Universally valid reduction of multiscale stochastic biochemical systems using simple non-elementary propensities

Yun Min Song[1,2◕], Hyukpyo Hong[1,2◕], Jae Kyoung Kim[1,2*]

1 Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
2 Biomedical Mathematics Group, Institute for Basic Science, Daejeon 34126, Republic of Korea

◕These authors contributed equally to this work.
* jaekkim@kaist.ac.kr

Abstract

Biochemical systems consist of numerous elementary reactions governed by the law of mass action. However, experimentally characterizing all the elementary reactions is nearly impossible. Thus, over a century, their deterministic models that typically contain rapid reversible bindings have been simplified with non-elementary reaction functions (e.g., Michaelis-Menten and Morrison equations). Although the non-elementary functions are derived by applying the quasi-steady-state approximation (QSSA) to deterministic systems, they have also been widely used to derive propensities for stochastic simulations due to computational efficiency and simplicity. However, the validity condition for this heuristic approach has not been identified even for the reversible binding between molecules, such as protein-DNA, enzyme-substrate, and receptor-ligand, which is the basis for living cells. Here, we find that the non-elementary propensities based on the deterministic total QSSA can accurately capture the stochastic dynamics of the reversible binding in general. However, serious errors occur when reactant molecules with similar levels tightly bind, unlike deterministic systems. In that case, the non-elementary propensities distort the stochastic dynamics of a bistable switch in the cell cycle and an oscillator in the circadian clock. Accordingly, we derive alternative non-elementary propensities with the stochastic low-state QSSA, developed in this study. This provides a universally valid framework for simplifying multiscale stochastic biochemical systems with rapid reversible bindings, critical for efficient stochastic simulations of cell signaling and gene regulation.

Introduction

To understand the complex dynamics of numerous molecular interactions in living cells, quantitative analysis using mathematical models is essential [1]. While elementary reactions in living cells can be modeled by the law of mass action, characterizing all their kinetics is challenging. Thus, over a century, the combined effect of a set of elementary reactions such as rapid reversible bindings has been described with non-elementary functions (e.g., Michaelis-Menten and Morrison equations) to simplify deterministic models [2–7]. Since the early 2000s, these deterministically driven non-elementary functions have also been widely used to derive propensity functions for stochastic simulations, which greatly reduces the computational cost [8–33]. This heuristic approach for efficient stochastic simulations was believed to be valid as long as the non-elementary reaction functions are accurate in the deterministic sense. However, unfortunately, this was not the case [33–39]. The reason for the discrepancy between the deterministic and stochastic simulations has been recently identified for some cases [37–39], but not for all [40]. Currently, guidelines for this popular but heuristic method for efficient stochastic simulations with non-elementary propensity functions are absent.

The non-elementary reaction functions are mainly the result of the reduction of deterministic models with the following reversible binding reaction:

$$\mathrm{A} + \mathrm{B} \underset{k_\mathrm{b}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} \mathrm{C}. \tag{1}$$

The reversible binding between molecules, such as enzyme-substrate, receptor-ligand, and protein-DNA, is the first step for nearly all biological functions of living cells [41]. However, rather than the reversible binding itself, its outcome is usually our major interest. For instance, rather than the binding between a transcription factor and DNA, we are interested in its outcome, the transcription. Furthermore, the transcription factor binding to DNA takes at most one second while transcription takes about 30 minutes in a mammalian gene [42], which causes stiffness in numerical simulations [43].

Fortunately, such rapid reversible binding reactions can be eliminated from models using the property that the level of the species (A, B, and C) regulated by the reversible bindings quickly equilibriate to their quasi-steady-states (QSSs). In deterministic models, their quasi-steady-state approximations (QSSAs), which are non-elementary reaction functions, can be obtained by finding the steady-state solution of the associated differential equation. Because the QSSAs are determined by the total concentrations of the bound and unbound species, which are not affected by the reversible binding, they are known as the "total" QSSA (tQSSA). After replacing the variables that represent the levels of A, B, and C with their tQSSAs, rapid reversible bindings have been successfully eliminated from various deterministic models describing enzyme catalysis, gene regulation, and cell cycle regulation [5–7, 23, 44–50].

In stochastic models, the QSSAs for the numbers of A, B, and C are their stationary average numbers (i.e., the first moment) conditioned on the total numbers of the bound and unbound species [27–30]. These stochastic QSSAs can be obtained by finding the steady-state solution of the chemical master equation (CME). However, unlike the deterministic tQSSA, the stochastic QSSA has a complex form (Eq. (4)), which does not provide any intuition and importantly increases computational cost. Thus, its approximation has been derived with the deterministic tQSSA. This approximation, often referred to as the stochastic tQSSA (stQSSA) [7, 31, 38, 39], leads to non-elementary propensity functions for stochastic simulations using the Gillespie algorithm [51]. In this way, the stochastic dynamics of various systems have been accurately captured with low computational cost [7, 30–33, 38, 39, 52, 53]. However, a recent study reported that the stQSSA can be inaccurate [40], which raises the question of validity conditions for the stQSSA.

Here, we identify the complete validity condition for using the stQSSA to simplify stochastic models containing rapid reversible bindings. Specifically, we find that the stQSSA is accurate for a wide range of conditions. However, when two species whose molar ratio is ∼1:1 tightly bind, the stQSSA highly overestimates the number of unbound species. In this case, using the stQSSA to simplify stochastic models distorts the stochastic dynamics of the transcriptional repression, the transcriptional negative feedback loop of the circadian clock, and the bistable switch for mitosis. Importantly, by using the fact that the number of the unbound species is low due to the tight binding when the stQSSA is inaccurate, we develop an alternative approach, stochastic "low-state" QSSA (slQSSA). In this way, when reversible bindings are tight and not tight, slQSSA and stQSSA can be used, respectively, which enables one to obtain accurately reduced stochastic models for any case. Our work provides a complete and simple guideline for the reduction of multiscale stochastic biochemical systems containing the fundamental elementary reaction, i.e., rapid reversible binding.

Results

stQSSA can overestimate the number of the unbound species

In the reversible binding reaction (Eq. (1)), the concentration of A, denoted by $\tilde{A}$, is governed by the following ODE:

$$\frac{d\tilde{A}}{dt} = -k_{\mathrm{f}}\tilde{A}\cdot\tilde{B} + k_{\mathrm{b}}\tilde{C} = -k_{\mathrm{f}}\tilde{A}\cdot(\tilde{B}_{\mathrm{T}} - \tilde{A}_{\mathrm{T}} + \tilde{A}) + k_{\mathrm{b}}(\tilde{A}_{\mathrm{T}} - \tilde{A}), \tag{2}$$

where $\tilde{A}_{\mathrm{T}} = \tilde{A} + \tilde{C}$ and $\tilde{B}_{\mathrm{T}} = \tilde{B} + \tilde{C}$ are the total concentrations of the bound and unbound species. By solving $\frac{d\tilde{A}}{dt} = 0$ in terms of $\tilde{A}_{\mathrm{T}}$ and $\tilde{B}_{\mathrm{T}}$, the tQSSA for $\tilde{A}$ can be obtained as follows:

$$\tilde{A}_{\mathrm{tq}} \coloneqq \frac{1}{2}\left\{(\tilde{A}_{\mathrm{T}} - \tilde{B}_{\mathrm{T}} - \tilde{K}_{\mathrm{d}}) + \sqrt{(\tilde{A}_{\mathrm{T}} - \tilde{B}_{\mathrm{T}} - \tilde{K}_{\mathrm{d}})^2 + 4\tilde{A}_{\mathrm{T}}\tilde{K}_{\mathrm{d}}}\right\}, \tag{3}$$

where the $\tilde{K}_{\mathrm{d}} = k_{\mathrm{b}}/k_{\mathrm{f}}$ is the dissociation constant. Note that if the reversible binding (Eq. (1)) is embedded in a larger system, there could be other reactions affecting the dynamics of $\tilde{A}$ and thus additional terms in Eq. (2). However, as long as the reversible binding is fast (i.e., k_{f} and k_{b} are much larger than the other reaction rates), $\tilde{A}_{\mathrm{tq}}$ is still an accurate tQSSA for $\tilde{A}$. Similarly, by solving $\frac{d\tilde{B}}{dt} = 0$ and $\frac{d\tilde{C}}{dt} = 0$, the tQSSAs for $\tilde{B}$ and $\tilde{C}$ can be obtained. These tQSSAs, also known as the Morrison equations [6], are generally valid, unlike the Michaelis-Menten type equations, which are valid only when the enzyme concentration is negligible [7, 46, 47, 49]. Thus, the tQSSAs have been used to simplify models containing not only interactions between metabolites but also proteins whose concentrations are typically comparable [7].

Unlike the deterministic QSSA (Eq. (3)), the stochastic QSSA, which is the stationary average number conditioned on the total numbers of the bound and unbound species, has a complex form [40, 54]. For instance, the stochastic QSSA for the number of A ($\langle A\rangle$) can be expressed in terms of the dimensionless variables and parameters, $X = \tilde{X}\Omega$, where Ω is the volume of a system (e.g., $A = \tilde{A}\Omega$, $K_{\mathrm{d}} = \tilde{K}_{\mathrm{d}}\Omega$) as follows (see Methods for details):

$$\langle A\rangle = \left(\sum_{l=A_0}^{A_{\mathrm{T}}} \frac{lK_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)\cdot\left(\sum_{l=A_0}^{A_{\mathrm{T}}} \frac{K_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)^{-1}, \tag{4}$$

where $A_0 = \max\{0, A_{\mathrm{T}} - B_{\mathrm{T}}\}$. This complex form of the stochastic QSSA does not provide any intuition and importantly increases computational cost. Thus, as an alternative to the stochastic QSSA, its approximation, stQSSA was derived with the deterministic tQSSA [7, 22–26, 31]. Specifically, the stQSSA for A (A_{tq}) can be derived from $\tilde{A}_{\mathrm{tq}}$ (Eq. (3)) after replacing the concentration-based variables and parameters ($\tilde{X}$) with dimensionless variables and parameters (X) as follows:

$$\langle A\rangle \approx A_{\mathrm{tq}} \coloneqq \frac{1}{2}\left\{(A_{\mathrm{T}} - B_{\mathrm{T}} - K_{\mathrm{d}}) + \sqrt{(A_{\mathrm{T}} - B_{\mathrm{T}} - K_{\mathrm{d}})^2 + 4A_{\mathrm{T}}K_{\mathrm{d}}}\right\}. \tag{5}$$

Similarly, the stQSSA for B and C (B_{tq} and C_{tq}) can be obtained from their deterministic tQSSAs as follows:

$$\begin{aligned} \langle B \rangle &\approx B_{\mathrm{tq}} := \frac{1}{2}\left\{(B_{\mathrm{T}} - A_{\mathrm{T}} - K_{\mathrm{d}}) + \sqrt{(A_{\mathrm{T}} - B_{\mathrm{T}} - K_{\mathrm{d}})^2 + 4A_{\mathrm{T}}K_{\mathrm{d}}}\right\}, \\ \langle C \rangle &\approx C_{\mathrm{tq}} := \frac{1}{2}\left\{(A_{\mathrm{T}} + B_{\mathrm{T}} + K_{\mathrm{d}}) - \sqrt{(A_{\mathrm{T}} - B_{\mathrm{T}} - K_{\mathrm{d}})^2 + 4A_{\mathrm{T}}K_{\mathrm{d}}}\right\}. \end{aligned} \tag{6}$$

To identify the validity conditions for these stQSSAs, we calculated the relative error ($R_{\mathrm{X}} := \left|\frac{X_{\mathrm{tq}} - \langle X \rangle}{\langle X \rangle}\right|$, X = A, B, C) of the stQSSA (X_{tq}) to the stochastic QSSA ($\langle X \rangle$) (Fig 1a-1c). The errors are nearly zero in most of the parameter regions, which explains why various stochastic models reduced with the stQSSA have been accurate in most previous studies [7, 30–33, 38, 39, 52, 53]. However, the relative errors of the unbound species (R_{A} and R_{B}) are high when $A_{\mathrm{T}} \approx B_{\mathrm{T}}$. Specifically, the relative error of the bound species (R_{C}) is at most $\sim$0.2 but that of the unbound species (R_{A}, R_{B}) can be $\sim$100.

To investigate why R_{A} is high when $A_{\mathrm{T}} \approx B_{\mathrm{T}}$, we derived the exact upper and lower bounds for R_{A} (see Methods for details):

$$F_{\mathrm{A}}S_{\mathrm{A}} \le R_{\mathrm{A}} \le 2F_{\mathrm{A}}S_{\mathrm{A}}, \tag{7}$$

where F_{A} is the Fano factor of A (i.e., $\frac{\mathrm{Var}(A)}{\langle A \rangle}$), and S_{A} is the relative sensitivity of A_{tq} with respect to B_{T} (i.e., $\frac{1}{A_{\mathrm{tq}}}\left|\frac{dA_{\mathrm{tq}}}{dB_{\mathrm{T}}}\right|$). Furthermore, we proved that the Fano factor (F_{A}) is less than 1 (i.e., A has a sub-Poissonian stationary distribution; see S1 Appendix for details). Therefore, R_{A}, especially its upper bound, mainly depends on S_{A} (Figs 1d, 1e, and S1) whose formula can be derived in the following simple form, unlike R_{A}:

$$S_{\mathrm{A}} = \frac{1}{A_{\mathrm{tq}}}\left|\frac{dA_{\mathrm{tq}}}{dB_{\mathrm{T}}}\right| = \frac{1}{\sqrt{(A_{\mathrm{T}} - B_{\mathrm{T}} - K_{\mathrm{d}})^2 + 4A_{\mathrm{T}}K_{\mathrm{d}}}}. \tag{8}$$

Because S_{A} attains the maximum value $\frac{1}{\sqrt{4A_{\mathrm{T}}K_{\mathrm{d}}}}$ at $B_{\mathrm{T}} = A_{\mathrm{T}} - K_{\mathrm{d}}$, S_{A} has a large maximum value when $K_{\mathrm{d}} \ll 1$ at $A_{\mathrm{T}} = B_{\mathrm{T}} + K_{\mathrm{d}} \approx B_{\mathrm{T}}$. This explains why R_{A}, whose upper bound is mainly determined by $2S_{\mathrm{A}}$, is large when the binding is tight ($K_{\mathrm{d}} \ll 1$) and the total numbers of the bound and unbound species are similar ($A_{\mathrm{T}} \approx B_{\mathrm{T}}$) (Fig 1d). In this case, the majority of A is bound with B, and thus $A = 0$ most of the time (Fig 1f left). That is, A rarely becomes 1 by the weak unbinding reaction and then immediately A becomes 0 by the strong binding reaction. As a result, the probability that $A = 1$ is approximately 1% (i.e., $\langle A \rangle \approx 0.01$), but the stQSSA for A (A_{tq}) overestimates it as 10%, which is 10 times larger (Fig 1f right). Since A and B are symmetric, the above analysis can be applied to B, analogously.

stQSSA can overestimate the transcriptional activity

We found that the stQSSA for the number of the unbound species is inaccurate if their molar ratio is $\sim$1:1 and their binding is tight (Fig 1d-1f). Thus, we expected that in such cases, using the stQSSA to eliminate a rapid reversible binding in a stochastic model can distort its dynamics. To illustrate this, we constructed a simple gene regulatory network where gene expressions are determined by a reversible binding between transcription factors and genes (Fig 2a left, Table S1); DNA (D) and a transcription factor (P) reversibly bind to form a complex (D:P). As P acts as a repressor of $\mathrm{M_R}$ transcription, the transcription rate of $\mathrm{M_R}$ is proportional to the number of the unbound DNA (D). On the other hand, as P acts as an activator of $\mathrm{M_A}$ transcription, the transcription rate of $\mathrm{M_A}$ is proportional to the number of the bound

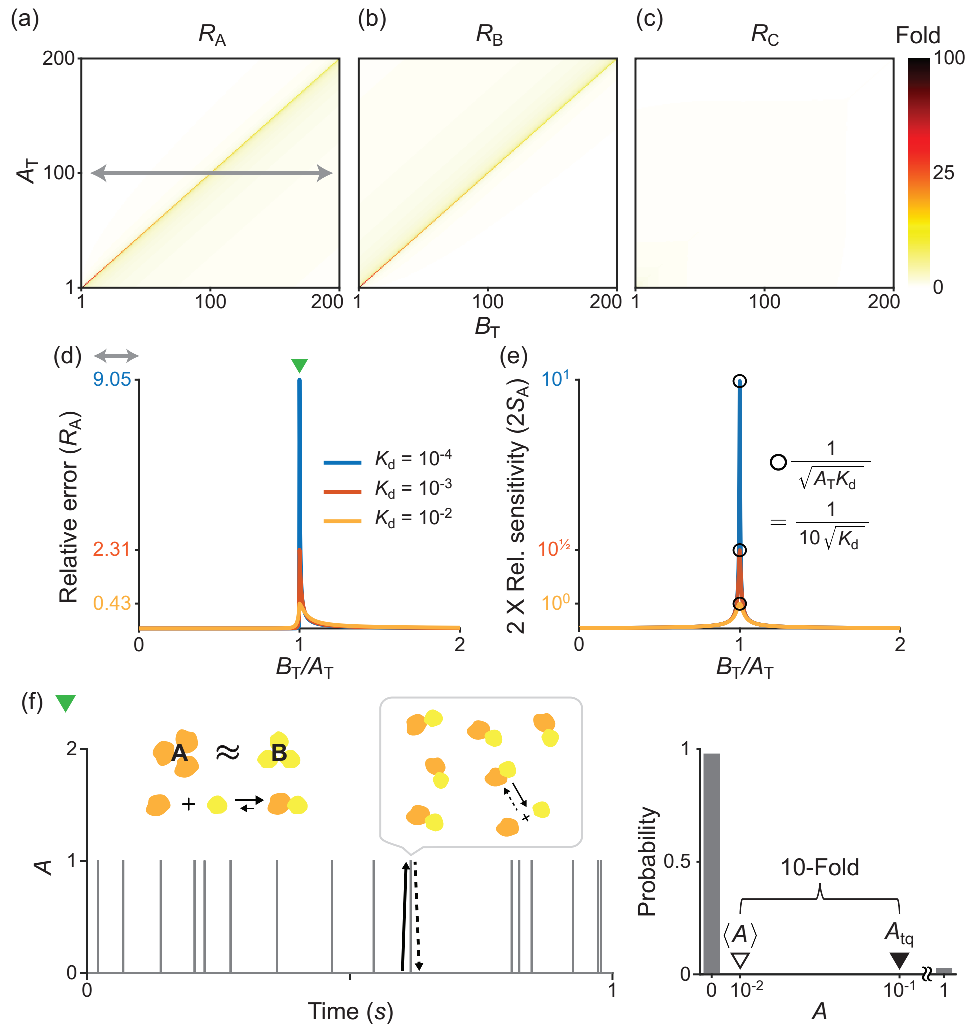

Fig 1. stQSSA overestimates the number of the unbound species when their molar ratio is ∼1:1 and binding is tight. (a-c) Heat maps of the relative errors ($R_X = \left|\frac{X_{tq}-\langle X\rangle}{\langle X\rangle}\right|$) of the stQSSA ($X_{tq}$) to the stochastic QSSA ($\langle X\rangle$) for $X = A, B, C$ in the reversible binding reaction (Eq. (1)). Color in the heat maps represents the maximum value of R_X calculated by varying K_d from 10^{-4} to 10^2 for each total number of the bound and unbound species ($A_T = A + C$ and $B_T = B + C$). R_A and R_B can be extremely large when $A_T \approx B_T$ while R_C is always small. **(d)** R_A calculated over B_T/A_T between 0 and 2 (gray arrow in a) for three fixed K_d values ($10^{-4}, 10^{-3}$ and 10^{-2}). R_A becomes larger as B_T/A_T is similar to 1 and the K_d becomes smaller (i.e., the binding becomes tighter). **(e)** R_A mainly depends on the relative sensitivity of A_{tq} (i.e., $2S_A$), which can be derived in a simple form, unlike R_A (Eq. (8)). The maximum value of $2S_A$ is given by $\frac{1}{\sqrt{A_T K_d}}$, which is achieved when B_T/A_T is similar to 1 as in the case of R_A. **(f)** A trajectory (left) and the stationary probability distribution (right) of A for a parameter set where R_A is large (green triangle in d, $A_T = B_T = 100$, $k_f/\Omega = 10^4 s^{-1}, k_b = 1 s^{-1}$), simulated using the Gillespie algorithm. Since $A_T = B_T$ and A binds with B tightly, $A = 0$ (i.e., every A is bound) most of the time, and it rarely becomes 1 by the weak unbinding reaction (solid arrow) and immediately comes back to 0 by the strong binding reaction (dotted arrow). As a result, when $K_d = 10^{-4}$, the probability that $A = 1$ is ∼0.01, but the stQSSA for A overestimates it as ∼0.1, which is 10 times larger (i.e., a 10-fold error)

DNA (D:P). Note that the number of unbound and bound DNA can be interpreted as the number of unbound and bound DNA binding sites. In this model, because the reversible binding reaction between D and P is much faster than the other reactions (i.e., the production and the decay of $\mathrm{M_R}$ and $\mathrm{M_A}$), the variables (D and D:P) rapidly reach their QSS. Thus, by replacing them with their stQSSAs (D_{tq} and D:P_{tq}), we can obtain a reduced model (Fig 2a right, Table S2). The reduced model consists of only the slow variables, M_{R} and M_{A}, because D_{tq} and D:P_{tq} are fully determined by the conserved total number of the DNA ($D_{\mathrm{T}} = D + D$:P) and the conserved total number of the transcription factor ($P_{\mathrm{T}} = P + D$:P), as illustrated in Table S2. This elimination of the fast variables, which are the major source of computational cost, greatly reduces the computation time of stochastic simulations [27–29].

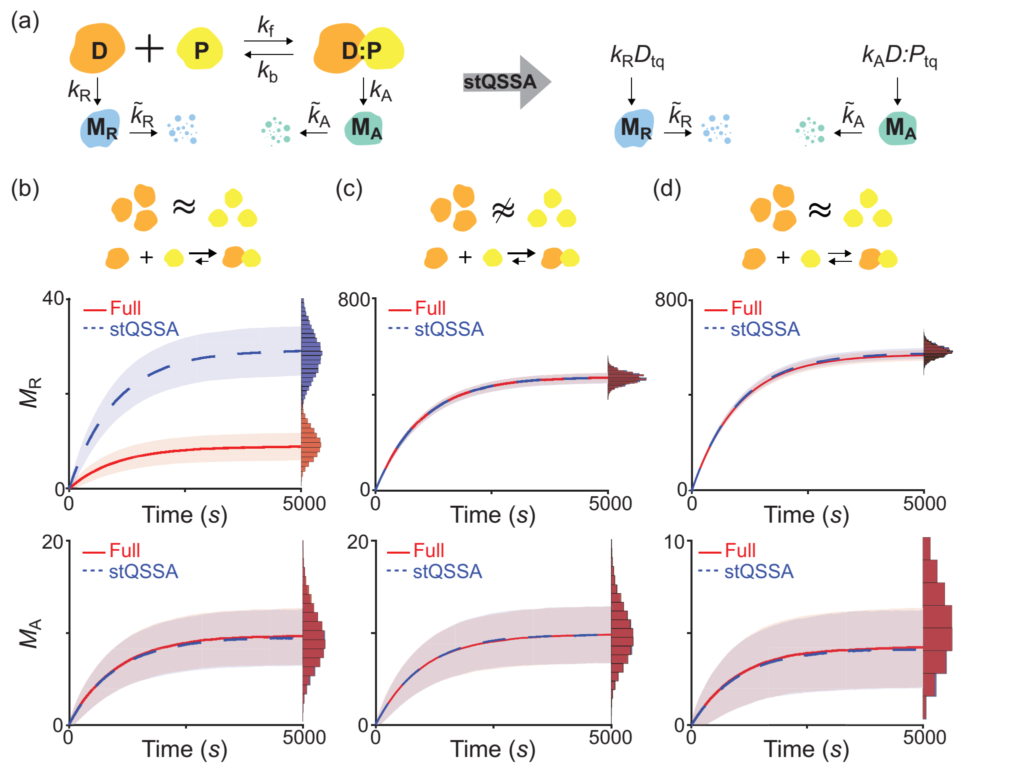

Fig 2. When DNA and a transcription factor bind tightly and their levels are similar, the stQSSA overestimates the number of the unbound DNA. (a) Full model diagram of a gene regulatory network containing a rapid reversible binding between DNA (D) and a transcription factor (P) to form a complex (D:P) (left, Table S1). The transcription rates of $\mathrm{M_R}$ and $\mathrm{M_A}$ are proportional to D and D:P, respectively. By replacing D and D:P with their stQSSAs (D_{tq} and D:P_{tq}), we can obtain a reduced model which consists of only slowly varing M_{R} and M_{A} (right, Table S2). **(b-d)** Trajectories of M_{R} (top) and M_{A} (bottom) from the full model (red) and the reduced model (blue) simulated using the Gillespie algorithm (see Tables S1 and S2 for propensity functions). The lines with colored ranges and the histograms represent the mean $\pm$ standard deviation and the stationary distribution of 10^4 trajectories, respectively. When D_{T} and P_{T} are the same ($D_{\mathrm{T}} = P_{\mathrm{T}} = 10$) and the binding is tight ($K_{\mathrm{d}} = 10^{-2}$), the M_{R} trajectories simulated with the reduced model largely exceed those simulated with the full model (b top) because D_{tq} overestimates the stochastic QSSA for D ($\langle D \rangle$). On the other hand, D:P_{tq} accurately approximates the stochastic QSSA for D:P ($\langle D{:}P \rangle$), and thus the reduced model accurately captures the dynamics of M_{A} (b bottom). If D_{T} is not similar to P_{T} ($D_{\mathrm{T}} = 15, P_{\mathrm{T}} = 10$) (c) or the binding is weak ($K_{\mathrm{d}} = 10$) (d), D_{tq} and D:P_{tq} accurately approximate $\langle D \rangle$ and $\langle D{:}P \rangle$, respectively, so that the reduced model accurately captures the dynamics of both M_{R} and M_{A} of the full model.

To test whether the reduced model accurately captures the dynamics of the full model, we compared their stochastic simulations with the Gillespie algorithm (see Tables S1 and S2 for propensity functions) [51]. When D_{T} and P_{T} are the same and the binding between D and P is tight, M_{R} simulated with the reduced model largely exceeds M_{R} simulated with the full model (Fig 2b top) because the stQSSA (D_{tq}) overestimates the stochastic QSSA for the number of the unbound DNA ($\langle D\rangle$), which determines the transcription rate of $\mathrm{M_R}$ (Fig 2a), as seen in Fig 1f. On the other hand, when D_{T} is not similar to P_{T} (Fig 2c top) or the binding is weak (Fig 2d top), D_{tq} accurately approximates $\langle D\rangle$ as seen in Fig 1d, and thus the reduced model accurately captures the dynamics of M_{R} in the full model.

Unlike M_{R} (Fig 2b top), the stochastic dynamics of M_{A} of the reduced model and the full model are identical (Fig 2b-2d bottom) because the stQSSA for $D{:}P$ ($D{:}P_{\mathrm{tq}}$) always accurately approximates the stochastic QSSA for the number of the bound DNA ($\langle D{:}P\rangle$), which determines the transcription of M_{A} (Fig 1c). Taken together, the stQSSA can be used to describe transcriptional activation depending on bound DNA under any conditions (Fig 2b-2d bottom). On the other hand, it needs to be restrictively used to describe transcriptional repression depending on unbound DNA (Fig 2b-2d top).

stQSSA can distort oscillatory dynamics

To illustrate how the stQSSA distorts the dynamics when the molar ratio between tightly binding species is ∼1:1, we investigated the simple model where the molar ratio is conserved (Fig 2). However, the molar ratio can be varied (e.g., oscillate) in a living cell due to other reactions in a larger system. This raises the question of whether the model reduction based on the stQSSA is accurate or not if the molar ratio is temporarily ∼1:1. To investigate this, we used a modified Kim-Forger model, which describes the transcriptional negative feedback loop of the mammalian circadian clock [24, 48, 50]. In this model (Fig 3a top, Table S3), free activator (A) promotes the transcription of mRNA (M), and the protein translated from M produces repressor (R) passing through several steps ($\mathrm{P}_i,\ i = 1, 2, 3$). Then R reversibly binds with A to form a complex (A:R), which no longer promotes the transcription, and thus represses its own transcription. In this model, the reversible binding between R and A is much faster than the other reactions (i.e., production and decay). Thus, by replacing the fast variable A, which determines the transcription rate of M, with its stQSSA (A_{tq}), we can obtain a reduced model (Fig 3a bottom, Table S4). The reduced model consists of only the slow variables R_{T}, M and P_i because A_{tq} is fully determined by the conserved total number of the activator ($A_{\mathrm{T}} = A + A{:}R$) and the slowly varying total number of the repressor ($R_{\mathrm{T}} = R + A{:}R$), as illustrated in Table S4.

In the model, because R tightly binds with A, when $R_{\mathrm{T}}/A_{\mathrm{T}} \approx 1$, A_{tq} overestimates the stochastic QSSA for A ($\langle A\rangle$) and thus the transcription rate of M. As a result, when the trajectory of $R_{\mathrm{T}}/A_{\mathrm{T}}$ reaches close to 1 (dashed lines in Fig 3b), the transcription more frequently occurs in the reduced model (Fig 3b bottom) compared to the full model (Fig 3b top). This overestimated transcriptional activity leads to the shorter peak-to-peak periods of the reduced model compared to the full model (Fig 3c). On the other hand, when the degradation rate of R increases and thus the trajectory of $R_{\mathrm{T}}/A_{\mathrm{T}}$ stays near 1 for an extremely short time (Fig 3d dashed lines), the reduced model accurately captures the dynamics of the full model (Fig 3e). Taken together, if ∼1:1 molar ratio between the tightly binding activator and repressor of the transcriptional negative feedback loop persists for a considerable time, using the stQSSA overestimates the transitional activity and thus the frequency of oscillation.

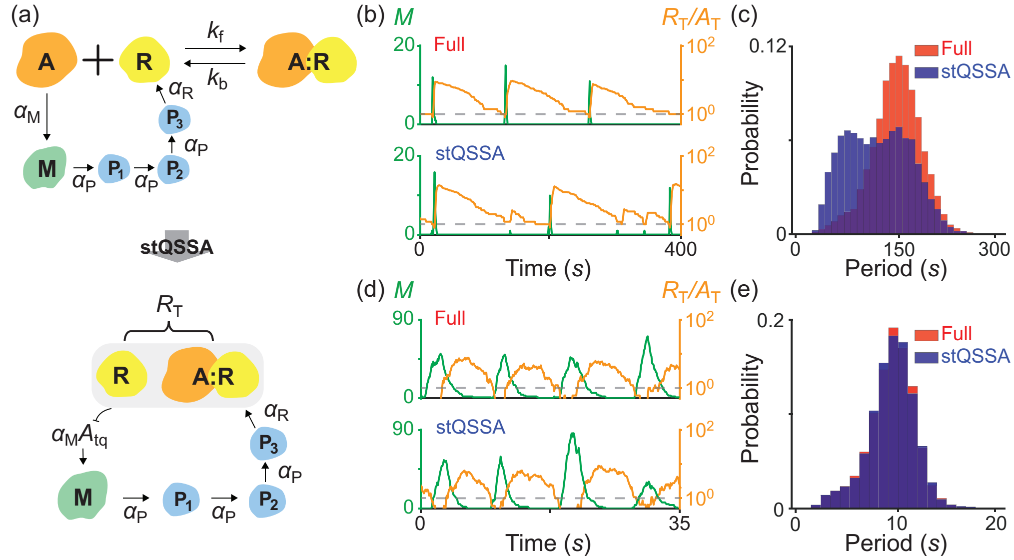

Fig 3. stQSSA can distort the dynamics of a biological oscillator. (a) Full model diagram of an oscillatory transcriptional negative feedback loop (top, Table S3). Unbound activator (A) promotes the transcription of mRNA (M), and the protein translated from M produces repressor (R) passing through several steps (P_i, $i = 1, 2, 3$). Then R binds with A to form a complex (A:R) which is transcriptionally inactive, and thus represses its own transcription. As the reversible binding between R and A is rapid, by replacing A with its stQSSA (A_{tq}), we can obtain a reduced model which consists of only slowly varying R_{T}, M, and P_i (bottom, Table S4). **(b-c)** Oscillatory trajectories of M (green) and $R_{\mathrm{T}}/A_{\mathrm{T}}$ (orange) simulated with the full model (b top) and the reduced model (b bottom), using the Gillespie algorithm (see Tables S3 and S4 for propensity functions). When R binds with A tightly ($K_{\mathrm{d}} = 10^{-4}$) both the full model and the reduced model show the oscillatory behaviors. However, when the trajectory of $R_{\mathrm{T}}/A_{\mathrm{T}}$ stays near 1 (dashed lines in b), A_{tq} overestimates the stochastic QSSA for A ($\langle A \rangle$), and thus the transcription more frequently occurs in the reduced model (b bottom) compared to the full model (b top). As a result, the reduced model predicts a shorter period than the full model (c). **(d-e)** On the other hand, when the degradation rate of R increases and thus the trajectory of $R_{\mathrm{T}}/A_{\mathrm{T}}$ stays near 1 for a short time (d; dashed lines), the reduced model accurately captures the dynamics of the full model (e).

stQSSA can distort the bistable dynamics

To investigate how the misuse of the stQSSA distorts the dynamics of a bistable switch, we used a previously developed bistable switch model for the maturation promoting factor, cyclin B/Cdc2, whose activation promotes mitosis (Fig 4a top, Table S5) [45, 55]. In the model, the inactive form of cyclin B/Cdc2 (P) is converted to an active form (M) by Cdc25 (D). Furthermore, as M activates D, which converts P to M, M promotes its own activation (i.e., form a positive feedback loop; see [45, 55] for details). The positive feedback loop is suppressed by Suc1 protein (B) as it binds with M to form a complex (M:B) which no longer activates D. The total activated cyclin B/Cdc2 (M and M:B) become P with the same constant rate. In this model, the reversible binding between M and B is much faster than the other reactions. Thus, by replacing the fast variable M with its stQSSA (M_{tq}) a reduced model can be derived (Fig 4a bottom, Table S6). The reduced model consists of only the slow variables, M_{T} and P, because M_{tq} is fully determined by the conserved total number of Suc1 ($B_{\mathrm{T}} = B + M{:}B$) and the slowly varying total number of the activated cyclin B/Cdc2 ($M_{\mathrm{T}} = M + M{:}B$), as illustrated in Table S6.

When M and B tightly bind, both the full model and the reduced model show the

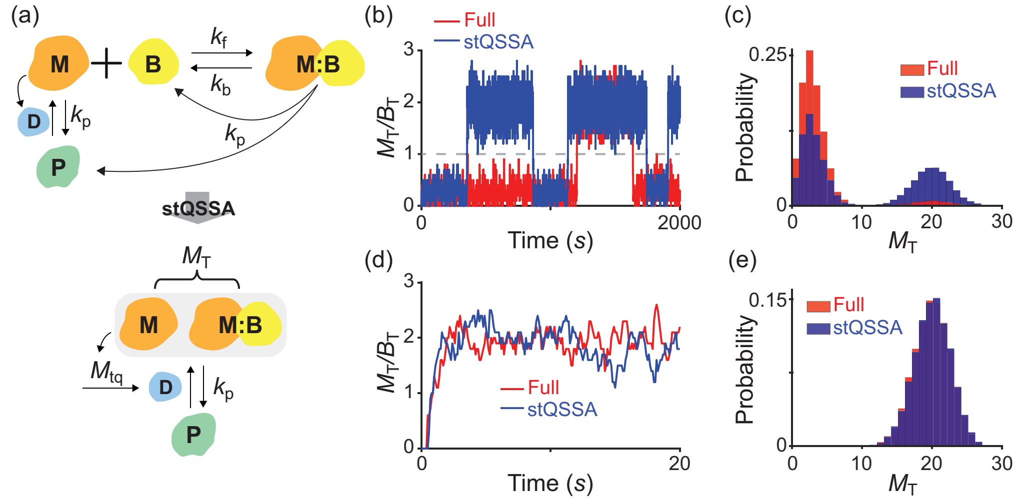

Fig 4. stQSSA can distort the dynamics of a bistable switch (a) Full model diagram of a bistable switch for mitosis (top, Table S5). The inactive form of cyclin B/Cdc2 (P) becomes an active form (M) by Cdc25 (D). In this process, M enhances its own activation by activating D, and thus forms a positive feedback loop (see [45, 55] for details). The positive feedback loop is suppressed as Suc1 protein (B) binds with M to form a complex (M:B) which does not activates D. The total activated cyclin B/Cdc2, M and M:B, becomes P with the same constant rate. As the reversible binding between M and B is rapid, by replacing M with its stQSSA (M_{tq}), we can obtain a reduced model which consists of only slowly varying M_{T} and P (bottom, Table S6). **(b-c)** Simulated trajectories (b) and the stationary distributions (c) of M_{T} from the full model and the reduced model using the Gillespie algorithm (see Tables S5 and S6 for propensity functions). When M binds with B tightly ($K_{\mathrm{d}} = 10^{-3}$), both the full model and the reduced model show the bistable behaviors between the upper and lower modes, which are separated by $M_{\mathrm{T}}/B_{\mathrm{T}} = 1$ (dashed line in b). However, because M_{tq} overestimates the stochastic QSSA for M ($\langle M\rangle$) when $M_{\mathrm{T}}/B_{\mathrm{T}}$ is close to 1, the trajectory from the reduced model is more attracted to the upper mode compared to the full model (b). As a result, the bimodal distribution of M_{T} from the reduced model is biased to the upper mode (c). **(d-e)** On the other hand, when the binding between M and B becomes weak ($K_{\mathrm{d}} = 10$), M_{tq} accurately estimates $\langle M\rangle$, and thus the reduced model accurately captures the dynamics of the full model, which no longer shows the bistable behavior.

bistable behaviors (i.e., bimodal stationary distributions) of M_{T} (Fig 4b). However, the trajectory of the reduced model is more attracted to the upper mode of M_{T} compared to the full model (Fig 4b and 4c). This dynamics biased to the upper mode occurs because M_{tq} overestimates the stochastic QSSA for M ($\langle M\rangle$) near the $M_{\mathrm{T}}/B_{\mathrm{T}} = 1$ region (Fig 4b dashed line), which separates the upper and lower modes. On the other hand, when the binding between M and B becomes weak, M_{tq} accurately approximates the stochastic QSSA for M even when M_{T} is similar to B_{T}. Thus, the reduced model accurately captures the dynamics of the full model, which no longer shows bistable behavior (Fig 4d and 4e). Taken together, when the binding between activated Cyclin B/Cdc2 and Suc1 protein is tight, which is essential to generate the bistable switch, using the stQSSA overestimates the activation of Cyclin B/Cdc2 and distorts the dynamics of the bistable switch.

An alternative approach when the stQSSA is not applicable

In the presence of a rapid and tight reversible binding between species whose molar ratio is ∼1:1, the reduction of stochastic models with the stQSSA for the number of the unbound species can cause errors (Figs 2b, 3c, and 4c). In such cases, due to the tight binding, the two species tend to bind until no molecules of one species left (Fig 1f). Specifically, if $A_\mathrm{T} \leq B_\mathrm{T}$ ($A_\mathrm{T} \geq B_\mathrm{T}$), the majority of the A (B) will be bound. Thus, in the presence of tight binding, we can assume that the stationary distributions of A or B are concentrated on 0 and 1. This low-state assumption allows us to derive the simple approximation for the stochastic QSSA ($\langle A \rangle$ in Eq. (4)) (see Methods for details):

$$\langle A \rangle \approx A_\mathrm{lq} = \begin{cases} \frac{(A_\mathrm{T}-B_\mathrm{T}+1)(A_\mathrm{T}-B_\mathrm{T}+B_\mathrm{T}K_\mathrm{d})}{A_\mathrm{T}-B_\mathrm{T}+B_\mathrm{T}K_\mathrm{d}+1} & \text{if } A_\mathrm{T} \geq B_\mathrm{T}, \\ \frac{A_\mathrm{T}K_\mathrm{d}}{B_\mathrm{T}-A_\mathrm{T}+A_\mathrm{T}K_\mathrm{d}+1} & \text{if } A_\mathrm{T} < B_\mathrm{T}. \end{cases} \tag{9}$$

We will refer to this approximation as the stochastic "low-state" QSSA (slQSSA).

The accuracy of the slQSSA for A (Eq. (9)) is expected to increase when $A_\mathrm{T}K_\mathrm{d}$ decreases because $A_\mathrm{T}K_\mathrm{d}$ is an approximated number of the unbound A. On the other hand, the accuracy of the stQSSA for A decreases as $A_\mathrm{T}K_\mathrm{d}$ decreases (Fig 1d). To investigate this, we calculated the maximum relative error of A_tq ($R_\mathrm{A} = \left|\frac{A_\mathrm{tq}-\langle A \rangle}{\langle A \rangle}\right|$) and A_lq ($R_\mathrm{A}^\mathrm{lq} = \left|\frac{A_\mathrm{lq}-\langle A \rangle}{\langle A \rangle}\right|$) to the stochastic QSSA for A ($\langle A \rangle$) for each $A_\mathrm{T}K_\mathrm{d}$ and K_d (Fig 5a and 5b). As expected, when $A_\mathrm{T}K_\mathrm{d}$ is low and high, the slQSSA and the stQSSA are accurate, respectively. In particular, when $A_\mathrm{T}K_\mathrm{d} < 10^{-1}$ and $A_\mathrm{T}K_\mathrm{d} > 10^{1}$, R_A^lq and R_A are less than 0.1 (i.e., the relative errors are less than 10%), respectively.

The parameters used in Figs 2b (triangle), 3b (circle), and 4b (square) are located in the region where the stQSSA is inaccurate (Fig 5a) but the slQSSA is accurate (Fig 5b). Therefore, with these parameters, the reduced models obtained by using the slQSSAs accurately capture the dynamics of the full models for the simple gene regulatory network (Fig 5c, Table S2), the transcriptional negative feedback loop (Fig 5d, Table S4), and the bistable switch for mitosis (Fig 5e, Table S6), unlike the stQSSA (Figs 2b, 3c, and 4c). Furthermore, by allowing A or B to reach more than two states (e.g., 0, 1, and 2), more accurate slQSSAs can be derived (see Methods for details). In particular, the relative errors of the slQSSAs derived by allowing the 3/4/5 states are less than 0.1 when $A_\mathrm{T}K_\mathrm{d}$ is less than 2/5/10, respectively (Fig S2). Consequently, if $A_\mathrm{T}K_\mathrm{d} < 10^1$ and thus the stQSSA is inaccurate, the slQSSA can be used to approximate the stochastic QSSA for A (Fig 5f). Taken together, by using either stQSSA or slQSSA depending on $A_\mathrm{T}K_\mathrm{d}$, we can always accurately reduce multiscale stochastic biochemical systems with rapid reversible bindings.

Discussion

Reversible binding between molecules—for example, between DNA and a transcription factor, a ligand and a receptor, and an enzyme and a substrate—is a fundamental reaction for numerous biological functions [41]. As the reversible binding reactions occur typically on a timescale of 1∼1000ms, which is much faster than the other reactions (e.g., 30min for a mammalian mRNA transcription or a protein translation and 10h for their typical lifetimes) [42], a system containing the rapid reversible binding becomes a multi-timescale system. In such multi-timescale systems, the rapid reversible binding prohibitively increases the computational cost of stochastic simulations. Accordingly, to accelerate stochastic simulations, various methods have been developed [43, 56]. In particular, the model reduction using the stQSSA has successfully simplified various stochastic models in numerous studies [7, 30–33, 38, 39]. Thus, it has been commonly believed that the stQSSA is generally accurate for any conditions, until a recent counterexample was identified [40]. In this work, we rigorously derived the validity conditions for using the stQSSA to reduce stochastic models with a rapid reversible binding. Specifically, we showed that the relative error of the stQSSA for the number of unbound species (R_A) mainly depends on the relative sensitivity of the stQSSA (S_A, Eq.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.08.438974; this version posted April 9, 2021.

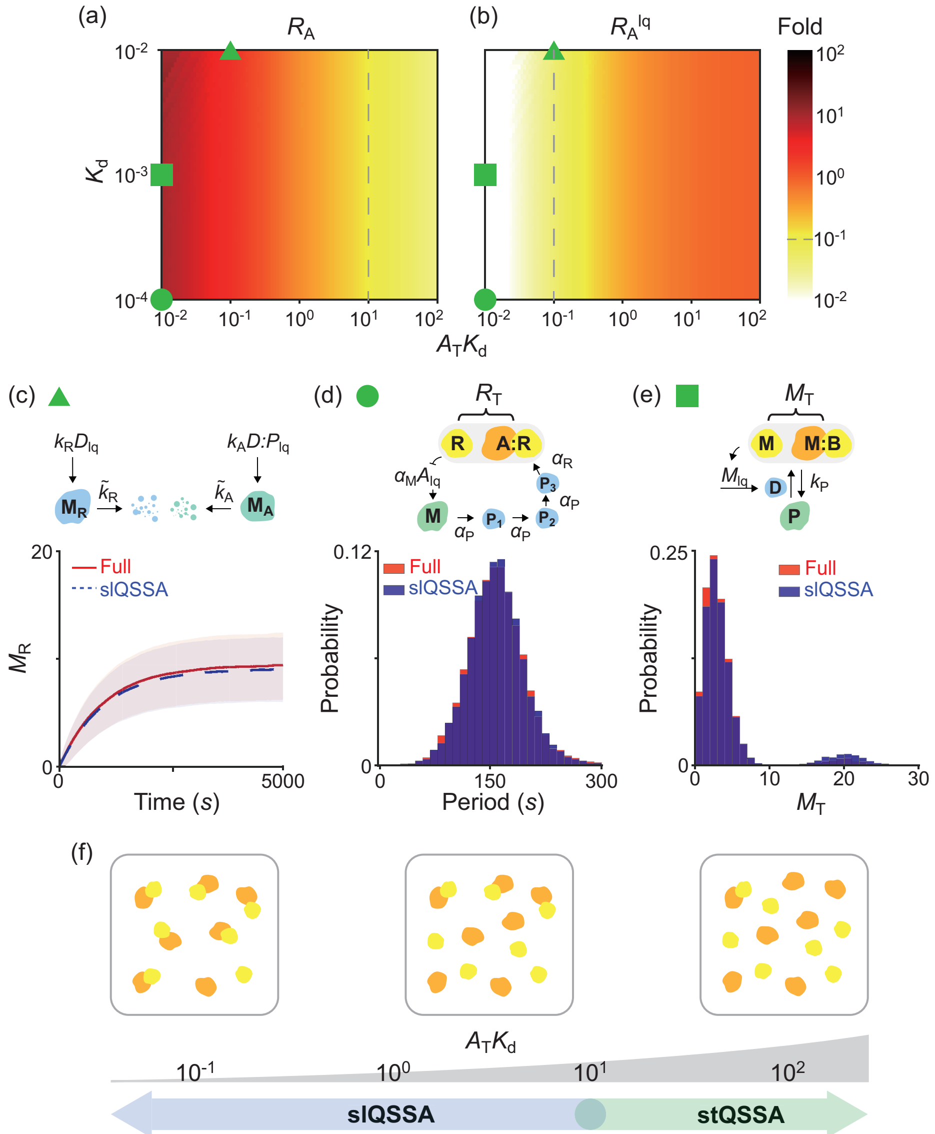

Fig 5. slQSSA can be used to reduce multiscale stochastic biochemical systems containing rapid reversible bindings when the stQSSA is not applicable. (a-b) Heat maps of the relative errors ($R_{\mathrm{A}} = \left|\frac{A_{\mathrm{tq}} - \langle A \rangle}{\langle A \rangle}\right|$ and $R_{\mathrm{A}}^{\mathrm{lq}} = \left|\frac{A_{\mathrm{lq}} - \langle A \rangle}{\langle A \rangle}\right|$) when the stQSSA ($A_{\mathrm{tq}}$) and the two-state slQSSA (A_{lq}) approximate the stochastic QSSA for A ($\langle A \rangle$) in the reversible binding reaction (Eq. (1)). Color represents the maximum value of R_{A} and $R_{\mathrm{A}}^{\mathrm{lq}}$ for each $A_{\mathrm{T}}K_{\mathrm{d}}$ and K_{d} when B_{T} varies, and the dashed lines represent when those values are 0.1. When $A_{\mathrm{T}}K_{\mathrm{d}}$ are high and low, the stQSSA and the slQSSA are accurate, respectively. The parameters used in Figs 2b (triangle), 3b (circle), and 4b (square) are located in the region where the slQSSA (b), but not the stQSSA (a), is accurate (the circle is actually located outside of the heat maps; $A_{\mathrm{T}}K_{\mathrm{d}} = 5 \times 10^{-4}$ and $K_{\mathrm{d}} = 10^{-4}$). **(c-e)** As a result, the full models are successfully reduced with the slQSSA (c-e) but not the stQSSA (Figs 2b, 3c, and 4c). See Tables S2, S4, and S6 for the propensity functions used for the simulations. **(f)** The adaptive use of the stQSSA and the slQSSA to approximate the stochastic QSSA for A when $A_{\mathrm{T}}K_{\mathrm{d}} > 10^1$ and otherwise, respectively, guarantees the successful reduction of stochastic models containing rapid reversible bindings. Note that when $10^{-1} < A_{\mathrm{T}}K_{\mathrm{d}} < 10^1$, the slQSSAs with more than two states need to be used (see Fig S2 for details).

(8)), which attains maximum value $\frac{1}{\sqrt{4A_{\mathrm{T}}K_{\mathrm{d}}}}$ at $A_{\mathrm{T}} = B_{\mathrm{T}} + K_{\mathrm{d}}$. This allowed us to find that the stQSSA for the number of the unbound species is inaccurate if their molar ratio is $\sim$1:1 and their binding is tight (Fig 1f). In that case, the stQSSA highly overestimates the number of the unbound species. Therefore, the reduced models obtained by using the stQSSA distort the dynamics of the gene regulatory model (Fig 2b), the transcriptional negative feedback loop model for circadian rhythms (Fig 3c), and the bistable switch model for mitosis (Fig 4c).

Interestingly, even in the invalid range of the stQSSA, identified in this study, the deterministic tQSSA is known to be accurate [7, 47, 49]. Indeed, for all examples considered in our work (Figs 2b, 3b, and 4b), the deterministic simulations with the tQSSA are accurate, unlike the stochastic simulations. This indicates that it is risky to investigate the validity conditions of the stQSSA solely based on the validity conditions of the deterministic tQSSA. Instead, direct derivation of the relative error of the stQSSA is needed as demonstrated in this study (Eq. (7)). It would be interesting in future work to perform such error analysis for more complex examples, such as coupled enzymatic networks with multiple rapid reversible bindings [26, 57, 58].

While the deterministic tQSSA (Eq. (3)) was used to approximate the stochastic QSSA for the number of reversibly binding species in this work, a simpler deterministic QSSA referred to as the "standard" QSSA (sQSSA) is more widely used to approximate the stochastic QSSA due to its simplicity [8–21, 27, 32]. For instance, the stochastic sQSSA for A in Eq. (1), which has the Michaelis-Menten type form:

$$A_{\mathrm{sq}} = \frac{A_{\mathrm{T}}B}{B + K_{\mathrm{d}}},$$

has been widely used as a propensity function for Gillespie algorithm. However, it is less accurate than the stQSSA (Eq. (5)) [38, 39]. This is why many examples showing the inaccuracy of the stochastic sQSSA have been reported [33–39], whereas only one example showing the inaccuracy of the stQSSA has been reported [40]. Furthermore, the "pre-factor" QSSA (pQSSA), which is more accurate than sQSSA, has also been used for stochastic simulations [59, 60]. However, recent studies have shown that the stQSSA is more accurate than the stochastic pQSSA (see [38, 39] for details).

The accuracy of the stQSSA for the number of the unbound species depends on both the molar ratio between reversibly binding species and the tightness of their binding (Fig 1d). However, as the molar ratio typically varies in larger models containing reversible binding, practically, the accuracy is mainly determined by the tightness of binding. Specifically, for the relative error of the stQSSA to be less than 0.1, $A_{\mathrm{T}}K_{\mathrm{d}}$ ($\approx$ the number of the unbound A) should be larger than 10 (Fig 5a dashed line). This $A_{\mathrm{T}}K_{\mathrm{d}}$ value-based criteria explains the controversy about the accuracy of the stQSSA in previous studies. That is, $A_{\mathrm{T}}K_{\mathrm{d}}$ was less than 10 in a previous study where the reduced model obtained by using the stQSSA was inaccurate [40]. On the other hand, $A_{\mathrm{T}}K_{\mathrm{d}}$ were much greater than 10 in all of the examples investigated in previous studies reporting the accuracy of the stQSSA [7, 33, 38, 39, 52, 53]. Furthermore, the stQSSA always accurately approximates the stochastic QSSA for the number of the bound species (Fig 1c). This explains why the stQSSA was accurate in previous studies where the stQSSA was used to approximate the number of enzyme-substrate complex [30–33].

In real biological systems, the validity condition for the stQSSA for the number of the unbound species ($A_{\mathrm{T}}K_{\mathrm{d}} > 10$) is not always guaranteed. Specifically, the range of $A_{\mathrm{T}}K_{\mathrm{d}}$ can span approximately from 10^{-2} to 10^{10} since the volume of the human cells is $10^{-15} \sim 10^{-14}m^3$, the protein-protein dissociation constant is $10fM \sim 1\mu M$ (i.e., $10^{12} \sim 10^{20}m^{-3}$), and the numbers of molecules is $10^0 \sim 10^4$ [42, 61]. Accordingly, as an alternative for the stQSSA, we derived the slQSSA, which accurately approximates the stochastic QSSA when $A_{\mathrm{T}}K_{\mathrm{d}}$ is less than 10. Specifically, the relative error of the slQSSA, unlike that of the stQSSA (Fig 5a and 5b), decreases as $A_\mathrm{T}K_\mathrm{d}$ decreases because the slQSSA relies on the assumption that the stationary distributions of the number of the unbound species ($\approx A_\mathrm{T}K_\mathrm{d}$) are concentrated on the few lowest states. Taken together, by using the stQSSA and the slQSSA when the $A_\mathrm{T}K_\mathrm{d}$ value is greater and less than 10, respectively, one can always accurately simplify stochastic models containing rapid reversible binding reactions to accelerate simulation and also facilitate stochastic analysis (Fig 5f).

Methods

Exact bounds for the relative error of the stQSSA to the stochastic QSSA

In this section, we derive the exact upper and lower bounds for $R_\mathrm{A} = \left|\frac{A_\mathrm{tq}-\langle A\rangle}{\langle A\rangle}\right|$ (Eq. (7)) where A_tq and $\langle A\rangle$ are the stQSSA and the stochastic QSSA for A, respectively. From the CME describing the reversible binding reaction (Eq. (1)), the following steady-state moment equation can be derived:

$$\langle k_\mathrm{f} A \cdot B/\Omega\rangle = \langle k_\mathrm{b} C\rangle, \tag{10}$$

where $\langle\cdot\rangle$ is the stationary expectation. Eq. (10) becomes $\langle A\cdot(B_\mathrm{T} - A_\mathrm{T} + A)\rangle = K_\mathrm{d}\langle A_\mathrm{T} - A\rangle$ by using the definitions $A_\mathrm{T} = A + C$, $B_\mathrm{T} = B + C$, and $K_\mathrm{d} = k_\mathrm{b}\Omega/k_\mathrm{f}$. Since A_T and B_T are invariant under the reversible binding reactions in Eq. (1), we obtain $\langle A^2\rangle - (A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})\langle A\rangle - A_\mathrm{T}K_\mathrm{d} = 0$, and by using the relation $\langle A^2\rangle = \mathrm{Var}(A) + \langle A\rangle^2$, we get the following quadratic equation:

$$\langle A\rangle^2 - (A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})\langle A\rangle - A_\mathrm{T}K_\mathrm{d} + \mathrm{Var}(A) = 0. \tag{11}$$

The non-negative root of this quadratic equation becomes $\langle A\rangle$:

$$\langle A\rangle = \frac{1}{2}\left\{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d}) + \sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d} - 4\mathrm{Var}(A)}\right\}. \tag{12}$$

By subtracting Eq. (12) from Eq. (5), we get

$$\begin{aligned} &A_\mathrm{tq} - \langle A\rangle &(13)\\ =&\frac{1}{2}\left\{\sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}} - \sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d} - 4\mathrm{Var}(A)}\right\}\\ =&\frac{2\mathrm{Var}(A)}{\sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}} + \sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d} - 4\mathrm{Var}(A)}}. &(14)\end{aligned}$$

Since $0 \le (A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d} - 4\mathrm{Var}(A) \le (A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}$, we get the bounds for $A_\mathrm{tq} - \langle A\rangle$ from Eq. (14):

$$\frac{\mathrm{Var}(A)}{\sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}}} \le A_\mathrm{tq} - \langle A\rangle \le \frac{2\mathrm{Var}(A)}{\sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}}}. \tag{15}$$

By dividing Eq. (15) by $\langle A\rangle$, we can get the bounds for the relative error, $R_\mathrm{A} = \left|\frac{A_\mathrm{tq}-\langle A\rangle}{\langle A\rangle}\right|$ as follows:

$$\frac{\mathrm{Var}(A)}{\langle A\rangle}\frac{1}{\sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}}} \le R_\mathrm{A} \le 2\frac{\mathrm{Var}(A)}{\langle A\rangle}\frac{1}{\sqrt{(A_\mathrm{T} - B_\mathrm{T} - K_\mathrm{d})^2 + 4A_\mathrm{T}K_\mathrm{d}}}.$$

This can be re-expressed as $F_{\mathrm{A}}S_{\mathrm{A}} \leq R_{\mathrm{A}} \leq 2F_{\mathrm{A}}S_{\mathrm{A}}$ (Eq. (7)) because $\frac{\mathrm{Var}(A)}{\langle A\rangle}$ is the Fano factor of A (F_{A}), and $\frac{1}{\sqrt{(A_{\mathrm{T}}-B_{\mathrm{T}}-K_{\mathrm{d}})^2+4A_{\mathrm{T}}K_{\mathrm{d}}}}$ is the relative sensitivity of A_{tq}, i.e., $S_{\mathrm{A}} = \frac{1}{A_{\mathrm{tq}}}\left|\frac{dA_{\mathrm{tq}}}{dB_{\mathrm{T}}}\right|$.

The relative sensitivity, S_{A}, attains the maximum value $\frac{1}{\sqrt{4A_{\mathrm{T}}K_{\mathrm{d}}}}$ when the term in the square root of the denominator has the minimum value, i.e., $B_{\mathrm{T}} = A_{\mathrm{T}} - K_{\mathrm{d}}$ (Eq. (8)). In particular, S_{A} has a large maximum value when $K_{\mathrm{d}} \ll 1$ at $A_{\mathrm{T}} = B_{\mathrm{T}} + K_{\mathrm{d}} \approx B_{\mathrm{T}}$. On the other hand, if $A_{\mathrm{T}} \ll B_{\mathrm{T}}$, $S_{\mathrm{A}} \approx 0$ because the majority of A presents in the bound state regardless of B_{T} (i.e.,$\frac{dA_{\mathrm{tq}}}{dB_{\mathrm{T}}} \approx 0$). When $A_{\mathrm{T}} \geq B_{\mathrm{T}}$, $\frac{dA_{\mathrm{tq}}}{dB_{\mathrm{T}}} \approx 1$ because as B_{T} decreases by one, approximately one A is released from the complex. In this case, if $A_{\mathrm{T}} \gg B_{\mathrm{T}}$, the majority of A are free and thus $\frac{1}{A_{\mathrm{tq}}} \approx \frac{1}{A_{\mathrm{T}}-B_{\mathrm{T}}} \approx 0$, leading to $S_{\mathrm{A}} \approx 0$. However, if $A_{\mathrm{T}} \approx B_{\mathrm{T}}$, the majority of A is sequestered by B, $A_{\mathrm{tq}} \approx 0$, leading to $S_{\mathrm{A}} \gg 1$. When binding is weak ($K_{\mathrm{d}} \gg 1$), $S_{\mathrm{A}} \approx 0$ because the number of A, which is approximated by A_{tq}, changes little as B_{T} changes (i.e., $\frac{dA_{\mathrm{tq}}}{dB_{\mathrm{T}}} \approx 0$). Taken together, S_{A} is large only when the binding reaction is tight ($K_{\mathrm{d}} \ll 1$) and the binding species are present with 1:1 molar ratio ($A_{\mathrm{T}} \approx B_{\mathrm{T}}$).

Derivation of the stochastic QSSA and the slQSSA

Here we derive the stochastic QSSA for A ($\langle A\rangle$, Eq. (4)). Let $p(l)$ be the probability that $A = l$ at its stationary distribution (i.e., the probability that $A(\infty) = l$). Then the following recurrence relation of $p(l)$ can be obtained from the steady-state CME:

$$(l+1)(B_{\mathrm{T}}-A_{\mathrm{T}}+l+1)p(l+1)-l(B_{\mathrm{T}}-A_{\mathrm{T}}+l)p(l)+K_{\mathrm{d}}(A_{\mathrm{T}}-l+1)p(l-1)-K_{\mathrm{d}}(A_{\mathrm{T}}-l)p(l) = 0. \tag{16}$$

Let $A_0 = \max\{A_{\mathrm{T}} - B_{\mathrm{T}}, 0\}$. Since A_0 is the lowest state that A can reach, $p(l) = 0$ for $l < A_0$. Then we can inductively prove that the following relation satisfies Eq. (16):

$$p(l + A_0) = \begin{cases} \pi(l + A_0)p(A_0) & \text{for } 0 \leq l \leq A_{\mathrm{T}} - A_0, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where $\pi(l) = \frac{K_{\mathrm{d}}^{l-A_0}\min(A_{\mathrm{T}},B_{\mathrm{T}})!|A_{\mathrm{T}}-B_{\mathrm{T}}|!}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}$. Then, because $\sum p(l) = 1$, $p(l) = \pi(l)\cdot\left(\sum_{l=A_0}^{A_{\mathrm{T}}}\pi(l)\right)^{-1}$ if $A_0 \leq l \leq A_{\mathrm{T}}$, and $p(l) = 0$ otherwise by Eq. (17). Therefore, we can obtain the stationary average number of A (Eq. (4)) as

$$\begin{aligned}\langle A\rangle &= \sum_{l=A_0}^{A_{\mathrm{T}}} l\pi(l)\cdot\left(\sum_{l=A_0}^{A_{\mathrm{T}}}\pi(l)\right)^{-1} \\ &= \left(\sum_{l=A_0}^{A_{\mathrm{T}}}\frac{lK_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)\cdot\left(\sum_{l=A_0}^{A_{\mathrm{T}}}\frac{K_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)^{-1}.\end{aligned}$$

Next we derive the slQSSA, which is the approximation for Eq. (4). In the presence of tight binding, we can assume that the stationary distributions of A and B are concentrated on the states $\{0, 1\}$ when $A_{\mathrm{T}} < B_{\mathrm{T}}$ and $A_{\mathrm{T}} \geq B_{\mathrm{T}}$, respectively. Since when the distribution of B is concentrated on 0 and 1, the distribution of A is concentrated on $A_{\mathrm{T}} - B_{\mathrm{T}}$ and $A_{\mathrm{T}} - B_{\mathrm{T}} + 1$, we can simply say that the distribution of A is concentrated on A_0 and $A_0 + 1$. Thus, by assuming that $p(l) = \pi(l)\cdot\left(\sum_{l=A_0}^{A_{\mathrm{T}}}\pi(l)\right)^{-1}$ is approximately zero for $l > A_0 + 1$ and $\sum_{l=A_0}^{A_{\mathrm{T}}}\pi(l) \approx \sum_{l=A_0}^{A_0+1}\pi(l)$, we can derive the two-state slQSSA for A (Eq. (9)) as follows:

$$
\begin{aligned}
\langle A\rangle &\approx \left(\sum_{l=A_0}^{A_0+1} \frac{lK_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)\cdot\left(\sum_{l=A_0}^{A_0+1} \frac{K_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)^{-1} \\
&= \begin{cases} (A_{\mathrm{T}}-B_{\mathrm{T}}+B_{\mathrm{T}}K_{\mathrm{d}})\cdot\left(1+\frac{B_{\mathrm{T}}K_{\mathrm{d}}}{A_{\mathrm{T}}-B_{\mathrm{T}}+1}\right)^{-1} & \text{if } A_{\mathrm{T}}\geq B_{\mathrm{T}} \\ \frac{A_{\mathrm{T}}K_{\mathrm{d}}}{B_{\mathrm{T}}-A_{\mathrm{T}}+1}\left(1+\frac{A_{\mathrm{T}}K_{\mathrm{d}}}{B_{\mathrm{T}}-A_{\mathrm{T}}+1}\right)^{-1} & \text{if } A_{\mathrm{T}}<B_{\mathrm{T}} \end{cases} \\
&= \begin{cases} \frac{(A_{\mathrm{T}}-B_{\mathrm{T}}+1)(A_{\mathrm{T}}-B_{\mathrm{T}}+B_{\mathrm{T}}K_{\mathrm{d}})}{A_{\mathrm{T}}-B_{\mathrm{T}}+B_{\mathrm{T}}K_{\mathrm{d}}+1} & \text{if } A_{\mathrm{T}}\geq B_{\mathrm{T}} \\ \frac{A_{\mathrm{T}}K_{\mathrm{d}}}{B_{\mathrm{T}}-A_{\mathrm{T}}+A_{\mathrm{T}}K_{\mathrm{d}}+1} & \text{if } A_{\mathrm{T}}<B_{\mathrm{T}} \end{cases}.
\end{aligned}
$$

In general, for any integer $k\geq 2$, we can derive the k-state slQSSA as

$$
A_{\mathrm{lq}}^k := \left(\sum_{l=A_0}^{A_0+k-1} \frac{lK_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)\cdot\left(\sum_{l=A_0}^{A_0+k-1} \frac{K_{\mathrm{d}}^l}{l!(A_{\mathrm{T}}-l)!(B_{\mathrm{T}}-A_{\mathrm{T}}+l)!}\right)^{-1}. \tag{18}
$$

We provide a Matlab code, LQSSA, that can be used to calculate Eq. (18).

Supporting information

S1 Appendix. Supplementary Methods, Tables S1-S6, and Figs S1-S2. (PDF)

Acknowledgments

We thank Krešimir Josić for valuable comments. This work was funded by NRF-2016 RICIB 3008468 (JKK), the Institute for Basic Science IBS-R029-C3 (JKK), and NRF-2019-Fostering Core Leaders of the Future Basic Science Program/Global Ph.D. Fellowship Program 2019H1A2A1075303 (HH).

Author Contributions

All authors designed the study and performed mathematical analysis. YS performed the computation and all authors analyzed the computation results. YS and JKK wrote the draft and all authors revised the manuscript.

References

1. Tyson JJ, Novak B. A dynamical paradigm for molecular cell biology. Trends Cell Biol. 2020;.

2. Hill AV. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J Physiol. 1910;40:4–7.

3. Michaelis L, Menten ML. Die kinetik der invertinwirkung. Biochem Z. 1913;49(333-369):352.

4. Gunawardena J. Time-scale separation–Michaelis and Menten's old idea, still bearing fruit. FEBS J. 2014;281(2):473–488.

5. Laidler KJ. Theory of the transient phase in kinetics, with special reference to enzyme systems. Can J Chem. 1955;33(10):1614–1624.

6. Morrison JF. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. Biochim Biophys Acta (BBA)-Enzymology. 1969;185(2):269–286.

7. Kim JK, Tyson JJ. Misuse of the Michaelis–Menten rate law for protein interaction networks and its remedy. PLoS Comput Biol. 2020;16(10):e1008258.

8. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000;403(6767):335–338.

9. Gonze D, Halloy J, Gaspard P. Biochemical clocks and molecular noise: Theoretical study of robustness factors. J Chem Phys. 2002;116(24):10997–11010.

10. Gonze D, Halloy J, Goldbeter A. Deterministic versus stochastic models for circadian rhythms. J Biol Phys. 2002;28(4):637–653.

11. Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. Science. 2005;307(5717):1965–1969.

12. Scott M, Ingalls B, Kærn M. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. Chaos. 2006;16(2):026107.

13. Tian T, Burrage K. Stochastic models for regulatory networks of the genetic toggle switch. Proc Natl Acad Sci U S A. 2006;103(22):8372–8377.

14. Çağatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Süel GM. Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell. 2009;139(3):512–522.

15. Ouattara DA, Abou-Jaoudé W, Kaufman M. From structure to dynamics: Frequency tuning in the p53-Mdm2 network. II: Differential and stochastic approaches. J Theor Biol. 2010;264(4):1177–1189.

16. Gonze D, Abou-Jaoudé W, Ouattara DA, Halloy J. How Molecular Should Your Molecular Model Be?: On the Level of Molecular Detail Required to Simulate Biological Networks in Systems and Synthetic Biology. Methods Enzymol. 2011;487:171–215.

17. Smadbeck P, Kaznessis Y. Stochastic model reduction using a modified Hill-type kinetic rate law. J Chem Phys. 2012;137(23):234109.

18. Riba A, Bosia C, El Baroudi M, Ollino L, Caselle M. A combination of transcriptional and microRNA regulation improves the stability of the relative concentrations of target genes. PLoS Comput Biol. 2014;10(2):e1003490.

19. Dovzhenok AA, Baek M, Lim S, Hong CI. Mathematical modeling and validation of glucose compensation of the neurospora circadian clock. Biophys J. 2015;108(7):1830–1839.

20. Zhang W, Tian T, Zou X. Negative feedback contributes to the stochastic expression of the interferon-β gene in virus-triggered type I interferon signaling pathways. Math Biosci. 2015;265:12–27.

21. Schuh L, Saint-Antoine M, Sanford EM, Emert BL, Singh A, Marr C, et al. Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer. Cell Syst. 2020;10(4):363–378.

22. Bersani AM, Bersani E, Mastroeni L. Deterministic and stochastic models of enzymatic networks—applications to pharmaceutical research. Computers & Mathematics with Applications. 2008;55(5):879–888.

23. Choi B, Rempala GA, Kim JK. Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. Sci Rep. 2017;7(1):1–11.

24. D'Alessandro M, Beesley S, Kim JK, Jones Z, Chen R, Wi J, et al. Stability of wake-sleep cycles requires robust degradation of the PERIOD protein. Curr Biol. 2017;27(22):3454–3467.

25. Beesley S, Kim DW, D'Alessandro M, Jin Y, Lee K, Joo H, et al. Wake-sleep cycles are severely disrupted by diseases affecting cytoplasmic homeostasis. Proc Natl Acad Sci U S A. 2020;117(45):28402–28411.

26. Bersani AM, Borri A, Carravetta F, Mavelli G, Palumbo P. On a stochastic approach to model the double phosphorylation/dephosphorylation cycle. Math Mech Complex Syst. 2020;8(4):261–285.

27. Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. J Chem Phys. 2003;118(11):4999–5010.

28. Cao Y, Gillespie DT, Petzold LR. The slow-scale stochastic simulation algorithm. J Chem Phys. 2005;122(1):014116.

29. Gómez-Uribe CA, Verghese GC, Tzafriri AR. Enhanced identification and exploitation of time scales for model reduction in stochastic chemical kinetics. J Chem Phys. 2008;129(24):244112.

30. Cao Y, Gillespie DT, Petzold LR. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. J Chem Phys. 2005;123(14):144917.

31. Barik D, Paul MR, Baumann WT, Cao Y, Tyson JJ. Stochastic simulation of enzyme-catalyzed reactions with disparate timescales. Biophys J. 2008;95(8):3563–3574.

32. Sanft KR, Gillespie DT, Petzold LR. Legitimacy of the stochastic Michaelis–Menten approximation. IET Syst Biol. 2011;5(1):58–69.

33. Kang HW, KhudaBukhsh WR, Koeppl H, Rempała GA. Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics. Bull Math Biol. 2019;81(5):1303–1336.

34. Thomas P, Straube AV, Grima R. Communication: limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks. J Chem Phys. 2011;135(18):181103.

35. Thomas P, Straube AV, Grima R. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. BMC Syst Biol. 2012;6(1):1–23.

36. Agarwal A, Adams R, Castellani GC, Shouval HZ. On the precision of quasi steady state assumptions in stochastic dynamics. J Chem Phys. 2012;137(4):044105.

37. Lawson MJ, Petzold L, Hellander A. Accuracy of the Michaelis–Menten approximation when analysing effects of molecular noise. J R Soc Interface. 2015;12(106):20150054.

38. Kim JK, Josić K, Bennett MR. The validity of quasi-steady-state approximations in discrete stochastic simulations. Biophys J. 2014;107(3):783–793.

39. Kim JK, Josić K, Bennett MR. The relationship between stochastic and deterministic quasi-steady state approximations. BMC Syst Biol. 2015;9(1):1–13.

40. Kim JK, Sontag ED. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. PLoS Comput Biol. 2017;13(6):e1005571.

41. Williams M, Daviter T. Protein-Ligand Interactions. Springer; 2016.

42. Milo R, Phillips R. Cell biology by the numbers. Garland Science; 2015.

43. Marchetti L, Priami C, Thanh VH. Simulation algorithms for computational systems biology. Springer; 2017.

44. Borghans JA, De Boer RJ, Segel LA. Extending the quasi-steady state approximation by changing variables. Bull Math Biol. 1996;58(1):43–63.

45. Thron C. A model for a bistable biochemical trigger of mitosis. Biophys Chem. 1996;57(2-3):239–251.

46. Schnell S, Maini PK. Enzyme kinetics far from the standard quasi-steady-state and equilibrium approximations. Math Comput Model. 2002;35(1-2):137–144.

47. Tzafriri AR. Michaelis-Menten kinetics at high enzyme concentrations. Bull Math Biol. 2003;65(6):1111–1129.

48. Kim JK, Forger DB. A mechanism for robust circadian timekeeping via stoichiometric balance. Mol Syst Biol. 2012;8(1):630.

49. Bersani AM, Bersani E, Dell'Acqua G, Pedersen MG. New trends and perspectives in nonlinear intracellular dynamics: one century from Michaelis–Menten paper. Contin Mech Thermodyn. 2015;27(4):659–684.

50. Kim JK. Protein sequestration versus Hill-type repression in circadian clock models. IET Syst Biol. 2016;10(4):125–135.

51. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Chem Phys. 1977;81(25):2340–2361.

52. Kim JK, Rempala GA, Kang HW. Reduction for stochastic biochemical reaction networks with multiscale conservations. Multiscale Model Simul. 2017;15(4):1376–1403.

53. Herath N, Del Vecchio D. Reduced linear noise approximation for biochemical reaction networks with time-scale separation: The stochastic tQSSA+. J Chem Phys. 2018;148(9):094108.

54. Sontag ED, Zeilberger D. A symbolic computation approach to a problem involving multivariate Poisson distributions. Adv Appl Math. 2010;44(4):359–377.

55. Thron CD. Bistable biochemical switching and the control of the events of the cell cycle. Oncogene. 1997;15(3):317–325.

56. Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. J Phys A Math Theor. 2017;50(9):093001.

57. Ciliberto A, Capuani F, Tyson JJ. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. PLoS Comput Biol. 2007;3(3):e45.

58. Kumar A, Josić K. Reduced models of networks of coupled enzymatic reactions. J Theor Biol. 2011;278(1):87–106.

59. Isaacs FJ, Hasty J, Cantor CR, Collins JJ. Prediction and measurement of an autoregulatory genetic module. Proc Natl Acad Sci U S A. 2003;100(13):7714–7719.

60. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J. 2001;81(6):3116–3136.

61. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, et al. A structure-based benchmark for protein–protein binding affinity. Protein Science. 2011;20(3):482–491.