Simplified yet highly accurate enzyme kinetics for cases of low substrate concentrations

Hanna M. Härdin[1,2], Antonios Zagaris[2,3], Klaas Krab[1] and Hans V. Westerhoff[1,4,5]
1 Department of Molecular Cell Physiology, VU University, Amsterdam, The Netherlands.
2 Modelling, Analysis and Simulation, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands.
3 Korteweg–de Vries Instituut, University of Amsterdam, Amsterdam, The Netherlands.
4 Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, The University of Manchester, Manchester, UK.
5 Netherlands Institute for Systems Biology, The Netherlands.

Correspondence
H. V. Westerhoff, Department of Molecular Cell Physiology, VU University, De Boelelaan 1085, NL-1081 HV Amsterdam, The Netherlands
Fax: +31 20 5987229
Tel: +31 20 5987228
E-mail: hans.westerhoff@manchester.ac.uk
http://www.siliconcell.net

Running title
Enzyme kinetics for low substrate concentrations

Abbreviations
QSSA, quasi-steady-state approximation; ZDP, zero-derivative principle; SIM, slow invariant manifold; ODE, ordinary differential equation; PTS, phosphotransferase system; PEP, phosphoenolpyruvate; Pyr, pyruvate; EI, enzyme I; HPr, histidine protein; EIIA, enzyme IIA; EIICB, enzyme IICB; Glc, glucose; P, phosphate group.

Keywords
biochemical system reduction; quasi-steady-state approximation (QSSA); zero-derivative principle (ZDP); slow invariant manifold (SIM); enzyme kinetics

Subdivision
Systems biology

Summary

Much of enzyme kinetics builds on simplifications enabled by the quasi-steady-state approximation (QSSA) and is highly useful when the concentrations of the enzyme is much lower than that of its substrate. However, *in vivo*, this condition is often violated. Here we show that under conditions of realistic yet high enzyme concentrations, the QSSA approach may readily be off by more than a factor of 4 when predicting concentrations. We then present a novel extension of the QSSA based on the *zero-derivative principle* (ZDP) which requires considerably less theoretical work than did previous such extensions. We show that the first order ZDP, already, describes much more accurately the true enzyme dynamics at enzyme concentrations close to the concentration of their substrates. This should be particularly relevant for enzyme kinetics where the substrate is an enzyme, such as in phosphorelay and MAP kinase pathways. We illustrate this for the important example of the phosphotransferase system involved in glucose uptake, metabolism, and signalling. We find that this system with a potential complexity of nine dimensions can be understood accurately using first order ZDP in terms of the behavior of a single variable with all other concentrations constrained to follow that behavior.

Introduction

The investigation of the function of molecular processes in cells, such as genetic networks, metabolic processes, and signal transduction pathways can benefit from the analysis of mathematical models of those systems. This analysis is essential for understanding the basis of the functional properties that the networks exhibit, and it is further used for drug development and experimental design. Due to the many molecular components involved in these systems, the models describing them often become large, *e.g.*, models with 499 and with 1343 dynamic variables are given in [1] and [2], respectively. The construction of such large models has become possible due to the advances in functional genomics which enable, in principle, the experimental determination of properties of virtually all molecules in living organisms [3]. Even larger models are expected to appear, possibly describing entire cells and organisms in detail.

The construction of perspicuous yet accurate biochemical models remains a challenge. First, considering that the smallest living cells already have a few hun-

dred genes, that each gene has its own transcription, splicing, and translation processes, and that the proteins corresponding to each gene may be part of metabolic and signalling networks, it becomes evident that the number of processes in a cell can readily exceed a few hundreds. Each one of these processes typically involves a large number of molecular components and, therefore, modeling the interactions between these requires the use of highly nonlinear rate laws. Furthermore, all of these processes are highly dependent on each other in nonlinear ways [4]. Due to these interdependencies, even the modeling of pathways apparently involving only a dozen of species becomes intricate, as the effect of the surrounding hundreds or thousands of molecules has to be summarized in a biologically meaningful way.

Because of the complexity of biochemical processes outlined above, which also reflects on the models describing them, their behavior becomes unintuitive and their function difficult to fathom [5, 6, 7, 8, 9]. However, precisely because much function is due to the very nonlinearities that cause these problems, the modeling and analysis of these systems in simple yet accurate ways become absolutely necessary for understanding the functions that the processes perform. To this end, a variety of different modeling approaches, as well as methods to simplify the models, have been developed, see, *e.g.*, the reviews [10, 11, 12]. Naturally, these approaches are approximate and subject to limitations, a fact which conveys an interest in further investigating and developing new modeling and simplification methods.

Several of the current modeling and simplification methods exploit the fact that the molecular processes within a cell are organized on a variety of spatial and temporal scales. In particular, although the complexity of biochemical systems (and, by extension, also of biochemical models) is necessary for biological function to arise from processes between 'dead' molecules, not all aspects of this complexity are relevant for all the functions of the living cell. In other words, although a given process performing a certain function within the cell may be employing a complex network of molecular interactions, there are also processes within this same cell whose effect can be effectively summarized (instead of modeled in detail) when studying this particular function. A prime example of this phenomenon is offered by an enzyme-catalyzed reaction where the function is the conversion of one metabolite into another: there, the formation and dispersion of the complex of the enzyme with its metabolites, which may be modeled by detailed mass action kinetics, occur on a faster timescale than the overall reaction of the metabolites and thus the dynamics of the overall reaction can be summarized by the simpler enzyme kinetics. Indeed, at the level of a metabolic pathway such as glycolysis, models employing enzyme kinetics (at each reaction) are suffi-

ciently accurate to describe the function of the entire pathway [13]. This practice allows the investigator to omit inessential complexity and to focus on the elements underlying the emergence of function of the pathway. The focus on those aspects of the cellular interactions that are indispensable to the biological function under study is necessary for understanding how function emerges from the molecular interactions. In this article, we revisit the use of timescale disparities present in complex biochemical systems in obtaining simplified models. Further, we present a family of methods which act as accurate extensions of the technique used to derive enzyme kinetics from mass action kinetics, and we demonstrate their use in obtaining accurate simplified models.

During the course of fast timescales (*i.e.*, over a short initial time span), certain processes are virtually stagnant while others proceed essentially independently of these. At slower timescales (*i.e.*, over longer time periods), the latter (*fast*) processes seem to evolve coherently with the former (*slower*) ones. In the example of the enzyme-catalyzed conversion of a substrate to a product, the fast timescale corresponds to an initial, short phase where the concentration of the enzyme–substrate complex saturates while the substrate concentration remains approximately constant, and the slow timescale to the subsequent, longer phase where both concentrations change slowly with that of the complex constrained to that of the substrate.

Approximations based on timescale separation have a long tradition in biochemistry, starting with the quasi-steady-state approximation (QSSA) dating back to the beginning of the previous century [14, 15, 16]—see the excellent review in [17]. The QSSA has been used to derive the tractable and abundantly used Michaelis–Menten kinetics from the more precise but more complex mass action kinetics, a clear indication of the important role it has played in biochemical modeling. A series of mathematical studies [17, 18, 19] have quantified its accuracy, proving it to be proportional to the timescale disparity present in the system to which it is applied. It follows that this approximation can be satisfactory for the enzyme catalysis example above, which may exhibit large timescale separation, while in signal transduction pathways, where the timescale separation is often relatively small, the quality of the approximation diminishes.

The QSSA has been extended to higher orders in [20, 21, 22]. Common to these extensions is the *explicit identification* of a small parameter, typically denoted by ε, which measures the timescale disparity. This identification requires a host of theoretical considerations (see [17], for example), and it readily becomes prohibitively complicated for the realistically complex systems of biology. In this article, we propose a sequence of increasingly accurate refinements of the QSSA

which are based on the zero-derivative principle (ZDP) [23, 24] and do not require the identification of such a parameter. The ZDP was pioneered by Kreiss and coworkers [25, 26, 27] in the applied mathematics/computational physics community. It has been employed to obtain accurate, yet simplified descriptions of complex models arising in meteorology [28], computational physics [29, 30], and more general multiscale systems [31, 32], but not yet in the current biochemical context. We apply the ZDP to two systems: first, to a prototypical example with a reversible enzymatic reaction and, second, to the substantially more complex phosphotransferase system (PTS), a signal transduction pathway regulating and catalyzing glucose uptake in enteric bacteria. In both cases, we demonstrate that our results are more accurate than those obtained by the QSSA.

We first revisit key ideas underlying the derivation of simplified models by exploiting the timescale separation present in biochemical systems and elucidate our discussion by working with the prototypical enzyme-catalyzed reaction discussed above. Subsequently, we briefly review the QSSA and then motivate and present the ZDP. We apply both of these to our prototypical example and discuss the similarities and differences between the results yielded by each one of them. Finally, we apply the QSSA and ZDP to the large, realistic PTS model.

Results

Timescale separation in biochemical systems

In this section, we briefly review how timescale separation leads to the emergence of constraining relations, and we demonstrate how these relations may be used to obtain simplified descriptions of dynamical systems. Our aim here is to provide a short, self-contained introduction to the subject of nonlinear multiscale reduction from a biochemical point of view. We refer to [33, 34, 35] for more detailed and broader introductions to this subject.

Timescale separation in an enzymatic reaction

For concreteness of presentation, we start with a specific mechanism, namely a reversible enzyme-catalyzed reaction. More specifically, we consider an enzyme E catalyzing the conversion of a substrate S to a product P by means of binding to

S to form a complex C,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P. \tag{1}$$

We assume that both the binding of S to E and the release of P are reversible reactions, and hence the conversion of substrate to product is also an overall reversible reaction. This mechanism has been analyzed in detail, see [36, 37] and references therein. Our presentation here summarizes certain key facts which we shall need below.

In what follows, we denote the concentrations of S, P, E, and C by s, p, e, and c, respectively. We regard the total concentration of (free and bound) enzyme $e_{tot} = e + c$ as constant, based on the fact that changes on the genetic level are slow compared to those on the metabolic one. We further assume that p is also kept constant—for example, by introducing another enzyme-catalyzed reaction in which P is consumed and where the enzyme has very high elasticity with respect to P. (This second assumption serves to reduce the number of variables so as not to clutter our model. It by no means pertains to the nature of our analysis.)

Under these assumptions, the state of the system is fully described by two state variables, either s and c or s and e; for historical reasons, we choose to employ s and c. The evolution in time of the state variables is given by the ordinary differential equations (ODEs)

$$\dot{s} = -v_1 \quad \text{and} \quad \dot{c} = v_1 - v_2, \tag{2}$$

together with the initial conditions $s(0) = s_0$ and $c(0) = c_0$. The reaction rates v_1 and v_2 are given by mass action kinetics; since $e = e_{tot} - c$, we find that

$$v_1 = k_1 (e_{tot} - c) s - k_{-1} c \quad \text{and} \quad v_2 = k_2 c - k_{-2} (e_{tot} - c) p, \tag{3}$$

where the rate constants $k_1, \ldots, k_{-2}$ are arbitrary but given.

The equilibrium of the enzymatic reaction (1), *i.e.* the state in which $v_1 = v_2 = 0$, is given by

$$(s^*, c^*) = \left(\frac{k_{-1} k_{-2} p}{k_1 k_2}, \frac{k_{-2} p e_{tot}}{k_2 + k_{-2} p} \right). \tag{4}$$

The concentrations $s(t)$ and $c(t)$ approach **the equilibrium** at a decreasing rate. Plotting these concentrations in the (s, c)-plane yields a *trajectory* (a curve) which is parameterized by time—every point on the curve corresponds to a value $(s(t), c(t))$,

for some time t, and *vice versa*, see Fig. 1. It becomes evident from this figure that the evolution of s and c towards their **equilibrium** values runs through two distinct phases. In the first phase, c increases (or decreases) while s remains essentially constant, corresponding to an initial rapid binding of S to E (or dissociation of C). In the second phase, both variables evolve at similar rates towards their **equilibrium** values, corresponding to the consumption of substrate by the enzyme. The duration of the first phase is far shorter than that of the second one, a fact which has led researchers to label the dynamics driving the former *fast* (or *transient*) and those driving the latter *slow*. This fact also suggests that, except for a short initial period, the evolution of the system is described by the part of the trajectory corresponding to the second, slow phase.

A related feature of model (2)—and one of central importance to this study—becomes apparent upon plotting the trajectories corresponding to several initial conditions. In particular, Fig. 1 shows that all trajectories approach a certain curve in the $(s,c)-$plane during the first phase and stay in a neighborhood of it during the second phase (see also [38] for the irreversible case). **This curve is called a *normally attracting, slow invariant manifold (SIM)*. The SIM serves to link the full to the fully relaxed dynamics, as the system dynamics follows a cascade from *full* (approach to the SIM) to *partially relaxed* (close to the SIM) to, eventually, *fully relaxed* (close to the equilibrium). In this sense, SIMs form the backbone on which the dynamics is organized at intermediate timescales.**

The SIM is the graph of a *constraining relation*—that is, of a relation $c = c(s)$ dictating that, past the transient phase, the complex concentration is approximately a function of the substrate concentration. Knowledge of the constraining relation $c = c(s)$ allows one to *reduce* (2)–(3) to the single ODE

$$\dot{s} = -k_1 \left(e_{tot} - c(s)\right) s + k_{-1}\, c(s). \tag{5}$$

This ODE, together with the constraining relation $c = c(s)$ and the conservation laws $e(t) + c(t) = e_0 + c_0$ and $p(t) = p$, describes the dynamics of the system at the slow timescale.

General multiscale systems

Here, we generalize the notions introduced above to more general multiscale systems. In what follows, we use the term *state variables* to denote those time-dependent variables in a biochemical system which fully describe the system at any given moment. (State variables are, typically but not exclusively, molecular concentrations. In certain models, they can also be linear combinations of such

concentrations or other time-dependent quantities such as pH or membrane potential.) First, we collect the values of all n state variables (where n is a natural number depending on the complexity of the system) at any time instant t in a column vector $z(t)$. The time evolution of the components of z is dictated by a set of *state equations* in the form of ODEs,

$$\dot{z}(t) = f(z(t)), \tag{6}$$

where f is a vector-valued function of n variables and with n components. In the case of the simple enzyme reaction model in the previous subsection, we have

$$n = 2, \quad z = \begin{pmatrix} s \\ c \end{pmatrix}, \quad \text{and} \quad f(s,c) = \begin{pmatrix} k_{-1}c - k_1(e_{tot} - c)s \\ (k_1 s + k_{-2}p)(e_{tot} - c) - (k_{-1} + k_2)c, \end{pmatrix},$$

see (2)–(3). The $n-$dimensional Euclidean space $\mathbb{R}^n$, which is where the state variables collected in z assume values, is called the *state space* (in the enzyme reaction example, this is the $(s,c)-$plane). A solution $z(t)$ of (6) corresponding to any given initial condition $z(0) = z_0$ and plotted in the state space for all t is a *trajectory*, whereas any value z^* satisfying $f(z^*) = 0$ is a *steady state*. **(In the example above, the condition $f(s^*, c^*) = 0$ is fulfilled when $v_1 = v_2 = 0$, cf. (3), and therefore the unique steady state of that specific system is the equilibrium (4) of the enzymatic reaction.)**

As we mentioned in the Introduction and demonstrated in the example above, the various processes in a biochemical system typically act at vastly disparate timescales, resulting in a separation of its dynamics into fast and slow. In this general case also, this behavior manifests itself in the state space by means of trajectories approaching a lower-dimensional SIM—**that is, a manifold which is invariant under the dynamics, attracts nearby orbits, and on which system evolution occurs on a slow timescale.**[1] In what follows, we write $n_x < n$ for the dimension of this SIM and use the shorthand $n_y = n - n_x$ (in the case of the enzyme reaction model above, this SIM is a curve and thus $n_x = n_y = 1$). This approach occurs along specific directions *transversal* to the SIM (*normal attractivity*) and corresponding to n_y **(possibly nonlinear)** combinations of molecular concentrations remaining approximately constant during the fast transient. (In the case of the enzyme reaction in Fig. 1, this approach is approximately vertical—$s \approx const.$—since s is approximately conserved in that phase.) Evolution on and

[1] SIMs are typically not unique—instead, there is an entire **continuous** *family* of SIMs **corresponding to trajectories with initial conditions in the slow region of the state space** and each member of which may be used to reduce the system (see, *e.g.*, [35]).

near the SIM occurs on a *slower* timescale, while trajectories starting on the SIM remain on it for all times (*invariance*)—**see [33, 35] for more technical definitions of these terms**.

It is typically the case that the state variables collected in z can be partitioned into two groups,

$$z = \begin{pmatrix} x \\ y \end{pmatrix}, \text{ where } x \text{ is } n_x\text{–dimensional and } y \text{ is } n_y\text{–dimensional,}$$

so that the SIM is the graph of a constraining relation $y = g(x)$, for some function g of n_x variables and with n_y components. In that case, one may rewrite (6) as

$$\dot{x} = f_x(x,y) \quad \text{and} \quad \dot{y} = f_y(x,y), \tag{7}$$

where f_x and f_y collect the vector field components of f corresponding to x and y, respectively. Thus, one obtains the reduced system

$$\dot{x} = f_x(x, g(x)), \quad \text{together with the constraining relation } y = g(x), \tag{8}$$

which employs the n_x variables x and describes the slow dynamics. This ODE describes the dynamics of the partially relaxed phase and is typically easier to analyze and interpret than the full model (6) (or, equivalently, (7)). Thus, this reduced dynamics is also easier to relate to the investigator's intuitive understanding in order to reinforce or correct intuition, as the case may be.

Remark. It often occurs that a given system has *many* timescales instead of only two (fast and slow). In the course of each timescale, a number of processes approximately balance, and thus the number of approximately balanced processes increases from one phase to the next. This behavior is manifested in the state space through a *hierarchy of SIMs* of decreasing dimensions and embedded in one another. In this setting, there are no unique transient and partially relaxed phases, but rather a cascade of as many phases as timescales, with each consecutive phase exhibiting slower and lower dimensional dynamics than its predecessor. At the end of each phase, trajectories have been attracted to the next SIM in the hierarchy so that the system dimensionality decreases further. Hence, the dimension of the reduced model depends on the timescale that is of interest to the investigator.

Approximating the slow behavior

The explicit determination of the constraining relations $y = g(x)$ is impossible for most biochemical systems. Indeed, the timescale separation in realistic systems is

always finite, and thus the transition from fast to slow dynamics described in the previous section is not instantaneous but gradual. As a result, the notions of fast and slow dynamics are not absolute but, rather, at an interplay with each other the assessment of which is a difficult task. To circumvent this difficulty, a collection of methods to *approximate* constraining relations has been developed. Among these, the QSSA is the best known and well-studied. It was developed to obtain an approximate reduced description of an enzymatic reaction valid over a slow timescale [16], and it is also the precursor to the ZDP. In the coming two sections, we review the QSSA and apply it to our enzyme reaction example. Then, we introduce the ZDP which extends the QSSA.

The quasi-steady-state approximation

In what follows, we assume the setting introduced in the previous section. In particular, we assume that the system under study is fully described by an n-dimensional vector z of state variables evolving under (6), for some function f, and also that it possesses a SIM of dimension $n_x < n$. The QSSA assumes that, during partial relaxation, certain of the variables (which we denote by $\bar{y}$, with $\dim(\bar{y}) = n_y = n - n_x$) are at quasi-steady-state with respect to the instantaneous values of the remaining state variables (which we denote by $\bar{x}$, with $\dim(\bar{x}) = n_x$). Mathematically, this assumption translates into the condition

$$f_{\bar{y}}(\bar{x},\bar{y}) = 0. \tag{9}$$

Here, the dimensionality n_x and the decomposition of z into an n_x-dimensional component $\bar{x}$ and an n_y-dimensional component $\bar{y}$ is to be *performed by the investigator*—typically on the basis of experience stemming from experimental results and possibly also from simulation or analysis of the model. The system of n_y equations in n unknowns collected in (9) constitutes the *QSSA constraining relation* (an approximation to the exact constraining relation), and its set of solutions describes, under generic conditions, an n_x-dimensional manifold called the *QSSA manifold* (an approximation to a SIM). Typically, (9) can be solved for n_y of the state variables, which we denote by y (see also the previous section), to yield the explicit reformulation $y = g_{qssa}(x)$ of the QSSA constraining relation; here, g_{qssa} is a vector function of n_x variables and with n_y components. In geometric terms, the QSSA manifold is the graph of $y = g_{qssa}(x)$, and we say that the QSSA manifold is *parameterized* by x. (It is often the case that $\bar{y} = y$, *i.e.*, that (9) may be solved for the same variables $\bar{y}$ that are at quasi-steady-state—see also our treatment of the enzyme reaction example below.)

Whenever (9) can be written as $y = g_{qssa}(x)$, one can obtain an approximation to the slow dynamics by substituting this expression into the state equation for x,

$$\dot{x} = f_x\left(x, g_{qssa}(x)\right). \tag{10}$$

This system of n_x ODEs describes the slow dynamics *on* the QSSA manifold and, together with the constraining relation $y = g_{qssa}(x)$, also the approximate state of the system during the partially relaxed phase.

Enzyme kinetics based on QSSA

We now discuss the application of QSSA to the reversible enzyme reaction (1) and demonstrate that the reduced system corresponds to the enzyme kinetic expression for the rate of reversible reactions known as the *reversible Michaelis–Menten equation*. We also identify a parameter regime for which the QSSA produces an inaccurate description of the system dynamics.

Recall the network (1) and the corresponding ODE system (2)–(3),

$$\dot{s} = k_{-1}\,c - k_1\,(e_{tot} - c)\,s \quad \text{and} \quad \dot{c} = (e_{tot} - c)\,(k_1\,s + k_{-2}\,p) - (k_{-1} + k_2)\,c. \tag{11}$$

In living cells, there is often a huge excess of substrate with respect to the total enzyme, and we write $s_0 \gg e_{tot}$. As a result, the concentration c of complex may assume its quasi-steady-state with respect to the initial value of s rapidly, whereas the effect of this process on s is marginal. Following the discussion above, it is natural to set $\bar{x} = s$ and $\bar{y} = c$, so that $n_x = n_y = 1$ and

$$f_{\bar{x}} = f_s = k_{-1}\,c - k_1\,(e_{tot} - c)\,s \quad \text{and} \quad f_{\bar{y}} = f_c = (e_{tot} - c)\,(k_1\,s + k_{-2}\,p) - (k_{-1} + k_2)\,c.$$

The QSSA (9) $f_c = 0$ can be solved for either c (case $x = \bar{x}$, $y = \bar{y}$) or s (case $x = \bar{y}$, $y = \bar{x}$). Here, we follow the conventional, former option to obtain the explicit form

$$c = g_{qssa}(s) = \frac{(k_1\,s + k_{-2}\,p)\,e_{tot}}{k_1\,s + k_{-2}\,p + k_{-1} + k_2} \tag{12}$$

for the QSSA constraining relation. The graph of g_{qssa} in the state space constitutes the QSSA manifold. Substitution from (12) into the first ODE in (11), together with the definitions

$$V_s = k_2\,e_{tot}, \quad V_p = k_{-1}\,e_{tot}, \quad K_s = (k_{-1} + k_2)/k_1, \quad \text{and} \quad K_p = (k_{-1} + k_2)/k_{-2}, \tag{13}$$

yields the reversible Michaelis–Menten form

$$\dot{s} = -\frac{\frac{V_s}{K_s}s - \frac{V_p}{K_p}p}{1 + \frac{s}{K_s} + \frac{p}{K_p}}. \tag{14}$$

This is the QSSA-reduced system (10) for the model (1).

In Fig. 2, we have plotted the QSSA manifolds given by (12) together with the time evolution of s and c, computed numerically using (11), for various initial conditions and for three different total enzyme concentrations. When the substrate concentration is much larger than the total enzyme concentration, as in Fig. 2A, the trajectories approach a curve which is virtually indistinguishable from the QSSA manifold, as expected. When the total enzyme concentration is comparable to or even higher than that of the substrate, as in Figs. 2B and 2C, respectively, the timescale separation is smaller but still sufficient to drive the trajectories onto a SIM. In those cases, the QSSA manifolds are poor approximations to the SIMs which are outlined by trajectories; this is to be expected, since the condition $s_0 \gg e_{tot}$ does not hold anymore. In what follows, we will see that the ZDP produces a more accurate approximation of the SIM than the QSSA manifold.

The zero-derivative principle

Here, we introduce the ZDP as an accurate generalization of the QSSA. The ZDP manifold of order m—where m can take the values $0, 1, 2, \ldots$—is defined to be the set of points that satisfy the algebraic condition

$$\frac{d^{m+1}\bar{y}}{dt^{m+1}} = 0 \tag{15}$$

and denoted by ZDP_m. As was the case with the QSSA, $\bar{y}$ denotes variables that can be assumed to be in partial relaxation, that is, variables which evolve in a fast timescale. The time derivative in the ZDP condition (15) is calculated using (7), so that this condition becomes

$$0 = \frac{d\bar{y}}{dt} = f_{\bar{y}} \quad \text{for} \quad m = 0, \tag{16}$$

$$0 = \frac{d^2\bar{y}}{dt^2} = \frac{\partial f_{\bar{y}}}{\partial \bar{x}} f_{\bar{x}} + \frac{\partial f_{\bar{y}}}{\partial \bar{y}} f_{\bar{y}} \quad \text{for} \quad m = 1, \tag{17}$$

and similarly for higher values of m, see also the Supporting information.

Plainly, the QSSA manifold and $\mathbf{ZDP}_0$ coincide, as the conditions (9) and (15)–(16) defining them are identical: the QSSA and the zeroth order ZDP yield the same approximate constraining relation. The ZDP manifolds of higher orders, in turn, do not coincide with the QSSA manifold in general—for example, $\mathbf{ZDP}_1$ generally differs from the QSSA manifold because of the presence of the first term in the right-hand side of (17). Instead, the ZDP conditions of higher orders are natural *extensions* of the QSSA: they, also, yield a system (15) of algebraic equations, and the $\mathbf{ZDP}_m$ is the locus of points satisfying them. The sole difference between the two approaches is that the ZDP replaces the first order time derivative employed by the QSSA with *higher order* time derivatives, see (15).

Although technically more involved, this approach has proven to perform well; in fact, the sequence of manifolds $\mathbf{ZDP}_0$, $\mathbf{ZDP}_1$, $\ldots$ limits to a SIM and hence serves to approximate an exact constraining relation with *arbitrary* accuracy [31]. To gain insight into this result, we recall that a SIM is the locus of points where system evolution is slow: the time derivatives of *all* orders of the state variables are small. On the QSSA manifold, $d\bar{y}/dt = 0$—nevertheless, the higher order time derivatives remain large on it. On $\mathbf{ZDP}_1$, in turn, $d^2\bar{y}/dt^2 = 0$ and, additionally, $d\bar{y}/dt$ is small—higher order derivatives are, here also, large. More generally, $d^{m+1}\bar{y}/dt^{m+1}$ is identically zero on $\mathbf{ZDP}_m$ and $d\bar{y}/dt, \ldots, d^m\bar{y}/dt^m$ are small on it, as long as the variables $\bar{y}$ evolve on a fast timescale and the matrix $\partial f_{\bar{y}}/\partial \bar{y}$ appearing in (17) is non-singular [23, 31]. Since the $\mathbf{ZDP}_m$ with $m > 1$ achieves to bound more time derivatives than the QSSA manifold, it is also typically closer to a SIM. Alternatively, each time differentiation of a solution to (6) amplifies its fast component, and hence higher order ZDP conditions filter out these fast dynamics to successively higher orders: points satisfying these conditions yield solutions with fast components of smaller magnitude, *i.e.*, these points lie closer to a SIM.

In biochemical terms, and focusing on our enzyme kinetics example to add concreteness to our exposition, if substrate is injected into an enzyme assay at time zero, one observes a rapid binding of substrate to enzyme; accordingly, the concentration c of complex increases rapidly. Subsequently, both c and the concentration s of the injected substrate decreases very slowly in time: it is this second phase that our simplified enzyme kinetics should describe accurately. As the change in c is slow compared to that during the initial transient, the most straightforward approach would be to neglect it—the SIM is then approximated by requiring c to be constant, $dc/dt = 0$. This approach corresponds to the zeroth order ZDP approach, which is iden-

tical to the well-known QSSA approach, and it cannot be exact as c *does* change—albeit slowly. The first order ZDP assumption is similar to that underlying QSSA: here, c is allowed to change in time, albeit at a constant rate of change—*i.e.*, it is the time derivative of $v_1 - v_2$ that is set to zero, $d(v_1 - v_2)/dt = d^2c/dt^2 = 0$. This assumption is also inexact, since it leads to linear temporal decay; nevertheless, it is more realistic than the QSSA, since the temporal evolution of $v_1 - v_2$ is slower (compared to its evolution over the initial transient) than that of c. This is precisely the amplification effect mentioned above, and it is plain to see in Fig. 3: as e_{tot} increases, the change in $v_1 - v_2$ during the fast transient becomes larger than that during the slow phase by whole orders of magnitude. A similar reasoning applies to higher order ZDP conditions.

When enzyme kinetics is analyzed in intact systems, the dynamic scenario will be more complex. Still higher order ZDP approaches can be expected to be closer to the true behavior than lower order ZDPs.

Accurate enzyme kinetics based on ZDP

In this section, we apply the first order ZDP to our enzyme reaction example (1) and derive the corresponding rate law which is comparable to the reversible Michaelis–Menten (14) albeit more accurate. Then, we demonstrate that the ZDP-reduced model remains accurate even when the QSSA-reduced model fails.

Recalling (11) and (17), we find that the condition defining ZDP_1 becomes

$$\frac{d^2c}{dt^2} = -v_1\frac{\partial(v_1 - v_2)}{\partial s} + (v_1 - v_2)\frac{\partial(v_1 - v_2)}{\partial c} = 0, \tag{18}$$

where v_1 and v_2 are given in (3). This equation can be solved for either s or c; we choose the latter so as to express c as a function of s (here again, then, $x = \bar{x}$ and $y = \bar{y}$, see also (12)). A tedious but direct calculation using (3) shows that (18) can be written in the quadratic form $\alpha(s)\,c^2 - \beta(s)\,c + \gamma(s) = 0$ where

$$\begin{aligned}
\alpha(s) &= k_1(k_1 s + k_{-1}),\\
\beta(s) &= (k_1 s + k_{-1} + k_2 + k_{-2}p)^2 + k_1 e_{tot}(2k_1 s + k_{-1}),\\
\gamma(s) &= k_1^2 e_{tot}^2 s + e_{tot}(k_1 s + k_{-2}p)(k_1 s + k_{-1} + k_2 + k_{-2}p).
\end{aligned} \tag{19}$$

The solutions to $\alpha(s)\,c^2 - \beta(s)\,c + \gamma(s) = 0$ are given by the standard formula $c_\pm(s) = [\beta(s) \pm \sqrt{[\beta(s)]^2 - 4\alpha(s)\gamma(s)}]/(2\alpha(s))$. The solution c_+, associated with the plus sign, is an artifact of the method and it must be discarded as it

does not admit physical interpretation. Indeed, the steady state (s^*, c^*) does not belong to this solution. Also, for large s, one can show that $c_+(s) \approx s$ and thus also $c > e_{tot}$; plainly, this is impossible since the concentration of enzyme bound in substrate cannot exceed that of the total enzyme. The solution c_- associated with the minus sign, on the other hand, can be recast in the form

$$c_- = g_{zdp_1}(s) = R_1(s)\,\frac{e_{tot}\,(k_1\,s + k_{-2}\,p)}{k_1\,s + k_{-1} + k_2 + k_{-2}\,p}, \tag{20}$$

where

$$R_1(s) = \frac{1 + \frac{k_1^2 e_{tot} s}{(k_1 s + k_{-2} p)\,(k_1 s + k_{-1} + k_2 + k_{-2} p)}}{1 + \frac{k_1 e_{tot}\,(2k_1 s + k_{-1})}{(k_1 s + k_{-1} + k_2 + k_{-2} p)^2}}\;\frac{2}{\left[1 + \sqrt{1 - 4\,\alpha(s)\,\gamma(s)/[\beta(s)]^2}\right]}.$$

The rightmost factor on the right-hand side of (20) is precisely the expression for the QSSA manifold, see (12). The coefficient $R_1(s)$, on the other hand, assumes moderate values and is close to 1 at large values of s, so that ZDP_1 lies close to the QSSA manifold for large s—this is plainly visible in Fig. 4. This figure also shows that, in the region where the two manifolds differ significantly, the former better approximates a SIM than the latter, as evidenced by the trajectories approaching it. When the enzyme concentration exceeds that of the substrate, the two manifolds differ by a factor as large as 4.1, see lower panel of Fig. 4B.

To obtain the reduced model corresponding to ZDP_1, we substitute from (20) into the first ODE in (11) and obtain

$$\dot{s} = -\frac{\frac{V_s}{K_s} s - \frac{V_p}{K_p} p + (R_1(s) - 1)\left[\left(\frac{s}{K_s} + \frac{p}{K_p} + \frac{V_p}{V_s}\left(1 + \frac{s}{K_s} + \frac{p}{K_p}\right)\right)\frac{V_s}{K_s} s - \frac{V_p}{K_p} p\right]}{1 + \frac{s}{K_s} + \frac{p}{K_p}}, \tag{21}$$

with $V_s, \ldots, K_p$ expressed in terms of $k_1, \ldots, k_{-2}$ via the parameter change (13). This is the precise analogue of (14). In Fig. 5, we have plotted the curves $(s, -\dot{s})$ corresponding to these two reduced equations against that corresponding to a simulation of the full mass action kinetic model (11). Plainly, the ZDP-derived reduced model performs better than the QSSA-derived one. In particular, the latter *overestimates* the decay rate $-\dot{s}$, an artifact which we now proceed to explain. First, in reality, c decreases ($\dot{c} < 0$) during the slow timescale; contrast this to the QSSA, $\dot{c} = 0$. Now, (11) reads

$$\dot{c} = e_{tot}\,(k_1\,s + k_{-2}\,p) - (k_1\,s + k_{-1} + k_2 + k_{-2}\,p)\,c,$$

and thus $\dot{c}$ decreases with c. Hence, to sustain the inequality $\dot{c} < 0$ during the partially relaxed phase, the *actual* partially equilibrated value $c = g(s)$ must be *higher* than the value $c = g_{qssa}(s)$ predicted by the QSSA and satisfying $\dot{c} = 0$. (Recall that g corresponds to the exact constraining relation.) In other words, the QSSA underestimates c, see also Fig. 4. Now, the ODE for s in (11) reads

$$-\dot{s} = k_1\, e_{tot}\, s - (k_{-1} + k_1\, s)\, c,$$

and hence $-\dot{s}$ decreases with c. Therefore, $-\dot{s}$ assumes a higher value if $c = g_{qssa}(s)$ is used instead of the exact $c = g(s)$, as Fig. 5 also shows. Naturally, the first order ZDP, $d^2c/dt^2 = 0$, is also inexact; nevertheless, Fig. 5 shows that it remains valid for modest timescale separations.

It became evident from this example that the analytic expressions for the approximate constraining relations provided by ZDP become increasingly complex as m increases. Additionally, since the number of equations in (15) equals $n_y < n$, and since n is much larger than 2 for most biochemical systems, one might wish to set $n_y > 1$, *i.e.*, eliminate *several* state variables. Such an elimination yields a *system* of nonlinear algebraic equations; analytic solutions of such systems are typically unattainable. Hence, high values of m and/or n_y imply that analytical solutions of (15) may be prohibitively complex or even unavailable. The obvious alternative to an analytical solution is a numerically computed approximation of it. In the next section, we demonstrate a method to calculate ZDP manifolds numerically.

ZDP for the phosphotransferase system in bacteria

In this section, we calculate numerically the one-dimensional ZDP_0 and ZDP_1 manifolds for the phosphotransferase system (PTS) as modeled in [8]. The PTS is a signal transduction pathway in enteric bacteria regulating the uptake of carbon sources and, in addition, it catalyzes the uptake of glucose. The model in [8] has thirteen state variables and all reaction rates are described by mass action kinetics. The reaction network is depicted in Fig. 6, with further details given in Materials and methods.

Calculation of ZDP manifolds for the PTS model

As preparation for the application of ZDP, we first identify all four conservation relations for our model corresponding to the conserved total concentrations of the

four proteins involved. This allows us to eliminate four state variables without any trade-off and in this way reduce the dimensionality of the state space to nine ($n = 9$), see Materials and methods.

As we remarked earlier, multiscale systems often possess a hierarchy of SIMs of decreasing dimension, embedded in one another, and corresponding to increasingly longer timescales. Since we aim to demonstrate ZDP, we restrict ourselves to one- and two-dimensional ZDP manifolds so as to be able to plot them. A simple timescale analysis using the eigenvalues of the Jacobian at the steady state shows that there is a considerable timescale difference between the least negative eigenvalue λ_1 and the second least negative eigenvalue λ_2 (in particular, $\lambda_2/\lambda_1 \approx 5.1$, see Materials and methods). By contrast, $\lambda_3/\lambda_2 \approx 1.5$ for the second and third least negative eigenvalues, and thus the corresponding timescale difference is relatively small. These calculations suggest, first, the existence of a one-dimensional SIM corresponding to the slowest timescale and, second, that the next manifold in the hierarchy is at least three-dimensional and thus not depictable. For these reasons, we focus on one-dimensional manifolds—that is, $n_x = 1$, and $n_y = 8$. We remark here that, first, more reliable methods to assess timescale disparities do exist and should be employed as needed (see also Supporting information); second, this timescale analysis is only valid locally. To address this latter issue, the timescale disparity could be monitored as the SIM is being tabulated.

Having settled on the dimensionality of the SIMs to be investigated, the investigator must select the single state variable x parameterizing these SIMs as well as the eight state variables constituting $\bar{y}$ which reach a partial equilibrium on a fast timescale and are used to formulate the ZDP conditions (15). Where biochemical intuition is present, it should guide this choice of $\bar{y}$ along the same lines as in the QSSA case; in this example, we identified the choices of $\bar{y}$ yielding manifolds which attract nearby trajectories (and which, then, are good candidates for SIMs). Having investigated also which choices of x lead to a fast tabulation of the SIMs, we settled on $x = \bar{x} = [\mathrm{EIIA} \cdot \mathrm{P}]$ for both ZDP_0 and ZDP_1; hence, $y = \bar{y}$ contains the remaining eight state variables.

Using this choice of x, we tabulate ZDP_0 (equivalently, the QSSA manifold) and ZDP_1 over a grid consisting of **3901** equidistant points on the interval [**0.4 , 39.4**], ***i.e.*** **almost the entire possible range of [EIIA·P] as [EIIA]$_{tot}$=40** (the steady state value of [EIIA·P] is 15.4 μM). For each point x_j on the grid, we solved the eight-dimensional, nonlinear system (15) using the Newton–Raphson method. This calculation over the entire grid takes less than 5 sec in **MATLAB** on an Intel Pentium 4 CPU at 2.80 GHz and with 512 MB of RAM. The algorithm is presented in detail in the Supporting information and our results are shown in

Fig. 7. Plainly, all trajectories approach a SIM and subsequently move along it towards the steady state. Further, the trajectories remain closer to the ZDP_1 than to the QSSA manifold on their way to the steady state—an indication that the former is closer to a SIM than the latter. Using these plots and having measured the concentration of EIIA·P, the investigator can read the **values** of the remaining 8 **concentrations** off the y-axes. For example, a concentration of 25 μM for EIIA·P yields a concentration of approximately 40 μM for HPr·P.

As we remarked earlier, an important assumption underlying both the QSSA and the ZDP_1 is that all variables collected in $\bar{y}$ evolve on a timescale which is fast relative to that of the behavior on the SIM. If this assumption is violated, then both methods yield erroneous results. For example, taking $\bar{x} = x = [\mathrm{EIICB} \cdot \mathrm{P} \cdot \mathrm{Glc}]$, we obtain QSSA and ZDP manifolds which are very bad approximations of a SIM, see Fig. 8. The reason for this is that the concentration of EIIA·P, which is now part of $\bar{y}$, evolves on a slow timescale compared to that of EIICB·P·Glc.

Using the tabulated ZDP_1 manifold to reduce the PTS model

As we saw, an explicit expression for a ZDP manifold—such as (20)—can be used to obtain a lower-dimensional model—such as (21). When a ZDP manifold is only available in tabulated form, though, such a reduced equation cannot be written out explicitly. Nevertheless, one can still employ it in a computational setting, as we now proceed to show.

In the case of the PTS, the reduction to a one-dimensional ZDP_1 effectuates a description of the long-term dynamics by an (analytically unavailable) ODE of the form (8), with g_1 replacing g and with $x = [\mathbf{EIIA} \cdot \mathbf{P}]$. The unknown quantity $g_1(x)$ may be approximated, at any point x in the domain, either by explicitly solving (15) (with $m = 1$) at that point in the way described above or, instead, by first tabulating g_1 over a fine grid and then using this tabulation and an interpolation technique to approximate g_1 at *any* point x. The two major advantages of this reduced ODE over the full ODE system are, first, that its dynamics are one-dimensional and thus transparent and, second, that only the slow timescale is present in it and thus it is both easier and faster to integrate numerically.

To demonstrate the validity of this last statement, we compared the performance of a simple integrator for the ZDP_1-reduced PTS system against that of a state-of-the-art integrator for the full PTS system. Our simple integrator was coded up in MATLAB and is the standard, explicit, fourth-order Runge–Kutta method RK4 [39] coupled with a fine grid of 1001 **points on the**

computational domain for x (which took 5sec to generate in MATLAB) and linear interpolation. The state-of-the-art integrator is MATLAB's implicit, stiff, fully automated integrator `ode23s` [40]. Normally speaking, explicit integrators are *prohibitively* costly when applied to stiff (*i.e.*, multiscale) problems [41]. In this case, though, and depending on the proximity of the initial condition to the steady state, our explicit integrator for the reduced system was between 5 and 25 times faster than the implicit integrator for the full model—a tangible indication of the degree to which the ZDP$_1-$reduced PTS system indeed describes the slow, non-stiff dynamics of the PTS model.

The behavior of the ZDP$_1-$reduced integrator is depicted in Fig. 9. To produce it, we have set the initial value of each state variable to one-half of its steady state value (thus obtaining a point off the SIM) and then used MATLAB's aforementioned stiff integrator to obtain numerically the corresponding trajectory over a time horizon of 50msec. At the same time, we projected this initial condition on ZDP$_1$ and used it to initialize the reduced integrator and obtain the corresponding trajectory for the reduced system. It becomes evident that, following a short transient during which the two solutions differ, the solutions enter a phase where they converge to each other and progress in unison towards the steady state: the reduced model matches the full one once the fast dynamics have been filtered out.

Discussion

The development of biochemical modeling for use in experimental design, drug development, and decipherment of cellular processes has accelerated in the last decade. Due to this, systems biology faces substantial challenges—most notably that of combining a large number of models of cellular processes to produce comprehensive quantitative descriptions of cellular function. Since these models—and hence also the resulting comprehensive descriptions—tend to be complex, the exploration of reduction methods designed to extract the core dynamics pertaining to cellular function is of great interest.

In this article, we have focused on the idea that biochemical systems may be reduced by exploiting the wide range of timescales typically present in them. In biochemistry, the most prominent reduction result is the Michaelis–Menten kinetics derived by employing QSSA to the mass action kinetic description of single enzymatic reactions. The Michaelis–Menten rate laws have proven extremely useful in describing the kinetics of reactions in which the enzyme concentrations are

much lower than those of the substrate.

Such conditions are often encountered in *in vitro* assays and in the many processes *in vivo* in which the substrates are low-molecular weight molecules, as is the case in *e.g.* metabolic pathways. However, as cell biology has developed, attention has shifted away from major metabolic pathways to pathways of gene expression and signal transduction. Hence, the substrates of enzymatic activity are no longer exclusively low-molecular weight substances but, instead, are often macromolecules (such as other enzymes). In certain cases, such as in the autokinase activity of growth factor receptors, the difference between substrate and enzyme is blurred. By consequence, the vast separation in the concentrations of enzymes and substrates disappears. This is also the case with enzymes acting on polynucleotides (such as DNA gyrase and ribosomes), where the concentrations of enzymes and binding sites are often of the same order of magnitude. In all of these cases, the accuracy of **Michaelis–Menten** kinetics is unsatisfactory due to the small timescale separation. Particular examples where the QSSA fails include the signal transduction routes such as the MAP kinase cascade, the EGF receptor transphosphorylation upon dimerization, and the regulation of processes through sequestration [42].

For mechanisms such as those mentioned above, where enzyme and substrate concentrations are comparable, modeling approaches offering higher accuracy are called for. Several approaches to develop rate laws for such cases have been taken. Specifically, for the example of phosphorylation cycles, the rapid-equilibrium approximation has been employed to derive such laws [43]. Moreover, several methods extending the QSSA for general biochemical systems have been explored in [20, 21, 22]. In this article, we have introduced a novel generalization of the QSSA for general biochemical systems which is based on the zero-derivative principle (ZDP) and, contrary to these previous attempts, requires little theoretical work.

We derived the rate expression based on the (first order) ZDP for a reversible enzyme-catalyzed reaction, and we compared it to the corresponding Michaelis–Menten rate law. We showed that these two expressions match except for an additional multiplicative factor present in the ZDP description and absent from the QSSA one. This factor compensates, to a very large extent, for the fact that the concentration of the enzyme-substrate complex changes with time instead of remaining constant as the QSSA dictates. In cases of vast timescale separation, this factor is close to one and thus inconsequential. For modest timescale separations, however, this factor comes into play and renders the first order ZDP approximation considerably more accurate than the QSSA. We therefore expect that the novel kinetic description developed here will be useful in the many cases dis-

cussed above, *i.e.*, when the concentration levels of enzymes and their substrates are comparable.

To illustrate the **usefulness** of ZDP in cases where analytic expressions cannot be derived (as is typically the case already for systems of any complexity), we used it to perform the same task in a numerical setting and for the PTS model (which has a total of nine state variables). Using a standard numerical procedure, we computed the first order ZDP approximation for this model and demonstrated its superior accuracy. This study shows that the nine-dimensional PTS model behaves as a one-dimensional system in the slowest and most relevant timescale: tracing the evolution of [EIIA·P] suffices for understanding the behavior of the system as a whole over that timescale. **We then also showed that calculation of time courses by numerical integration based on the $\mathbf{ZDP}_1$ manifold is between 5 and 25 times faster than using a standard stiff integrator in MATLAB to integrate the nine-dimensional PTS model, depending on the initial condition used.**

An important reason for the ubiquitous use of enzyme kinetics of QSSA type has been its mathematical simplicity. In contrast, the ZDP methodology is quite complex and often defies analytical solutions, as is the case for the PTS model above, for example. In the past, this analytic intractability would have detracted greatly from the use of the method. Presently, however, the utilization of numerical mathematics in biochemistry has become so much more frequent that this limitation is retreating while the importance of accurate modeling and analysis approaches aiming at understanding the complex interactions in living cells is increasing.

Materials and methods

The PTS model

Here we present the PTS model [8] in detail, list the linear conservation relations associated with it, and report numerical data related to the dimensionality and the choice of a parameterizing variable x for the ZDP manifolds.

The PTS is a mixed signal transduction and transport pathway involved in transporting various sugars into enteric bacteria, and the model we consider here deals particularly with glucose uptake. The source of free energy in this pathway is the phosphate group on phosphoenolpyruvate (PEP) which can be translocated by successive phosphorylations of pyruvate (Pyr), enzyme I (EI), histidine protein

(HPr), enzyme IIA (EIIA), enzyme IICB (EIICB), and finally glucose (Glc), see Fig. 6. The last phosphorylation prevents the glucose transporter from recognizing it and in this manner enables further glucose import into the cell. Consequently, the PTS enables the cell to maintain a glucose concentration gradient through the membrane. The PTS also regulates the uptake of various carbon sources depending on their availability, a phenomenon known as *carbon catabolite repression*. The model we consider, however, focuses on the uptake of the most common carbon source, *i.e.* glucose, and hence does not deal with this particular regulation.

The model in [8] (*original model*) has 13 state variables representing concentrations of macromolecules; these are listed in Table I. The dynamics of the model is determined by the ODE system $\dot{\tilde{z}} = \tilde{N}v(\tilde{z})$, where the 13×10 stoichiometric matrix $\tilde{N}$ and the 10 reaction rates collected in $v(\tilde{z})$ are given in Table II. For the values of the kinetic parameters and of the constant concentrations, we refer the reader to the original article and remark that we used the values determined *in vivo* for the latter. All concentrations are measured in micromolar (μM) and time is measured in minutes. Since $\tilde{N}$ is of rank 9, there are 4 linear conservation relations. These can be determined as in [44], and they express mathematically the fact that the total concentration of each one of the four proteins is conserved. In particular, they are

$$\begin{array}{rclrcl} [\mathrm{EI}]_{tot} &=& \tilde{z}_1+\tilde{z}_2+\tilde{z}_6+\tilde{z}_7, & [\mathrm{HPr}]_{tot} &=& \tilde{z}_2+\tilde{z}_3+\tilde{z}_8+\tilde{z}_9, \\ [\mathrm{EIIA}]_{tot} &=& \tilde{z}_3+\tilde{z}_4+\tilde{z}_{10}+\tilde{z}_{11}, & [\mathrm{EIICB}]_{tot} &=& \tilde{z}_4+\tilde{z}_5+\tilde{z}_{12}+\tilde{z}_{13}. \end{array} \tag{1}$$

Using these, we can express $\tilde{z}_6$, $\tilde{z}_8$, $\tilde{z}_{10}$, and $\tilde{z}_{12}$ in terms of the remaining state variables, substitute these expressions in the original model, and obtain the nine-dimensional ODE system $\dot{z} = Nv(z)$ (*final model*). Here, z is the vector of the new state variables (see Table I) and the 9×10 stoichiometric matrix N is obtained from $\tilde{N}$ by deleting its sixth, eighth, tenth, and twelfth rows. We also note that we used the *in vivo* values from [8] for the conserved moieties collected in (1).

To select the dimension of the SIM and apply ZDP to the PTS system, we used the eigenvalues of the Jacobian $N\partial v(z)/\partial z|_{z^*}$, where $z^* = (3.05, 0.49, 18.45, 5.47, 2.99, 1.19, 29.78, 15.44, 0.12)$ is the steady state. Here, we report these eigenvalues: $\lambda_1 = -1732$, $\lambda_2 = -8851$, $\lambda_3 = -14109$, $\lambda_4 = -68060$, $\lambda_5 = -115750$, $\lambda_6 = -131042$, $\lambda_7 = -379063$, $\lambda_8 = -690343$, and $\lambda_9 = -6070420$.

Acknowledgements

This research was funded by the Netherlands Organisation for Scientific Research, NWO (project number NWO/EW/635.100.007 for H. M. H. and H. V. W. and NWO/EW/639.031.617 for A. Z.) and by the BBSRC. For more information about further funding, see http://www.bio.vu.nl/microb/.

References

[1] Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA & Sorger PK (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol Syst Biol* **5**, 239.

[2] Nordling TEM, Hiroi N, Funahashi A & Kitano H (2007) Deduction of intracellular sub-systems from a topological description of the network. *Mol BioSyst* **3**, 523–529.

[3] Westerhoff HV & Palsson BO (2004) The evolution of molecular biology into systems biology. *Nat Biotech* **22**, 1249–1252.

[4] Westerhoff HV & Kell DB (2007) The methodologies of systems biology. In *Systems biology philosophical foundations* (Boogerd FC, Bruggeman FJ, Hofmeyr J-HS & Westerhoff HV, eds), pp. 23–70. Elsevier, Amsterdam.

[5] Bakker BM, Michels PAM, Opperdoes FR & Westerhoff HV (1997) Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. *J Biol Chem* **272**, 3207–3215.

[6] Heinrich R & Schuster S (1996) *The regulation of cellular systems*. Chapman and Hall, New York.

[7] Leloup J-C & Goldbeter A (2003) Toward a detailed computational model for the mammalian circadian clock. *PNAS* **100**(12), 7051–7056.

[8] Rohwer JM, Meadow ND, Roseman S, Westerhoff HV & Postma PW (2000) Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements *in vitro*. *J Biol Chem* **275**(10), 34909–34921.

[9] Teusink B, Walsh MC, van Dam K & Westerhoff HV (1998) The danger of metabolic pathways with turbo design. *Trends Biochem Sci* **23**, 162–169.

[10] Westerhoff HV, Kolodkin A, Conradie R, Wilkinson SJ, Bruggeman FJ, Krab K, van Schuppen JH, Hardin H, Bakker BM, Moné MJ, Rybakova KN, Eijken M, van Leeuwen HJ & Snoep JL (2009) Systems biology towards life in silico: mathematics of the control of living cells. *J Math Biol* **58**(1–2), 7–34.

[11] Vallabhajosyula RR & Sauro HM (2006) Complexity Reduction of Biochemical Networks. *Proc Wint Sim Conf* 1690–1697.

[12] Surovtsova I, Sahle S, Pahle J & Kummer U (2006) Approaches to complexity reduction in a systems biology research environment (SYCAMORE). *Proc Wint Sim Conf* 1683–1689.

[13] Teusink B *et al* (2000) Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* **267**, 5313–5329.

[14] Bodenstein M (1913) Eine Theorie der photochemischen Reaktionsgeschwindigkeiten. *Z physik Chem* **85**, 329–397.

[15] Bodenstein M & Lütkemeyer H (1924) Die photochemische Bildung von Bromwasserstoff und die Bildungsgeschwindigkeit der Brommolekul aus den Atomen. *Z physik Chem* **114**, 208–236.

[16] Briggs GE & Haldane JB (1925) A Note on the Kinetics of Enzyme Action. *Biochem J* **19**, 338–339.

[17] Segel LA & Slemrod M (1989) The quasi-steady-state assumption: A case study in perturbation. *SIAM Review* **31**, 446–477.

[18] Bowen JR, Acrivos A & Oppenheim AK (1963) Singular perturbation refinement to quasi-steady state approximation in chemical kinetics. *Chem Eng Sci* **18**, 177–188.

[19] Turányi T, Tomlin AS & Pilling MJ (1993) On the error of the quasi-steady-state approximation. *J Phys Chem* **97**, 163–172.

[20] Frenzen CL & Maini PK (1988) Enzyme kinetics for a two-step enzymic reaction with comparable initial enzyme–substrate ratios. *J Math Biol* **26**, 689–703.

[21] Segel LA (1988) On the validity of the steady state assumption of enzyme kinetics. *Bull Math Biol* **50**, 579–593.

[22] Schneider KR & Wilhelm T (2000) Model reduction by extended quasi-steady-state approximation. *J Math Biol* **40**, 443–450.

[23] Gear CW, Kaper TJ, Kevrekidis IG & Zagaris A (2005) Projecting to a slow manifold: Singularly perturbed systems and legacy codes. *SIAM J Appl Dyn Syst* **4**, 711–732.

[24] Zagaris A, Kaper HG & Kaper TJ (2005) Two perspectives on reduction of ordinary differential equations. *Mathematische Nachrichten* **278**(12-13), 1629–1642.

[25] Browning G & Kreiss H-O (1982) Problems with different time scales for nonlinear partial differential equations. *SIAM J Appl Math* **42**(4), 704–718.

[26] Kreiss H-O (1979) Problems with different time scales for ordinary differential equations. *SIAM J Numer Anal* **16**(6), 980–998.

[27] Kreiss H-O (1985) Problems with different time scales. In *Multiple Time Scales* (Brackbill, J. H. and Cohen, B. I., eds), pp. 29–57. Academic Press, New York.

[28] Lorenz EN (1980) Attractor sets and quasi-geostrophic equilibrium. *J Atmos Sci* **37**, 1685–1699.

[29] van Leemput P, Vanroose W & Roose D (2007) Mesoscale analysis of the equation-free constrained runs initialization scheme. *Multiscale Model Simul* **6**(4), 1234–1255.

[30] Vandekerckhove C, van Leemput P & Roose D (2008) Accuracy and stability of the coarse time-stepper for a lattice Boltzmann model. *J Algor Comp Tech* **2**(2), 249–273.

[31] Zagaris A, Gear CW, Kaper TJ & Kevrekidis IG (2008) Analysis of the accuracy and convergence of equation-free projection to a slow manifold, *ESAIM: M2AN*, in press.

[32] Zagaris A, Vandekerckhove C, Gear CW, Kaper TJ & Kevrekidis IG (2008) Stability and stabilization of the constrained runs schemes for equation-free projection to a slow manifold. Submitted.

[33] Gorban AN & Karlin IV (2005) *Invariant Manifolds for Physical and Chemical Kinetics*. Springer, Berlin.

[34] Reich JG & Sel'kov EE (1981) *Energy metabolism of the cell.* Academic Press, Bristol, UK.

[35] Kaper TJ (1999) An introduction to geometric methods and dynamical systems theory for singular perturbation problems. *Proc Sympos Appl Math* **56**, 85–131.

[36] O'Malley RE Jr (1991) *Singular Perturbation Methods for Ordinary Differential Equations.* Springer, New York.

[37] Segel IH (1993) *Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady-state enzyme systems.* Wiley, New York.

[38] Calder MS & Siegel D (2008) Properties of the Michaelis–Menten mechanism in phase space. *J Math Anal Appl* **339**(2), 1044–64.

[39] Hairer E, Nørsett SPN & Wanner G (1987) *Solving ordinary differential equations I: Nonstiff problems.* Springer, Berlin.

[40] Shampine LF & Reichelt MW (1997) The MATLAB ODE suite. *SIAM J Sci Comput* **18**(1), 1–22.

[41] Hairer E & Wanner G (1991) *Solving ordinary differential equations II: Stiff and differential-algebraic problems.* Springer, Berlin.

[42] Blüthgen N, Bruggeman FJ, Legewie S, Herzel H, Westerhoff HV & Kholodenko BN (2006) Effects of sequestration on signal transduction cascades. *FEBS J* **273**(5), 895–906.

[43] Salazar C & Höfer T (2000) Kinetic models of phosphorylation cycles: A systematic approach using the rapid-equilibrium approximation for protein–protein interactions. *Biosystems* **83**(2-3), 195–206.

[44] Reder C (1988) Metabolic control theory: A structural approach. *J Theor Biol* **135**(2), 175–201.

Supporting information

The following supplementary material is available online:
Doc S1. An algorithm for the numerical calculation of constraining relations based on the zero-derivative principle.

This material is available as part of the online article from http://blackwell-synergy.com

Tables

Compound	O	F	Compound	O	F	Compound	O	F
EI·P·Pyr	$\tilde{z}_1$	z_1	EI	$\tilde{z}_6$	–	EIIA	$\tilde{z}_{10}$	–
EI·P·HPr	$\tilde{z}_2$	z_2	EI·P	$\tilde{z}_7$	z_6	EIIA·P	$\tilde{z}_{11}$	z_8
HPr·P·EIIA	$\tilde{z}_3$	z_3	HPr	$\tilde{z}_8$	–	EIICB	$\tilde{z}_{12}$	–
EIIA·P·EIICB	$\tilde{z}_4$	z_4	HPr·P	$\tilde{z}_9$	z_7	EIICB·P	$\tilde{z}_{13}$	z_9
EIICB·P·Glc	$\tilde{z}_5$	z_5						

Table I: The state variables of the original (O) and the final (F) model.

$v_1 = k_{1f}\tilde{z}_6[\mathrm{PEP}] - k_{1r}\tilde{z}_1$	$v_4 = k_{4f}\tilde{z}_2 - k_{4r}\tilde{z}_6\tilde{z}_9$	$v_7 = k_{7f}\tilde{z}_{11}\tilde{z}_{12} - k_{7r}\tilde{z}_4$
$v_2 = k_{2f}\tilde{z}_1 - k_{2r}\tilde{z}_7[\mathrm{Pyr}]$	$v_5 = k_{5f}\tilde{z}_9\tilde{z}_{10} - k_{5r}\tilde{z}_3$	$v_8 = k_{8f}\tilde{z}_4 - k_{8r}\tilde{z}_{10}\tilde{z}_{13}$
$v_3 = k_{3f}\tilde{z}_7\tilde{z}_8 - k_{3r}\tilde{z}_2$	$v_6 = k_{6f}\tilde{z}_3 - k_{6r}\tilde{z}_8\tilde{z}_{11}$	$v_9 = k_{9f}\tilde{z}_{13}[\mathrm{Glc}] - k_{9r}\tilde{z}_5$
		$v_{10} = k_{10f}\tilde{z}_5 - k_{10r}\tilde{z}_{12}[\mathrm{Glc{\cdot}P}]$

$$
\begin{aligned}
1 &= \tilde{N}[1,1] = \tilde{N}[2,3] = \tilde{N}[3,5] = \tilde{N}[4,7] = \tilde{N}[5,9] = \tilde{N}[6,4] = \tilde{N}[7,2] \\
&= \tilde{N}[8,6] = \tilde{N}[9,4] = \tilde{N}[10,8] = \tilde{N}[11,6] = \tilde{N}[12,10] = \tilde{N}[13,8], \\
-1 &= \tilde{N}[1,2] = \tilde{N}[2,4] = \tilde{N}[3,6] = \tilde{N}[4,8] = \tilde{N}[5,10] = \tilde{N}[6,1] = \tilde{N}[7,3] \\
&= \tilde{N}[8,3] = \tilde{N}[9,5] = \tilde{N}[10,5] = \tilde{N}[11,7] = \tilde{N}[12,7] = \tilde{N}[13,9].
\end{aligned}
$$

Table II: Reaction rates (top) and nonzero entries of the matrix $\tilde{N}$ (bottom). The model contains four boundary metabolite concentrations taken to be constant ([PEP], [Pyr], [Glc], and [Glc·P]) and twenty kinetic parameter values (k_{1f}, …, k_{10f} and k_{1r}, …, k_{10r}).

Legends

Figure 1: Graph of the $(s,c)-$plane for (2)–(3) with several trajectories corresponding to different initial conditions (round dots) and the steady state $(s^*,c^*) = (0.003,0.0043)$ (square dot). The rate constants here are $k_1 = 1.833$, $k_{-1} = 0.25$, $k_2 = 2.5$, and $k_{-2} = 0.55$, while $e_{tot} = 0.2$ and $p = 0.1$.

Figure 2: Trajectories of the system (11) together with QSSA manifolds (12). The parameter values of k_1, k_{-1}, k_2, k_{-2}, and p are the same as in Fig. 1 and the total enzyme concentration is $e_{tot} = 0.2$ **in panel A,** $e_{tot} = 4$ **in B, and** $e_{tot} = 40$ **in C.**

Figure 3: The time evolution of c and $\dot{c}$ for the system (11). The parameter values of k_1, k_{-1}, k_2, k_{-2}, and p and the total enzyme concentrations in each panel are the same as in Fig. 2.

Figure 4: Upper panels: trajectories of the system (11) together with the ZDP_1 (20) and the QSSA (12) manifolds; parameter values in A and B are as in Fig. 2B and 2C, respectively. Lower panels: the ratio $g_{zdp_1}(s)/g_{qssa}(s)$ for the corresponding parameter sets.

Figure 5: The curves $(s(t),-\dot{s}(t))$ given by the mass action kinetic model (11) (solid line), the QSSA-reduced model (14) (dotted line), and the ZDP_1-reduced model (21) (dashed line); the initial condition used was $s(0) = 4$ for the latter two systems and for the former system the additional initial condition used was $c(0) = 33.6$ (*i.e.*, the initial point is close to the SIM). The parameter values are the same as in Fig. 4B.

Figure 6: Reaction scheme for the PTS. The concentrations of the molecules depicted in boxes are the state variables in the model [8] while the concentrations of the remaining molecules are modeled as constants. Molecular names containing dots correspond to molecular complexes and P's denote phosphate groups. For explanations of the molecules involved, see Materials and methods.

Figure 7: QSSA manifold (dashed), ZDP_1 (solid black), and trajectories (solid gray) for the PTS model. The one-dimensional manifolds are embedded in the nine-dimensional state space and are therefore depicted in eight plots: in each one of these, one of the state variables collected in y is plotted against the parameterizing variable $x =$[EIIA·P]. One of the plots is enlarged to show more detail. The steady state is indicated, in each plot, by a black dot.

Figure 8: The analogue of Fig. 7 for $x = \bar{x} =$[EIICB·P·Glc]. Three of the eight plots are shown.

Figure 9: **Time courses of the scaled state variables** $(z_i - z_i^*)/z_i^*$**, with** $i = 1,\ldots,9$**, obtained by integrating the nine-dimensional PTS model with MATLAB's** `ode23s` **ODE suite (solid curves) and with the time measured in msec. The dashed curves are the corresponding solutions to the one-dimensional,** $\mathbf{ZDP}_1$**-reduced model. The initial condition for all scaled variables was set to** 0.5 **(one half of the steady state, in terms of the unscaled variables** z_i**) and all time trajectories approach zero (as the unscaled variables tend to the steady state).**