Communication

Communication: Limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks

Cite as: J. Chem. Phys. **135**, 181103 (2011); https://doi.org/10.1063/1.3661156
Submitted: 04 September 2011 . Accepted: 25 October 2011 . Published Online: 14 November 2011

Philipp Thomas, Arthur V. Straube, and Ramon Grima

Communication: Limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks

Philipp Thomas,[1,2] Arthur V. Straube,[1] and Ramon Grima[2,a)]
[1]*Department of Physics, Humboldt University of Berlin, Newtonstr. 15, D-12489 Berlin, Germany*
[2]*School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom*

(Received 4 September 2011; accepted 25 October 2011; published online 14 November 2011)

It is commonly believed that, whenever timescale separation holds, the predictions of reduced chemical master equations obtained using the stochastic quasi-steady-state approximation are in very good agreement with the predictions of the full master equations. We use the linear noise approximation to obtain a simple formula for the relative error between the predictions of the two master equations for the Michaelis-Menten reaction with substrate input. The reduced approach is predicted to overestimate the variance of the substrate concentration fluctuations by as much as 30%. The theoretical results are validated by stochastic simulations using experimental parameter values for enzymes involved in proteolysis, gluconeogenesis, and fermentation. © *2011 American Institute of Physics*. [doi:10.1063/1.3661156]

It is well known that whenever transients in the concentration of a substrate species decay over a much slower timescale than those of the enzyme species, one can invoke the quasi-steady-state approximation (QSSA) to considerably simplify the deterministic (macroscopic) rate equations.[1,2] The study by Rao and Arkin[3] pioneered the use of the same approximation but on a mesoscopic level, i.e., applying a stochastic version of the approximation to obtain reduced chemical master equations. This approximation has since become ubiquitous in stochastic simulations of large biochemical reaction networks inside cells (see, for example, Refs. 4–7) although its range of validity is presently unknown. A plausible hypothesis is that the stochastic QSSA is valid in the same regions of parameter space where the deterministic QSSA is known to be valid. A handful of numerical studies[8,9] have shown that for some choices of rate constants which are compatible with the deterministic QSSA, the differences between the reduced and full master equation approaches are practically negligible. However none of these studies exclude the possibility that there exist regions of parameter space where the deterministic QSSA is valid but the stochastic QSSA exhibits large systematic errors in its predictions. In particular one is interested in knowing how accurate are the predictions of the stochastic QSSA for the size of intrinsic noise, i.e., the size of fluctuations in concentrations, since such noise is known to play important functional roles in biochemical circuits.[10] Numerical approaches cannot easily answer such questions because the stochastic simulation algorithm, the standard method which exactly samples the trajectories of master equations,[11] is computationally expensive.[12]

In this communication we seek to develop a theoretical approach to answer the following question: Given that the rate constants are chosen such that the deterministic QSSA is valid, what are the differences between the predictions of the reduced and full master equations for the variance of the fluctuations about the mean concentrations? We obtain a formula estimating the size of these differences for the simplest biochemical circuit which embeds the Michaelis–Menten reaction and confirm its accuracy using stochastic simulations. We find using physiological parameter values that the reduced master equation approach can overestimate the variance of the fluctuations by as much as ∼30%.

We start by considering the Michaelis–Menten reaction with substrate input

$$\xrightarrow{k_{in}} X_S, \quad X_S + X_E \underset{k_1}{\overset{k_0}{\rightleftharpoons}} X_C \xrightarrow{k_2} X_E + X_P, \tag{1}$$

where X_i denotes chemical species i and the k's denote the associated macroscopic rate constants. The reaction can be described as follows. Substrate molecules (species S) are pumped into some compartment at a constant rate, they bind to free enzyme molecules (species E) to form substrate-enzyme complexes (species C) which then either decay back to the original substrate and free enzyme molecules or else decay into free enzyme and product molecules (species P). The first reaction in Eq. (1) could equally represent the production of substrate by a first-order chemical reaction provided the species transforming into substrate exists in concentrations large enough such that fluctuations in its concentration can be ignored. The sum of the concentrations of free enzyme and complex is a constant since the enzyme can only be in one of these two forms. Hence all mathematical descriptions of the Michaelis–Menten reaction can be expressed in terms of just complex and substrate variables. On the macroscopic level, the QSSA proceeds by considering the case in which transients in the complex concentration decay much faster than those of the substrate. This condition of timescale separation is imposed by setting the time derivative of the macroscopic complex concentration to zero, solving for the steady-state complex concentration and substituting the latter into the rate equation for the substrate concentration which leads to the

a)Author to whom correspondence should be addressed. Electronic mail: ramon.grima@ed.ac.uk.

new rate equation

$$\frac{\partial}{\partial t}[X_S(t)] = k_{in} - \frac{k_2[E_T][X_S(t)]}{K_M + [X_S(t)]}, \tag{2}$$

where $[X_S(t)]$ is the substrate concentration at time t, $K_M = (k_1 + k_2)/k_0$ is the Michaelis–Menten constant, and $[E_T]$ is the total enzyme concentration, i.e., the sum of the concentration of free enzyme, $[X_E(t)]$, and of the concentration of complex, $[X_C(t)]$, which is a constant as previously mentioned. Note that the notation $[X_i]$ (without explicit dependence on t) will be reserved for the steady-state concentration of species i. We assume that at $t = 0$, we are in steady-state conditions; note that the system (1) is guaranteed to have a stable steady-state if the condition $k_{in}/k_2[E_T] < 1$ is satisfied. Linear stability analysis of the full rate equations describing (1) shows that the timescale for the decay of transients in the substrate concentrations is $\tau_s = (k_0[X_E])^{-1}$ while the timescale for the decay of transients in the complex concentrations is $\tau_c = (k_0[X_S] + k_1 + k_2)^{-1}$. Hence the criterion for the validity of the QSSA on the macroscopic rate equations (the deterministic QSSA), i.e., for the validity of Eq. (2), reads $\tau_s/\tau_c = \gamma = ([X_S] + K_M)/[X_E] > K_M/[E_T] \gg 1$ (see also Ref. 13).

The question that we address in the rest of this communication is the following: Given that the condition $\gamma \gg 1$ is satisfied, what is the variance of the noise about the mean concentrations as predicted by the reduced and full master equations?

The stochastic QSSA method implicitly starts by deducing that Eq. (2) is effectively the rate equation one would associate with a system of two chemical processes

$$\xrightarrow{k_{in}} X_S, \quad X_S \xrightarrow{k'} X_P, \tag{3}$$

where k' is an effective (time-dependent) rate constant equal to $k_2[E_T]/(K_M + [X_S(t)]$. Note that while the first reaction is elementary, the second is clearly not since it can be broken down into a set of more fundamental constituent reactions. Given the reduced set of reactions (3) one can then construct a reduced master equation for the set of reactions (1) (see Ref. 14 and supplementary material[15] for the construction of master equations)

$$\frac{\partial}{\partial t}P(n_S, t) = \Omega k_{in}\left(E_S^{-1} - 1\right)P(n_S, t) + \left(E_S^{+1} - 1\right)\frac{k_2[E_T]n_S}{K_M + n_S/\Omega}P(n_S, t), \tag{4}$$

where Ω is the compartment volume in which the reactions are occurring, n_S is the absolute number of substrate molecules, $P(n_S, t)$ is the probability that the system has n_S substrate molecules at time t and E_S^m is the step operator which upon acting on a function of n_S changes it into a function of $n_S + m$.[14] We note and emphasize that the physical basis of this master equation is not clear because such equations have been derived from first principles for elementary reactions[16,17] while Eq. (3) involves a non-elementary reaction. Equation (4) is simply written by analogy to what one would write down for Eq. (3) if both reactions were elementary and hence its legitimacy is *a priori* doubtful.

Now we want to use this master equation to deduce the variance of the noise in the macroscopic substrate concentrations. It is well known that in the macroscopic limit, the master equation for monostable chemical systems can be approximated by a linear Langevin equation, an approximation called the linear noise approximation (LNA),[14,19] from which all noise statistics can be estimated analytically. For systems with absorbing states or exhibiting multimodality, the LNA will not usually give accurate results (see, for example, Ref. 18) but its application to our example, the Michaelis–Menten reaction with substrate input, is not problematic since this reaction is only capable of monostable behavior. The steps to construct the LNA for a general monostable chemical reaction system are summarized in the supplementary material.[15] Here we will simply state the results of this recipe when applied to the master equation (Eq. (4)). Note that the transient decay timescales in the LNA are the same as those obtained from linear stability analysis of the rate equations. The variance of the substrate fluctuations in steady-state conditions is given by

$$\sigma_{\text{sLNA}} = \frac{[X_S]}{\Omega}\left(1 + \frac{[X_S]}{K_M}\right), \tag{5}$$

where the subscript "sLNA" stands for "LNA of the master equation reduced using the Stochastic QSSA." Note that $[X_S]$ in Eq. (5) is the steady-state substrate concentration obtained by solving for $[X_S]$ from Eq. (2) with time derivative set equal to zero.

We will now derive expressions for the variance of substrate concentration fluctuations using the full master equation approach. The steps of this method are as follows: (1) one writes down the master equation for the elementary chemical processes (1), (2) the two Langevin equations for the complex and substrate fluctuations are obtained using the LNA of the master equation, (3) expressions are found for the variance of complex and substrate concentration fluctuations in steady-state conditions, (4) the limit $\gamma \gg 1$ is taken of the expressions derived in step (3) leading to the final expressions for the variance of substrate fluctuations about the steady-state substrate concentration solution of the rate equation, Eq. (2). We note that this method, unlike the first one, does not make any assumptions about the validity of an *ad hoc* reduced master equation since it is based on the master equation for elementary processes and hence is guaranteed to be correct (see supplementary material[15]). We now proceed to put this systematic recipe in practice.

The master equation for the four elementary chemical processes given by Eq. (1) is

$$\partial_t P(n_S, n_C, t) = \left[\frac{k_0}{\Omega}\left(E_S^{+1}E_C^{-1} - 1\right)n_S(n_T - n_C) + \Omega k_{in}\left(E_S^{-1} - 1\right) + k_1\left(E_S^{-1}E_C^{+1} - 1\right)n_C + k_2\left(E_C^{+1} - 1\right)n_C\right]P(n_S, n_C, t), \tag{6}$$

where n_C is the absolute number of complex molecules and n_T is the absolute total number of molecules of enzyme in free and complex form. Note that n_T is a constant equal to $[E_T]\Omega$. In the macroscopic limit, the master equation, Eq. (6), can be

approximated by a pair of Langevin equations as given by the LNA. The variance of the substrate fluctuations in steady-state conditions is given by (see supplementary material[15])

$$\sigma_{\rm LNA} = \frac{[X_S]}{\Omega}\left(1+\frac{[X_S]}{K_M}\frac{K_1+[X_S]}{K_M+[X_S]}\frac{\gamma}{1+\gamma}\right) \xrightarrow{\gamma\gg 1} \frac{[X_S]}{\Omega}\left(1+\frac{[X_S]}{K_M}\frac{K_1+[X_S]}{K_M+[X_S]}\right), \quad (7)$$

where we have defined $K_1 = k_1/k_0$ and in the last step have taken the limit of $\gamma \gg 1$ which corresponds to the condition in which the deterministic QSSA Eq. (2) is valid.

Comparing Eqs. (5) and (7), we see that the two are not generally equal to each other except in the case $\beta = k_2/k_1 \ll 1$. From a fundamental point of view, this disagreement implies that the reduced master equation does not obey the generalized fluctuation-dissipation theorem of non-equilibrium physics[14,20] and that hence it is flawed. More importantly, we observe that the condition $\beta = k_2/k_1 \ll 1$ is not equivalent to the quasi-steady-state condition $\gamma \gg 1$. The former condition is consistent with the enzyme-substrate complex being in thermodynamic equilibrium with free enzyme and substrate, a condition which is difficult to uphold in open systems since they are characterized by non-equilibrium steady states. While the quasi-steady-state condition can easily be satisfied in open systems since it is only required that the total enzyme concentration is much less than the Michaelis–Menten constant. Hence we can conclude that for open systems, the stochastic QSSA based on Eqs. (3) and (4) is NOT the legitimate stochastic equivalent of the deterministic QSSA.

There are two possible hypothetical scenarios which would imply that the stochastic QSSA is perhaps still a very good general method to estimate the size of the concentration fluctuations. The first case would be if experimental evidence showed that for many enzymes it just happens that $\beta \ll 1$. The second case would be if experimental evidence showed no such restriction on β but nevertheless the difference between the variance prediction of the reduced and full master equations is so small as to be negligible. We now consider each case.

A perusal of the experimental data available in the literature shows that there are very few studies which simultaneously report values of k_1 and k_2, the data required to estimate β. The vast majority of studies report values for K_M, a considerable number report k_2 and a small percentage report both k_2 and K_M.[21] Now the ratio k_2/K_M, frequently called the enzyme efficiency,[22] is defined as

$$\Theta = \frac{k_2}{K_M} = \frac{\beta}{1+\beta}k_0. \quad (8)$$

The recent study by Bar-Even *et al.*[21] based on mining the Brenda[24] and KEGG databases[25] concluded that for most enzymes Θ lies in the range $10^3-10^6\,{\rm M}^{-1}\,{\rm s}^{-1}$. It is also known that the association constant k_0 takes values in the range $10^6-10^9\,{\rm M}^{-1}{\rm s}^{-1}$.[22] We can conclude from these two pieces of data and using Eq. (8) that the range of β for most enzymes is between 0 and some number which is much greater than 1 and that hence on the basis of experimental data one cannot argue for the general validity of the stochastic QSSA.

Of course as previously mentioned, it could still happen that even though there is no restriction on β, that the variance as predicted by the stochastic QSSA and the true variance are negligibly small. We can test this hypothesis quantitatively by using Eqs. (5) and (7) to derive the fractional relative error ϵ in the variance prediction of the stochastic QSSA

$$\epsilon = \frac{\sigma_{\rm LNA}-\sigma_{\rm sLNA}}{\sigma_{\rm LNA}} = \frac{-(1-\alpha)\alpha\beta}{1+\beta(1-\alpha(1-\alpha))}, \quad (9)$$

where $\alpha = k_{in}/(k_2[E_T])$, a non-dimensional quantity which can take values between 0 and 1 as previously mentioned in the discussion after Eq. (2). Furthermore, it can be shown using Eq. (2) that at steady-state one has $[X_E] = [E_T](1-\alpha)$ and $[X_C] = [E_T]\alpha$, from which we can deduce that α is a measure of how saturated is the enzyme with substrate. Note that Eq. (9) shows that the relative error tends to zero as $\alpha \to 0$ and $\alpha \to 1$ and that hence the reduced master equation provides a correct prediction of the size of the substrate fluctuations whenever the free enzyme or complex concentrations are very small (similar results have been obtained by Mastny *et al.*[23] for the Michaelis–Menten reaction with no substrate input; however their results are not for general α and β and do not enforce the validity of the deterministic QSSA; see later for discussion). In Fig. 1, the solid lines illustrate the predictions of Eq. (9) for three different values of β: (1) 1, (2) 2.8, and (3) 10. Case (1) utilizes experimental data for the enzymes Chymotrypsin and Malate dehydrogenase with respective substrates acetyl-L-tryptophan and NADH (Refs. 26 and 27) while case (2) is based on data for the enzyme Lactate dehydrogenase with substrate pyruvate.[28] These enzymes are respectively involved in proteolysis, gluconeogenesis and the conversion of pyruvate (the final product of glycolysis) to lactate in anaerobic conditions. Case (3) showcases the largest possible error made by the stochastic QSSA; this is consistent with a highly efficient enzyme such as β-Lactamase for which Θ is of the same order of magnitude as the maximum possible association rate constant $k_0 \sim 10^8 - 10^9\,{\rm s}^{-1}\,{\rm M}^{-1}$.[22] The theoretical predictions of our LNA based method are confirmed by stochastic simulations of the master equations, Eqs. (4) and (6), using Gillespie's algorithm[11] (data points in Fig. 1). Note that the maximum possible percentage error is about 30% which is significant. Also note that the maximum error in all cases is reached at $\alpha = 1/2$ namely when the enzyme is half saturated with substrate which occurs when the substrate concentrations are equal to the Michaelis–Menten constant K_M (this is the case for most enzymes of the glycolytic pathway[30]); for substrate concentrations much smaller or larger than K_M, the error is negligible.

The LNA is strictly speaking valid for large volumes or equivalently in the limit of large number of molecules[14,29] and hence one could argue that our theoretical formula Eq. (9) is of limited validity inside cells where molecule numbers can be quite small.[31] Figure S1 (see supplementary material[15]) shows the results of stochastic simulations for the case $\alpha = 0.5$ and $\beta = 10$ using a total number of enzyme molecules n_T varying between 1 and 100 molecules. Note that the error ϵ is practically constant at 30%, the value predicted by the LNA and shown in Fig. 1. This suggests that the estimates provided

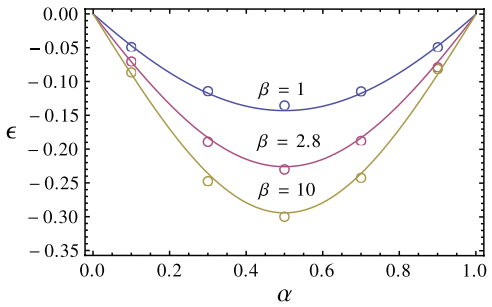

FIG. 1. Plot of the fractional relative error, ϵ, in the variance of fluctuations as predicted by the reduced master equation obtained from the stochastic QSSA versus the enzyme saturation parameter α. The solid lines show the theoretical prediction Eq. (9) for three different values of β: 1 (blue), 2.8 (magenta), and 10 (yellow-green). These three cases are, respectively, consistent with the enzymes being Chymotrypsin or Malate dehydrogenase, Lactate dehydrogenase and a highly efficient enzyme such as β-Lactamase. The data points show the results of stochastic simulations using the Gillespie algorithm for reactions involving a total of 100 enzyme molecules and total enzyme concentrations in the nano and millimolar range. The ratio of substrate and complex decay timescales, γ, is greater than 10, which enforces the validity of the deterministic QSSA. Details in the supplementary material (Ref. 15).

by our method are accurate even for low copy number conditions.

Our study has focused on the most common type of stochastic QSSA in the literature which is heuristic in nature and hence the question regarding its validity. There are a class of alternative model reduction techniques[23,32] based on singular-perturbation analysis (sQSPA and sQSPA-Ω) which are rigorous and whose validity is not under question. For the Michaelis–Menten reaction without substrate input, these methods lead to a reduced master equation of the same form as the heuristic stochastic QSSA whenever the free enzyme or complex concentrations are very small (see Table II of Ref. 23). This implies that for such conditions the error in the predictions of the stochastic QSSA should be zero, a result which is also reproduced by our method. However, note that though these concentration conditions can be compatible with the deterministic QSSA they are not synonymous with it. The sQSPA methods do not lead to a reduced master equation for parameters consistent with the deterministic QSSA and hence cannot make statements regarding the accuracy of the heuristic stochastic QSSA in such conditions. Our contribution fills this important gap by deriving an explicit formula for the error in the predictions of the stochastic QSSA, i.e., Eq. (9), for all parameter values consistent with the deterministic QSSA. We finish by noting that a recent study by Gonze *et al.*[8] also studied the reaction system (1) using numerical simulations and found little difference between the predictions of the stochastic QSSA and the full master equation. The study used values of $\beta = 0.1$ (see Table 7.2 in Ref. 8) and hence in the light of our results, it is clear why they observed high accuracy of the stochastic QSSA. However, as we have shown, this is not the general case: many enzymes have large β and hence discrepancies of the order of few tens of percent between the predictions of the reduced and full approaches will be visible whenever substrate concentrations are approximately equal to the Michaelis–Menten constant.

[1] L. A. Segel and M. Slemrod, SIAM Rev. **31**, 446 (1989).
[2] S. Schnell and P. K. Maini, Comments Theor. Biol. **8**, 169 (2003).
[3] C. V. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[4] D. Gonze, M. Jacquet, and A. Goldbeter, J. R. Soc., Interface **5**, S95 (2008).
[5] J. Paulsson, O. G. Berg, and M. Ehrenberg, Proc. Nat. Acad. Sci. U.S.A. **97**, 7148 (2000).
[6] M. L. Guerriero, A. Pokhilko, A. Pinas-Fernandez, K. J. Halliday, A. J. Millar, and J. Hillston, J. R. Soc. Interface (in press).
[7] G. Dupont, A. Abou-Lovergne, and L. Combettes, Biophys. J. **95**, 2193 (2008).
[8] D. Gonze, W. Abou-Jaoude, D. A. Ouattara, and J. Halloy, Methods Enzymol. **487**, 171 (2011).
[9] K. R. Sanft, D. T. Gillespie, and L. R. Petzold, IET Syst. Biol. **5**, 58 (2011).
[10] A. Eldar and M. B. Elowitz, Nature **467**, 167 (2010).
[11] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[12] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[13] I. Stoleriu, F. A. Davidson, and J. L. Liu, J. Math. Biol. **48**, 82 (2004).
[14] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).
[15] See supplementary material at http://dx.doi.org/10.1063/1.3661156 for details of calculations.
[16] D. T. Gillespie, Physica A **188**, 404 (1992).
[17] D. T. Gillespie, J. Chem. Phys. **131**, 164109 (2009).
[18] W. Hortshemke and L. Brenig, Z. Physik B **27**, 341 (1977); W. Hortshemke, M. Malek-Mansour, and L. Brenig, *ibid.* **28**, 135 (1977); T. T. Marquez-Lago and J. Stelling, Biophys. J. **98**, 1742 (2010).
[19] J. Elf and M. Ehrenberg, Genome Res. **13**, 2475 (2003).
[20] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, Berlin, 1987).
[21] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo, Biochemistry **50**, 4402 (2011).
[22] A. Fersht, *Structure and Mechanism in Protein Science: Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
[23] E. A. Mastny, E. L. Haseltine, and J. B. Rawlings, J. Chem. Phys. **127**, 094106 (2007).
[24] P. Pharkya, E. V. Nikolaev, and C. D. Maranas, Metab. Eng. **5**, 71 (2003).
[25] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi, Nucleic Acids Res. **36**, D480 (2008).
[26] A. Lodola, J. D. Shore, M. D. Parker, and J. Holbrook, Biochemical J. **175**, 987 (1978).
[27] M. Renard and A. R. Fersht, Biochemistry **12**, 4713 (1973).
[28] M. J. Boland and H. Gutfreund, Biochem. J. **151**, 715 (1975).
[29] R. Grima, J. Chem. Phys. **133**, 035101 (2010).
[30] R. A. Copeland, *Evaluation of Enzyme Inhibitors in Drug Discovery* (Wiley-Interscience, Hoboken, NJ, 2005).
[31] R. Grima and S. Schnell, Essays Biochem. **45**, 41 (2008).
[32] R. Srivastava, E. L. Haseltine, E. A. Mastny, and J. B. Rawlings, J. Chem. Phys. **134**, 154109 (2011).