Supplementary Material: 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]

[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*

Supplement A: General formulation of master equations for elementary and non-elementary processes

Consider a general chemical system confined in a compartment of volume Ω and consisting of a number N of distinct chemical species interacting via R chemical reactions of the type

$$s_{1j}X_1 + \ldots + s_{Nj}X_N \xrightarrow{k_j} r_{1j}X_1 + \ldots + r_{Nj}X_N. \quad \text{(A1)}$$

Here j is an index running from 1 to R, X_i denotes chemical species i, s_{ij} and r_{ij} are the stoichiometric coefficients, and k_j is the macroscopic rate of reaction. Note that these reactions are not necessarily elementary (unimolecular or bimolecular reactions). If the jth reaction is elementary then its rate k_j is a constant while if it is non-elementary k_j is a function of macroscopic concentrations. The general form of the master equation for both cases is [1]

$$\frac{\partial P(\vec{n},t)}{\partial t} = \Omega \sum_{j=1}^{R}\left(\prod_{i=1}^{N} E_i^{-S_{ij}} - 1\right)\hat{f}_j(\vec{n},\Omega)P(\vec{n},t), \quad \text{(A2)}$$

where $P(\vec{n},t)$ is the probability that the system is in a particular mesoscopic state $\vec{n} = (n_1,...,n_N)^T$ and n_i is the number of molecules of the ith species. Note that E_i^x is a step operator – when it acts on some function of the absolute number of molecules, it gives back the same function but with n_i replaced by $n_i + x$. The chemical reaction details are encapsulated in the stoichiometric matrix $S_{ij} = r_{ij} - s_{ij}$ and in the microscopic rate functions $\hat{f}_j(\vec{n},\Omega)$. The probability that the jth reaction occurs in the time interval $[t, t+dt)$ is given by $\Omega\hat{f}_j(\vec{n},\Omega)dt$.

For elementary reactions, the microscopic rate function takes one of four different forms, depending on the order of the jth reaction: (i) a zeroth-order reaction by which a species is input into a compartment gives $\hat{f}_j(\vec{n},\Omega) = k_j$; (ii) a first-order unimolecular reaction involving the decay of some species h gives $\hat{f}_j(\vec{n},\Omega) = k_j n_h \Omega^{-1}$; (iii) a second-order bimolecular reaction between two molecules of the same species h gives $\hat{f}_j(\vec{n},\Omega) = k_j n_h(n_h-1)\Omega^{-2}$; (iv) a second-order bimolecular reaction between two molecules of different species, h and v, gives $\hat{f}_j(\vec{n},\Omega) = k_j n_h n_v \Omega^{-2}$. Note that these forms for the microscopic rate functions have been rigorously derived from microscopic physics [2, 3] and hence the validity of Eq. (A2) for elementary reactions is guaranteed [13].

For non-elementary reactions, the form of the microscopic rate function has to be basically guessed by analogy with the prescription for elementary reactions. For example, for the set of reactions (3) in the main text, the second reaction is a non-elementary first-order reaction with a time-dependent macroscopic rate constant $k'(t) = k_2[E_T]/(K_M + [X_S(t)])$, where $[E_T]$ is the constant macroscopic total enzyme concentration and $[X_S(t)]$ is the instantaneous macroscopic concentration of species S. Hence one would use the microscopic rate function $\hat{f}(\vec{n},\Omega) = k_2[E_T](n_S/\Omega)/(K_M + n_S/\Omega)$ based on the formula stated above for an elementary first-order reaction. Of course master equations based on microscopic rate functions obtained from this procedure are ad-hoc and have no fundamental basis.

Supplement B: General formulation of the linear noise approximation in steady-state conditions

Here we provide a step by step recipe to construct the linear noise approximation (LNA) of the master equation, Eq. (A2), for the set of reactions (A1). We note that this approximation is only valid for a monostable system (the condition is formally given by Eq. 3.4 in Ch X of the book by van Kampen [1]). Let the macroscopic steady-state concentration of species i be given by $[X_i]$ and the derivative with respect to this variable be denoted by ∇_i. Furthermore we shall distinguish matrices by underlining them. The five steps to constructing the LNA in steady-state conditions (for both elementary and non-elementary reactions) are then as follows:

1. Construct the $N \times R$ stoichiometric matrix, $\underline{\mathrm{S}}$, whose $i-j$ element is given by $r_{ij} - s_{ij}$.

2. Construct the macroscopic rate function vector $\vec{f}$ with elements $f_j = k_j \prod_{m=1}^{N}([X_m])^{s_{mj}}$ and the diagonal matrix $\underline{\mathrm{F}}$ with elements $F_{ii} = f_i$.

3. Construct the Jacobian matrix $\underline{\mathrm{J}}$ whose $i-j$ element is given by $\nabla_j(\underline{\mathrm{S}}.\vec{f})_i$. Construct the diffusion matrix $\underline{\mathrm{D}} = \underline{\mathrm{S}} \cdot \underline{\mathrm{F}} \cdot \underline{\mathrm{S}}^T$.

4. The stochastic differential equations (linear Langevin equations) approximating the chemical master equation for the set of reactions (A1) in the limit of large molecule numbers are given by [4]

$$\frac{\partial}{\partial t}\vec{\eta}(t) = \underline{\mathrm{J}} \cdot \vec{\eta}(t) + \Omega^{-1/2}\,\underline{\mathrm{S}} \cdot \sqrt{\underline{\mathrm{F}}}\,\vec{\Gamma}(t), \quad \text{(B1)}$$

where $\eta_i(t)$, the ith entry of the vector $\vec{\eta}(t)$, denotes the fluctuations about the macroscopic steady-state concentration of species i, i.e., $\eta_i(t) = (n_i(t)/\Omega) - [X_i]$. The R dimensional vector $\vec{\Gamma}(t)$ is Gaussian white noise defined by $\langle\Gamma_i(t)\rangle = 0$ and $\langle\Gamma_i(t)\Gamma_j(t')\rangle = \delta_{i,j}\delta(t-t')$.

5. The covariance matrix $\underline{\sigma}$ of the fluctuations in Eq. (B1) is obtained by solving the Lyapunov equation [4, 5]

$$\underline{\mathrm{J}} \cdot \underline{\sigma} + \underline{\sigma} \cdot \underline{\mathrm{J}}^T + \underline{\mathrm{D}}/\Omega = 0, \tag{B2}$$

where $\sigma_{ij} = \langle\eta_i\eta_j\rangle$. The variance of the fluctuations is hence given by the diagonal elements of $\underline{\sigma}$.

Following this recipe we can explicitly construct the linear Langevin equations for the full set of elementary reactions of the Michaelis-Menten reaction, as given by Scheme (1) in the main text. The macroscopic rate function vector is given by $\vec{f} = (k_{\mathrm{in}}, k_0[X_S][X_E], k_1[X_C], k_2[X_C])$, where $[X_E]$, $[X_C]$ and $[X_S]$ denote the macroscopic concentrations of free enzyme, complex and substrate species, respectively. The stoichiometric matrix reads

$$\underline{\mathrm{S}} = \begin{pmatrix} 0 & +1 & -1 & -1 \\ +1 & -1 & +1 & 0 \end{pmatrix}. \tag{B3}$$

Note that the order of columns in $\underline{\mathrm{S}}$ reflects the order of reactions in $\vec{f}$ while the rows are related to the species type (row 1 is for the complex species and row 2 is for the substrate species). Note also that the free enzyme species has been removed by conservation of total enzyme number. The diagonal matrix $\underline{\mathrm{F}}$ and the Jacobian $\underline{\mathrm{J}}$ can then be calculated by steps 2 and 3. Finally we obtain by step 4, that in the macroscopic limit, the master equation, Eq. (6) in the main text, can be approximated by a pair of Langevin equations

$$\begin{aligned}\frac{\partial}{\partial t}\begin{pmatrix}\eta_C(t)\\ \eta_S(t)\end{pmatrix} =& k_0 \begin{pmatrix} -(K_M+[X_S]) & [X_E] \\ K_1+[X_S] & -[X_E]\end{pmatrix}\cdot\begin{pmatrix}\eta_C(t)\\ \eta_S(t)\end{pmatrix} \\ &+ \Omega^{-1/2}\begin{pmatrix} 0 & \sqrt{k_0[X_E][X_S]} & -\sqrt{k_1[X_C]} & -\sqrt{k_2[X_C]} \\ \sqrt{k_{in}} & -\sqrt{k_0[X_E][X_S]} & \sqrt{k_1[X_C]} & 0\end{pmatrix}\cdot\vec{\Gamma}(t),\end{aligned} \tag{B4}$$

where $\eta_S(t)$ and $\eta_C(t)$ denote the fluctuations about the macroscopic steady-state substrate and complex concentrations and $\vec{\Gamma}(t)$ is a four-dimensional vector whose entries are white Gaussian noise with the properties $\langle\Gamma_i(t)\rangle = 0$ and $\langle\Gamma_i(t)\Gamma_j(t')\rangle = \delta_{i,j}\delta(t-t')$.

The Langevin equations approximating the reduced chemical master equation, Eq. (4) in the main text, can be constructed in a similar manner. The non-elementary set of reactions are here given by Scheme (3) in the main text. Hence the macroscopic rate function vector is $\vec{f} = (k_{\mathrm{in}}, k_2[E_T][X_S]/(K_M + [X_S]))$, where $K_M = (k_1 + k_2)/k_0$ while the stoichiometric matrix is $\underline{\mathrm{S}} = (+1, -1)$. As before these can be used to compute the Langevin approximation which reads

$$\frac{\partial}{\partial t}\tilde{\eta}_S(t) = -\frac{k_2}{\gamma}\tilde{\eta}_S(t) + \sqrt{\frac{2k_2[X_S]}{\Omega\gamma}\left(1+\frac{[X_S]}{K_M}\right)}\Gamma(t), \tag{B5}$$

where $\tilde{\eta}_S(t)$ denotes the substrate fluctuations as predicted by the reduced master equation obtained from the stochastic quasi-steady state approximation.

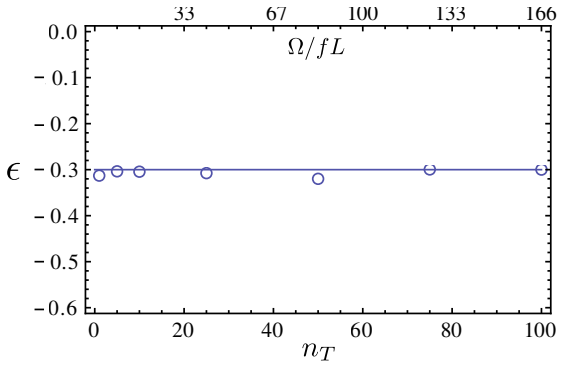

FIG. S1. Plot of the fractional relative error ϵ versus the total number of enzyme molecules n_T for the case $\alpha = 0.5$ and $\beta = 10$ (compatible with a highly efficient enzyme, parameters as given in Supp. C). The data points are obtained from stochastic simulations. The solid line is simply a guide to the eye. The total number of enzymes is varied at constant total enzyme concentration.

Supplement C: Parameter values used in stochastic simulations

Parameter values for the stochastic simulations shown in Fig. 1 in the main text are as follows. The enzymes, Malate dehydrogenase and Chymotrypsin, both have $\beta =$

1; we have simulated only the first of these enzymes by using $\Omega = 17\,fL$, $[E_T] = 10\,nM$, $k_2 = 5\,s^{-1}$, $k_1 = 5\,s^{-1}$, and $k_0 = 5 \times 10^7\,M^{-1}s^{-1}$. For the case $\beta = 2.8$ (the enzyme Lactate dehydrogenase), we used $\Omega = 0.017\,fL$, $[E_T] = 10\,\mu M$, $k_2 = 210\,s^{-1}$, $k_1 = 75\,s^{-1}$, and $k_0 = 3 \times 10^6\,M^{-1}s^{-1}$. For the case $\beta = 10$ (a case compatible with a highly efficient enzyme), we used $\Omega = 170\,fL$, $[E_T] = 1\,nM$, $k_2 = 1\,s^{-1}$, $k_1 = 0.1\,s^{-1}$, and $k_0 = 10^8\,M^{-1}s^{-1}$. In all cases the total number of enzyme molecules n_T was 100. Parameter values for Fig. S1 are $\beta = 10$, $k_2 = 1\,s^{-1}$, $k_1 = 0.1\,s^{-1}$, $[E_T] = 1\,nM$ and $k_0 = 10^8\,M^{-1}s^{-1}$.

The rate constants for the cases $\beta = 1$ and $\beta = 2.8$ were obtained from the experimental studies [6–8]. The rate constants for $\beta = 10$ were not for a specific enzyme and hence were chosen from the known physiological ranges: for k_2 the range is $1 - 10^4\,s^{-1}$ [9], for k_0 the range is $10^6 - 10^8\,s^{-1}M^{-1}$ [10] and for K_M the range is $10^{-1} - 10^{-7}\,M$ [9]. Similarly, the total enzyme concentrations were chosen from the physiological ranges: nano- to millimolar concentrations [11]. The compartment volumes for the data in Fig. 1 in the main text were chosen such that the total number of enzyme molecules n_T was 100 in all cases; for Fig. S1 the volumes were chosen such that n_T could be varied over the range 1 to 100.

[1] N. G. van Kampen, *Stochastic processes in physics and chemistry* (Elsevier, Amsterdam, 2007).
[2] D. T. Gillespie, Physica A **188**, 404 (1992).
[3] D. T. Gillespie, J. Chem. Phys. **131**, 164109 (2009).
[4] J. Elf and M. Ehrenberg, Genome Res. **13**, 2475 (2003).
[5] J. Keizer, *Statistical thermodynamics of nonequilibrium processes* (Springer-Verlag, Berlin, 1987).
[6] A. Lodola, J. D. Shore, M. D. Parker, and J. Holbrook, Biochemical J. **175**, 987 (1978).
[7] M. Renard and A. R. Fersht, Biochemistry **12**, 4713 (1973).
[8] M. J. Boland and H. Gutfreund, Biochem. J. **151**, 715 (1975).
[9] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry* (Freeman, New York, 2002).
[10] A. Fersht, *Structure and mechanism in protein science: Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
[11] R. Grima and S. Schnell, Essays in Biochemistry **45**, 41 (2008).
[12] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[13] We note that the general form of the master equation, Eq. (4) in Ref. 12, is constructed from the laws of probability given two simple premises: (i) the chemical system, at any point in time, can be in one of a number of possible states, each state described by the number of molecules of each species, and (ii) when a reaction occurs, the state of the system changes to a new one. Due to premise (ii), the master equation depends on the probability that a reaction occurs in a short time interval. Assuming well mixed conditions, the form of these probabilities have been rigorously derived from the well established laws of microscopic physics for elementary reactions (those involving the simultaneous interaction of at most two molecules) in the gas-phase [2] and in liquids [3].