JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 15, Number 7, 2008

Pp. 779–792
DOI: 10.1089/cmb.2007.0135

Production-Passage-Time Approximation: A New Approximation Method to Accelerate the Simulation Process of Enzymatic Reactions

HIROYUKI KUWAHARA[1,3] and CHRIS J. MYERS[2]

ABSTRACT

Given the substantial computational requirements of stochastic simulation, approximation is essential for efficient analysis of any realistic biochemical system. This paper introduces a new approximation method to reduce the computational cost of stochastic simulations of an enzymatic reaction scheme which in biochemical systems often includes rapidly changing fast reactions with enzyme and enzyme-substrate complex molecules present in very small counts. Our new method removes the substrate dissociation reaction by approximating the passage time of the formation of each enzyme-substrate complex molecule which is destined to a production reaction. This approach skips the firings of unimportant yet expensive reaction events, resulting in a substantial acceleration in the stochastic simulations of enzymatic reactions. Additionally, since all the parameters used in our new approach can be derived by the Michaelis-Menten parameters which can actually be measured from experimental data, applications of this approximation can be practical even without having full knowledge of the underlying enzymatic reaction. Here, we apply this new method to various enzymatic reaction systems, resulting in a speedup of orders of magnitude in temporal behavior analysis without any significant loss in accuracy. Furthermore, we show that our new method can perform better than some of the best existing approximation methods for enzymatic reactions in terms of accuracy and efficiency.

Key words: biochemical networks, stochastic processes.

1. INTRODUCTION

THIS PAPER CONSIDERS a well-stirred chemically reacting system with the following enzymatic reaction scheme:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{1}$$

[1]School of Computing and [2]Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah.
[3]The Microsoft Research–University of Trento Centre for Computational and Systems Biology, Trento, Italy.

where E, S, C, and P represent an enzyme, a substrate, an enzyme-substrate complex, and a product, respectively, and k_1, k_{-1}, and k_2 represent non-zero rate constants for the reaction channels, R_1, R_{-1}, and R_2, respectively. This enzymatic reaction scheme specifies the transformation of S into P catalyzed by E, where E has one active site to which S can bind to form C. This type of enzymatic reaction scheme can be found in many biochemical pathways such as metabolic pathways, and therefore, abstracting away low-level details found in enzymatic reaction schemes (1) may have a significant computational benefit in analyzing the overall system behavior.

Traditionally, biochemical systems—including the enzymatic reaction system that is considered in this paper—are modeled and analyzed typically within the continuous-deterministic, *classical chemical kinetics* (CCK) framework based on the *law of mass action* where the dynamics of a well-stirred system is described by a set of ordinary differential equations (ODEs). However, the limitations of the CCK analysis have been broadly accepted (Arkin et al., 1998; Elowitz et al., 2002; Rao et al., 2002; Samoilov et al., 2005). In particular, given the same initial condition, the CCK analysis of biochemical systems always produces the same result as it neglects molecular fluctuations. Such treatment, nevertheless, can be justified when the molecular populations are very large, and hence a CCK analysis may provide the most efficient approach to determine the time evolution of a system in such cases. However, many regulatory components in biological systems are often present in amounts too small to simply neglect the effects of the inherent fluctuations (Cai et al., 2006; Golding et al., 2005; McAdams and Arkin, 1997; Newman et al., 2006; Pedraza and van Oudenaarden, 2005). Moreover, if a system being analyzed has multiple steady states, the traditional ODE approach may not be able to provide the accurate time evolution of a system since it cannot capture spontaneous transitions between steady states (Gillespie, 1992a, 2000).

In order to more accurately predict the temporal behavior of biochemical systems without acquiring more information on a biological system such as the positions and the velocities of every molecule, the discrete-stochastic, *stochastic chemical kinetics* (SCK) framework can be used (Gillespie, 2005). SCK describes the time evolution of a well-stirred biochemical system as a discrete-state jump Markov process that is analytically governed by the *chemical master equation* (CME) (Gillespie, 1992b). The CME is derived from the *state-change vector*, specifying the change in each molecular species population for each reaction, and a *propensity function* for each reaction. For example, the enzymatic reaction scheme (1) contains the following propensity functions for each reaction R_i:

$$R_1 : a_1(\mathbf{x}) = k_1 x_E x_S, \quad R_{-1} : a_{-1}(\mathbf{x}) = k_{-1} x_C, \quad R_2 : a_2(\mathbf{x}) = k_2 x_C$$

where $\mathbf{x} = (x_E, x_S, x_C, x_P)$, and each x_* is the value of random variable $X_*(t)$ representing the molecular population of the species subscripted. Thus, the vector of these random variables: $\mathbf{X}(t) = (X_E(t), X_S(t), X_C(t), X_P(t))$ represents the system state at time t. This SCK approach describes, with the spatially homogeneous assumption, the time evolution of a biochemical system at the individual reaction level by exactly tracking the quantities of each molecular species and by treating each reaction as a separate random event. However, directly obtaining the solution of the CME of any realistic system, either analytically or numerically, is not feasible due to its intrinsic complexity. Note that, though it is possible to numerically solve the CME of the enzymatic reaction scheme (1) as the system state is bounded albeit with potentially substantial computational demands, if systems also contain other reactions and species as is the case for many realistic biological systems, then the space complexity of CMEs of such systems often inevitably becomes too large to be tractable, making the numerical solutions of such CMEs infeasible.

Instead of attempting to solve the CME, exact discrete-stochastic numerical realizations of a system via Gillespie's *stochastic simulation algorithm* (SSA) (Gillespie, 1976), which is derived from the same premise as the CME, are often used to infer the temporal system behavior with a much smaller memory footprint. This Monte Carlo simulation approach is useful to intuitively observe the trend of system dynamics, which may be possible with just as few as tens of numerical realizations. Furthermore, *in silico* experiments via Monte Carlo simulation come with potentially unlimited controlling capabilities and abilities to capture virtually any dynamical properties of the system, making a number of qualitative and quantitative analyses which cannot be done in *wet-lab* experiments possible. Unfortunately, the computational requirements of the SSA—even with the Gibson and Bruck (2000) optimization, which, among other things, reduces the generations of the random numbers by reusing them—can be substantial. This is due largely to the fact that it not only requires a potentially large number of simulation runs in order to estimate the system behavior at a reasonable degree of statistical confidence, but it also requires every single reaction event

to be simulated one at a time. For example, if $k_2 \ll k_{-1}$ in the enzymatic reaction scheme (1), then C dissociates into S much more often than into P, and thus, much of the computation is allocated for this *substrate-complex loop*.

Several approximation methods have been proposed to accelerate the simulation process of the SSA by sacrificing the exactness. For example, the explicit τ-leaping method approximates the number of firings of each reaction in a pre-defined interval rather than executing each reaction individually (Gillespie, 2001). While this and similar methods (Cao et al., 2005; Gillespie and Petzold, 2003; Rathinam et al., 2003) are very promising, they may not perform well for an enzymatic reaction which includes rapidly changing fast reactions driven by the enzyme and enzyme-substrate complex molecules present in very small counts.

Some acceleration methods for the stochastic simulations of enzymatic reactions have been proposed that perform well even when the enzyme is present in a very small count by eliminating the undesirable substrate-complex loop in the enzymatic reaction scheme. For example, Rao and Arkin (2003) have performed model abstraction by using biochemical insight in combination with the *quasi-steady-state approximation* (QSSA) to remove the expensive substrate-complex loop. While this method provides a framework to reduce the complexity of a biochemical model and the runtime of the simulation by removing fast reactions, making the quasi-steady-state assumption may not be valid for some enzymatic reactions. In such situations, this method may not give a good approximation. Cao et al. (2005a) have demonstrated how the substrate-complex loop can be removed by applying the enzyme substrate reaction system to their slow-scale SSA approach, which explicitly simulates the firings of only the slow reaction events (Cao et al., 2005b). Since this is fundamentally the QSSA-based approximation, it may not perform well when reactions are far from equilibrium.

In general, both approximation methods require the use of special simulation procedures which may require in-depth analytical analysis of the biochemical network prior to its simulation in order to obtain the necessary probability distribution function. Thus, there might be cases where one finds the use of these approximation methods inconvenient when it comes to the analysis of a system containing enzymatic reactions along with other types of reactions. Such cases occur, for example, when a biochemical system is represented in the *Systems Biology Markup Language* (SBML), the emerging standard format to represent models of biochemical reaction networks (Finney and Hucka, 2003). SBML level 2 version 1 contains reactions only in the generic type, and it cannot specify any specific reaction types without a use of a proprietary annotation. Thus, in order for SBML compliant SSA tools to know when to use specially tailored Monte Carlo simulation procedures, the tools must either understand the semantics of proprietary fields that specify reaction types or perform structural analysis to find reaction types.

This paper introduces a new approximation approach to accelerate the process of the stochastic simulations of enzymatic reactions. Our new approach, which we call *production-passage-time approximation* (PPTA), approximates the passage time of C which is destined to turn into P, and only keeps track of such instances of the formations of C. Thus, this approach eliminates the substrate-complex loop by removing R_{-1}, allowing a substantial acceleration in the stochastic simulations of enzymatic reactions. Furthermore, since our approach does not require a customized simulation procedure, it allows a biochemical system comprising the PPTA reactions along with other types of reactions to still be modeled using a SBML modeling tool such as PathwayBuilder from BioSPICE (2008), and analyzed by using any SBML compliant SSA tools.

This paper first describes the PPTA method in Section 2. Section 3 demonstrates how our approach can help analyze the temporal behaviors of enzymatic one-substrate reaction models efficiently while keeping reasonable accuracy. This is shown by applying our new approximation method to various systems and comparing the full models with the corresponding PPTA models in terms of their accuracy—by calculating means and standard deviations—as well as runtime. Finally, this paper concludes in Section 4 by discussing the benefits gained by the PPTA.

2. PRODUCTION-PASSAGE-TIME APPROXIMATION

To describe the PPTA method, the enzymatic reaction scheme (1) is first considered to hold in the initial condition: $\mathbf{X}(t_0) = \mathbf{x}_{\mathbf{t_0}}$, where $\mathbf{x}_{\mathbf{t_0}} = (e_{tot}, s_{tot}, 0, 0)$, $e_{tot} \geq 1$, and $s_{tot} \geq 1$. Let $\mathbf{x}_\infty = (e_{tot}, 0, 0, s_{tot})$, then the probability that $\mathbf{X}(t) = \mathbf{x}_\infty$ given $\mathbf{X}(t_0) = \mathbf{x}_{\mathbf{t_0}}$ approaches 1, as $t \to \infty$. In other words, in any

simulation run, the enzymatic reaction process always reaches $\mathbf{x}_\infty$, eventually. In order for each numerical realization of $\mathbf{X}(t)$ to transition from $\mathbf{x}_{\mathbf{t_0}}$ to $\mathbf{x}_\infty$, S must be transformed into C at least s_{tot} times and C must be converted into P exactly s_{tot} times. Thus, let $\mathbf{x}^{(i)}(t)$ be the ith sample trajectory of $\mathbf{X}(t)$ given that $\mathbf{X}(t_0) = \mathbf{x}_{\mathbf{t_0}}$ and $\mathbf{T_i}$ be a set of time instances such that each time instance t_j^i represents the time point where the jth reaction event occurs in $\mathbf{x}^{(i)}(t)$. Then, the statement $\forall i.\ |\mathbf{T_i}| \in [2s_{tot}, \infty)$ must be true. Intuitively, if $k_{-1} \ll k_2$, then C tends to be consumed by R_2 rather than R_{-1}, making the size of each $\mathbf{T_i}$ close to the lower bound $2s_{tot}$. On the other hand, if $k_{-1} \gg k_2$, then C is more likely to be consumed by R_{-1}, and in consequence $|\mathbf{T_i}|$ can be much greater than $2s_{tot}$ with a very high likelihood, making the computational cost of simulations significantly higher.

Our new PPTA approach minimizes the number of reaction events that fire through the passage of each $\mathbf{x}^{(i)}(t)$ to $\mathbf{x}_\infty$ by preventing each $\mathbf{x}^{(i)}(t)$ from revisiting the same state. Thus, it guarantees that $\forall i.\ |\mathbf{T_i}| = 2s_{tot}$. This is achieved by eliminating R_{-1} and approximating transitions of each $\mathbf{x}^{(i)}(t)$ using only complex-formation and production reactions. In other words, the PPTA approximates the passage time of the formation of each C molecule which leads to a production of P, and only keeps track of such instances of the formations of C, rather than explicitly also simulating the formation of C molecules that are destined to dissociate into E and S molecules. Therefore, the PPTA can accelerate the stochastic simulations of the enzymatic reaction scheme (1), especially when $k_{-1} \gg k_2$ where the reduction in $|\mathbf{T_i}|$ by this new approach is substantial.

Let us first consider the special case where the total molecular count of the enzyme is 1 (i.e., $e_{tot} = 1$), and describe the derivation of the PPTA model. This section then extends this special case to more general cases where the total molecular count of the enzyme is greater than 1 (i.e., $e_{tot} > 1$).

When e_{tot} is 1, the enzyme state for all $t \geq 0$ is defined by $X_E(t) = 1 - X_C(t)$. Also, R_1 is only enabled when E is active (i.e., $X_E(t) = 1$), and R_{-1} and R_2 are only enabled when C is active (i.e., $X_C(t) = 1$). In this case, $\mathbf{X}(t)$ can be seen as a temporal-homogeneous birth-death Markov process $\mathbf{Y}(t)$ with $2s_{tot} + 1$ states as shown in Figure 1. Each state $s \in [0, 2s_{tot}]$ of $\mathbf{Y}(t)$ can then be mapped onto a system state $\mathbf{x_s}$ of $\mathbf{X}(t)$ by the relationship: $\mathbf{x_s} \equiv ((s+1) \bmod 2, s_{tot} - \lceil s/2 \rceil, s \bmod 2, \lfloor s/2 \rfloor)$. Thus, for all $t > t_0$, the probability that $\mathbf{Y}(t) = s$ given that $\mathbf{Y}(t_0) = 0$ is the same as the probability that $\mathbf{X}(t) = \mathbf{x_s}$ given that $\mathbf{X}(t) = \mathbf{x}_{\mathbf{t_0}}$, and with the initial condition $\mathbf{X}(t_0) = \mathbf{x}_{\mathbf{t_0}}$, each simulation run of $\mathbf{Y}(t)$ starts in state 0, and eventually ends up in state $2s_{tot}$. Since E is active only in even number states in this process, R_1 can fire only in these states except in state $2s_{tot}$. Similarly, C is active only in odd number states, so R_{-1} and R_2 can fire in these states. Thus, let $\mathbf{S_e}$ be a set of even number states $\{2m \mid 0 \leq m \leq s_{tot}\}$, $\mathbf{S_{e'}}$ be a set of states $\mathbf{S_e} \setminus \{2s_{tot}\}$, and $\mathbf{S_o}$ be a set of odd number states $\{2m+1 \mid 0 \leq m < s_{tot}\}$. Then, the $s \to s+1$ transition rate λ_s is $a_1(\mathbf{x_s})$ if $s \in \mathbf{S_{e'}}$, and $a_2(\mathbf{x_s})$ if $s \in \mathbf{S_o}$, whereas the $s \to s-1$ transition rate μ_s is $a_{-1}(\mathbf{x_s})$ if $s \in \mathbf{S_o}$ and 0 if $s \in \mathbf{S_e}$.

Suppose $\mathbf{Y}(t)$ starts in state s_0 where $s_0 \in \mathbf{S_{e'}}$. Then, the average waiting times that $\mathbf{Y}(t)$ spends in states s_0 and $s_0 + 1$ for each simulation run are equivalent to $t(s_0; s_0 \to s_0 + 2)$ and $t(s_0 + 1; s_0 \to s_0 + 2)$, respectively, where $t(s_j; s_i \to s_k)$ is the mean time that $\mathbf{Y}(t)$ spends in state s_j in the course of a (first) passage from s_i to s_k. In other words, using the variable $t(s_j; s_i \to s_k)$,

$$t(s_0; s_0 \to s_0 + 2) \equiv t(s_0; 0 \to 2s_{tot}),$$

$$t(s_0 + 1; s_0 \to s_0 + 2) \equiv t(s_0 + 1; 0 \to 2s_{tot}),$$

since the transitions: $s_0 \to s_0 - 1$ and $s_0 + 2 \to s_0 + 1$ are not allowed in $\mathbf{Y}(t)$. To determine the mean waiting times in states s_0 and $s_0 + 1$ using the *pedestrian approach* (Gillespie, 1992a), then, variables: $v(s)$ and $v_+(s)$ are defined. The variable $v(s)$ is defined as the average number of visits by $\mathbf{Y}(t)$ to state s

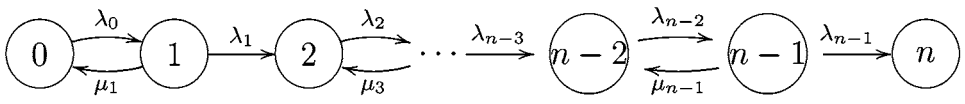

FIG. 1. The state graph of the birth-death process of the enzymatic reaction scheme (1) when $e_{tot} = 1$. This birth-death process has $n+1$ states where $n = 2s_{tot}$, and each state s can be mapped onto a system state of $\mathbf{X}(t)$ by the relationship $\mathbf{x_s} \equiv ((s+1) \bmod 2, s_{tot} - \lceil s/2 \rceil, s \bmod 2, \lfloor s/2 \rfloor)$. Transition rate λ_s is $a_1(\mathbf{x_s})$ if s is an even number, and $a_2(\mathbf{x_s})$ if s is an odd number. Transition rate μ_s is $a_{-1}(\mathbf{x_s})$ if s is an odd number and 0 otherwise.

in the course of a first passage from state 0 to state $2s_{tot}$ while $v_+(s)$ is defined as the average number of transitions $s \to s+1$ taken by $\mathbf{Y}(t)$ in the course of a first passage from state 0 to state $2s_{tot}$. Using these variables, the probability that $\mathbf{Y}(t)$ moves to state s_0+2 from state s_0+1 at the very next jump can be expressed as $v_+(s_0+1)/v(s_0+1)$. Since this probability can also be expressed as $\lambda_{s_0+1}/(\lambda_{s_0+1}+\mu_{s_0+1})$, and since $v_+(s_0+1)$ is 1, we can say

$$v(s_0+1) = \frac{(\lambda_{s_0+1}+\mu_{s_0+1})}{\lambda_{s_0+1}}.$$

Because state s_0+1 can only be visited from state s_0 in $\mathbf{Y}(t)$, $v_+(s_0)$ must be equal to $v(s_0+1)$. Furthermore, since the transition from state s_0 to state s_0-1 cannot occur in $\mathbf{Y}(t)$, $v(s_0)$ must be equivalent to $v(s_0+1)$. Therefore,

$$v(s_0) = \frac{(\lambda_{s_0+1}+\mu_{s_0+1})}{\lambda_{s_0+1}}.$$

Now, let $T(s)$ be a random variable which represents the pausing time in state s in $\mathbf{Y}(t)$. Then, since $\mathbf{Y}(t)$ is a temporally homogeneous birth-death Markov process, $T(s)$ must be a random variable which is necessarily exponentially distributed with parameter $(\lambda_s+\mu_s)$. Then, the mean pausing times in states s_0 and s_0+1 can be expressed, respectively, as:

$$\langle T(s_0)\rangle = \int_0^\infty t\lambda_{s_0}\exp(-\lambda_{s_0}t)dt = \frac{1}{\lambda_{s_0}}.$$

$$\langle T(s_0+1)\rangle = \int_0^\infty t\,(\lambda_{s_0+1}+\mu_{s_0+1})\exp\left(-(\lambda_{s_0+1}+\mu_{s_0+1})\,t\right)dt$$

$$= \frac{1}{\lambda_{s_0+1}+\mu_{s_0+1}}.$$

Since $t(s_j; s_i \to s_k)$ can be formulated as the product of $\langle T(s_j)\rangle$ and $v(s_j)$, the mean waiting times that $\mathbf{Y}(t)$ spends in states s_0 and s_0+1 can be expressed as:

$$t(s_0; 0 \to 2s_{tot}) = \frac{\lambda_{s_0+1}+\mu_{s_0+1}}{\lambda_{s_0+1}\lambda_{s_0}} = \frac{a_2(\mathbf{x_{s_0+1}})+a_{-1}(\mathbf{x_{s_0+1}})}{a_2(\mathbf{x_{s_0+1}})a_1(\mathbf{x_{s_0}})}.$$

$$t(s_0+1; 0 \to 2s_{tot}) = \frac{1}{\lambda_{s_0+1}} = \frac{1}{a_2(\mathbf{x_{s_0+1}})}.$$

Using this information, $\mathbf{Y}(t)$ can be approximated by creating a temporally homogeneous birth Markov process $\mathbf{Y}'(t)$ with the same state space where the mean waiting time in each state s is $t(s; 0 \to 2s_{tot})$ derived from $\mathbf{Y}(t)$. Figure 2 shows the state graph of $\mathbf{Y}'(t)$. Since the waiting time in each state s in $\mathbf{Y}'(t)$ is exponentially distributed, the $s \to s+1$ transition rate λ'_s is the reciprocal of $t(s; 0 \to 2s_{tot})$. Thus, λ'_s is $a_1(\mathbf{x_s})a_2(\mathbf{x_{s+1}})/(a_{-1}(\mathbf{x_{s+1}})+a_2(\mathbf{x_{s+1}}))$ if $s \in \mathbf{S_o}$ and $a_2(\mathbf{x_s})$ if $s \in \mathbf{S_{e'}}$. Therefore, using the PPTA, the enzymatic reaction scheme (1) with e_{tot} being 1 is approximated by a new reaction scheme:

$$\mathrm{E} + \mathrm{S} \xrightarrow{k_{1'}} \mathrm{C} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{2}$$

where $k_{1'} = k_1k_2/(k_{-1}+k_2)$.

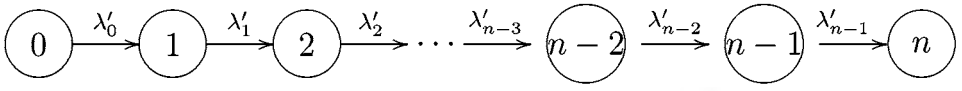

FIG. 2. The state graph of the pure birth process of the PPTA model when $e_{tot} = 1$. This birth process has the same state space as the birth-death process in Figure 1. Transition rate λ'_s is $a_1(\mathbf{x_s})a_2(\mathbf{x_{s+1}})/(a_{-1}(\mathbf{x_{s+1}})+a_2(\mathbf{x_{s+1}}))$ if $s \in \mathbf{S_o}$ and $a_2(\mathbf{x_s})$ if $s \in \mathbf{S_{e'}}$.

When $e_{tot} > 1$, the enzymatic reaction scheme (1) is considered as a set of the enzymatic reactions as follows:

$$\mathrm{E_i} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C_i} \xrightarrow{k_2} \mathrm{E_i} + \mathrm{P}, \quad 1 \leq i \leq e_{tot}$$

where $X_{E_i}(t_0) = 1$ and $X_{C_i}(t_0) = 0$ for each i. Although simulations of this process is definitely slower than that of $\mathbf{X}(t)$, this transformation itself does not require any approximation as, when $\mathbf{X}(t) = \mathbf{x}$, $k_1 x_E x_S \equiv \sum_{i=1}^{e_{tot}} k_1 x_{E_i} x_S$, $k_{-1} x_C \equiv \sum_{i=1}^{e_{tot}} k_{-1} x_{C_i}$, and $k_2 x_C \equiv \sum_{i=1}^{e_{tot}} k_2 x_{C_i}$. Thus, by applying the PPTA to each of the transformed enzymatic reactions, the enzymatic reaction scheme (1) can be approximated by

$$\mathrm{E_i} + \mathrm{S} \xrightarrow{k_{1'}} \mathrm{C_i} \xrightarrow{k_2} \mathrm{E_i} + \mathrm{P}, \quad 1 \leq i \leq e_{tot},$$

which can now be represented using reaction scheme (2).

The two parameters in a PPTA model: $k_{1'}$ and k_2 can be derived from K_M and V_{max}, the maximal reaction rate as follows:

$$k_{1'} = \frac{V_{max}}{K_M e_{tot}} \quad \text{and} \quad k_2 = \frac{V_{max}}{e_{tot}}.$$

Unlike the parameters: k_1 and k_{-1}, the parameters K_M and V_{max} can actually be measured experimentally. Thus, a PPTA model can be constructed and simulated even when full knowledge of the underlying enzymatic reaction is not available and the enzymatic reaction cannot be analyzed quantitatively at that level of detail. This is also true for a QSSA model as its MM form only requires K_M and V_{max} parameters; however, since a PPTA model does not assume that the intermediate species is in quasi-steady state, a PPTA model may perform better than a QSSA model in terms of accuracy, especially in the pre-steady state phase.

3. CASE STUDIES

This section describes the benefits gained by the PPTA method by applying it to various systems containing enzymatic reaction scheme. This section first considers three models of the single enzymatic reaction scheme (1). It then considers the *enzymatic futile cycle* motif which can be ubiquitously seen in biological systems including GTPase cycles, mitogen-activated protein kinase cascades, and glucose mobilization (Samoilov et al., 2005). Finally, it considers a more complex competitive enzymatic reaction. Each model is encoded in SBML (Finney and Hucka, 2003) and simulated for 1000 runs using the same stochastic simulator, an optimized SSA implementation within our modeling and analysis tool, `reb2sac` (Kuwahara et al., 2006). Accuracy of a PPTA model is measured by comparing the time evolution of means and standard deviations.

3.1. Single enzymatic reaction

The enzymatic reaction scheme (1) is first used to analyze the results of our PPTA method. The three different single enzymatic reaction systems are simulated by applying both the QSSA and the PPTA methods to compare their results with those from the original model. In addition, in order to compare the speedup gained by the PPTA with that by the ssSSA, the first two models are chosen to be the ones that are used to help illustrate the application of the slow-scale SSA in enzymatic reaction systems (Cao et al., 2005a).

The first system of the enzymatic reaction scheme (1) has the following initial condition and the reaction rate constants:

$$\mathbf{x_{t_0}} = (220, 3000, 0, 0), \; k_1 = 0.01, \; k_{-1} = 100.0, \; k_2 = 0.01.$$

This system is simulated for 20,000 time units, and each data point is plotted every 100 time units. Figure 3 shows the results from the original model, the PPTA model, and the QSSA model of this system.

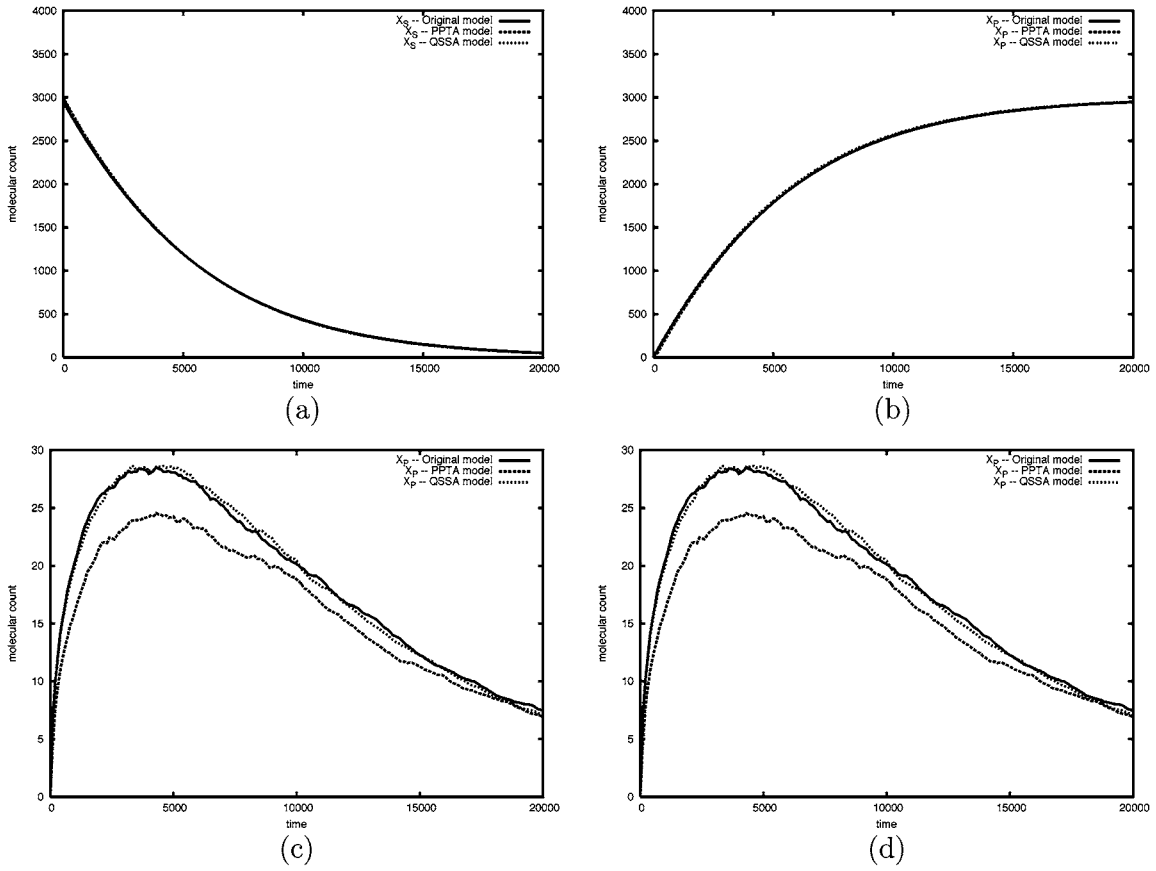

FIG. 3. Comparison of the original enzymatic reaction model, its PPTA model, and its QSSA model with initial conditions: $\mathbf{x}_{t_0} = (220, 3000, 0, 0)$ and the rate constants: $k_1 = 0.01$, $k_{-1} = 100.0$, $k_2 = 0.01$. **(a)** Mean of X_S. **(b)** Mean of X_P. **(c)** Standard deviation of X_S. **(d)** Standard deviation of X_P.

The estimated means X_S and X_P are shown in Figure 3a,b, and the estimated standard deviations of X_S and X_P are shown in Figure 3c,d, respectively.

The simulation results from the QSSA model are in a very close agreement with those from the original model. The average time evolution of the PPTA model is also in a very close agreement with that of the original model. The standard deviation produced via the simulation of the PPTA is slightly lower than that of both the original model and the QSSA model throughout; however, considering the ratio of the standard deviations—which are relatively low—and the average molecular counts—which are very high—the results from the PPTA model are still very accurate. Both the QSSA and the PPTA results in substantial speedup as shown in Table 1. While the entire simulation of the original model takes 68.58 hours, that of the PPTA model only takes 22.8 seconds, achieving 10,800 times speedup. The QSSA model produces an even higher speedup. It requires only 9.2 seconds for the simulation, resulting in 26,765 times speedup. Furthermore, since the speedup gained by the ssSSA is 950 on this model, both the PPTA and the QSSA methods are able to outperform the slow-scale SSA by an order of magnitude while maintaining a high degree of accuracy.

The second enzymatic reaction system has the following initial conditions and reaction rate constants:

$$\mathbf{x}_{t_0} = (10, 3000, 0, 0),\ k_1 = 0.01,\ k_{-1} = 600.0,\ k_2 = 0.1.$$

This system illustrates a case where the average of $X_C(t)$ remains less than 1 as the maximum reaction rate of R_1 (i.e., $k_1 e_{tot} s_{tot}$) is less than k_{-1}. This system is simulated for 80,000 time units and each data point is again plotted every 100 time units. Figure 4a,b shows the estimated means of X_S and X_P, and Figure 4c,d shows the estimated standard deviations of X_S and X_P, respectively.

TABLE 1. SPEEDUP GAINED BY THE ssSSA, THE QSSA, AND THE PPTA ON VARIOUS SYSTEMS INVOLVING ENZYMATIC REACTIONS

		Original	*QSSA*	*PPTA*	*ssSSA*
Single enzymatic reaction 1	Time	68.58 h	9.2 sec	22.8 sec	—
	Speedup	1	26,765	10,800	950
Single enzymatic reaction 2	Time	27.63 h	8.5 sec	17.9 sec	—
	Speedup	1	11,582	5500	400
Single enzymatic reaction 3	Time	34.69 sec	1.07 sec	1.72 sec	—
	Speedup	1	32	20	—
Enzymatic futile cycle	Time	17.73 h	53.43 sec	87.51 sec	—
	Speedup	1	1194	729	—
Competitive enzymatic reaction	Time	65.16 min	63.24 sec	35.78 sec	—
	Speedup	1	62	109	—

The results of the ssSSA are from Cao et al. (2005).

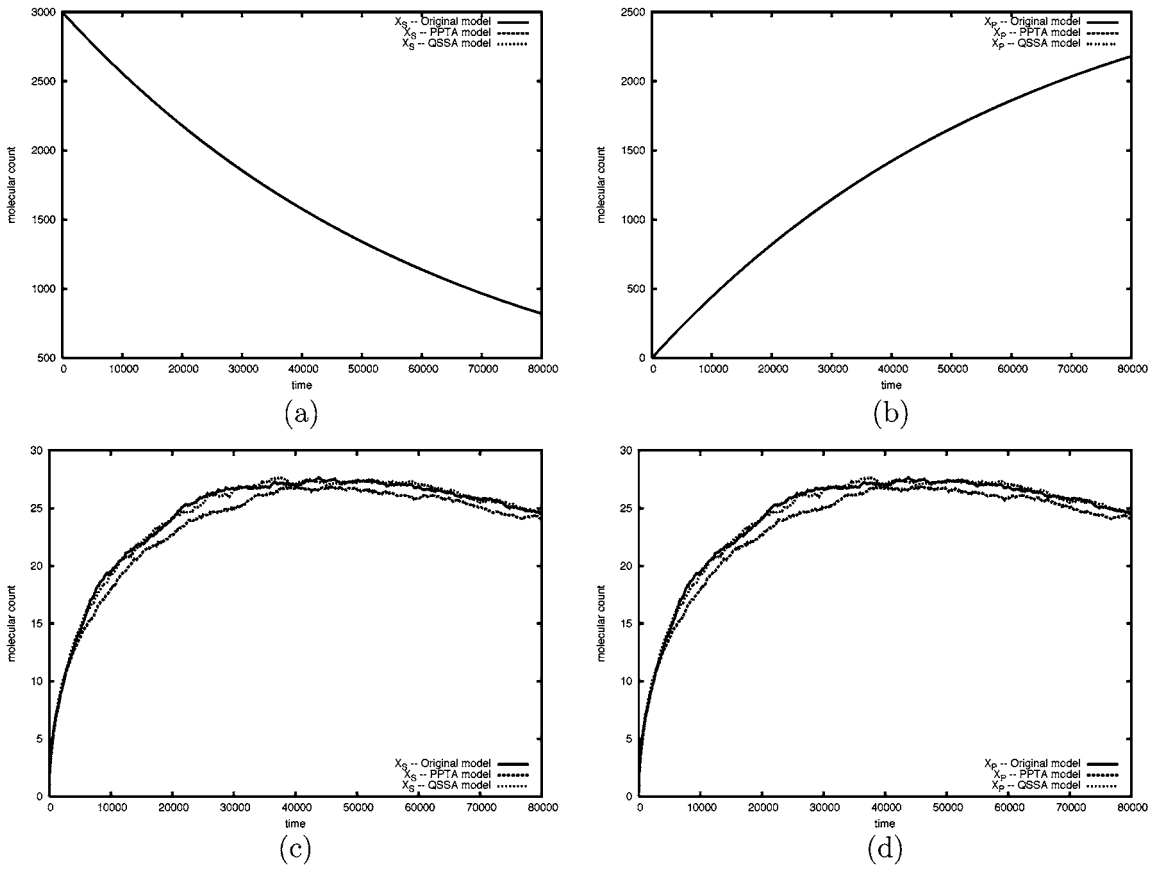

FIG. 4. Comparison of the original enzymatic reaction model, its PPTA, and its QSSA model model with initial conditions: $\mathbf{x}_{t_0} = (10, 3000, 0, 0)$ and the rate constants: $k_1 = 0.01$, $k_{-1} = 600.0$, $k_2 = 0.1$. **(a)** Mean of X_S. **(b)** Mean of X_P. **(c)** Standard deviation of X_S. **(d)** Standard deviation of X_P.

Both the means and the standard deviations from the QSSA model as well as the PPTA model track those from the original model very well while, at the same time, the simulation time of those abstractions are substantially reduced compared with that of the original model as shown in Table 1. Whereas the simulation of the original model takes 27.63 hours, that of the PPTA model and the QSSA model only takes 17.9 seconds and 8.5 seconds, respectively. Thus, the QSSA and the PPTA methods are able to improve the computation performance by a factor of 11,582 and 5500, respectively. Furthermore, since the speedup of the ssSSA is only 400 on this model, both methods are once again able to outperform the ssSSA by an order of magnitude in terms of acceleration. Therefore, for the first two models of the single enzymatic reaction, the QSSA would be the most efficient and effective abstraction as it achieves the highest speedup while maintaining accuracy.

However, since a PPTA model does not assume that the intermediate species are in quasi-steady state, a PPTA model may perform better than a QSSA model in terms of accuracy, especially in a case where the pre-steady state transition is crucial for a prediction of system behavior. For example, suppose the enzymatic reaction scheme (1) has the conditions:

$$\mathbf{x}_{t_0} = (25, 50, 0, 0),\ k_1 = 100.0,\ k_{-1} = 10.0,\ k_2 = 0.1.$$

Then, since e_{tot} is smaller than s_{tot}, the QSSA and the PPTA *could* be applied to safely approximate the temporal behavior of the underlying enzymatic reaction. However, in this system, the propagation effects of the pre-steady state dynamics are rather important, making any QSSA-based models unable to describe the temporal behavior well as shown in Figure 5. While the QSSA achieves slightly higher speedup compared with the PPTA as shown in Table 1, the difference is very minor as both model abstractions result in an

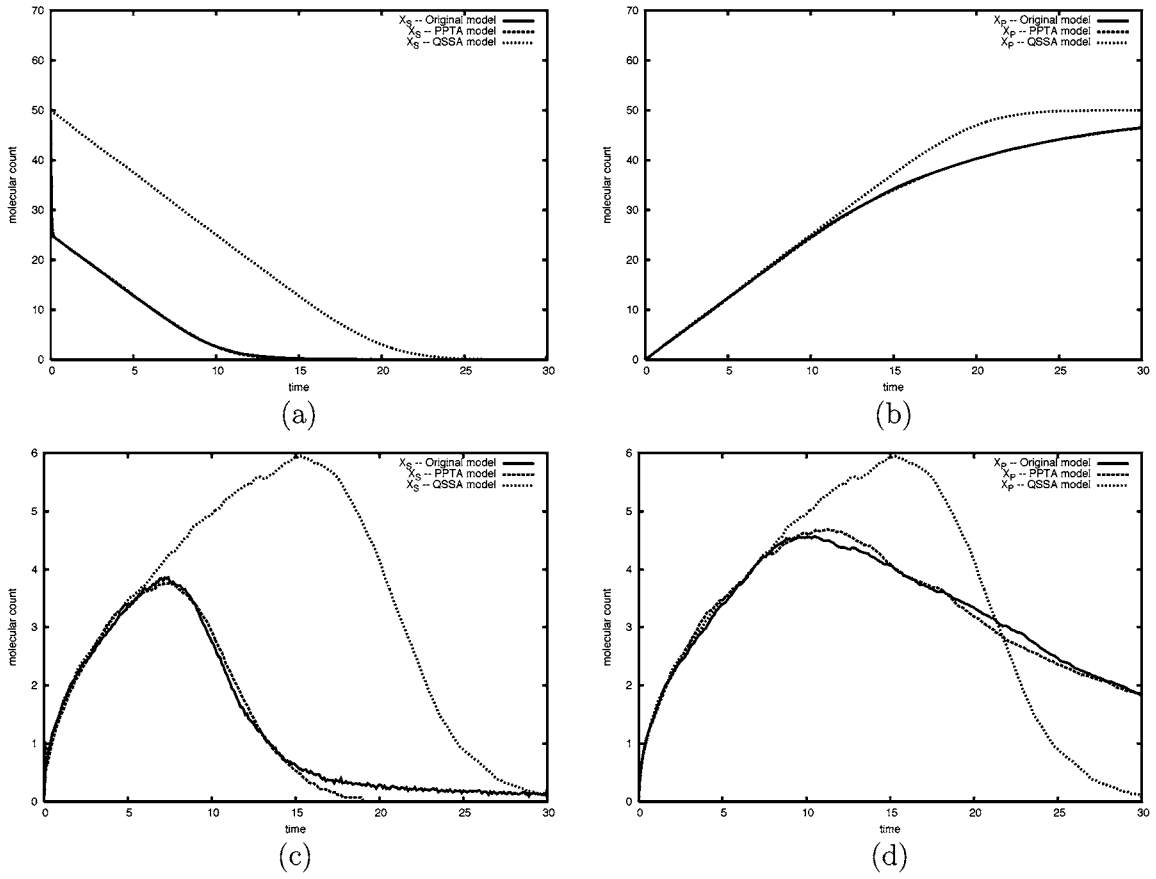

FIG. 5. Comparison of the average time evolutions of the original enzymatic reaction model, its PPTA model, and its QSSA model with initial conditions: $\mathbf{x}_{t_0} = (25, 50, 0, 0)$ and the rate constants: $k_1 = 100.0,\ k_{-1} = 10.0,\ k_2 = 0.01$. **(a)** Mean of X_S. **(b)** Mean of X_P. **(c)** Standard deviation of X_S. **(d)** Standard deviation of X_P.

order of magnitude speedup compared with that of the original model. Thus, considering the accuracy gained, the PPTA is probably the better model abstraction for this system for most analyses.

3.2. *Enzymatic futile cycle*

The enzymatic futile cycle motif consists of two instances of the enzymatic reaction scheme (1) as follows:

$$\mathrm{E}_1 + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C}_1 \xrightarrow{k_2} \mathrm{E}_1 + \mathrm{P}, \quad \mathrm{E}_2 + \mathrm{P} \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} \mathrm{C}_2 \xrightarrow{k_4} \mathrm{E}_2 + \mathrm{S}. \tag{3}$$

One is to transform S into P catalyzed by E_1, and the other one is to transform P into S catalyzed by E_2. This motif is found in many biological systems (Samoilov et al., 2005), abstracting away low-level detail of the motif such as unproductive substrate-complex cycles may provide a significant improvement in performance of the overall system behavior analysis. With the PPTA method, unproductive dissociation reactions are removed, transforming the enzymatic futile cycle model into the following PPTA model:

$$\mathrm{E}_1 + \mathrm{S} \xrightarrow{k_{1'}} \mathrm{C}_1 \xrightarrow{k_2} \mathrm{E}_1 + \mathrm{P}, \quad \mathrm{E}_2 + \mathrm{P} \xrightarrow{k_{3'}} \mathrm{C}_2 \xrightarrow{k_4} \mathrm{E}_2 + \mathrm{S}, \tag{4}$$

where $k_{1'} = k_1 k_2/(k_{-1} + k_2)$ and $k_{3'} = k_3 k_4/(k_{-3} + k_4)$.

The original enzymatic futile cycle model and its PPTA model are simulated for 300 time units with one time unit plot-interval to analyze the accuracy as well as the performance gain of the PPTA model with the initial conditions:

$$(X_S(0), X_P(0), X_{E_1}(0), X_{E_2}(0), X_{C_1}(0), X_{C_2}(0)) = (0, 100, 10, 20, 0, 0),$$

and the rate constants:

$$k_1 = 10^3; k_{-1} = 1.5 \times 10^3; k_2 = 2; k_3 = 10^3; k_{-3} = 5 \times 10^2; \text{ and } k_4 = 1.$$

Since each numerical simulation of the two models starts with no copies of S and 10 copies of E_1, this system illustrates a case where substrate is initially lower than the catalyzing enzyme. Furthermore, since $X_S(t) + X_P(t) + X_{C_1}(t) + X_{C_2}(t)$ is fixed at 100 for all $t \geq 0$, this enzymatic futile cycle system illustrates an applicability of the PPTA model when the numbers of both substrate and enzyme molecules are very low.

Figure 6 shows the results from the original model, the QSSA model, and the PPTA model of this enzymatic futile cycle system. The time evolutions of the estimated means of X_S and X_P are shown in Figure 6a,b, while the estimated standard deviations of X_S and X_P are shown in Figure 6c,d, respectively. While the temporal behavior of the system estimated via the QSSA model does not match that of the original model well, both the means and the standard deviations of X_S and X_P from the PPTA model are able to approximate those from the original model very well. Furthermore, the simulation time of the PPTA model is substantially shortened as shown in Table 1. While the simulation of the original enzymatic futile cycle model takes 17.73 hours, that of the PPTA model only takes 87.51 seconds, achieving more than 729 times speedup. Although, once again, the runtime of the QSSA model simulation is shorter than that of the PPTA model simulation, their speedup factors are still in the same order of magnitude and thus comparable.

3.3. *Competitive enzymatic reaction*

To further demonstrate the usefulness of the PPTA, the following competitive enzymatic reaction system is considered:

$$\begin{aligned} &\mathrm{S}_1 \xrightarrow{k_{b1}} \mathrm{P}_1, \quad \mathrm{E} + \mathrm{S}_1 \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C}_1 \xrightarrow{k_2} \mathrm{E} + \mathrm{P}_1, \quad \xrightarrow{k_{p1}} \mathrm{S}_1, \quad \mathrm{S}_1 \xrightarrow{k_{d1}}, \\ &\mathrm{S}_2 \xrightarrow{k_{b2}} \mathrm{P}_2, \quad \mathrm{E} + \mathrm{S}_2 \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} \mathrm{C}_2 \xrightarrow{k_4} \mathrm{E} + \mathrm{P}_2, \quad \xrightarrow{k_{p2}} \mathrm{S}_2, \quad \mathrm{S}_2 \xrightarrow{k_{d2}}. \end{aligned} \tag{5}$$

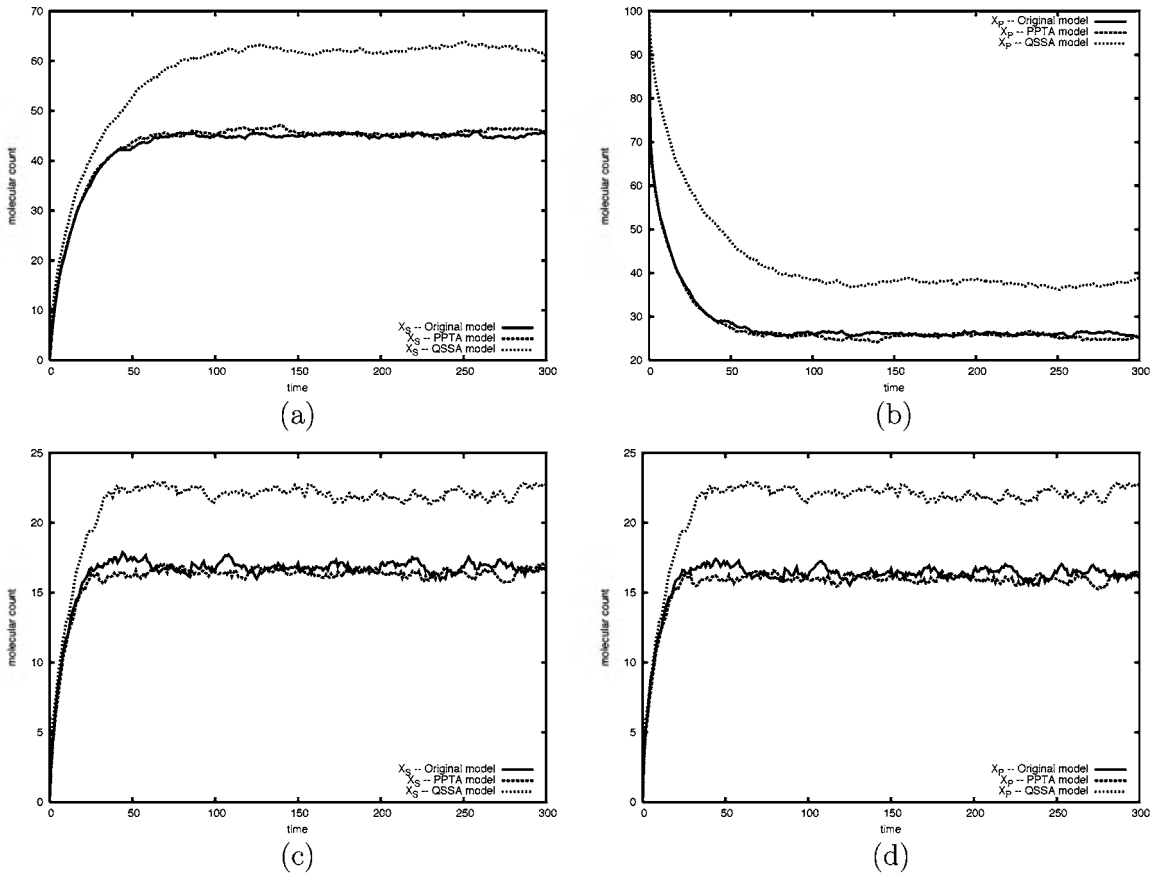

FIG. 6. Comparison of the original enzymatic futile cycle model, its PPTA model, and its QSSA model with initial conditions: $X_S(0) = 0$, $X_P(0) = 100$, $X_{E_1}(0) = 10$, $X_{E_2}(0) = 20$, $X_{C_1}(0) = 0$, $X_{C_2}(0) = 0$, and the rate constants: $k_1 = 1.0 \times 10^3$, $k_{-1} = 1.5 \times 10^3$, $k_2 = 2.0$, $k_3 = 1.0 \times 10^3$, $k_{-3} = 5.0 \times 10^2$, $k_4 = 1.0$. **(a)** Mean of X_S. **(b)** Mean of X_P. **(c)** Standard deviation of X_S. **(d)** Standard deviation of X_P.

In this system, both S_1 and S_2 compete to bind to E to produce P_1 and P_2, respectively. Also, this scheme contains basal reactions to transform S_1 and S_2 into P_1 and P_2, respectively, without being catalyzed by E. Moreover, since substrates S_1 and S_2 are often produced and consumed via various reactions, reaction scheme (5) also contains reactions to model productions and consumptions of S_1 and S_2.

The PPTA model of the competitive enzymatic reaction model (5) removes the substrate-dissociation reactions from C_1 and C_2, resulting in the following model:

$$\begin{aligned} &S_1 \xrightarrow{k_{b1}} P_1, \quad E + S_1 \xrightarrow{k_1'} C_1 \xrightarrow{k_2} E + P_1, \quad \xrightarrow{k_{p1}} S_1, \quad S_1 \xrightarrow{k_{d1}}, \\ &S_2 \xrightarrow{k_{b2}} P_2, \quad E + S_2 \xrightarrow{k_3'} C_2 \xrightarrow{k_4} E + P_2, \quad \xrightarrow{k_{p2}} S_2, \quad S_2 \xrightarrow{k_{d2}}. \end{aligned} \tag{6}$$

To analyze the accuracy of this PPTA model, the following initial conditions:

$$(X_E(0), X_{S_1}(0), X_{S_2}(0), X_{P_1}(0), X_{P_2}(0), X_{C_1}(0), X_{C_2}(0)) = (10, 0, 0, 0, 0, 0, 0),$$

and the rate constants:

$$k_{b1} = 2 \cdot 10^{-5}; k_1 = 10^2; k_{-1} = 10^2; k_2 = 0.1; k_{p1} = 10; k_{d1} = 0.2;$$

$$k_{b2} = 10^{-5}; k_3 = 200; k_{-3} = 10^2; k_4 = 0.15; k_{p2} = 10; \text{ and } k_{d2} = 0.2,$$

are used for the simulations. The values of rate constants for the productions and consumptions of S_1 and S_2 are chosen so that the consumption rate constants are relatively high to capture isolation of substrates from binding to the enzyme and that both substrates are present in low counts throughout the simulations (i.e., $\forall t \geq 0.\ \langle X_{S_1}(t) \rangle \leq 100 \wedge \langle X_{S_2}(t) \rangle \leq 100$). The values of basal transformation rate constants k_{b1} and k_{b2} are chosen so that basal transformation rates are much smaller than those from the catalyzed reactions when the substrates are present in low counts (i.e., $k_2 \cdot e_{tot} \gg 100k_{b1}$ and $k_4 \cdot e_{tot} \gg 100k_{b2}$).

Figure 7 shows the results from the simulations of the three models. The estimated means of X_{S_1} and X_{S_2} are shown in Figure 7a,b, while the estimated standard deviations of X_{S_1} and X_{S_2} are shown in Figure 7c,d. In this system, both the means and the standard deviations of X_{S_1} and X_{S_2} from the PPTA model and the QSSA model track those from the original model very well with a substantial improvement in simulation time as shown in Table 1. Unlike the other enzymatic reaction systems presented so far in this section, the PPTA model achieves a higher speedup compared with the QSSA model in this system. While the simulation of the the QSSA model takes 63.24 seconds achieving 62 times speedup compared with that of the original model, the simulation of the PPTA model only takes 35.78 seconds achieving 109 times speedup. Thus, for this competitive enzymatic reaction model, the PPTA is able to outperform the QSSA model in terms of speedup. One interpretation of this result is that, while the QSSA model can advance the time step further in each reaction event than the the PPTA model, the evaluations of the kinetic laws of the reduced competitive enzymatic reactions in the QSSA require higher computational costs due

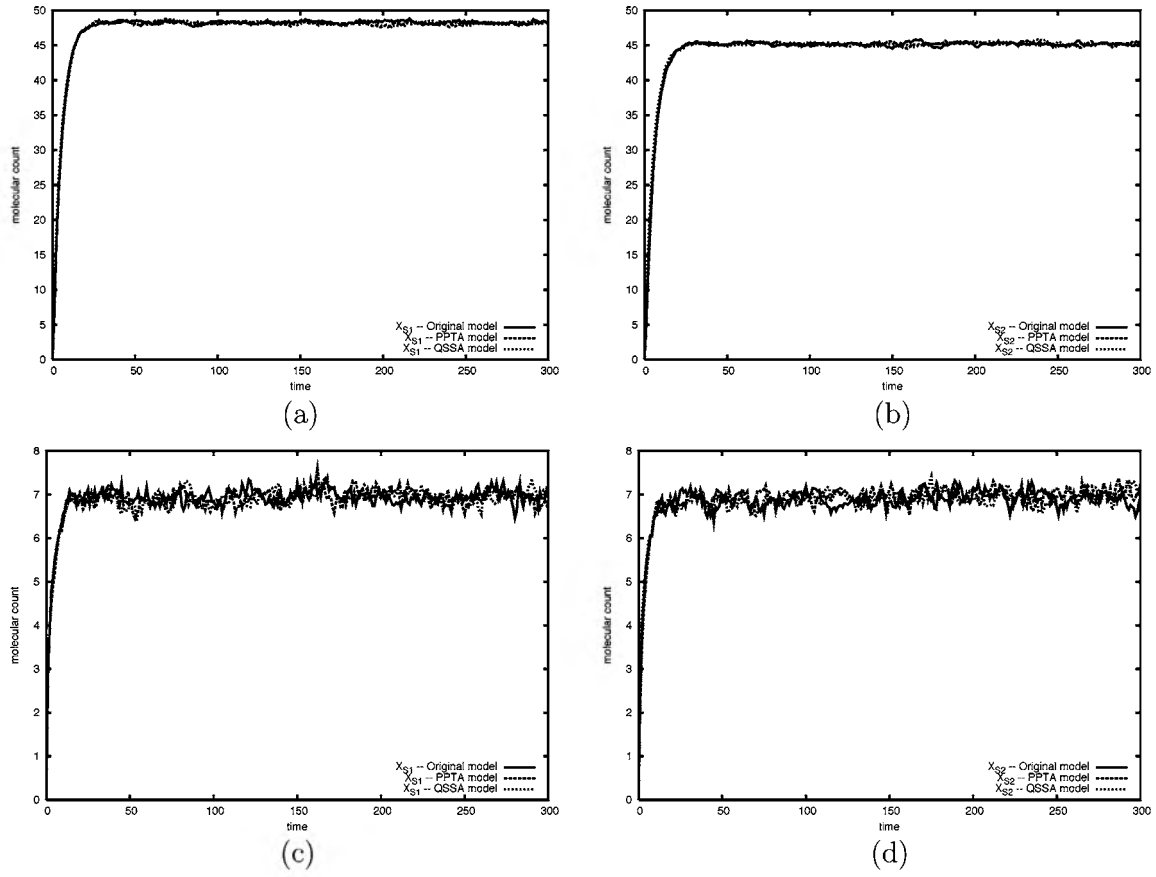

FIG. 7. Comparison of the original model of competitive enzymatic reaction model, its PPTA model, and its QSSA model with initial conditions: $X_E(0) = 10$ and 0 molecule for the rest of the species, and the rate constants: $k_{b1} = 2 \cdot 10^{-5}$, $k_1 = 10^2$, $k_{-1} = 10^2$, $k_2 = 0.1$, $k_{p1} = 10$, $k_{d1} = 0.2$, $k_{b2} = 10^{-5}$, $k_3 = 200$, $k_{-3} = 10^2$, $k_4 = 0.15$, $k_{p2} = 10$, $k_{d2} = 0.2$. **(a)** Mean of X_{S_1}. **(b)** Mean of X_{S_2}. **(c)** Standard deviation of X_{S_1}. **(d)** Standard deviation of X_{S_2}.

to an increase in the complexity, resulting in just enough overhead for the PPTA model to outperform the QSSA model in this example.

4. CONCLUSION

This paper introduces a new model abstraction method, *production-passage-time approximation* (PPTA), that can significantly improve the temporal behavior analysis time of enzymatic reaction systems. As a case study, we have applied the PPTA method to various systems, and compared the accuracy as well as the runtime between the original model and the PPTA model. The preliminary results are promising. This paper has shown that the PPTA model can make stochastic simulations of the single enzymatic reaction system orders of magnitude faster while maintaining accuracy. Moreover, this paper has shown that the PPTA can be utilized to efficiently approximate more complex systems, exemplified here using an enzymatic futile cycle model and a competitive enzymatic reaction model. At the same time, it has also demonstrated that our new method can perform better than some of the best existing enzymatic reaction approximation methods. This paper has demonstrated that the PPTA method achieves an acceleration of an order of magnitude over the slow-scale SSA for the two enzymatic reaction systems from Cao et al. (2005a). Furthermore, it has illustrated that the PPTA can perform better than the QSSA-based abstraction in terms of accuracy as well as the acceleration factors. Additionally, our approach can also be used within a continuous, deterministic framework to remove the computationally challenging stiff condition often found in enzymatic reactions with $k_{-1} \gg k_2$. An additional noteworthy benefit of the PPTA is that, since it does not require a customized simulation procedure for enzymatic reactions, it allows biochemical systems comprising such reactions along with other types of reactions to still take advantage of utilizing general stochastic simulation tools for the standard Gillespie stochastic simulation algorithm.

Future work includes comprehensive analysis of the PPTA errors under various conditions and analysis of efficiency-versus-accuracy among the PPTA and other approximations such as the QSSA based on the initial conditions, that is, a static and systematic approach to determine when to use the PPTA.

NOTE ADDED IN PROOF

More recent tests comparing PPTA and ssSSA show that the ssSSA with the appropriate approximations described in Sections V and VI of Cao et al. (2005) can be faster than the PPTA for comparable simulation accuracy.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (grant no. 0331270). The authors would like to thank Min Roh and Daniel Gillespie for providing the latest ssSSA suite and having stimulating discussions.

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Arkin, A., Ross, J., and McAdams, H. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648.

BioSPICE. 2008. Available at: *www.biospice.org/*. Accessed July 6, 2008.

Cai, L., Friedman, N., and Xie, X.S. 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, 358–362.

Cao, Y., Gillespie, D., and Petzold, L. 2005a. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *J. Phys. Chem.* 123, 144917.

Cao, Y., Gillespie, D., and Petzold, L. 2005b. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.* 122, abstract 014116.

Cao, Y., Gillespie, D.T., and Petzold, L.D. 2005. Avoiding negative populations in explicit tau leaping. *J. Chem. Phys.* 123, abstract 054104.

Elowitz, M.B., Levine, A.J., Siggia, E.D., et al. 2002. Stochastic gene expression in a single cell. *Science* 297, 1183–1186.

Finney, A., and Hucka, M. 2003. Systems Biology Markup Language (SBML): level 2 and beyond. *Biochem. Soc. Trans.* 31, 1472–1473.

Gibson, M., and Bruck, J. 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* 104, 1876–1889.

Gillespie, D.T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403–434.

Gillespie, D.T. 1992a. *Markov Processes: An Introduction for Physical Scientists.* Academic Press, San Diego.

Gillespie, D.T. 1992b. A rigorous derivation of the chemical master equation. *Physica A* 188, 404–425.

Gillespie, D.T. 2000. The chemical Langevin equation. *J. Chem. Phys.* 113, 297–306.

Gillespie, D.T. 2001. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* 115, 1716–1733.

Gillespie, D.T. 2005. Chapter 5.11, 1735–1752. In: *Handbook of Materials Modeling.* Springer, New York.

Gillespie, D.T., and Petzold, L.R. 2003. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.* 119, 8229–8234.

Golding, I., Paulsson, J., Zawilski, S.M., et al. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123, 1025–1036.

Kuwahara, H., Myers, C.J., Samoilov, M.S., et al. 2006. Automated abstraction methodology for genetic regulatory networks. *Trans. Computat. Syst. Biol.* 4, 150–175.

McAdams, H.H., and Arkin, A. 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* 94, 814–819.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., et al. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846.

Pedraza, J.M., and van Oudenaarden, A. 2005. Noise propagation in gene networks. *Science* 307, 1965–1969.

Rao, C.V., and Arkin, A.P. 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *J. Phys. Chem.* 118, 4999–5010.

Rao, C.V., Wolf, D.M., and Arkin, A.P. 2002. Control, exploitation and tolerance of intracellular noise. *Nature* 420, 231–238.

Rathinam, M., Cao, Y., Petzold, L., et al. 2003. Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *J. Chem. Phys.* 119, 12784–12794.

Samoilov, M., Plyasunov, S., and Arkin, A.P. 2005. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proc. Natl. Acad. Sci. USA* 102, 2310–2315.

Address reprint requests to:
Dr. Hiroyuki Kuwahara
The Microsoft Research–University of Trento
Centre for Computational and Systems Biology
Piazza Manci, 17
Trento 38100
Italy

E-mail: kuwahara@cosbi.eu