

Published in IET Systems Biology
 Received on 20th November 2008
 Revised on 4th February 2009
 doi: 10.1049/iet-syb.2008.0167

Special Issue – Selected papers from the Second q-bio
 Conference on Cellular Information Processing



Mesoscopic statistical properties of multistep enzyme-mediated reactions

W.H. de Ronde¹ B.C. Daniels² A. Mugler³ N.A. Sinitsyn⁴
 I. Nemenman⁴

¹FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ, Amsterdam, The Netherlands

²Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA

³Department of Physics, Columbia University, New York, NY 10027, USA

⁴Computer, Computational and Statistical Sciences Division, Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

E-mail: deronde@amolf.nl

Abstract: Enzyme-mediated reactions may proceed through multiple intermediate conformational states before creating a final product molecule, and one often wishes to identify such intermediate structures from observations of the product creation. In this study, the authors address this problem by solving the chemical master equations for various enzymatic reactions. A perturbation theory analogous to that used in quantum mechanics allows the determination of the first ($\langle n \rangle$) and the second (σ^2) cumulants of the distribution of created product molecules as a function of the substrate concentration and the kinetic rates of the intermediate processes. The mean product flux $V = d\langle n \rangle / dt$ (or ‘dose–response’ curve) and the Fano factor $F = \sigma^2 / \langle n \rangle$ are both realistically measurable quantities, and whereas the mean flux can often appear the same for different reaction types, the Fano factor can be quite different. This suggests both qualitative and quantitative ways to discriminate between different reaction schemes, and the authors explore this possibility in the context of four sample multistep enzymatic reactions. Measuring both the mean flux and the Fano factor can not only discriminate between reaction types, but can also provide some detailed information about the internal, unobserved kinetic rates, and this can be done without measuring single-molecule transition events.

Enzyme-mediated reactions are ubiquitous in biology. Traditionally, they have been described as a two-step Michaelis–Menten (MM) process [1], in which the enzyme and the substrate form a complex that can decay either back into the enzyme and the substrate, or forward into the enzyme and the product (see Fig. 1a). The latter step is usually assumed to be irreversible, leaving three kinetic rates that specify the reaction. To determine these kinetic rates, a typical experiment measures the average rate of product formation (or product ‘flux’) V as a function of substrate concentration S (also called a ‘dose–response’ curve), producing a plot as in Fig. 2a. Two pieces of information can be extracted from this plot: the saturating reaction rate V_{\max} and the Michaelis constant K (the substrate concentration at

half of the maximum rate). Importantly, these two measurements do not specify the three underlying kinetic rates; thus they do not allow for a full identification of the reaction processes.

The MM mechanism is not entirely general: many enzyme-mediated reactions consist of multiple intermediate internal steps (such as conformational changes of either the enzyme or the substrate, enzymes that occur in active and inactive states etc.), each with its own forward and backward reaction rates. While measurements of substrate–enzyme complex formation and product releases are possible even on a single-molecule level in enzymatic kinetics [2] and in ion channel transport [3, 4], typical

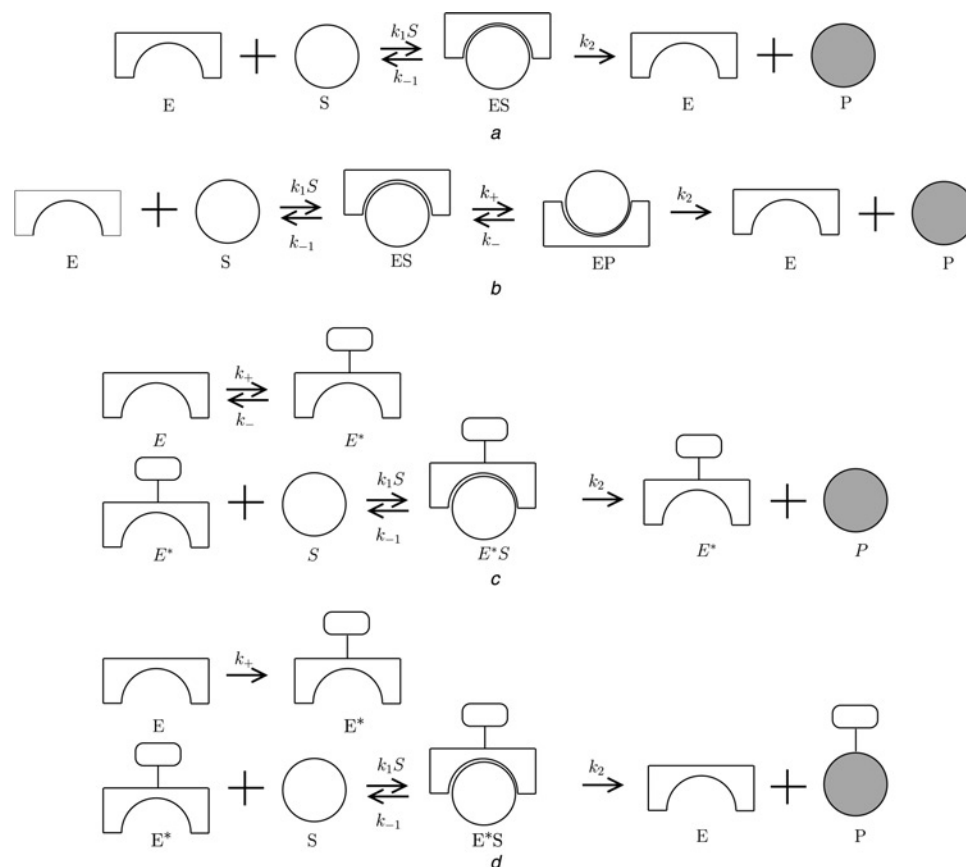


Figure 1 Potential schemes for an enzyme-mediated reaction, in which substrate S is converted to product P

a A simple Michaelis-Menten (MM) reaction

b A MM reaction with an additional intermediate state (e.g. if the complex undergoes a conformational change before creating the product)

c A scheme in which the enzyme must become active (e.g. through phosphorylation) before mediating the reaction

d A scheme in which the enzyme must become active before mediating the reaction, and the reaction leaves the enzyme inactive

experiments cannot resolve intermediate steps when measuring only the average reaction rate since they produce qualitatively similar curves for $V(S)$. For example, the mean flux through an arbitrary complex ion channel that holds at most one large transported molecule at a time is indistinguishable from that through a simple channel with just two internal states [5].

An interesting problem then is to determine which experimental measurements could identify the multistep nature of an enzyme-mediated reaction. One can consider three different types of experiments for which this could be done: single-molecule experiments, which measure single reaction events; mesoscopic experiments, where single events are unobservable, and yet the product flux can be measured repeatedly and precisely, allowing estimation of the mean flux and its fluctuations; and macroscopic experiments, where fluctuations of such fluxes (with a typical relative magnitude of one over the square root of the number of molecules) are not empirically observable. The theory for reconstruction of enzymatic kinetic schemes for the first type of experiments has been worked out quite well, and has been applied experimentally [2, 6]. However, not much has been

said about mesoscopic experiments. The main goal of this paper, therefore, is to develop a simple computational technique that could be used to reconstruct kinetic diagrams based on such data, and to apply it to a few toy model examples to argue its experimental applicability. The third type of experiments, the macroscopic ones, will be discussed in a subsequent publication.

Specifically, we present here a general perturbative approach for calculating the cumulants of a product molecule flux for a given enzymatic reaction scheme. We suggest that this method could be used in experiments that measure not only the mean rate but also the variance in the rate of the creation of product molecules; modern experiments can clearly perform this task in different experimental systems [2, 7]. Measurements of higher order product formation cumulants, if experimentally possible, would allow one to constrain properties of the reaction even more strongly.

To illustrate the method, we first apply it to the usual MM reaction (Fig. 1a). In addition to recovering the well-known result for the mean rate of product formation as a function of substrate concentration, we derive the dependence on

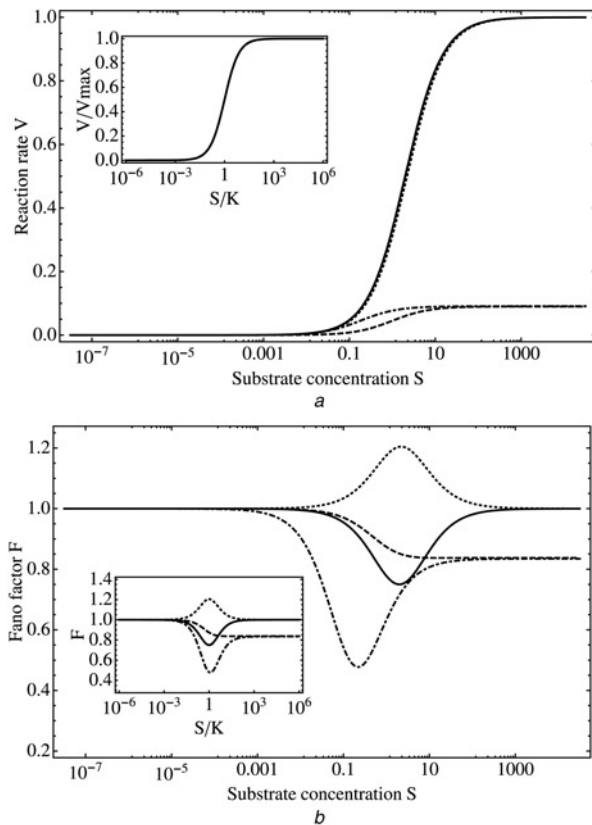


Figure 2 Mean product flux (also called dose–response curve) V and the Fano factor F against substrate concentration S for the four cases in Fig. 1. A: solid, B: dashed, C: dotted and D: dot-dashed

Plots are of (20), (26), (32) and (38) for V and (21), (27), (33) and (39) for F , with $k_1 = 1$, $k_{-1} = 1$, $k_2 = 1$, $k_+ = 0.1$ and $k_- = 0.01$. Note that while there are no qualitative differences in V (and in fact all curves collapse when V is normalised by V^{\max} and S by K , as seen in the inset), features can appear in F that signify that a process is more complicated than the single-intermediate case A

substrate of the Fano factor, the ratio of the variance in the number of product molecules to the mean. Importantly, our approach is extendible, at least in principle, to an arbitrary enzyme-mediated reaction scheme, and we demonstrate this by analysing three more complex reaction schemes, shown in Figs. 1b–d. In the context of these reactions, we show that the dependence of the Fano factor on the substrate concentration can produce qualitatively different results for different reaction types, allowing one to distinguish them experimentally. In addition, we argue that quantitative features of the Fano factor measurements can constrain the values of the underlying kinetic rate constants more tightly than the mean rate measurements alone.

1 Methods: the MM model

Going beyond a simple description of the mean production of a particular molecule and making predictions about the intrinsic noise requires a stochastic description, such as the chemical master equation (CME) [8]. The CME describes

the evolution in time of the joint probability distribution for the copy numbers of all species involved in a reaction scheme. For the enzyme-mediated reactions we consider, we make the assumption that each enzyme acts independently, that is, the substrate concentration is much larger than the enzyme concentration. This is equivalent to treating the process as if only one enzyme was present at a time. Furthermore, we assume that the concentration of the substrate is constant during each experimental measurement, and thus our master equation needs only to keep track of the enzyme's state and the number of created product molecules n . In other words, we are still solving the master equation, but treating, however, the substrate concentration as a parameter in it. We note that both of these assumptions can be relaxed using recently developed techniques [9, 10]. Finally, we only search for the distribution of the number of product molecules at times much longer than a typical enzymatic turnover time.

We begin by demonstrating our method on the simple MM reaction in Fig. 1a. In the MM reaction, the enzyme will be in either a free state E or a bound state ES. Therefore we partition the joint probability distribution into two parts: P_n^E , the probability that n product molecules have been created and the enzyme is free, and P_n^{ES} , the probability that n product molecules have been created and the enzyme is bound, yielding the CME [8]

$$\frac{dP_n^E}{dt} = -k_1 S P_n^E + k_{-1} P_n^{ES} + k_2 P_{n-1}^{ES} \quad (1)$$

$$\frac{dP_n^{ES}}{dt} = k_1 S P_n^E - (k_{-1} + k_2) P_n^{ES} \quad (2)$$

where the rates are defined in Fig. 1a, and S is the number of substrate molecules. (Note that S can equivalently be thought of as the concentration of substrate as long as one appropriately rescales the rates.) The total probability of having n product molecules is then $P_n = P_n^E + P_n^{ES}$.

We note that the situation where the product molecules are created and never destroyed or transformed back into the substrate is not physical, and additional reactions that degrade the product in some way are needed. However, as long as we are interested in how many product molecules have been created, rather than are present at a given time, the creation, (1) and (2), and the decay reactions can be considered independently.

Similar to [9–13] and others, we begin our solution of (1) and (2) by defining the generating function

$$G^z(\chi) = \sum_{n=0}^{\infty} P_n^z e^{i\chi n} \quad (3)$$

with $z \in \{E, ES\}$. Defining the vector $|\mathbf{G}\rangle = (G^E, G^{ES})^T$,

we may write the total generating function as

$$G(\chi) = \langle \mathbf{1} | \mathbf{G} \rangle = G^E + G^{ES} \quad (4)$$

where $\langle \mathbf{1} | = (1, 1)$ (note that we are adopting “bra-ket” vector notation commonly used in quantum mechanics literature). The advantage of this formalism is that the mean $\langle n \rangle$ and variance σ^2 of the distribution of product molecules P_n can be calculated from $G(\chi)$ via

$$\langle n \rangle = \left. \frac{d(\ln G)}{d(i\chi)} \right|_{\chi=0}, \quad \sigma^2 = \left. \frac{d^2(\ln G)}{d(i\chi)^2} \right|_{\chi=0} \quad (5)$$

Furthermore, we note that having N (independently acting) enzymes is equivalent to taking G to G^N , so that extension to larger number of enzymes is straightforward.

Now multiplying (1) and (2) by $e^{i\chi n}$ and summing over n produces

$$\frac{d|\mathbf{G}\rangle}{dt} = \hat{\mathbf{H}}|\mathbf{G}\rangle \quad (6)$$

where, for the MM reaction,

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}_A = \begin{pmatrix} -k_1 S & k_{-1} + k_2 e^{i\chi} \\ k_1 S & -(k_{-1} + k_2) \end{pmatrix} \quad (7)$$

Equation (6) is formally solved by

$$|\mathbf{G}(t)\rangle = e^{\hat{\mathbf{H}}t} |\mathbf{G}_0\rangle \quad (8)$$

with an initial condition $|\mathbf{G}_0\rangle$. If we write the matrix exponential in terms of the eigenvalues λ_j and eigenvectors $|\mathbf{u}_j\rangle$ of $\hat{\mathbf{H}}$ as

$$e^{\hat{\mathbf{H}}t} = \sum_j e^{\lambda_j t} |\mathbf{u}_j\rangle \langle \mathbf{u}_j| \quad (9)$$

(note that since $\hat{\mathbf{H}}$ is not symmetric, the eigenvectors do not satisfy $|\mathbf{u}_j\rangle = \langle \mathbf{u}_j|^T$, but rather they solve $\hat{\mathbf{H}}|\mathbf{u}_j\rangle = \lambda_j|\mathbf{u}_j\rangle$ and $\langle \mathbf{u}_j|\hat{\mathbf{H}} = \lambda_j\langle \mathbf{u}_j|$, respectively) then, at t much larger than the typical enzyme turnover time, $G(\chi)$ becomes

$$G(\chi) = \sum_j e^{\lambda_j t} \langle \mathbf{1} | \mathbf{u}_j \rangle \langle \mathbf{u}_j | \mathbf{G}_0 \rangle \simeq e^{\lambda_0 t} \langle \mathbf{1} | \mathbf{u}_0 \rangle \langle \mathbf{u}_0 | \mathbf{G}_0 \rangle \quad (10)$$

where λ_0 is the eigenvalue with the least negative real part. Taking the log, we obtain

$$\ln G(\chi) = \lambda_0 t + \ln(\langle \mathbf{1} | \mathbf{u}_0 \rangle \langle \mathbf{u}_0 | \mathbf{G}_0 \rangle) \simeq \lambda_0 t \quad (11)$$

since again, in the long-time limit, the first term dominates the second (for any bounded \mathbf{G}_0), and the initial number of product molecules is forgotten. Recalling (5), it is clear now that one only needs to find the χ -dependence of the least negative eigenvalue λ_0 of the matrix $\hat{\mathbf{H}}_A$ in order to compute the cumulants of the product molecule distribution. In fact,

writing λ_0 as a power series,

$$\lambda_0 = \sum_{m=0}^{\infty} \lambda_0^{(m)} \frac{(i\chi)^m}{m!} \quad (12)$$

it is clear that one only needs to know the coefficients up to $m = 2$ in order to compute the mean and variance of the distribution. Specifically

$$\langle n \rangle = \lambda_0^{(1)} t \quad (13)$$

$$\sigma^2 = \lambda_0^{(2)} t \quad (14)$$

and higher order terms are needed for higher cumulants only. Since (5) takes $\chi \rightarrow 0$, this permits a perturbative approach similar to that used in quantum mechanics [14], with χ treated as a small parameter.

Specifically, we write $\hat{\mathbf{H}} = \hat{\mathbf{H}}^{(0)} + \hat{\mathbf{H}}^{(1)} \sum_{m=1}^{\infty} (i\chi)^m / m!$ where (for the MM case)

$$\hat{\mathbf{H}}_A^{(0)} = \begin{pmatrix} -k_1 S & k_{-1} + k_2 \\ k_1 S & -(k_{-1} + k_2) \end{pmatrix}, \quad \hat{\mathbf{H}}_A^{(1)} = k_2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (15)$$

and we truncate at $m = 2$. We emphasise that this truncation does not introduce any further approximation if one is interested only in the first and second moments of the product molecule distribution. The least negative eigenvalue of $\hat{\mathbf{H}}^{(0)}$ is $\lambda_0^{(0)} = 0$ (more precisely, $\hat{\mathbf{H}}_0$ is a propensity matrix whose columns sum to zero, which means one of its eigenvalues is zero and the rest are negative [8]), and the higher order corrections are given by [14]

$$\lambda_0^{(1)} = \langle \mathbf{u}_0^{(0)} | \hat{\mathbf{H}}^{(1)} | \mathbf{u}_0^{(0)} \rangle \quad (16)$$

$$\lambda_0^{(2)} = \lambda_0^{(1)} - 2 \sum_{j \neq 0} \frac{1}{\lambda_j^{(0)}} |\langle \mathbf{u}_j^{(0)} | \hat{\mathbf{H}}^{(1)} | \mathbf{u}_0^{(0)} \rangle|^2 \quad (17)$$

Noting (13) and (14), the rate of product formation $V = d\langle n \rangle / dt$ and the Fano factor $F = \sigma^2 / \langle n \rangle$ can now be written as

$$V = \lambda_0^{(1)} \quad (18)$$

$$F = \lambda_0^{(2)} / \lambda_0^{(1)} \quad (19)$$

For the MM case (Fig. 1a), this gives

$$V_A = V_A^{\max} \frac{S}{S + K_A} \quad (20)$$

$$F_A = 1 - \alpha_A \frac{S}{(S + K_A)^2} \quad (21)$$

where $V_A^{\max} = k_2$, $K_A = (k_2 + k_{-1}) / k_1$ and $\alpha_A = 2k_2 / k_1$. The expression for mean flux V_A is well known [1], and K_A

is called the Michaelis constant; the expression for the Fano factor F_A is less familiar.

This procedure is fully extendible to other more complicated enzyme-mediated reactions. The reaction scheme determines the master equation and thus $\hat{H}^{(0)}$ and $\hat{H}^{(1)}$. Specifically, $\hat{H}^{(0)}$ is given by the Markov transition matrix for the enzymatic states (disregarding the n variable), and $\hat{H}^{(1)}$ has a 1 marking every rate where the product gets created, and a -1 where it is destroyed. Then (18) and (19) give the product formation rate and the Fano factor, and higher orders in perturbation theory would provide more cumulants. To illustrate the breadth of the method, in the next section, we apply this procedure to three reaction schemes that include multiple intermediate reaction steps.

2 Results: complex enzymatic reactions

2.1 Product distribution statistics

Many enzyme-mediated reactions involve intermediate steps, and it is instructive to illustrate our approach with three prototypical examples, shown in Figs. 1b–d.

2.1.1 Reaction scheme B: Fig. 1b depicts a case in which the complex undergoes an intermediate step, such as a conformational change, before creating the product [15]. This kinetic scheme is also equivalent to certain ion channels [5]. Such multistep enzymatic reactions have been shown (including via our method here) to reduce noise in chemical reactions [16]. The master equation describing this system is

$$\frac{dP_n^E}{dt} = -k_1 S P_n^E + k_{-1} P_n^{ES} + k_2 P_{n-1}^{EP} \quad (22)$$

$$\frac{dP_n^{ES}}{dt} = k_1 S P_n^E - (k_{-1} + k_+) P_n^{ES} + k_- P_n^{EP} \quad (23)$$

$$\frac{dP_n^{EP}}{dt} = k_+ P_n^{ES} - (k_- + k_2) P_n^{EP} \quad (24)$$

which yields

$$\hat{H}_B^{(0)} = \begin{pmatrix} -k_1 S & k_{-1} & k_2 \\ k_1 S & -(k_{-1} + k_+) & k_- \\ 0 & k_+ & -(k_- + k_2) \end{pmatrix}, \quad \hat{H}_B^{(1)} = k_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (25)$$

The product flux and Fano factor are then

$$V_B = V_B^{\max} \frac{S}{S + K_B} \quad (26)$$

$$F_B = 1 - \alpha_B \frac{S(S + K_B')}{(S + K_B)^2} \quad (27)$$

where $V_B^{\max} = k_2 k_+ / (k_2 + k_+ + k_-)$, $K_B = (k_2 k_+ + k_2 k_{-1} + k_{-1} k_-) / (k_1 (k_2 + k_+ + k_-))$, $\alpha_B = 2k_2 k_+ / (k_2 + k_+ + k_-)^2$ and $K_B' = (k_2 + k_+ + k_- + k_{-1}) / k_1$.

2.1.2 Reaction scheme C: Fig. 1c depicts a case in which the enzyme exists in an inactive and an active state. The enzyme switches autonomously between these states, but can only react with the substrate in its active form. Note that in this case we have two isolated reactions, since the enzyme remains in the active state when a product is produced. This scheme can be interpreted as a toy model for a voltage-gated ion channel that can only transmit a single molecule at a time [17]. Alternatively, this scheme could be a model for the production–degradation and subsequent translation of mRNA (E^*) by ribosomes (S) into protein (P). Finally, this is also an extreme model of an enzyme that has internal states with different rates of product formation, such as studied in [2]. For this scheme we can write the following master equation

$$\frac{dP_n^E}{dt} = -k_+ P_n^E + k_- P_n^{E^*} \quad (28)$$

$$\frac{dP_n^{E^*}}{dt} = k_+ P_n^E - k_- P_n^{E^*} + k_2 P_{n-1}^{E^*S} - k_1 S P_n^{E^*} + k_{-1} P_n^{E^*S} \quad (29)$$

$$\frac{dP_n^{E^*S}}{dt} = -k_2 P_n^{E^*S} - k_{-1} P_n^{E^*S} + k_1 S P_n^{E^*} \quad (30)$$

which yields

$$\hat{H}_C^{(0)} = \begin{pmatrix} -k_+ & k_- & 0 \\ k_+ & -(k_- + k_1 S) & k_{-1} + k_2 \\ 0 & k_1 S & -(k_{-1} + k_2) \end{pmatrix}, \quad \hat{H}_C^{(1)} = k_2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad (31)$$

The product flux and Fano factor are then

$$V_C = V_C^{\max} \frac{S}{S + K_C} \quad (32)$$

$$F_C = 1 - \alpha_C \frac{S}{(S + K_C)^2} \quad (33)$$

where $V_C^{\max} = k_2$, $K_C = (k_+ + k_-)(k_2 + k_{-1}) / (k_+ k_1)$ and $\alpha_C = 2k_2 [1 + k_- (k_+ - k_2 - k_{-1}) / k_+^2] / k_1$. Note that these

expressions reduce to those for the MM reaction [(20) and (21)] for $k_- \rightarrow 0$, since this limit corresponds to the enzyme always being in the active state. Note also that since α_C can be negative, F_C can be greater than 1 (and in fact it is infinite in the limit of rare activation $k_+ \rightarrow 0$) because of the compounded noise from the independent stochastic processes of enzyme activation and complex formation. Under the interpretation of this scheme as protein translation, $F \gg 1$ corresponds to many proteins in a translation burst from a single rare mRNA.

2.1.3 Reaction scheme D: Fig. 1d shows a third example of a more complex reaction scheme, in which an active enzyme transforms a substrate into a product and, in contrast to scheme C, returns to its inactive state in the process. The enzyme must switch back to its active state for a new reaction to occur. Similar dynamics have been found for the β -galactosidase enzyme [2]. Alternatively, this can be a model for an enzyme that transfers a phosphate group to a substrate, and needs to reacquire a new phosphate group before continuing to function as an enzyme. For this scheme, we can write the following master equation

$$\frac{dP_n^E}{dt} = -k_+ P_n^E + k_2 P_{n-1}^{E^*S} \quad (34)$$

$$\frac{dP_n^{E^*}}{dt} = k_+ P_n^E - k_1 S P_n^{E^*} + k_{-1} P_n^{E^*S} \quad (35)$$

$$\frac{dP_n^{E^*S}}{dt} = k_1 S P_n^{E^*} - k_{-1} P_n^{E^*S} - k_2 P_n^{E^*S} \quad (36)$$

which yields

$$\hat{H}_D^{(0)} = \begin{pmatrix} -k_+ & 0 & k_2 \\ k_+ & -k_1 S & k_{-1} \\ 0 & k_1 S & -(k_{-1} + k_2) \end{pmatrix},$$

$$\hat{H}_D^{(1)} = k_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (37)$$

Table 1 Bounds on experimentally measurable quantities that are useful in distinguishing among schemes for enzyme-mediated reactions

	A	B	C	D
$F(S \rightarrow \infty)$	1	[1/2, 1]	1	[1/2, 1]
F^*	[1/2, 1]	[1/3, 1]	[1/2, ∞)	[1/3, 1]
S^*/K	1	[1, ∞)	1	[1, ∞)

A, B, C and D refer to reaction schemes in Fig. 1. Star (*) denotes the extremum of the Fano factor, such that F^* is the minimum or maximum value and S^* is the substrate concentration at which it occurs. K is the substrate concentration at which product formation rate V is half-maximal. Generally speaking, minimum bounds on all three quantities occur when forward reaction rates dominate backward rates, and maximum bounds occur when backward rates dominate forward rates; see text for more details.

The product flux and the Fano factor are then

$$V_D = V_D^{\max} \frac{S}{S + K_D} \quad (38)$$

$$F_D = 1 - \alpha_D \frac{S(S + K_D')}{(S + K_D)^2} \quad (39)$$

where $V_D^{\max} = k_2 k_+ / (k_2 + k_+)$, $K_D = k_+(k_2 + k_{-1}) / (k_1(k_2 + k_+))$, $\alpha_D = 2k_2 k_+ / (k_2 + k_+)^2$ and $K_D' = (k_2 + k_+ + k_{-1}) / k_1$. Note that these expressions reduce to those for the MM reaction [(20) and (21)] for $k_+ \rightarrow \infty$, since this limit corresponds to the immediate reversion of the enzyme to its active state following a product formation.

All four reactions in Fig. 1 use an enzyme to convert a substrate into a product, but as we have derived using the present method, the statistical properties of the product molecule distributions differ among the cases.

2.2 Measurable differences between reaction schemes

Since different reactions have different statistical properties, it should be possible to use our methods and results to differentiate among the underlying reactions based on experimental observations. Here we demonstrate how basic measurements can differentiate among the four reaction schemes presented above.

The mean product formation rates V for all four reaction schemes A, B, C and D shown in Fig. 1, (20), (26), (32) and (38), are qualitatively similar functions of substrate concentration S , and it would not be possible to differentiate the schemes based on mean data alone (see Fig. 2). Measurement of the Fano factor F [(21), (27), (33) and (39)], however, can reveal qualitative and quantitative features that can differentiate among these schemes, which we outline here and summarise in Table 1.

First, a distinction is possible based on the asymptotic value of F as the substrate concentration S saturates. For

reaction schemes A and C,

$$F_{A,C}(S \rightarrow \infty) = 1 \quad (40)$$

whereas for reaction schemes B and D

$$F_{B,D}(S \rightarrow \infty) = 1 - \alpha_{B,D} \quad (41)$$

where α_B and α_D are defined following (27) and (39), respectively. This expression has a minimum value 1/2 in the limits $k_2 = k_+ \gg k_-$ for B and $k_2 = k_+$ for D. Thus a saturation value of F that is significantly less than 1 offers evidence for reaction scheme B or D over A or C (see Fig. 2).

Second, distinctions are possible based on the value F^* at the extremum of the Fano factor as a function of substrate concentration S . For a MM reaction (case A), there is a minimum

$$F_A^* = 1 - \frac{\alpha_A}{4K_A} = 1 - \frac{1}{2} \frac{k_2}{k_2 + k_{-1}} \quad (42)$$

which is always between 1/2 (for $k_2 \gg k_{-1}$) and 1 (for $k_{-1} \gg k_2$). Similarly, for reaction scheme C, we obtain

$$F_C^* = 1 - \frac{\alpha_C}{4K_C} \quad (43)$$

where α_C and K_C are defined following (33). This expression also has a minimum value of 1/2 (for $k_+ \gg k_-$ and $k_2 \gg k_{-1}$), but, unlike in the MM case, it can become larger than 1 if $k_+(k_+ + k_-) < k_-(k_2 + k_{-1})$ (see Fig. 2). Indeed, as mentioned, in the limit of rare activation $k_+ \rightarrow 0$, we find $F^* \rightarrow \infty$.

Depending on the kinetic rates, reaction schemes B and D may or may not have a minimum of F for positive S (see Fig. 2 for an example of each). When a minimum exists

$$F_{B,D}^* = 1 - \frac{\alpha_{B,D}}{4} \frac{K_{B,D}^{\prime 2}}{K_{B,D}(K_{B,D}' - K_{B,D})} \quad (44)$$

where α_B , K_B and K_B' are defined following (27) and α_D , K_D , and K_D' are defined following (39). This expression has the minimum value 1/3 in the limit $k_+ = k_2 \gg k_{-1}$ for both schemes (and additionally $k_+ \gg k_-$ for B). In the reaction scheme B, these limits reduce the system to a linear irreversible three-step cascade; an L -step irreversible cascade has minimum F^* of $1/L$ in the analogous limits [16]. Comparing with the MM minimum value of $F^* = 1/2$, it is clear that a measured value of F^* less than 1/2 is a strong indication that more than one intermediate step is present. This is a commonly known result, and it may be used by nature to suppress noise in natural signalling systems such as phototransduction [16].

Lastly, distinctions can be made based on measurement of S^* , the substrate concentration at which an extremum in F

occurs. For cases A and C

$$\frac{S_{A,C}^*}{K_{A,C}} = 1 \quad (45)$$

where K_A and K_C are defined following (21) and (33), respectively, and, as in all four cases, K is the concentration at which V is half-maximal. For cases B and D, on the other hand (when there is a minimum)

$$\frac{S_{B,D}^*}{K_{B,D}} = \frac{K_{B,D}'}{K_{B,D}' - 2K_{B,D}} \quad (46)$$

where K_B and K_B' are defined following (27) and K_D and K_D' are defined following (39). This expression is bounded from below by 1 (e.g. for $k_+ \gg \{k_-, k_2, k_{-1}\}$ for B, or for $k_+ \gg \{k_2, k_{-1}\}$ for D), but can potentially be infinite (e.g. for $k_- = k_{-1} \gg \{k_2, k_+\}$ for B, or for $k_{-1} \gg k_2 = k_+$ for D). This implies that if an extremum of the Fano factor occurs at a substrate concentration significantly different from that at which the mean product formation rate is half-maximal, it is a strong indication that more than one intermediate step is present.

Table 1 summarises these distinctions, and Fig. 2 showcases the qualitative differences in the Fano factor curves among the four reaction schemes caused by differences in the underlying kinetics. For more complicated reaction schemes, such as multiple substrate binding by the enzyme, modelled by a high Hill coefficient, the Fano factor curve would gain even more distinguishing features, such as additional extrema and/or inflection points. We leave this as an exercise for future q-bio Summer School students.

2.3 Extracting reaction rates from data

In addition to helping one distinguish among competing reaction schemes, experimental measurement of the dose–response curve $V(S)$ and the Fano curve $F(S)$ can be used to determine the kinetic rates of the underlying biochemical reactions. If the structure of the biochemical reaction is known, analytical expressions for both curves in terms of the kinetic rates and the substrate concentration can be obtained using our method [see e.g. (20), (21), (26), (27), (32), (33), (38), (39)] and can be fit to experimental data. Often times, measurements of the qualitative features of both curves (such as those highlighted in Table 1) are enough to extract the kinetic rates; for more complex reactions a full fit to the data would be necessary. Additionally, we note that performing full fits of experimental data to the analytical expressions may also help in the original task of distinguishing among (or at least eliminating) different biochemical reaction schemes.

The MM reaction is an example of a case in which measurement of the qualitative features is enough to extract all kinetic rates. However, it is important to note that in order to do this, one needs both the dose–response curve

and the Fano curve. In particular, one needs only to measure the reaction rate at saturation V_A^{\max} , the substrate concentration K_A at which the rate is half maximal, and the minimum value of the Fano curve F_A^* . Then, from (42) and the expressions following (21), one obtains

$$k_2 = V_A^{\max} \quad (47)$$

$$k_{-1} = \frac{F_A^* - 1/2}{1 - F_A^*} V_A^{\max} \quad (48)$$

$$k_1 = \frac{V_A^{\max}}{2K_A(1 - F_A^*)} \quad (49)$$

Instead of obtaining only k_2 and a combination of k_1 and k_{-1} by measuring only the dose–response curve (as is traditionally done for MM reactions), we now have analytical expressions for all three rates.

For more complex reaction schemes, a similar analysis can be performed to obtain analytical expressions for the kinetic rates in terms of the experimental data. However, it can be the case that not all rates can be determined unambiguously from measurements of V and F (for the reaction scheme B, for example, symmetries in the inverted expression imply that measurements of V and F do not always uniquely determine the five unknown kinetic rates). When experimentally feasible, one may also compare higher moments of the measured product molecule distributions with those calculated via our method.

3 Discussion

In this paper, we have developed a novel, simple method for calculating the first-order cumulants of the number of complex chemical reactions that occur over a long time. We have used this method to calculate these cumulants for a handful of simple enzymatic schemes, providing results for the reaction number averages and Fano factors that have not, to our knowledge, appeared in the literature before. Finally, we have argued that these results can be used by experimentalists with access to measurements of fluctuations of mesoscopic particle fluxes to disambiguate different kinetic schemes, even if single-molecule experiments, traditionally used for this type of work, are impossible to perform.

To allow disambiguation of kinetic schemes by our methods, one must be able to measure variances of product fluxes. As our analysis shows, for simple schemes, the variance is of the same order as the mean. That is, if n product molecules are created on average, one would need to measure fluctuations on the order of \sqrt{n} , for a relative measurement error of $1/\sqrt{n}$. For just 10^2 molecules, this requires an experimental precision of better than 10%, and this number falls quickly as the system size grows. Thus, for disambiguation of kinetic schemes, we have tried to

point out qualitative rather than quantitative distinctions, which might be easier to measure experimentally.

Our method uses a perturbation theory approach analogous to that encountered in quantum mechanics. Under this approach, calculating cumulants of the distribution of product molecules is equivalent to diagonalising a matrix with dimensionality equal to the number of internal states in the kinetic diagram of the reaction. Then obtaining the first m cumulants of the reaction can be done by solving the perturbation theory to m th order, which is straightforward. In particular, the first two moments $\langle n \rangle$ and σ^2 together define the dose–response curve $V = d\langle n \rangle/dt$ and the Fano factor $F = \sigma^2/\langle n \rangle$. As both are currently measurable in a variety of systems, comparing the calculated F to experimental data can be used to identify the underlying structure of molecular reactions.

We have applied this perturbation theory approach to four different enzymatic reaction schemes, starting with the simplest MM kinetics, and progressing to more complicated kinetic schemes with internal states. We calculated the dose–response curve and the Fano factor for each as functions of the substrate concentration. Importantly, while the dose–response curves for all of the considered reactions are qualitatively similar, prominent qualitative features of the Fano factor curve (such as its values at large substrate concentrations, as well as the position and values at its extremum) allow us to disambiguate the considered reaction schemes. Performing detailed fits of the curves to experimental data (when feasible) can be an ultimate test for whether the underlying kinetic structure is known.

For the MM reaction, knowing just a handful of features of the $F(S)$ curve allows us to derive all three rates that completely define the kinetic scheme, while the entire dose–response curve is insufficient for this purpose. Similar results hold for the reactions with intermediate steps, but here the analytical treatment is more difficult, and often qualitative properties of F alone do not define all of the underlying kinetic parameters. Instead, a quantitative fit of derived expressions for $F(S)$ to experimental data would be required.

We stress that the kinetic schemes analysed in this paper are simple toy models only. However, extending our analysis to more complicated schemes to derive the first few cumulants of the product number distribution is not difficult, and it can be automated with just a simple linear-algebra solver. In particular, calculation of the Fano factor for a signalling cascade as in [16] or for a complex network of single protein conformations [6] is straightforward. It should be noted, however, that generating sufficient experimental data to distinguish minute details of competing kinetic schemes is not easy. Our approach simplifies the problem somewhat since it does not require single-molecule kinetic data, as in [2, 6], but it is based on measuring a mesoscopic, fluctuating flux. Still, qualitative

differences would ideally dominate the disambiguation task, as emphasised with the toy models considered here.

4 Acknowledgments

The authors would like to thank the organisers, the lecturers, the participants and the sponsors of The Second q-bio Summer School on Cellular Information Processing in Los Alamos, NM, where this work was started. We are also thankful to Michael Wall for careful reading of the manuscript. WdR was supported by the research programme of the 'Stichting voor Fundamenteel Onderzoek der Materie', which is financially supported by The Netherlands Organization for Scientific Research. BCD was supported by NSF Grant DMR-0705167. AM was supported by NSF Grant DGE-0742450. NAS and IN were supported by DOE under Contract No. DE-AC52-06NA25396. IN was further supported by NSF Grant No. ECS-0425850. WdR, BCD and AM contributed equally to this work.

5 References

- [1] MICHAELIS L., MENTEN M.: 'Die kinetik der invertinwirkung', *Biochem. Z.*, 1913, 49, pp. 333–369
- [2] ENGLISH B., MIN W., VAN OIJEN A., ET AL.: 'Ever-fluctuating single enzyme molecules: Michaelis–Menten equation revisited', *Nat. Chem. Biol.*, 2006, 2, pp. 87–94
- [3] ROSTOVITSEVA T., BEZRUKOV S.: 'ATP transport through a single mitochondrial channel, VDAC, studied by current fluctuation analysis', *Biophys. J.*, 1998, 74, pp. 2365–2373
- [4] NESTOROVICH E., DANOLON C., WINTERHALTER M., BEZRUKOV S.: 'Designed to penetrate: time-resolved interaction of single antibiotic molecules with bacterial pores', *Proc. Natl. Acad. Sci. (USA)*, 2002, 99, pp. 9789–9794
- [5] BEZRUKOV S., BEREZHKOVSII A., SZABO A.: 'Diffusion model of solute dynamics in a membrane channel: mapping onto the two-site model and optimizing the flux', *J. Chem. Phys.*, 2007, 127, p. 115101
- [6] LI C.-B., YANG H., KOMATSUZAKI T.: 'Multiscale complex network of protein conformational fluctuations in single-molecule time series', *Proc. Natl. Acad. Sci. (USA)*, 2008, 105, p. 536
- [7] GOLDING I., PAULSSON J., ZAWILSKI S., COX E.: 'Real-time kinetics of gene activity in individual bacteria', *Cell*, 2005, 123, pp. 1025–1036
- [8] VAN KAMPEN N.: 'Stochastic processes in physics and chemistry' (North-Holland, Amsterdam, 1992)
- [9] SINITSYN N., NEMENMAN I.: 'The Berry phase and the pump flux in stochastic chemical kinetics', *EPL-Europhys. Lett.*, 2007, 77, p. 58001
- [10] SINITSYN N., NEMENMAN I.: 'Universal geometric theory of mesoscopic stochastic pumps and reversible ratchets', *Phys. Rev. Lett.*, 2007, 99, p. 220408
- [11] BAGRETS D., NAZAROV Y.: 'Full counting statistics of charge transfer in coulomb blockade systems', *Phys. Rev. B*, 2003, 67, p. 085316
- [12] GOPICH I., SZABO A.: 'Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis', *J. Chem. Phys.*, 2006, 124, p. 4712
- [13] HORNOS J., SCHULTZ D., INNOCENTINI G., ET AL.: 'Self-regulating gene: an exact solution', *Phys. Rev. E*, 2005, 72, p. 51907
- [14] GRIFFITHS D.: 'Introduction to quantum mechanics.' (Prentice-Hall, 1994)
- [15] FRENZEN C., MAINI P.: 'Enzyme kinetics for a two-step enzymic reaction with comparable initial enzyme-substrate ratios', *J. Math. Biol.*, 1988, 26, p. 689
- [16] DOAN T., MENDEZ A., DETWILER P., CHEN J., RIEKE F.: 'Multiple phosphorylation sites confer reproducibility of the rod's single-photon responses', *Science*, 2006, 313, pp. 530–533
- [17] HODGKIN A., HUXLEY A.: 'A quantitative of membrane current and its application to conduction and excitation in nerve', *J. Physiol.*, 1952, 117, p. 500