SCIENTIFIC REPORTS

OPEN

Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters

Received: 5 May 2017
Accepted: 22 November 2017
Published online: 05 December 2017

Boseung Choi[1], Grzegorz A. Rempala[2] & Jae Kyoung Kim[3]

Examining enzyme kinetics is critical for understanding cellular systems and for using enzymes in industry. The Michaelis-Menten equation has been widely used for over a century to estimate the enzyme kinetic parameters from reaction progress curves of substrates, which is known as the progress curve assay. However, this canonical approach works in limited conditions, such as when there is a large excess of substrate over enzyme. Even when this condition is satisfied, the identifiability of parameters is not always guaranteed, and often not verifiable in practice. To overcome such limitations of the canonical approach for the progress curve assay, here we propose a Bayesian approach based on an equation derived with the total quasi-steady-state approximation. In contrast to the canonical approach, estimates obtained with this proposed approach exhibit little bias for any combination of enzyme and substrate concentrations. Importantly, unlike the canonical approach, an optimal experiment to identify parameters with certainty can be easily designed without any prior information. Indeed, with this proposed design, the kinetic parameters of diverse enzymes with disparate catalytic efficiencies, such as chymotrypsin, fumarase, and urease, can be accurately and precisely estimated from a minimal amount of timecourse data. A publicly accessible computational package performing such accurate and efficient Bayesian inference for enzyme kinetics is provided.

Because enzymes can modulate biochemical reaction rates by selectively catalyzing specific substrates[1], they play fundamental roles in metabolism, signal transduction, and cell regulation, and their malfunction can cause serious diseases[2,3]. Furthermore, enzymes have been used as extremely specific catalysts in diverse industrial fields such as drug development, biofuel production, and food processing[4]. A canonical approach used to understand enzyme kinetics for a century has been based on the Michaelis-Menten equation (MM equation), which was developed by Michaelis and Menten[5] and then was more rigorously derived by Briggs and Haldane[6] using the standard quasi-steady-state approximation (sQSSA)[7]. The equation describes the dependence of enzyme-catalyzed reaction rates on the concentration of substrate by using two parameters, the catalytic constant, k_{cat} and the Michaelis-Menten constant, K_M (see below for details). The k_{cat} determines the maximum rate of the reaction at saturating substrate concentrations, $V_{max} = k_{cat}E_T$, where E_T is total enzyme concentration, and the K_M is the substrate concentration at which the reaction rate is half of V_{max}.

There are two major assays to estimate k_{cat} and K_M from a measured accumulation of product over time (i.e. progress curve): the initial velocity assay (initial rate analysis) and the reaction progress curve assay (progress curve analysis)[8–12]. For the initial velocity assay, initial rates of the reaction are measured for a range of substrate concentrations. Then, by using a linear transform of these data, such as Lineweaver-Burk plots, the two parameters can be easily estimated without use of any computational tools[8,9]. Recent advances in computational tools have led to an alternative approach: the reaction progress curve assay. In this assay, the entire timecourse (i.e. progress curve) is fitted to the solution of a differential equation or integrated rate equation, and thus the data

[1]Korea University Sejong campus, Division of Economics and Statistics, Department of National Statistics, Sejong, 30019, Korea. [2]The Ohio State University, Division of Biostatistics and Mathematical Biosciences Institute, Columbus, OH, 43210, USA. [3]Korea Advanced Institute of Science and Technology, Department of Mathematical Sciences, Daejeon, 34141, Korea. Correspondence and requests for materials should be addressed to J.K.K. (email: jaekkim@kaist.ac.kr)

is used more efficiently than in the initial velocity assay[10,11,13]. Albeit more technically challenging, the progress curve assay requires less data to estimate parameters than the initial velocity assay does.

Since both assays are based on the MM equation, they should be performed only when the MM equation is valid, that is, when the enzyme concentration is a much lower than the sum of the substrate concentration and the K_M[7,14] (see below for more details). Because the value of K_M is usually not known a priori, to ensure the validity of the MM equation, *in vitro* experiments are typically performed with a much lower enzyme concentration than substrate concentration[15]. However, such conditions cannot be guaranteed *in vivo*, because endogenous enzyme concentrations are much higher than those used in a typical *in vitro* assay[16,17]. It is therefore risky to use the MM equation to analyze *in vivo* data and to predict *in vivo* enzyme activity by using parameters estimated from an *in vitro* assay[15,18]. Furthermore, even when the MM equation is valid, precise estimation is not guaranteed, because of the highly correlated structure and unidentifiability of the parameters[19–23]. That is, even though estimated parameters can fit the data accurately, the estimates can differ greatly from the actual values of k_{cat} and K_M. Because of the identifiability issue, experimental designs to infer the maximum possible information about the parameters have been investigated[12,13,20–24]. For instance, to ensure that the parameters can be identified from the initial velocity assay, the initial concentration of substrate needs to be increased from a low level to a higher level until the reaction velocity is saturated. For the saturation, generally the initial substrate concentration needs to be larger than $10\,K_M$, but often such high concentrations cannot be achieved[24]. For the progress curve assay, the initial substrate concentration is recommended to be at a similar level to K_M[23,25]. Note that both assays require prior knowledge of K_M, which gives rise to the conundrum that, in order to estimate K_M, the approximate value of K_M needs to be known.

To overcome such limits on the inference using the model based on the MM equation, which is referred to as the sQ (standard QSSA) model, here we propose an alternative approach. In our approach, we use a different approximate model that is derived with the total QSSA and is referred to as the tQ (total QSSA) model[26–29]. By applying the Bayesian inference based on either the sQ or the tQ model to the product progress curve, we found that the estimates obtained with the sQ model were considerably biased when the enzyme concentration was not low. On the other hand, the estimates obtained with the tQ model were not biased for any combination of enzyme and substrate concentrations. Thus, with the tQ model, the experimental data from various conditions can be pooled without any restrictions to improve the accuracy and precision of the estimation. For instance, when two sets of timecourse data obtained under low and high enzyme concentrations are used together, the tQ model, but not the sQ model, leads to accurate and precise estimation. Another advantage of our approach is that, by analyzing the scatter plots of current estimates, the next optimal experiment to ensure the parameter identifiability can be easily designed without requiring any prior knowledge of the k_{cat} and K_M values. The proposed optimized design yields accurate and precise estimation from a minimal amount of data simulated based on the kinetics of various enzymes: chymotrypsin, fumarase and urease, which have disparate catalytic efficiencies (k_{cat}/K_M). We provide a publicly accessible computational package that performs the Bayesian inference based on the tQ model, thus leading to accurate and efficient estimation of enzyme kinetics.

Results

Two types of models describing enzyme kinetics: The sQ and tQ models. A fundamental enzyme reaction consists of a single enzyme and a single substrate, where the free enzyme (E) reversibly binds with the substrate (S) to form the complex (C), and the complex irreversibly dissociates into the product (P) and the free enzyme:

$$E + S \underset{k_b}{\overset{k_f}{\rightleftharpoons}} C \xrightarrow{k_{cat}} E + P,$$

where the total enzyme concentration ($E_T \equiv C + E$) and the total substrate and product concentration ($S_T \equiv S + C + P$) are conserved. A popular model describing the accumulation of the product over time is based on the MM equation, as follows (see Supplementary Method for detailed derivation):

$$\dot{P} = k_{cat}\frac{E_T(S_T - P)}{K_M + S_T - P}, \tag{1}$$

where $K_M = (k_b + k_{cat})/k_f$ is the Michaelis-Menten constant and k_{cat} is the catalytic constant. This sQ model derived with the standard QSSA has been widely used to estimate the kinetic parameters, K_M and k_{cat} from the progress curve of the product[8–11,23,25]. Another model describing the accumulation of the product is derived with the total QSSA; it was developed later than the sQ model and thus has received less attention for parameter estimation[26–29]:

$$\dot{P} = k_{cat}\frac{E_T + K_M + S_T - P - \sqrt{(E_T + K_M + S_T - P)^2 - 4E_T(S_T - P)}}{2}. \tag{2}$$

Although this tQ model is more complicated than the sQ model, it is accurate over wider ranges than the sQ model. Specifically, the sQ model is accurate when

$$\frac{E_T}{K_M + S_T} \ll 1, \tag{3}$$

which requires a low enzyme concentration[7,14]. On the other hand, the tQ model is accurate when

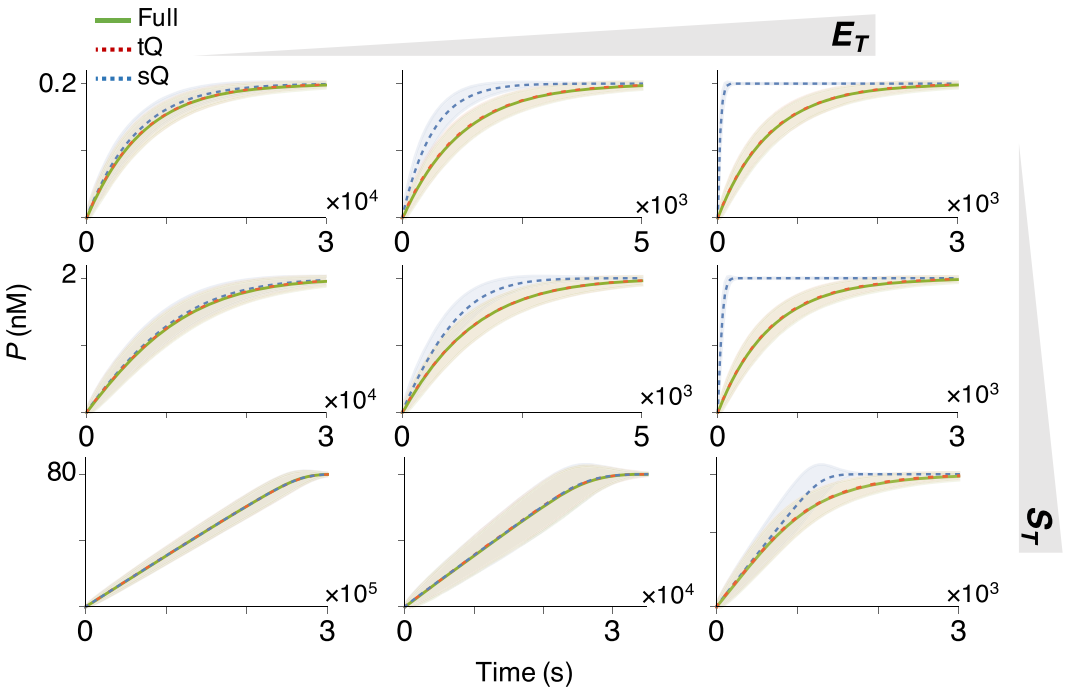

Figure 1. Whereas the sQ model fails to approximate the original full model as E_T increases, the tQ model is accurate regardless of E_T. Stochastic simulations of the original full model (Table S1), the sQ model (Table S2), and the tQ model (Table S3) were performed with $S_T = 0.2$, 2, or 80 nM, and $E_T = 0.2$, 2, or 40 nM. Note that these concentrations are either lower than, similar to, or higher than $K_M \approx 2$ nM. Here, the lines and colored ranges represent a mean trajectory and fluctuation range ($\pm 2\sigma$ from the mean) of 10^4 stochastic simulations.

$$\frac{K}{2S_T}\frac{E_T + K_M + S_T}{\sqrt{(E_T + K_M + S_T - P)^2 - 4E_T(S_T - P)}} \ll 1, \tag{4}$$

where $K = k_b/k_f$ is the dissociation constant[27–29]. Importantly, this condition is generally valid and thus the tQ model, unlike the sQ model, is accurate even when the enzyme is in excess. See[14,30] for more details.

Next, we investigated the accuracy of the *stochastic* simulations performed with both models. Specifically, we compared the stochastic simulations using the Gillespie algorithm based on the propensity functions from either the original full model (described in Table S1), the sQ model (Table S2), or the tQ model (Table S3) for 9 different conditions[31–36]: E_T is either lower than, similar to, or higher than K_M, and S_T is also either lower than, similar to, or higher than K_M (Fig. 1). The stochastic simulations of the sQ model fail to approximate those of the original full model when E_T is not low (i.e., E_T is lower than neither S_T nor K_M). On the other hand, stochastic simulations using the tQ model are accurate for all conditions (Fig. 1), as is consistent with a recent study showing that stochastic simulations with the sQ and the tQ models are accurate when their deterministic validity conditions hold (Eqs (3) and (4))[37,38]. Taken together, the tQ model is valid for a wider range of conditions than the sQ model is in both the deterministic and the stochastic sense.

Estimation with the tQ model is unbiased for any combination of enzyme and substrate concentrations. Because the tQ model is accurate for a wider range of conditions than the sQ model is (Fig. 1), we hypothesized that the parameter estimation based on the tQ model is also accurate for more general conditions. To investigate this hypothesis, we first generated 10^2 noisy progress curves of P from the stochastic simulations of the original full model (Fig. S1). Then, we inferred parameters (k_{cat} and K_M) from these simulated data sets by applying the Bayesian inference with the likelihood functions based on either the sQ or the tQ model, under weakly informative gamma priors (Fig. S2) (see Methods for details). Note that throughout this study, we have used the simulated product progress curves (e.g. Fig. S1) because we need to know the true values of parameters for the accurate comparison of the estimations based on the sQ model and the tQ model.

We first focused on the estimation of the k_{cat} under the assumption that the value of K_M is known. When E_T is low, so that both the sQ and the tQ models are accurate (Fig. 1 left), posterior samples obtained with both models are similar and successfully capture the true value of k_{cat} (Fig. 2a left). The posterior samples obtained with the two models are similar because, when E_T is low and thus $E_T \ll S_T + K_M$, both models (Eqs 1 and 2) are approximately equivalent as follows:

$$\frac{E_T + K_M + S_T - P - \sqrt{(E_T + K_M + S_T - P)^2 - 4E_T(S_T - P)}}{2} \approx \frac{E_T(S_T - P)}{K_M + E_T + S_T - P} \approx \frac{E_T(S_T - P)}{K_M + S_T - P}, \tag{5}$$

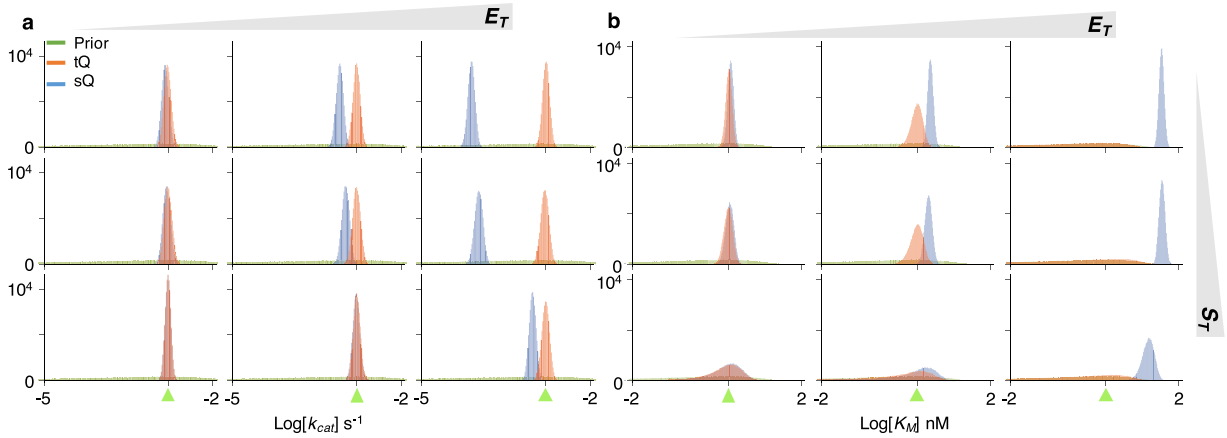

Figure 2. The estimation of a single parameter (k_{cat} or K_M) with either the sQ or the tQ model. For each condition ($S_T = 0.2$, 2, or 80 nM, and $E_T = 0.2$, 2, or 40 nM), 10^5 posterior samples of either k_{cat} (**a**) or K_M (**b**) were obtained by applying the Bayesian inference to 10^2 noisy data sets (Fig. S1) (see Methods for details). When the k_{cat} is sampled, the K_M is fixed at its true value (**a**) and vice versa (**b**). Here, green triangles indicate the true values of the parameters. Whereas the estimates of k_{cat} and K_M obtained with the sQ model are biased as E_T increases, those obtained with the tQ model have negligible bias regardless of conditions (See Fig. S3 for box plots of estimates). As E_T or S_T increases, the posterior variance of K_M increases when the tQ model is used.

where the first approximation comes from the Taylor expansion in terms of $E_T(S_T - P)/(E_T + K_M + S_T - P) \ll 1$ (see[27–29] for details). Therefore, when $E_T \ll S_T + K_M$ and thus the sQ model is accurate, estimations with the sQ and the tQ models should be similar. On the other hand, when E_T is high, they show clear differences (Fig. 2a right): the posterior samples obtained with the sQ model show large errors, while those obtained with the tQ model accurately capture the true value of k_{cat}.

Similar results are also observed in the box plots of posterior means and posterior coefficient of variations (CVs) (Fig. S3a,b). Whereas posterior means obtained with the sQ model are biased when E_T is high, those obtained with the tQ model are accurate for all conditions (Fig. S3a). In particular, narrow distributions of posterior means indicate that the estimation of k_{cat} with the tQ model is robust aginst the noise in the data (Fig. S1). Furthermore, posterior CVs are much smaller than prior CVs (Fig. S3b), indicating precise estimation of k_{cat} with the tQ model.

Next, K_M was estimated under the assumption that the value of k_{cat} is known (Fig. 2b). Posterior samples of the K_M obtained with the sQ model again show errors that grow with increasing E_T. Note that the estimates of the K_M are biased upward, which implies that using the posterior estimates of K_M to validate the MM equation ($K_M \gg E_T$) can be misleading. On the other hand, the estimates of K_M obtained with the tQ model are little biased for all conditions. However, unlike the narrow posterior distributions of k_{cat} (Fig. 2a), those of K_M obtained with the tQ model become wider; so precision decreases as E_T or S_T increases (Fig. 2b). These patterns are also observed in the box plots of posterior means and posterior CVs (Fig. S3c,d). The identifiability problem arises because, when $E_T \gg K_M$ or $S_T \gg K_M$ and thus $E_T + S_T \gg K_M$, the K_M is negligible in the tQ model (Eq. 2), as follows:

$$\frac{E_T + K_M + S_T - P - \sqrt{(E_T + K_M + S_T - P)^2 - 4E_T(S_T - P)}}{2} \approx \frac{E_T + S_T - P - \sqrt{(E_T + S_T - P)^2 - 4E_T(S_T - P)}}{2}. \quad (6)$$

Specifically, when K_M is too low, the value of K_M has little effect on the dynamics of the tQ model and thus the K_M is structurally unidentifiable. Taken together, the estimations of K_M with both the sQ and the tQ models are not satisfactory, although for different reasons: estimations with the sQ model can be biased and those with the tQ model can be structurally unidentifiable (Fig. 2b). Similar patterns were also observed when a more informative prior was given (Fig. S4). In particular, even with the informative prior, estimates obtained with the sQ model still show considerable error as E_T increases.

Simultaneous estimation of k_{cat} and K_M suffers from the lack of identifiability. Next, we considered simultaneous estimation of two parameters, k_{cat} and K_M, which is the typical goal of enzyme kinetics. For the same gamma priors used in the single-parameter estimation (Fig. 2), the distributions of posterior samples obtained with both models became wider overall (Fig. 3). To find the reason for such imprecise estimation, we analysed the scatter plots of posterior k_{cat} and K_M samples (Fig. 4). When $S_T \ll K_M$ (Fig. 4a–c), the posterior samples of k_{cat} and K_M obtained with the sQ model exhibited a strong correlation, because the dynamics of the sQ model depend only on the ratio k_{cat}/K_M, as seen in the following approximation:

$$k_{cat}\frac{E_T(S_T - P)}{K_M + S_T - P} \approx k_{cat}\frac{E_T(S_T - P)}{K_M},$$

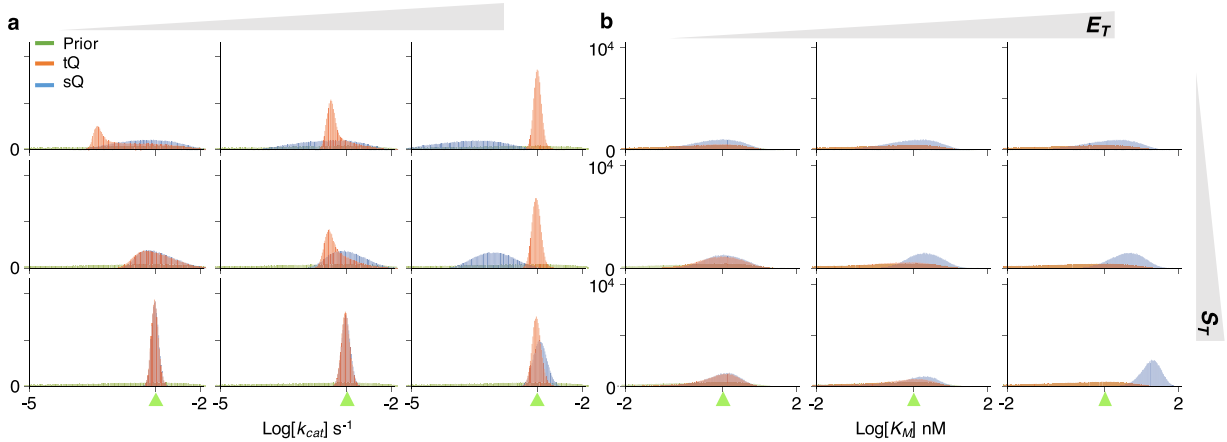

Figure 3. Simultaneous estimation of two parameters (k_{cat} and K_M) with either the sQ or the tQ model. From the same 10^2 data sets (Fig. S1) used in the single-parameter estimation (Fig. 2), 10^5 posterior samples of the k_{cat} (**a**) and the K_M (**b**) were obtained together. Although the same prior is given, the posterior distributions become wider than the single-parameter estimation (Fig. 2). Here, green triangles indicate the true values of k_{cat} or K_M.

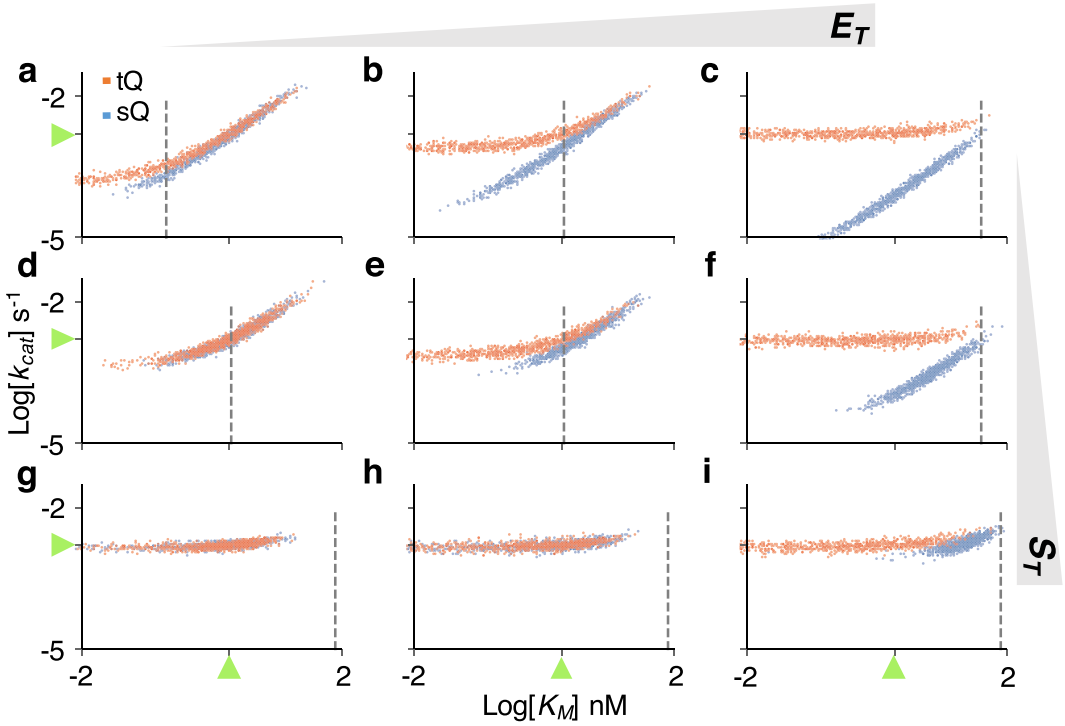

Figure 4. The scatter plots of posterior samples obtained with the two-parameter estimation (Fig. 3). The scatter plots imply two types of structure unidentifiability: strong correlation between k_{cat} and K_M, and unidentifiability of K_M, which is represented as a horizontal plot. Positively correlated scatter plots of the tQ model are changed to horizontal ones when the sampled K_M is much lower than $S_T + E_T$ (dashed gray lines). Here, green triangles represent the true values of parameters.

where $K_M \gg S_T \geq S_T - P$ is used. On the other hand, when $S_T \gg K_M$ (Fig. 4g–i), the scatter plot of the sQ model becomes horizontal, indicating the structure unidentifiability of the K_M. Indeed, the value of K_M has nearly no effect on the dynamics of the sQ model, as seen in the following approximation:

$$k_{cat}\frac{E_T(S_T - P)}{K_M + S_T - P} \approx k_{cat}E_T,$$

where $K_M + S_T \approx S_T$ is used as $S_T \gg K_M$. Such lack of parameter identifiability when $S_T \ll K_M$ or $S_T \gg K_M$ is consistent with previous studies, which recommend using $S_T \approx K_M$ for more precise estimation[22,23]. However, even

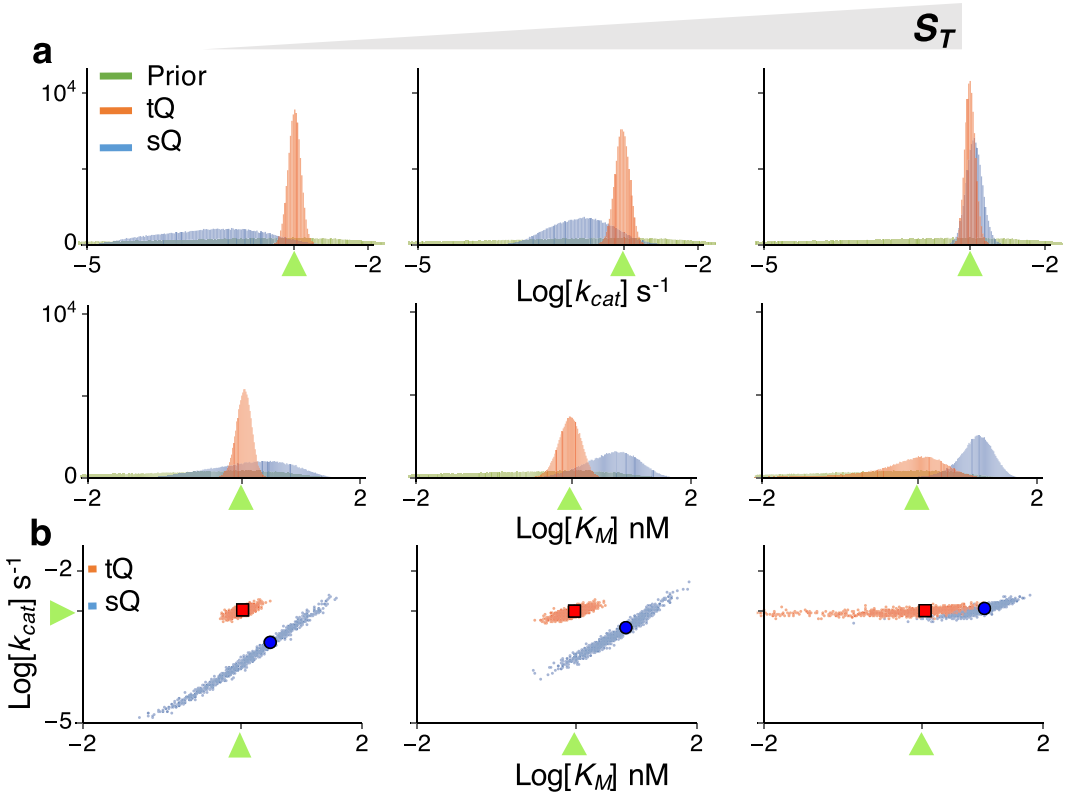

Figure 5. When data obtained under low E_T and high E_T are used together, the accuracy and precision of estimaties obtained with the tQ model, but not with the sQ model, are enhanced. (**a**) Posterior samples are inferred using data sets from $E_T = 0.2$ nM (Fig. S1 top) and $E_T = 40$ nM (Fig. S1 bottom) together for either $S_T = 0.2$, 2, or 80 nM. The posterior variance of the tQ model dramatically decreases to the level of the single-parameter estimation (Fig. 2). However, the estimates of the sQ model show considerable bias. Here, green triangles represent the true values of k_{cat} or K_M. (**b**) The scatter plots of the posterior samples. Here green triangles, blue circles, and red squares represent true values, posterior means of the sQ model, and those of the tQ model, respectively.

when $S_T \approx K_M$, estimates are still imprecise (Fig. 3a and b middle). Furthermore, as E_T increases, the estimates obtained with the sQ model are biased (Fig. 3) like in the single-parameter estimation (Fig. 2). Based on this analysis it appears that the simultaneous estimation of k_{cat} and K_M with the sQ model is challenging because of both identifiability and bias problems.

When $E_T \gg K_M$ or $S_T \gg K_M$, the K_M has a negligible effect on the dynamics of the tQ model (Eq. 6), and thus only k_{cat} was identifiable in the single-parameter estimation (Fig. 2a and b right or bottom). Similarly, when both k_{cat} and K_M are inferred simultaneously with the tQ model, estimation of only k_{cat} is accurate and precise (Fig. 3a and b right or bottom), as is shown by the horizontal scatter plots along the true value of k_{cat} (Fig. 4c,f,g–i). In other cases (when neither $E_T \gg K_M$ nor $S_T \gg K_M$), posterior variance of both parameters dramatically increases compared to the single-parameter estimation (Figs 2 and 3 left and top). Such imprecise estimation stems from two sources, according to the scatter plots (Fig. 4a,b,d,e). When k_{cat} and K_M decrease together, the behavior of the tQ model changes little as the SQ model (Eq. 5), which leads to the strong correlation between posterior samples of k_{cat} and K_M. As the estimates of K_M keep decreasing together with those of k_{cat}, so that they become much less than $E_T + S_T$ (dashed vertical line of Fig. 4), the tQ model no longer depends on the value of K_M, as shown in Eq. 6, and thus the scatter plots become horizontal.

Combined data from different experiments allow accurate and precise estimation with the tQ model. As shown above, the estimation of both k_{cat} and K_M using a single progress curve suffers from considerable bias and lack of identifiability (Figs 3 and 4), which is consistent with previous studies reporting that a progress curve obtained from a single experiment is not enough to identify both parameters simultaneously[19]. Thus, here, we investigate whether using multiple timecourse data sets obtained under different experimental conditions can improve the estimation.

In typical *in vitro* assays, progress curves are measured with either a fixed S_T and varied E_T or a fixed E_T and varied S_T[8–11,39]. We first consider the case when progress curves are measured with a fixed S_T and a varied E_T. Specifically, progress curves from both low and high E_T are used to estimate parameters for a fixed S_T at different levels (Fig. S1 top and bottom). In this case, posterior samples obtained with the sQ model show considerable errors as the data from high E_T is used (Figs 5a and S5). On the other hand, the posterior samples obtained with the tQ model accurately capture the true values of both k_{cat} and K_M with low variance (Figs 5a and S5). Such

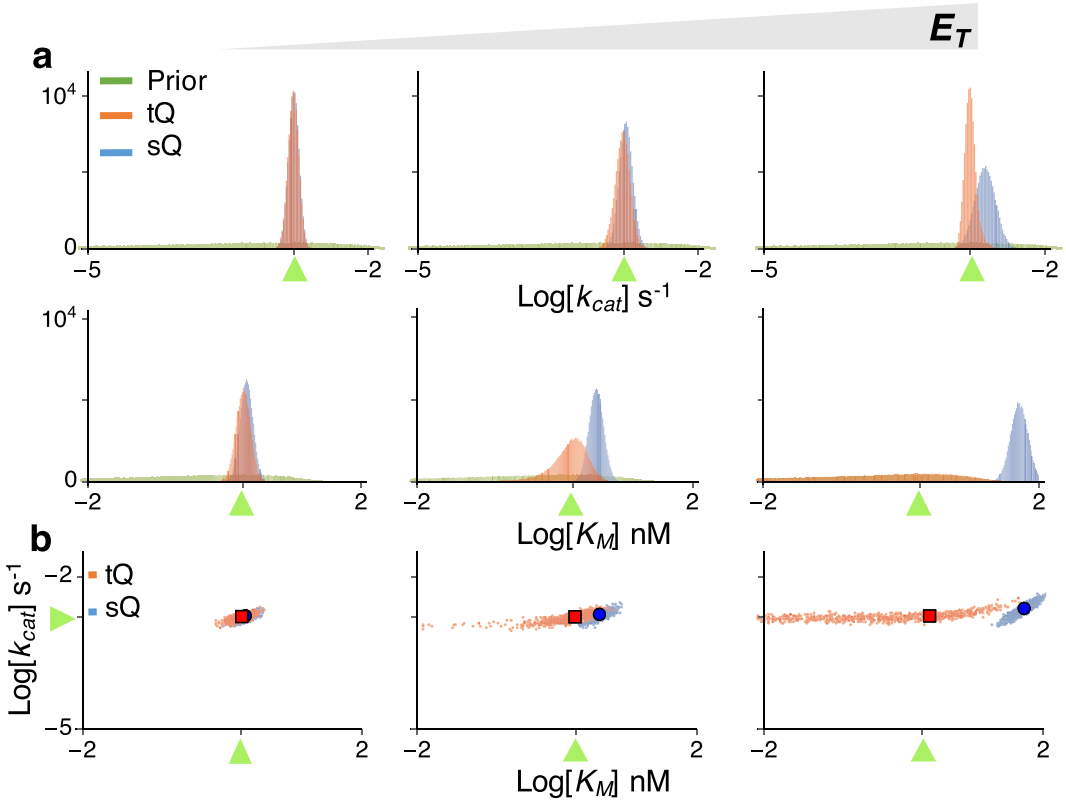

Figure 6. Estimation using the data obtained under low S_T and high S_T together. (**a**) Posterior samples are inferred using data sets from $S_T = 0.2$ nM (Fig. S1 left) and $S_T = 80$ nM (Fig. S1 right) together for either $E_T = 0.2$, 2, or 40 nM. When E_T is low, both the sQ and the tQ models allow accurate and precise estimation. As E_T increases, the estimates obtained with the sQ model become inaccurate, and the estimates of K_M obtained with the tQ model become less precise, similar to the single-parameter estimation (Fig. 2). Here, green triangles represent the true values of k_{cat} or K_M. (**b**) The scatter plots of the posterior samples. Here green triangles, blue circles, and red squares represent true values, posterior means of the sQ model, and those of the tQ model, respectively.

improvement stems from the fact that data obtained under the low and high E_T provide different types of information for parameter estimation. Specifically, from the high E_T data, although the K_M is not identifiable, the k_{cat} can be accurately estimated with the tQ model (Fig. 4c,f,i). Such accurate estimation of k_{cat} from the high E_T data can prevent the correlation between the k_{cat} and the K_M when they are estimated from the low E_T data (Fig. 4a,d). Indeed, the narrow scatter plots of the tQ model (Fig. 5b left and middle) are the intersection of two scatter plots, a horizontal one obtained with the high E_T data (Fig. 4c,f) and a nonhorizontal one obtained with the low E_T data (Fig. 4a,d). However, when S_T is high, the scatter plot from the low E_T also becomes horizontal (Fig. 4c), and thus the synergistic effect of using combined data decreases (Fig. 5a,b right). Taken together, the tQ model can accurately estimate both parameters from the combination of low E_T and high E_T data when S_T is not much larger than K_M. Note that such low S_T is preferred for *in vitro* experiments[24,39–41] and is the case for most physiological conditions[24].

Next, we consider the case when progress curves are measured with a fixed E_T and a varied S_T. Specifically, the combination of two progress curves from low and high S_T is used to infer parameters for a fixed E_T at different levels (Fig. S1 left and right). When E_T is low, and thus the sQ and the tQ models behave similarly (Eq. 5), posterior samples obtained with both models accurately capture the true values of k_{cat} and K_M (Figs 6a left and S6). Again, the narrow scatter plot (Fig. 6b left) is obtained as the intersection of a nonhorizontal scatter plot of low S_T (Fig. 4a) and a horizontal scatter plot of high S_T (Fig. 4g). However, as E_T increases, and thus the sQ model becomes less accurate, those obtained with the sQ model are biased, as expected (Figs 6a right and S6). Whereas such biases are not observed in those obtained with the tQ model, the precision of K_M estimates decreases as E_T increases, as in the single-parameter estimation (Fig. 2 and Eq. 6).

Optimal design of experiments for accurate and efficient estimation with the tQ model. When a progress curve obtained from a single experiment is used, the posterior scatter plots of the tQ model can be categorized as a correlated type (Fig. 4a,b,d,e) and a horizontal type (Fig. 4c,f,g–i). The intersections of these two different types of scatter plots tend to be narrowly distributed near the true value (Figs 5b and 6b). Thus, combining two such data sets allows accurate estimation of both k_{cat} and K_M (Figs 5a and 6a). Specifically, a progress curve measured under $E_T \ll K_M$ and $S_T \ll K_M$ (Fig. 4a,b,d,e) and one measured under $E_T \gg K_M$ or $S_T \gg K_M$ (Fig. 4c,f,g–i) provide different types of information for parameter estimation; so using both data sets leads to

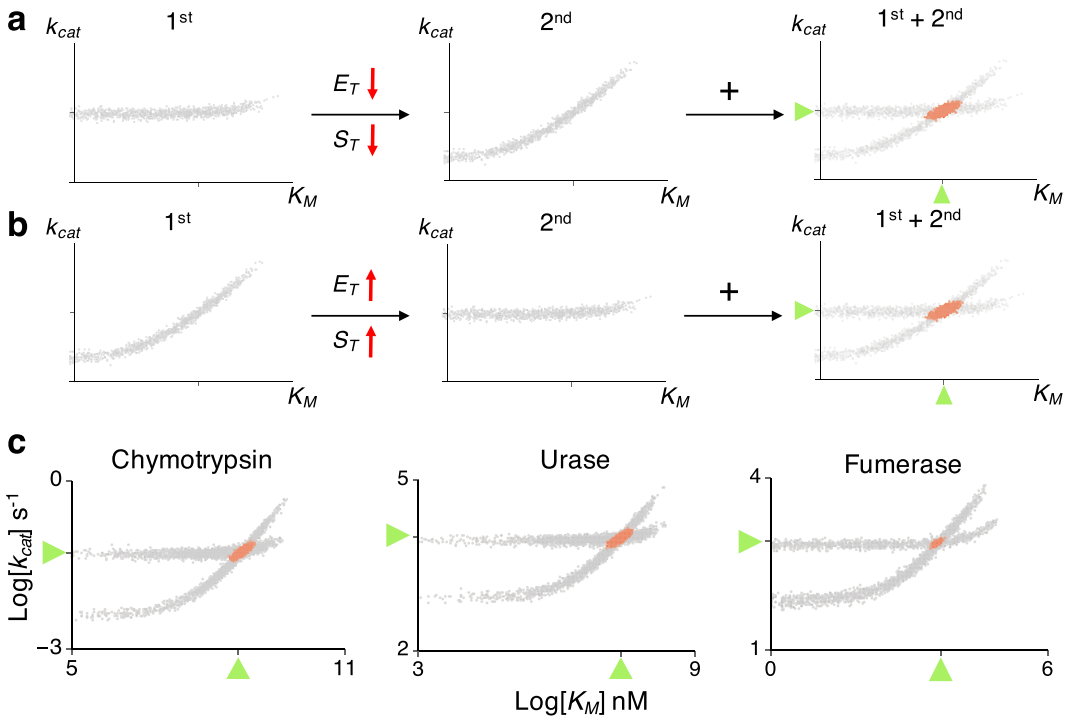

Figure 7. The optimal experimental design for accurate and precise estimation with the tQ model. (**a**) When the scatter plot of posterior samples from the first experiment is horizontal, E_T and S_T need to be decreased to obtain the nonhorizontal scatter plot in the next experiment. Then, using the combination of the two experiments leads to accurate and precise estimation (red scatter plots). (**b**) When the scatter plot from the first experiment is nonhorizontal, E_T or S_T need to be increased in the next experiment to obtain a horizontal scatter plot. (**c**) Inference with a single progress curve from the low E_T (0.1 K_M) and the high E_T (10 K_M) leads to nonhorizontal and horizontal scatter plots, respectively, for chymotrypsin, urease, and fumarase (gray scatter plots). When both data sets were used together, accurate estimates were obtained for all enzymes (red scatter plots). Here, low S_T (0.1 K_M) is used. Here, green triangles represent the true values of the parameters.

successful estimation. However, it is hard to compare the values of S_T, E_T, and K_M in practice, because the value of K_M is usually unknown a priori. This problem can be easily resolved by using the scatter plot. That is, if the posterior scatter plot obtained from the first experiment is horizontal, then both E_T and S_T should be decreased for the next experiment, so that the nonhorizontal scatter plot can be obtained (Fig. 7a). On the other hand, if the scatter plot from the first experiment shows a strong correlation between K_M and k_{cat}, then either S_T or E_T should be increased in the next experiment (Fig. 7b). Basically, without any prior information of the value of K_M and k_{cat}, the shape of the scatter plots of the current estimates determines the next optimal experimental design, which ensures accurate and precise estimation. However, this approach cannot be used with the sQ model, because estimation with the sQ model can be biased, depending on the relationship between E_T or S_T and K_M, which is unknown a priori. That is, unlike the tQ model, precise estimation does not always guarantee accurate estimation with the sQ model, as seen above (e.g. Fig. 5a right).

We test whether the proposed approach with the tQ model can accurately estimate k_{cat} and K_M for catalysis of the N-acetylglycine ethyl ester, fumarate, and urea by the enzymes the chymotrypsin, urease, and fumarase, respectively (Fig. 7c). These three enzymes were chosen because they have disparate catalytic efficiencies (k_{cat}/K_M)[1]: 0.12, $4 \cdot 10^5$, and $1.6 \cdot 10^8\, s^{-1} M^{-1}$, respectively. For each enzyme, 10^2 noisy timecourse data sets were generated using stochastic simulations based on known enzyme kinetic parameters[1]. When progress curves obtained with low E_T and low S_T are used, as expected, nonhorizontal scatter plots of posterior samples were obtained for all three enzymes (Fig. 7c). This indicates that either E_T or S_T should be increased in the next experiment to obtain a horizontal scatter plot. Indeed, when a progress curve with a 100-fold increase of E_T was used, horizontal scatter plots were obtained for all enzymes (Fig. 7c). Therefore, when these two progress curves are used together, both k_{cat} and K_M can be accurately estimated (Fig. 7c red dots). These results support that such two-step optimized experimental design (Fig. 7a,b) to get two different types of scatter plots allows accurate and efficient estimation of enzyme kinetics with the tQ model. The computational package performing such estimation is provided (see Method for the details).

Discussion

The standard approach for estimating enzyme kinetic parameters even today continues to be based on the 100-year old MM equation (Eq. 1)[5,6]. However, when enzyme concentration is high, this approach can lead to biased estimation (Fig. 2). Even when enzyme concentration is relatively low, it may not be possible to identify kinetic parameters (Figs 3 and 4). To overcome the limitations of the canonical approach, we proposed an

estimation method based on an alternative to the MM equation: the tQ model (Eq. 2), which is derived with the total QSSA[26–29]. Because the estimation procedure with the tQ model is not biased regardless of enzyme or substrate concentrations (Fig. 2), more accurate and precise estimations can be made when pooled data from different experimental conditions are used, unlike the canonical approach (Figs 5 and 6). It appears thus that the tQ model is especially appropriate for creating a consistent Bayesian inferential framework, which becomes more accurate as more data is used.

The canonical enzyme kinetic assay based on the MM equation generally requires a large excess of substrate over enzyme[42]. However, such conditions impose experimental limitations and cannot be always guaranteed and verified[15]. For instance, it is hard to generate a high concentration of barely soluble substrate[24], and a low concentration of substrate is required for sensitive kinetic analysis, e.g., in the case of QD-FRET-based probes[39–41]. Importantly, to analyze *in vivo* enzyme kinetics, where enzyme concentration is often high[16–18], our approach, but not the canonical approach, can be used. For example, one needs to estimate the kinetic parameters underlying drug metabolism by CYP enzymes in the liver in order to predict the effects of drugs, as is essential for drug development[43]. Because of dosing requirements for potent drugs, the amount of CYP enzyme can greatly exceed the drug amount in the liver[44,45]. Another large area where our estimation method can be applied is in the development of nanobiosensors, which measure *in vivo* activity of a specific enzyme for precise diagnostics, because such enzymes are often in large excess over biosensors[46,47].

KinTek Explorer has been widely used to estimate enzyme kinetic parameters from the progress curves[48–50]. This software provides the confidence contours, which reveal the relationships between the estimated parameters. This approach recommends using multiple data sets to narrow down the confidence contours and thus improve precision of estimates and resolve the unidentifiability issue. Our finding (Fig. 7a,b) can provide the specific type of data sets required for the identifiability of k_{cat} and K_M, so that the KinTek Explore could perform parameter estimation more efficiently for the Michales-Menten type of enzyme reactions.

Since the initial velocity estimation with the MM equation is not accurate when enzyme concentration is high (Fig. 1), the standard initial velocity based on the MM equation would also be inaccurate[15]. On the other hand, the tQ model accurately captures the initial velocity for all conditions, and thus the modified initial velocity assay based on the tQ model is likely to be accurate over a wider range of conditions. To simplify such estimation procedures, an interesting future study could derive an analogous Lineweaver-Burk plot or the Hanes-Woolf plot[8,9,15,42] for the tQ model.

Even with relatively large noise in the data (Fig. S1), our proposed method leads to accurate estimation (Figs 5, 6 and 7), indicating its robustness against experimental noise and some minor inaccuracy of the tQ model in certain ranges of parameter observed in[14,30]. Furthermore, if there are departures from simple non-inhibitory enzyme kinetics (e.g. inhibition of enzyme by product)[51,52], our method can be easily adjusted by modifying the tQ model (see[53,54] for the tQ model for other enzyme kinetics). Our work can also be used to improve the estimation of the kinetics underlying diverse biological functions, such as gene regulation[55,56], cellular rhythms[57–59], quorum sensing[60,61], signal cascade[62,63] and membrane transport[64,65], where the MM equation has been widely used.

Methods

Simulated Data. To obtain timecourse data (Fig. S1) for Bayesian inference, stochastic simulations of the original full model (Table S1) were performed with the Gillespie algorithm[66]. $E(0) = E_T$, $S(0) = S_T$, $C(0) = 0$, and $P(0) = 0$ are used as initial conditions following the typical *in vitro* enzyme kinetics protocol.

Description of the Bayesian inference approach. The Bayesian inference approach is used to estimate the catalytic constant k_{cat} and the Michaelis-Menten constant K_M, based on the hazard function, with respective rates described in Eq. 1 for the sQ model and in Eq. 2 for the tQ model (Fig. S2). The likelihood functions are constructed based on an approximation to the underlying Markov model[66] as follows.

$$L(k_{cat}, K_M | P_0, P_1, P_2, \ldots, P_m, t_1, t_2, \ldots, t_m) \propto \prod_{i=1}^{m} \lambda_i^{n_i} [t_i - t_{i-1}]^{n_i - 1} \exp\{-\lambda_i [t_i - t_{i-1}]\},$$

where λ_i is given by

$$\lambda_i = k_{cat} \frac{E_T(S_T - P_{i-1})}{K_M + S_T - P_{i-1}}, \text{ for the sQ model}$$

or

$$\lambda_i = k_{cat} \frac{E_T + K_M + S_T - P_{i-1} - \sqrt{(E_T + K_M + S_T - P_{i-1})^2 - 4E_T(S_T - P_{i-1})}}{2}, \text{ for the tQ model,}$$

where P_i is the scaled number of product molecules observed at time point t_i over $[0, T] = [t_0, t_m]$ and $n_i = P_i - P_{i-1}$ is an observed increment of P_i. With these likelihood functions, the usual independent gamma priors[67] are assigned to k_{cat} and K_M in order to get their posterior distributions with the help of the Markov Chain Monte Carlo (MCMC) method. Weakly informative gamma priors are used for both k_{cat} and K_M: their prior means are the same as their true values, and their prior variance is 10 times larger than the prior mean, which covers orders of magnitude (e.g. Fig. 2). The estimation of a single parameter, i.e., either k_{cat} or K_M is done conditionally on the other parameter. For estimating the two parameters simultaneously, the Gibbs sampler method is used. In order to draw the sample for K_M, we also use the Metropolis-Hastings algorithm within the Gibbs sampler step. See Supplementary material for further details.

Computational code. The R package that performs the Bayesian inference based on the tQ model is available on the CRAN repository (https://cran.r-project.org/web/packages/EKMCMC).

References

1. Chang, R. *Physical chemistry for the chemical and biological sciences* (University Science Books, 2000).
2. Cooper, G. *The Cell: A Molecular approach* (USA: Sinauer Associates, 2000).
3. Griffiths, A. J. *Modern genetic analysis: integrating genes and genomes* (Macmillan, 2002).
4. Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351 (2002).
5. Michaelis, L. & Menten, M. L. Die kinetik der invertinwirkung. *Biochem. z* **49**, 352 (1913).
6. Briggs, G. E. & Haldane, J. B. S. A note on the kinetics of enzyme action. *Biochem. J.* **19**, 338 (1925).
7. Segel, L. A. & Slemrod, M. The quasi-steady-state assumption - a case-study in perturbation. *SIAM Rev.* **31**, 446–477 (1989).
8. Tummler, K., Lubitz, T., Schelker, M. & Klipp, E. New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. *FEBS J.* **281**, 549–571 (2014).
9. Johnson, K. A. A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett* **587**, 2753–2766 (2013).
10. Duggleby, R. G. Analysis of enzyme progress curves by nonlinear regression. *Methods Enzymol* **249**, 61–90 (1995).
11. Duggleby, R. G. & Wood, C. Analysis of progress curves for enzyme-catalysed reactions. automatic construction of computer programs for fitting integrated rate equations. *Biochem. J.* **258**, 397–402 (1989).
12. Chen, W. W., Niepel, M. & Sorger, P. K. Classic and contemporary approaches to modeling biochemical reactions. *Genes & Dev* **24**, 1861–1875 (2010).
13. Varón, R. *et al.* An alternative analysis of enzyme systems based on the whole reaction time: evaluation of the kinetic parameters and initial enzyme concentration. *J. Math. Chem.* **42**, 789–813 (2007).
14. Schnell, S. & Maini, P. K. A century of enzyme kinetics. should we believe in the km and vmax estimates? *Comments Theor. Biol.* **8**, 169–187 (2003).
15. Pinto, M. F. *et al.* Enzyme kinetics: the whole picture reveals hidden meanings. *The FEBS J.* **282**, 2309–2316 (2015).
16. Albe, K. R., Butler, M. H. & Wright, B. E. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.* **143**, 163–195 (1990).
17. Srere, P. A. Enzyme concentrations in tissues. *Science.* **158**, 936–937 (1967).
18. Eunen, K. V. & Bakker, B. M. The importance and challenges of *in vivo*-like enzyme kinetics. *Perspectives Sci.* **1**, 126–130 (2014).
19. Nikolova, N., Tenekedjiev, K. & Kolev, K. Uses and misuses of progress curve analysis in enzyme kinetics. *Cent. Eur. J. Biol* **3**, 345–350 (2008).
20. Cornish-Bowden, A. One hundred years of michaelis–menten kinetics. *Perspectives Sci.* **4**, 3–9 (2015).
21. Yang, X., Long, G., Jiang, H., Liao, P. & Liao, F. Integration of kinetic analysis of reaction curve with a proper classical approach for enzymatic analysis. *The Sci. World J.* **2012** (2012).
22. Duggleby, R. G. Experimental designs for estimating the kinetic parameters for enzyme-catalysed reactions. *J. Theor. Biol.* **81**, 671–684 (1979).
23. Stroberg, W. & Schnell, S. On the estimation errors of k m and v from time-course experiments using the michaelis–menten equation. *Biophys. Chem.* **219**, 17–27 (2016).
24. Bisswanger, H. Enzyme assays. *Perspectives Sci.* **1**, 41–55 (2014).
25. Duggleby, R. G. & Clarke, R. B. Experimental designs for estimating the parameters of the michaelis-menten equation from progress curves of enzyme-catalyzed reactions. *Biochimica et Biophys. Acta (BBA)-Protein Struct. Mol. Enzymol.* **1080**, 231–236 (1991).
26. Cha, S. Kinetic behavior at high enzyme concentrations magnitude of errors of michaelis-menten and other approximations. *J. Biol. Chem.* **245**, 4814–4818 (1970).
27. Tzafriri, A. R. Michaelis-menten kinetics at high enzyme concentrations. *Bull. Math. Biol.* **65**, 1111–1129 (2003).
28. Bersani, A. M., Bersani, E., DellAcqua, G. & Pedersen, M. G. New trends and perspectives in nonlinear intracellular dynamics: one century from michaelis–menten paper. *Continuum Mech. Thermodyn.* **27**, 659–684 (2015).
29. Borghans, J. M., De Boer, R. J. & Segel, L. A. Extending the quasi-steady state approximation by changing variables. *Bull. Math. Biol.* **58**, 43–63 (1996).
30. Schnell, S. & Maini, P. Enzyme kinetics far from the standard quasi-steady-state and equilibrium approximations. *Math. Comput. Model.* **35**, 137–144 (2002).
31. Rao, C. V. & Arkin, A. P. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.* **118**, 4999–5010 (2003).
32. Barik, D., Paul, M. R., Baumann, W. T., Cao, Y. & Tyson, J. J. Stochastic simulation of enzyme-catalyzed reactions with disparate timescales. *Biophys. J.* **95**, 3563–3574 (2008).
33. Thomas, P., Straube, A. V. & Grima, R. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Syst. Biol.* **6** (2012).
34. Kim, J. K., Josić, K. & Bennett, M. R. The validity of quasi-steady-state approximations in discrete stochastic simulations. *Biophys. J.* **107**, 783–793 (2014).
35. Kim, J. K. & Sontag, E. D. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. *PLoS Comput. Biol.* **13**, e1005571 (2017).
36. Kim, J. K., Rempala, G. A. & Kang, H.-W. Reduction for stochastic biochemical reaction networks with multiscale conservations. *Multiscale Modeling & Simulation,* **15**(4), 1376–1403 (2017).
37. Sanft, K. R., Gillespie, D. T. & Petzold, L. R. Legitimacy of the stochastic michaelis-menten approximation. *IET Syst. Biol.* **5**, 58–69 (2011).
38. Kim, J. K., Josić, K. & Bennett, M. R. The relationship between stochastic and deterministic quasi-steady state approximations. *BMC Syst. Biol.* **9**, 87 (2015).
39. Sapsford, K. E. *et al.* Monitoring of enzymatic proteolysis on a electroluminescent-ccd microchip platform using quantum dot-peptide substrates. *Sensors Actuators B: Chem.* **139**, 13–21 (2009).
40. Algar, W. R. *et al.* Proteolytic activity at quantum dot-conjugates: Kinetic analysis reveals enhanced enzyme activity and localized interfacial hopping. *Nano Lett.* **12**, 3793–3802 (2012).
41. Algar, W. R. *et al.* Multiplexed tracking of protease activity using a single color of quantum dot vector and a time-gated forster resonance energy transfer relay. *Anal. Chem.* **84**, 10136–10146 (2012).
42. Singh, N. *et al.* A safe lithium mimetic for bipolar disorder. *Nat. Commun.* **4**, 1332 (2013).
43. Obach, R. S. & Reed-Hagen, A. E. Measurement of michaelis constants for cytochrome p450-mediated biotransformation reactions using a substrate depletion approach. *Drug Metab. Dispos.* **30**, 831–837 (2002).
44. Wienkers, L. C. & Heath, T. G. Predicting *in vivo* drug interactions from *in vitro* drug discovery data. *Nat. Rev. Drug Discov.* **4**, 825–833 (2005).
45. Houston, J. B. & Kenworthy, K. E. *In vitro-in vivo* scaling of cyp kinetic data not consistent with the classical michaelis-menten model. *Drug Metab. Dispos.* **28**, 246–254 (2000).
46. Turk, B. E., Huang, L. L., Piro, E. T. & Cantley, L. C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol.* **19**, 661–667 (2001).
47. Kwong, G. A. *et al.* Mathematical framework for activity-based cancer biomarkers. *Proc. Natl. Acad. Sci.* **112**, 12627–12632 (2015).

48. Johnson, K. A. Fitting enzyme kinetic data with kintek global kinetic explorer. *Methods Enzymol.* **467**, 601–626 (2009).
49. Johnson, K. A., Simpson, Z. B. & Blom, T. Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387**, 20–29 (2009).
50. Johnson, K. A., Simpson, Z. B. & Blom, T. Fitspace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem.* **387**, 30–41 (2009).
51. Cao, W. & Enrique, M. Quantitative full time course analysis of nonlinear enzyme cycling kinetics. *Sci. Reports* **3**, 2658 (2013).
52. Duggleby, R. G. Quantitative analysis of the time courses of enzyme-catalyzed reactions. *Methods* **24**, 168–174 (2001).
53. Pedersen, M. G., Bersani, A. M., Bersani, E. & Cortese, G. The total quasi-steady-state approximation for complex enzyme reactions. *Math. Comput. Simul.* **79**, 1010–1019 (2008).
54. Ciliberto, A., Capuani, F. & Tyson, J. J. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Comput. Biol* **3**, e45 (2007).
55. Ronen, M., Rosenberg, R., Shraiman, B. I. & Alon, U. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci.* **99**, 10555–10560 (2002).
56. Del Vecchio, D., Abdallah, H., Qian, Y. & Collins, J. J. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Syst.* (2017).
57. Pigolotti, S., Krishna, S. & Jensen, M. H. Oscillation patterns in negative feedback loops. *Proc. Natl. Acad. Sci.* **104**, 6533–6537 (2007).
58. Kim, J. K. Protein sequestration versus hill-type repression in circadian clock models. *IET Syst. Biol.* **10**, 125–135(10) (2016).
59. Gotoh, T. *et al*. Model-driven experimental approach reveals the complex regulatory distribution of p53 by the circadian factor period 2. *Proc. Natl. Acad. Sci.* **113**, 13516–13521 (2016).
60. Dockery, J. D. & Keener, J. P. A mathematical model for quorum sensing in pseudomonas aeruginosa. *Bull. Math. Biol.* **63**, 95–116 (2001).
61. Chen, Y., Kim, J. K., Hirning, A. J., Josić, K. & Bennett, M. R. Emergent genetic oscillations in a synthetic microbial consortium. *Sci.* **349**, 986–989 (2015).
62. Ossareh, H. R., Ventura, A. C., Merajver, S. D. & Del Vecchio, D. Long signaling cascades tend to attenuate retroactivity. *Biophys. J.* **100**, 1617–1626 (2011).
63. Chen, M., Wang, L., Liu, C. C. & Nie, Q. Noise attenuation in the on and off states of biological switches. *ACS Synth. Biol.* **2**, 587–593 (2013).
64. Knight, M. J., Senior, L., Nancolas, B., Ratcliffe, S. & Curnow, P. Direct evidence of the molecular basis for biological silicon transport. *Nat. Commun.* **7** (2016).
65. Lawson, M. J., Drawert, B., Khammash, M., Petzold, L. & Yi, T.-M. Spatial stochastic dynamics enable robust cell polarization. *PLoS Comput. Biol* **9**, e1003139 (2013).
66. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The J. Phys. Chem.* 2340–2361 (1977).
67. Choi, B. & Rempala, G. A. Inference for discretely observed stochastic kinetic networks with applications to epidemic modeling. *Biostat.* **13**, 153–165 (2012).

Acknowledgements

The research was initiated when all authors were visiting The Mathematical Biosciences Institute at the Ohio State University. This work was supported by the National Research Foundation of Korea grant NRF-2017R1D1A3B03031008 (B.C.), US National Science Foundation (NSF) Grant DMS-1440386 (G.A.R.) and DMS-1318886 (G.A.R.), the National Research Foundation of Korea grant N01160447 (J.K.K.), and the TJ Park Science Fellowship of POSCO TJ Park Foundation (J.K.K.).

Author Contributions

J.K.K. designed the research. B.C. and J.K.K. performed simulations and analysis. All authors discussed the results and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-17072-z.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017