Summary of BR_1R_2 formation and its impact on signaling noise.

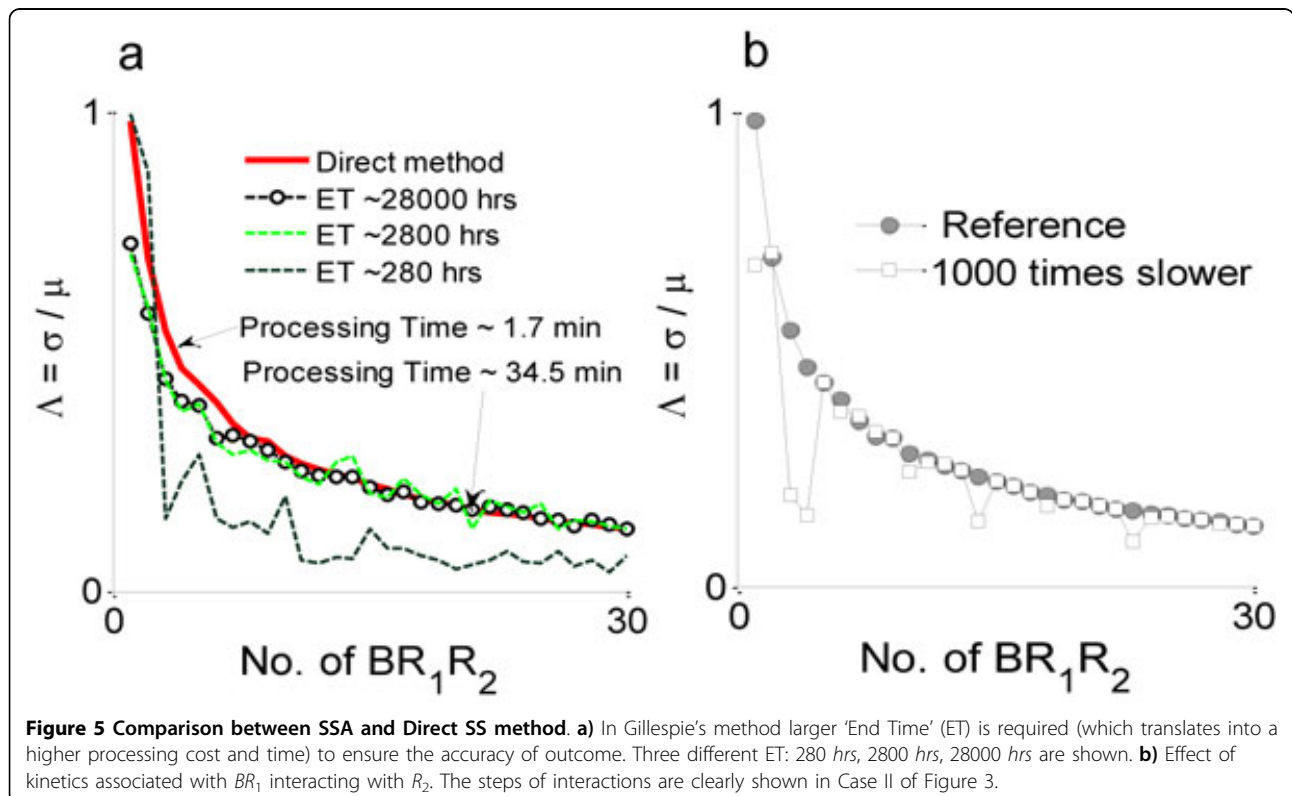


Figure 5 Comparison between SSA and Direct SS method. **a)** In Gillespie's method larger 'End Time' (ET) is required (which translates into a higher processing cost and time) to ensure the accuracy of outcome. Three different ET: 280 hrs, 2800 hrs, 28000 hrs are shown. **b)** Effect of kinetics associated with BR_1 interacting with R_2 . The steps of interactions are clearly shown in Case II of Figure 3.

summarized in Figure 4d, where it is shown that regardless of the different cases as shown in Figure 3 the coefficient of variation (Λ) of BR_1R_2 is approximately equal to that of BR_1 .

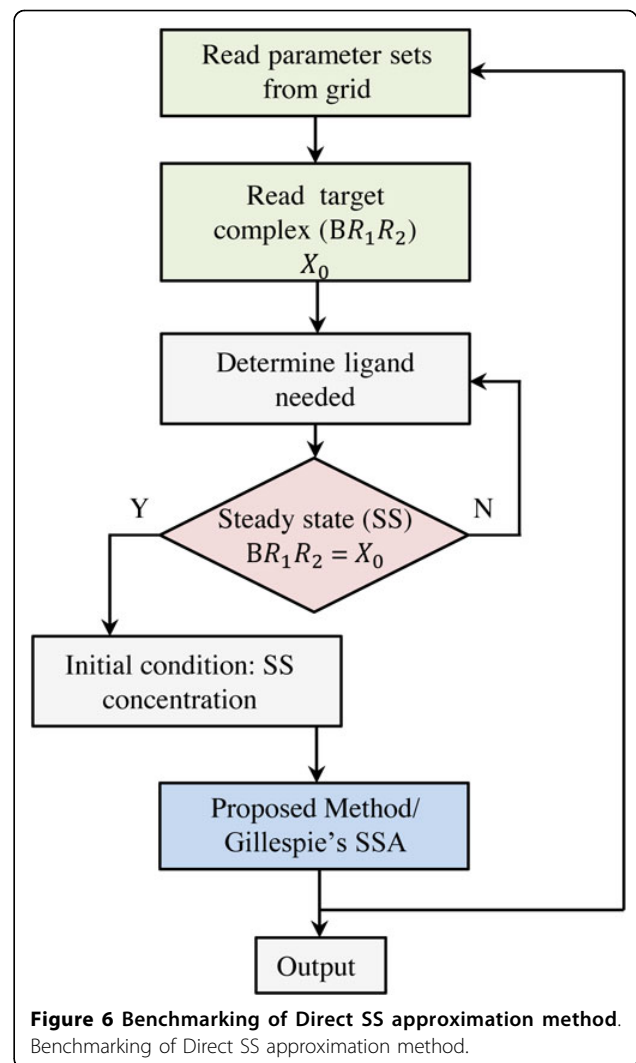
Additionally, it is also found from the simulated result that the rate at which the BMP:Type I recruits Type II receptor (Case II in Figure 3) also decides the effect of Type II recruitment process on the stochastic variability of the system. With a comparatively slower rate, the coefficient of variation tends to oscillate as observed in Figure 5b. When the recruitment rate is slower than the formation rate of BMP:Type I complex, free Type II receptors fail to get frequent access to BMPs via the BMP:Type I:Type II tripartite complex, and can cause the concentration of BMP:Type I:Type II complex to oscillate more than the case with a comparatively faster dynamics. Thus, mitigating noise is not a natural output of receptor oligomerization + transduction and instead, requires another co-factor such as Cv-2 [4].

Benchmarking of Direct SS approximation method

Carrying out large-scale stochastic simulation can be time consuming but calculation of the approximate solution via a truncated state-space can greatly improve the speed. In order to show the performance improvement in terms of computational cost and time of direct SS approximation in the analysis of stochastic biochemical networks, we benchmarked the method by comparing it with Gillespie's stochastic simulation algorithm (SSA) method [9] for numerical calculations of stochastic biochemical networks. In the benchmarking, the processing time taken by each method was calculated based on the steps in the blue box as mentioned in the flow chart diagram (Figure 6). The sample problem was calculated for both methods on the same hardware and software configuration: *Processor*: Intel(R) Xeon (R) CPU E5405, 2.00 GHz (quad-core), *RAM*: 16 GB, SBTOOLBOX2 [27] and Matlab R2010a with SimBiology 3.0.

The processing time and computed Λ values for a target $BR_1R_2 = 20$ for Case II, Figure 3, is provided in Table 2 to show the accuracy and time gain that can be obtained if the proposed direct SS distribution approximation method is used. Gillespie's SSA takes longer to generate an output that contains enough information to calculate the distribution as compared to the time taken by Direct SS approximation method. The problem becomes severe when continuation of a multiple parameters are necessary to explore the system's parameter dependency as done previously in [4].

In Table 2, the term 'End time (ET) in Gillespie's SSA' corresponds to the amount of time the system dynamics were allowed to evolve. The accuracy of the Gillespie's SSA approach depends on the 'End time in Gillespie's SSA'(directly contributes to the processing time) set in the



model simulation, and is shown clearly in Figure 5a and Table 2. Very low propensities require long simulation times in Gillespie's SSA due to the infrequency of events. Accuracy of Gillespie's method for the sample example increases as the 'End time in Gillespie's SSA' is increased. This large simulation time in turn directly impacts the processing time, resulting in a large computational cost to achieve the desired accuracy (Table 2).

Conclusions

In this study, we illustrate an approach of determining the steady state probability distribution efficiently to carry on continuation in multiple variables within a large-scale parameter screen. The approach is demonstrated further with a couple of applications, where we investigated 1) the dynamic dependency of a class of proteins, known as SBPs, in the regulation of BMP signaling, and 2) the binding of Type II receptor in BMP signaling. The results suggest that the recruitment of a

Table 2 Benchmarking: Gillespie's SSA and Direct SS approximation for a target $BR_1R_2 = 20$

Method	End Time (ET) in Gillespie's SSA(hrs)	Processing Time (sec)	Λ
Direct SS approximation	Not Applicable	0.4 -0.6	0.1707
Gillespie's SSA	28000	90-95	0.1705
	2800	8-10	0.1878
	1390	4-5	0.2254

type II receptor in BMP signaling doesn't affect the stochasticity of the system over the range of concentration and parameters investigated. Direct calculation of the SS probability distribution can be successfully applied to systems with a unique deterministic SS solution, and future work will investigate similar approaches for other biochemical systems.

Acknowledgements

Based on "Steady state probability approximation applied to stochastic model of biological network", by Shahriar Karim, David M Umulis and Greger T Buzzard which appeared in *Genomic Signal Processing and Statistics (GENSIPS), 2011 IEEE International Workshop on*. © 2011 IEEE [30]. We would like to thank Purdue University, West Lafayette, IN 47907, for financial support.

This article has been published as part of *BMC Genomics* Volume 13 Supplement 6, 2012: Selected articles from the IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS) 2011. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/13/S6>.

Author details

¹Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, USA. ²Department of Electrical and Computer Engineering, Purdue University, West Lafayette, USA. ³Department of Mathematics, Purdue University, West Lafayette, USA.

Authors' contributions

MSK designed the research, contributed to algorithm development, performed simulation and data analysis, and wrote the manuscript. GB contributed to algorithm development, discussed the results and wrote the manuscript. DU contributed in research design, discussed the results and wrote the paper. All authors contributed to replying to reviewers' comments and approved the final manuscript of the paper.

Competing interests

The authors declare that they have no competing interests.

Published: 26 October 2012

References

- Umulis D, O'Connor M, Blair S: **The extracellular regulation of bone morphogenetic protein signaling.** *Development* 2009, **136**(22):3715.
- Fox J: **Human iPSC and ESC translation potential debated.** *Nature Biotechnology* 2011, **29**(5):375-376.
- Lander A, Lo W, Nie Q, Wan F: **The measure of success: constraints, objectives, and tradeoffs in morphogen-mediated patterning.** *Cold Spring Harbor Perspectives in Biology* 2009, **1**:a002022.
- Karim M, Buzzard G, Umulis D: **Secreted, receptor-associated bone morphogenetic protein regulators reduce stochastic noise intrinsic to many extracellular morphogen distributions.** *J R Soc Interface* 2012, **9**:1073-1083.
- Raj A, van Oudenaarden A: **Nature, nurture, or chance: stochastic gene expression and its consequences.** *Cell* 2008, **135**(2):216-226.
- Samoilov M, Price G, Arkin A: **From fluctuations to phenotypes: the physiology of noise.** *Science's STKE* 2006, **2006**(366).
- Thattai M, Van Oudenaarden A: **Intrinsic noise in gene regulatory networks.** *proceedings of the national academy of sciences of the united states of America* 2001, **98**(15):8614.
- Gillespie D: **A rigorous derivation of the chemical master equation.** *Physica A: Statistical Mechanics and its Applications* 1992, **188**(1-3):404-425.
- Gillespie D: **A general method for numerically simulating the stochastic time evolution of coupled chemical reactions.** *Journal of computational physics* 1976, **22**(4):403-434.
- van Kampen N: *Stochastic processes in physics and chemistry* North Holland; 2007.
- Peleš S, Munsky B, Khammash M: **Reduction and solution of the chemical master equation using time scale separation and finite state projection.** *The Journal of chemical physics* 2006, **125**:204104.
- Gillespie D, Petzold L: **Improved leap-size selection for accelerated stochastic simulation.** *The Journal of Chemical Physics* 2003, **119**:8229.
- Gillespie D: *Stochastic simulation of chemical kinetics* 2007.
- Gillespie D: **Exact stochastic simulation of coupled chemical reactions.** *The journal of physical chemistry* 1977, **81**(25):2340-2361.
- Hegland M, Burden C, Santoso L, MacNamara S, Booth H: **A solver for the stochastic master equation applied to gene regulatory networks.** *Journal of Computational and Applied Mathematics* 2007, **205**(2):708-724.
- Goutsias J: **Classical versus stochastic kinetics modeling of biochemical reaction systems.** *Biophysical journal* 2007, **92**(7):2350-2365.
- Serpe M, Umulis D, Ralston A, Chen J, Olson D, Avanesov A, Othmer H, O'Connor M, Blair S: **The BMP-binding protein Crossveinless 2 is a short-range, concentration-dependent, biphasic modulator of BMP signaling in Drosophila.** *Developmental cell* 2008, **14**(6):940-953.
- Gillespie D: **Approximate accelerated stochastic simulation of chemically reacting systems.** *The Journal of Chemical Physics* 2001, **115**:1716.
- Hegland M, Hellander A, Lötstedt P: **Sparse grids and hybrid methods for the chemical master equation.** *BIT Numerical Mathematics* 2008, **48**(2):265-283.
- Munsky B, Khammash M: **The finite state projection algorithm for the solution of the chemical master equation.** *The Journal of chemical physics* 2006, **124**:044104.
- Wolpert L: *Principles of Development* UK: Oxford University Press; 2006.
- Schmierer B, Tournier A, Bates P, Hill C: **Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system.** *Proc Natl Acad Sci U S A* 2008, **105**(18):6608-6613.
- Ben-Zvi D, Shilo B, Fainsod A, Barkai N: **Scaling of the BMP activation gradient in Xenopus embryos.** *Nature* 2008, **453**:1205-1211.
- Mizutani CM, Nie Q, Wan FY, Zhang YT, Vilmos P, Sousa-Neves R, Bier E, Marsh JL, Lander AD: **Formation of the BMP activity gradient in the Drosophila embryo.** *Dev Cell* 2005, **8**(6):915-24.
- Craciun G, Helton J, Williams R: **Homotopy methods for counting reaction network equilibria.** *Mathematical biosciences* 2008, **216**(2):140-149.
- Lee T, Li T, Tsai C: **HOM4PS-2.0: a software package for solving polynomial systems by the polyhedral homotopy continuation method.** *Computing* 2008, **83**(2):109-133.
- Schmidt H, Jirstrand M: **Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology.** *Bioinformatics* 2006, **22**(4):514-515.
- Kirsch T, Nickel J, Sebald W: **BMP-2 antagonists emerge from alterations in the low-affinity binding epitope for receptor BMPRII.** *The EMBO Journal* 2000, **19**(13):3314-3324.
- Rentzsch F, Zhang J, Kramer C, Sebald W, Hammerschmidt M: **Crossveinless 2 is an essential positive feedback regulator of Bmp signaling during zebrafish gastrulation.** *Development* 2006, **133**(5):801.
- Karim S, Umulis DM, Buzzard GT: **Steady state probability approximation applied to stochastic model of biological network.** *Genomic Signal*

Processing and Statistics (GENSIPS), 2011 IEEE International Workshop on: 4-6 December 2011 2011, 56-59.

doi:10.1186/1471-2164-13-S6-S10

Cite this article as: Karim *et al.*: Efficient calculation of steady state probability distribution for stochastic biochemical reaction network. *BMC Genomics* 2012 **13**(Suppl 6):S10.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

