BIOINFORMATICS ORIGINAL PAPER Vol. 21 no. 6 2005, pages 774–780
doi:10.1093/bioinformatics/bti068

Systems biology

An enzyme mechanism language for the mathematical modeling of metabolic pathways

Chin-Rang Yang[1,3], Bruce E. Shapiro[4], Eric D. Mjolsness[2,3] and G. Wesley Hatfield[1,3,*]

[1]Department of Microbiology and Molecular Genetics, College of Medicine, [2]School of Information and Computer Science, University of California, Irvine, CA 92697, USA, [3]Institute for Genomics and Bioinformatics, University of California, Irvine, CA 92697, USA and [4]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Received on February 23, 2004; revised and accepted on September 16, 2004
Advance Access publication October 27, 2004

ABSTRACT

Motivation: As a first step toward the elucidation of the systems biology of complex biological systems, it was our goal to mathematically model common enzyme catalytic and regulatory mechanisms that repeatedly appear in biological processes such as signal transduction and metabolic pathways.

Results: We describe kMech, a Cellerator language extension that describes a suite of enzyme mechanisms. Each enzyme mechanism is parsed by kMech into a set of fundamental association–dissociation reactions that are translated by Cellerator into ordinary differential equations that are numerically solved by Mathematica™. In addition, we present methods that use commonly available kinetic measurements to estimate rate constants required to solve these differential equations.

Availability: A Mathematica™ executable kMech.m file is available at the University of California, Irvine, Institute for Genomics and Bioinformatics website, http://www.igb.uci.edu/servers/sb.html. Cellerator, free of charge to academic, US government, and other non-profit organizations, can be obtained at the Cellerator website, (http://www-aig.jpl.nasa.gov/public/mls/cellerator/feedback.html).

Contact: Biology correspondence should be addressed to gwhatfie@uci.edu. Computation correspondence should be addressed to emj@uci.edu.

Supplementary information: http://www.igb.uci.edu/servers/sb.html.

INTRODUCTION

Systems biology may be broadly defined as the integration of diverse data into useful biological models that allow scientists to easily observe complex cellular behaviors and to predict the outcomes of metabolic and genetic perturbations. In this report we present kMech, a Cellerator (Shapiro *et al*., 2003) language extension that describes a suite of enzyme mechanisms suitable for the mathematical modeling of enzyme-related pathways. Cellerator is a tool for generating reaction network models of cellular processes. It contains reaction mechanisms based on mass action kinetics for elementary catalytic and non-catalytic reactions, along with other regulatory relationships (Shapiro *et al*., 2003). kMech supplements Cellerator with a suite of catalytic reaction mechanisms involving multiple substrates and products and regulatory mechanisms. In kMech, each mechanism has been codified to generate a set of elementary reactions that can be translated by Cellerator into ordinary differential equations (ODEs) solvable by Mathematica™, a widely used commercial computer algebra system that integrates numeric and symbolic computational engines with a graphical output and a programming language. Alternatively, Cellerator can generate ODEs in systems biology markup language (SBML) for other simulators (Hucka *et al*., 2003).

Traditional enzyme modeling approaches, for example, GEPASI (Mendes, 1997) or JARNAC (Sauro *et al*., 2003), use the Michaelis–Menten kinetic equation for one substrate/ one product reactions and the King–Altman method to derive equations for more complex multiple reactant reactions. These types of equations are called steady-state velocity equations since the derivatives of the concentration of each reactant in the model over time are set to zero to simplify a set of non-linear differential equations to linear algebric equations (Falk *et al*., 1998). Therefore, the kinetic model based on this approach has embedded in it the steady-state hypothesis. In contrast, the model generated by kMech/Cellerator consists of non-simplified, non-linear, differential equations that describe the rates of change of each reactant in the model over time. For the solution of these differential equations, users can input the rate constants (k_f, k_r) if they have been experimentally measured. Alternatively, these rate constants can be approximated from easily measured kinetic constants (K_m, k_{cat}) using the Lambda and Omega approximation methods presented in this paper. Moreover, if the assumption of equilibrium between substrate and enzyme is not favored, values of Lambda and Omega can be adjusted to fit the experimental data.

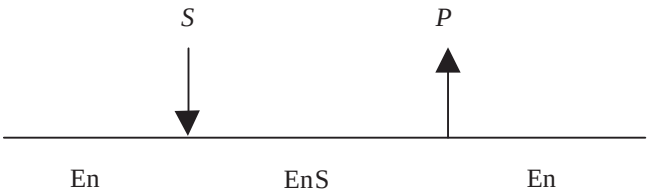
kMech is built in a modular manner such that complex enzyme mechanisms may be constructed from simpler ones. This design is sufficiently versatile to accommodate enzyme mechanisms such as single and multiple substrate (e.g. Ping Pong and Bi Bi) enzyme kinetic reactions, and feedback inhibition (e.g. allosteric, competitive and non-competitive) mechanisms.

*To whom correspondence should be addressed.

SYSTEMS AND METHODS

The development of kMech models for the simulation of enzyme reaction mechanisms

Simple catalytic model The simple catalytic model describes a single-substrate, single-product, non-reversible catalytic mechanism that involves no cofactors or enzyme intermediates, is implemented in Cellerator.



The Cellerator input for this reaction called 'catalytic' is

$$\{S \overset{\text{En}}{\rightleftharpoons} P, k_\text{f}, k_\text{r}, k_\text{cat}\}$$

(Note: $\rightleftharpoons$ is the Cellerator notation for an enzyme-catalyzed reaction.) where S is the substrate, P is the product, En is the free enzyme, EnS is the enzyme–substrate complex, k_f is the rate constant for enzyme–substrate association, k_r is the rate constant for enzyme–substrate dissociation and k_cat is the rate constant for the formation and dissociation of product P.

Cellerator converts this reaction input into its basic association–dissociation reactions (Note: Mathematica syntax for all association–dissociation reactions and differential equations is shown in Supplementary Materials.):

$$S + \text{En} \underset{k_\text{r}}{\overset{k_\text{f}}{\Leftrightarrow}} \text{EnS} \overset{k_\text{cat}}{\rightarrow} \text{En} + P$$

The first reaction represents the formation of the EnS. This reaction is reversible with forward and reverse rate constants, k_f and k_r, respectively. The second reaction represents the catalysis of S to P and the release of En with a rate constant, k_cat. Given these reactions, the Cellerator 'interpret' function uses the law of mass action to generate the following differential equations for each reactant:

$$\begin{aligned}
\frac{\text{d}[S]}{\text{d}t} &= -k_\text{f}[S][\text{En}] + k_\text{r}[\text{EnS}],\\
\frac{\text{d}[P]}{\text{d}t} &= k_\text{cat}[\text{EnS}],\\
\frac{\text{d}[\text{En}]}{\text{d}t} &= -k_\text{f}[S][\text{En}] + k_\text{r}[\text{EnS}] + k_\text{cat}[\text{EnS}],\\
\frac{\text{d}[\text{EnS}]}{\text{d}t} &= k_\text{f}[S][\text{En}] - k_\text{r}[\text{EnS}] - k_\text{cat}[\text{EnS}].
\end{aligned}$$

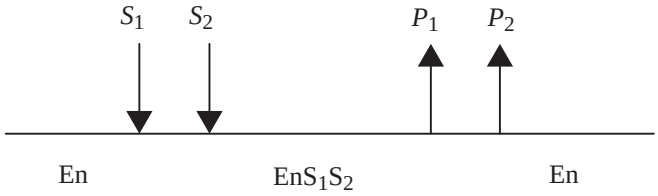
A Lambda approximation method (described in the Lambda approximation method for enzyme rate constants section) and the measured K_m values for each substrate can be used to estimate these rate constants. These differential equations and variable definitions are passed to Mathematica™ where they are solved by its numeric solver (NDSolve) function and time versus concentration plots are generated.

For a reversible catalytic reaction, Cellerator employs two simple catalytic reactions, one for the forward reaction and one for the reverse reaction. For example, the Cellerator input for a reversible simple catalytic reaction is

$$\{S \overset{\text{En}}{\rightleftharpoons} P, k_{\text{f}1}, k_{\text{r}1}, k_{\text{cat}1}\}, \quad \{P \overset{\text{En}}{\rightleftharpoons} S, k_{\text{f}2}, k_{\text{r}2}, k_{\text{cat}2}\}.$$

However, most enzyme mechanisms are more complex and cannot be described by this simple Cellerator encoded catalytic model. Thus, kMech was developed to extend the functionality of Cellerator to model more complex enzyme mechanisms. These models are described below.

Bi Bi model The Bi Bi model describes a generalized two-substrate two-product (Bi Bi) reactions model where product formation occurs only after the formation of an enzyme–two substrate complex.



The kMech input for this Bi Bi mechanism is

$$\text{Enz}[\{S_1, S_2\} \overset{\text{En}}{\rightleftharpoons} \{P_1, P_2\}, \text{Bi Bi}[k_\text{f}, k_\text{r}, k_\text{cat}]]$$

where S_1, S_2 are the substrates, P_1, P_2 are the products, and EnS_1S_2 is the enzyme–two substrate complex. This mechanism is modeled with single forward and reverse rate constants for the association and dissociation of the enzyme–substrate complex. A Lambda approximation method (described in the Lambda approximation method for enzyme rate constants section) and the measured K_m values for each substrate can be used to estimate these rate constants. Enz[...] denotes the kMech-defined function that provides additional capabilities to Cellerator.

kMech converts this input into association–dissociation reactions defined in Cellerator as follows:

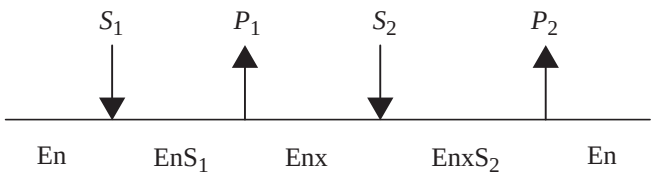
$$S_1 + S_2 + \text{En} \underset{k_\text{r}}{\overset{k_\text{f}}{\Leftrightarrow}} \text{EnS}_1\text{S}_2 \overset{k_\text{cat}}{\rightarrow} \text{En} + P_1 + P_2$$

The first reaction represents the formation of the EnS_1S_2. The second reaction represents the release of products and free enzyme from the complex. Given these reactions, the Cellerator 'interpret' function generates the following differential equations for each reactant:

$$\begin{aligned}
\frac{\text{d}[S_1]}{\text{d}t} &= -k_\text{f}[S_1][S_2][\text{En}] + k_\text{r}[\text{EnS}_1\text{S}_2],\\
\frac{\text{d}[S_2]}{\text{d}t} &= -k_\text{f}[S_2][S_2][\text{En}] + k_\text{r}[\text{EnS}_1\text{S}_2],\\
\frac{\text{d}[P_1]}{\text{d}t} &= k_\text{cat}[\text{EnS}_1\text{S}_2],\\
\frac{\text{d}[P_2]}{\text{d}t} &= k_\text{cat}[\text{EnS}_1\text{S}_2],\\
\frac{\text{d}[\text{En}]}{\text{d}t} &= -k_\text{f}[S_1][S_2][\text{En}] + k_\text{r}[\text{EnS}_1\text{S}_2]\\
&\quad + k_\text{cat}[\text{EnS}_1\text{S}_2],\\
\frac{\text{d}[\text{EnS}_1\text{S}_2]}{\text{d}t} &= k_\text{f}[S_1][S_2][\text{En}] - k_\text{r}[\text{EnS}_1\text{S}_2]\\
&\quad - k_\text{cat}[\text{EnS}_1\text{S}_2].
\end{aligned}$$

These differential equations and variable definitions are passed to Mathematica™ where they are solved by the NDSolve function and time versus concentration plots are generated.

Ping Pong Bi Bi model The Ping Pong Bi Bi model describes a specialized Bi Bi mechanism in which the binding of substrates and release of products is ordered. It is a Ping Pong mechanism because the enzyme shuttles between a free and a substrate-modified intermediate state:



The kMech input for this Ping Pong Bi Bi model is

$$\mathrm{Enz}[\{S_1, S_2\} \overset{\mathrm{En,\ Enx}}{\rightleftharpoons} \{P_1, P_2\},\ \mathrm{Ping\ Pong}[k_{f1}, k_{r1}, k_{cat1}, k_{f2}, k_{r2}, k_{cat2}]]$$

where Enx is the modified enzyme intermediate; k_{f1} and k_{f2} are rate constants of the enzyme–substrate associations for S_1 and S_2, respectively; k_{r1} and k_{r2} are rate constants of the enzyme–substrate dissociations for S_1 and S_2, respectively; k_{cat1} and k_{cat2} are the catalytic rate constants for the formation of products P_1 and P_2, respectively. kMech interprets this input and converts it into association–dissociation reactions (in this case, two single-substrate single-product reactions) defined in Cellerator as follows:

$$S_1 + \mathrm{En} \underset{k_{r1}}{\overset{k_{f1}}{\Leftrightarrow}} \mathrm{EnS_1} \overset{k_{cat1}}{\rightarrow} \mathrm{Enx} + P_1$$

$$S_2 + \mathrm{Enx} \underset{k_{r2}}{\overset{k_{f2}}{\Leftrightarrow}} \mathrm{EnxS_2} \overset{k_{cat2}}{\rightarrow} \mathrm{En} + P_2$$

The first reaction represents the formation of the enzyme–substrate complex with S_1(EnS$_1$), the release of product P_1 from the complex and the formation of the enzyme intermediate Enx. The second reaction represents the formation of the intermediate enzyme–substrate complex with S_2(EnxS$_2$) and the release of product P_2 and free enzyme, En.

Given these reactions, Cellerator generates the following differential equations for each reactant:

$$\frac{\mathrm{d}[S_1]}{\mathrm{d}t} = -k_{f1}[S_1][\mathrm{En}] + k_{r1}[\mathrm{EnS_1}],$$

$$\frac{\mathrm{d}[S_2]}{\mathrm{d}t} = -k_{f2}[S_2][\mathrm{Enx}] + k_{r2}[\mathrm{EnxS_2}],$$

$$\frac{\mathrm{d}[P_1]}{\mathrm{d}t} = k_{cat1}[\mathrm{EnS_1}],$$

$$\frac{\mathrm{d}[P_2]}{\mathrm{d}t} = k_{cat2}[\mathrm{EnxS_2}],$$

$$\frac{\mathrm{d}[\mathrm{En}]}{\mathrm{d}t} = -k_{f1}[S_1][\mathrm{En}] + k_{r1}[\mathrm{EnS_1}] + k_{cat2}[\mathrm{EnxS_2}],$$

$$\frac{\mathrm{d}[\mathrm{Enx}]}{\mathrm{d}t} = -k_{f2}[S_2][\mathrm{Enx}] + k_{r2}[\mathrm{EnxS_2}] + k_{cat1}[\mathrm{EnS_1}],$$

$$\frac{\mathrm{d}[\mathrm{EnS_1}]}{\mathrm{d}t} = k_{f1}[S_1][\mathrm{En}] - k_{r1}[\mathrm{EnS_1}] - k_{cat1}[\mathrm{EnS_1}],$$

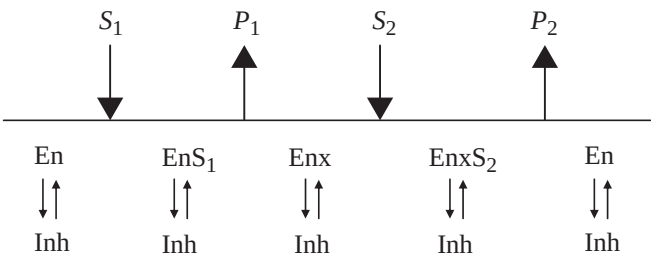
$$\frac{\mathrm{d}[\mathrm{EnxS_2}]}{\mathrm{d}t} = k_{f2}[S_2][\mathrm{Enx}] - k_{r2}[\mathrm{EnxS_2}] - k_{cat2}[\mathrm{EnxS_2}].$$

These differential equations and variable definitions are passed to Mathematica™ where they are solved by the NDSolve function and time plots are generated.

Extensions of enzyme reaction models with regulatory circuits

The enzyme that catalyzes the first step in a biological pathway is usually feedback inhibited by its end product (Umbarger, 1992). Non-competitive, competitive, and allosteric inhibition mechanisms are implemented here.

Non-competitive inhibition (NCI) indicates an inhibitor can bind all enzyme states including free enzyme, intermediate enzyme and enzyme–substrate complex (Segel, 1993).

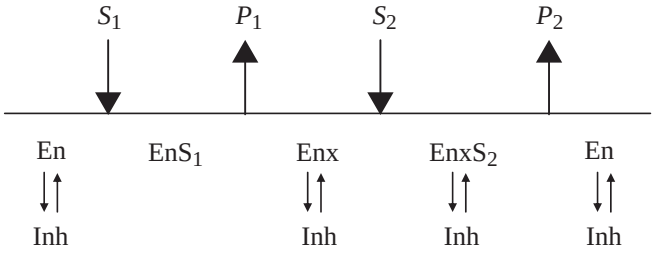


This mechanism can be described by the following Ping Pong Bi Bi model kMech input:

$$\mathrm{Enz}[\{S_1, S_2\} \overset{\mathrm{En,\ Enx}}{\Leftrightarrow} \{P_1, P_2\}, \mathrm{Ping\ Pong}[k_{f1}, k_{r1}, k_{cat1}, k_{f2}, k_{r2}, k_{cat2}], \mathrm{NCI}[\mathrm{inh}, k_{fi1}, k_{ri1}, k_{fi2}, k_{ri2}, \mathrm{residualRate}]]$$

where inh is the inhibitor name, k_{fi1}, k_{ri1}, and k_{fi2}, k_{ri2} are inhibitor association and dissociation rate constants for the first and second substrate reactions, respectively; residualRate is the fraction of maximal enzyme activity (V_{max}) remaining at saturating inhibitor concentrations. Other notations are as described above. The detailed reactions and equations that include these non-competitive inhibition and residual activity parameters can be found in Supplementary Figure 1 online.

Competitive inhibition indicates an inhibitor can only bind to the free enzyme state (Segel, 1993).



The kMech input for this Ping Pong Bi Bi mixed inhibition model (competitive inhibition for the first substrate reaction and NCI for the second substrate reaction) is:

$$\mathrm{Enz}[\{S_1, S_2\} \overset{\mathrm{En,\ Enx}}{\Leftrightarrow} \{P_1, P_2\}, \mathrm{Ping\ Pong}[k_{r1}, k_{r1}, k_{cat1}, k_{f2}, k_{r2}, k_{cat2}], \mathrm{CI}[\mathrm{inh}, k_{fi1}, k_{ri1}], \mathrm{NCI}[\mathrm{inh}, k_{fi2}, k_{ri2}]],$$

where CI stands for competitive inhibition. Other notations are as described above. The detailed reactions and equations for competitive and non-competitive reactions are included in the Supplementary Figure 1 online.

Allosteric regulation

Allosteric regulation is simulated by the two-state, concerted transition, allosteric model of Monod, Wyman and Changeux (MWC) where the enzyme can exist in either an active state, R, or an inactive state, T (Monod *et al*., 1965). This model is described by two equations:

$$R = \frac{(1+\alpha)^n}{L(1+c\alpha)^n + (1+\alpha)^n}$$

and

$$Y_f = \frac{Lc\alpha(1+c\alpha)^{n-1} + \alpha(1+\alpha)^{n-1}}{L(1+c\alpha)^n + (1+\alpha)^n},$$

where

$$L = L_0\frac{(1+\beta)^n}{(1+\gamma)^n}, \quad \alpha = \frac{S}{K_m}, \quad \beta = \frac{I}{K_i}, \quad \gamma = \frac{A}{K_a}.$$

The first equation describes the fraction of the enzyme in the catalytically active state (R) as a function of substrate and effector concentrations. The second equation describes the fractional saturation (Y_f) of the enzyme occupied by substrate as a function of substrate and effector concentrations.

In order to simulate the MWC model, several parameters such as substrate (S), activator (A) and inhibitor (I) concentrations and their respective dissociation constants K_m, K_a and K_i must be known. With these values, usually available in the literature, the values of α, β and γ are calculated. The value of n, the number of substrate and effector ligand binding sites, is

also usually available (e.g. $n = 4$ for *Escherichia coli* threonine deaminase); however, with few exceptions the values of c, the ratio of the affinity of the substrate for the catalytically active R state and the inhibited T state, and L_0, the equilibrium constant (allosteric constant) for the R and T states in the absence of ligands, are unavailable. Nevertheless, these values can be readily derived by fitting generally available substrate saturation curves generated in the presence of several inhibitor concentrations (Hatfield, 1970) as described in the Discussion section.

Lambda approximation method for enzyme rate constants

In order to solve the Cellerator-generated ODEs, values for forward and reverse kinetic rate constants k_f, k_r are required that often are not available. To deal with this problem, we developed the Lambda (Λ) approximation method. Briefly, this method calculates these constants from the easily determined, and usually available, enzyme parameters, K_m (Michaelis–Menten constant) and k_{cat} (catalytic constant or enzyme turnover number). This approach results in simple mathematical relations that can be used to estimate rate constants of differential equations for enzyme reactions. For example, the single substrate–single product 'catalytic' enzyme reaction considered earlier can be represented as:

$$A + \mathrm{En} \underset{k_r}{\overset{k_f}{\Leftrightarrow}} \mathrm{EnA} \xrightarrow{k_{cat}} \mathrm{En} + C \tag{1}$$

where A is the substrate, En is the free enzyme, EnA is the enzyme–substrate complex, C is the product, k_f is the forward (association) rate constant, k_r is the reverse (dissociation) rate constant and k_{cat} is the catalytic rate constant. Again, the ODEs that describe the above reactions by the law of mass action are as follows:

$$\begin{aligned}
\frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_f[\mathrm{En}][A] + k_r[\mathrm{EnA}] \\
\frac{\mathrm{d}[\mathrm{En}]}{\mathrm{d}t} &= -k_f[\mathrm{En}][A] + k_r[\mathrm{EnA}] + k_{cat}[\mathrm{EnA}] \\
\frac{\mathrm{d}[\mathrm{EnA}]}{\mathrm{d}t} &= k_f[\mathrm{En}][A] - k_r[\mathrm{EnA}] - k_{cat}[\mathrm{EnA}] \\
\frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_{cat}[\mathrm{EnA}]
\end{aligned}$$

To solve this system of ODEs, we apply a rapid equilibrium approximation which assumes that En, A and EnA reach equilibrium fast relative to the rate of catalysis. Thus,

$$\begin{aligned}
k_f[\mathrm{En}][A] &= k_r[\mathrm{EnA}] \\
k_f[A][\mathrm{En}] &\gg k_{cat}[\mathrm{EnA}]
\end{aligned} \tag{2}$$

In this case, the ratio of $k_f[A][\mathrm{En}]/k_{cat}[\mathrm{EnA}]$ represents a large number Q that varies with time as the reaction approaches steady state,

$$Q = \frac{k_f[A][\mathrm{En}]}{k_{cat}[\mathrm{EnA}]} \tag{3}$$

In the special condition (denoted by *) where $[A]^* = K_m$, 50% of En is saturated with A at equilibrium. In other words, under this condition, the amounts of free and substrate-bound enzyme are equal,

$$K_m = \frac{k_r + k_{cat}}{k_f} = [A]^* \quad \text{and} \quad [\mathrm{En}]^* = [\mathrm{EnA}]^*. \tag{4}$$

Consequently, the large quantity Q becomes a constant.

$$\begin{aligned}
Q^* &= \frac{k_f}{k_{cat}} \frac{[A]^*[\mathrm{En}]^*}{[\mathrm{EnA}]^*} \\
&= \frac{k_f \cdot K_m}{k_{cat}}
\end{aligned} \tag{5}$$

Therefore, we define Λ as a special case of Q

$$\Lambda = \frac{k_f \cdot K_m}{k_{cat}} \tag{6}$$

$$k_f[A]^*[\mathrm{En}]^* = \Lambda \cdot k_{cat}[\mathrm{EnA}]^* \tag{7}$$

where Λ is a large time-invariant constant number, and the larger Λ is, the faster En, A and EnA reach equilibrium. By rearrangement of (6), we can approximate k_f as,

$$k_f = \frac{\Lambda \cdot k_{cat}}{K_m} \tag{8}$$

Now, by substitution of (8) into (4), we can approximate k_r as

$$k_r = (\Lambda - 1) \cdot k_{cat} \tag{9}$$

Note that rearrangement of Equations (8) and (9) can be expressed as (4), which defines the Michaelis constant, K_m. The advantage of the Λ approximation is that two unknown parameters, k_f and k_r, are reduced to a single unknown parameter Λ.

By substitution of Equations (3) and (9) into the differential equation describing the state of EnA, we obtain,

$$\begin{aligned}
\frac{\mathrm{d}[\mathrm{EnA}]}{\mathrm{d}t} &= k_f[A][\mathrm{En}] - k_r[\mathrm{EnA}] - k_{cat}[\mathrm{EnA}] \\
&= k_f[A][\mathrm{En}] - (\Lambda - 1) \cdot k_{cat}[\mathrm{EnA}] - k_{cat}[\mathrm{EnA}] \\
&= k_f[A][\mathrm{En}] - \Lambda \cdot k_{cat}[\mathrm{EnA}] \\
&= k_f[A][\mathrm{En}] - \Lambda \cdot \frac{k_f[A][\mathrm{En}]}{Q} \\
&= \left(1 - \frac{\Lambda}{Q}\right) \cdot k_f[A][\mathrm{En}]
\end{aligned} \tag{10}$$

If the value of Λ is large and close to Q, that is $\Lambda/Q \approx 1$, then Equation (10) ≈ 0 (EnA reaches steady state). In other words, when Λ is large, there is a very fast approach to steady state, which is the same as the Michaelis–Menten pseudo steady-state assumption (Murray, 1993). Under these conditions, when Λ is large, k_f/k_f approaches $1/K_m$ [Equation (4)] which is observable. To determine the usable range of values of Λ that satisfy this criterion, we varied Λ from 10 to 1 000 000 in simulations with no significant changes in the steady levels of enzyme intermediates and end products. Therefore, the value of Λ can be reasonably set to a value of 100. Otherwise, the values of rate constants directly measured or empirically adjusted to fit experimental data can be used in the model.

K_m values can be obtained from the literature, and k_{cat} (μmol/min/μmol) values can be calculated from specific activities (μmol/min/mg) and molecular weights of purified enzymes. However, because of uncertainties about the percentages of purified enzymes that are active, k_{cat} values often require adjustments to fit experimental data.

The same Λ approximation methods, (8) and (9) are also used for multiple substrate and product enzymes that bind and release reactants in a Ping Pong fashion. In the general case of two-substrate, two-product enzymes that bind and release reactants in a Bi Bi fashion, k_f and k_r are approximated according to Equations (11) and (12).

$$k_f = \frac{\Lambda \cdot k_{cat}}{K_m^A \cdot K_m^B}, \tag{11}$$

$$k_r = (\Lambda - 1) \cdot k_{cat}, \tag{12}$$

where K_m^A is the K_m for substrate A, and K_m^B is the K_m for substrate B.

Omega approximation method for enzyme inhibition rate constants

In addition to needing methods to approximate forward and reverse reaction rates of enzyme substrate binding reactions, k_f, k_r, we need to be able to estimate the forward and reverse rate constants for enzyme–inhibitor binding reactions, k_{fi}, k_{ri} if their values are not known experimentally. To accomplish this we define a value Ω that approximates the rate that an enzyme binds to its substrate relative to its inhibitor. Similar to the derivation of Λ, the derivation

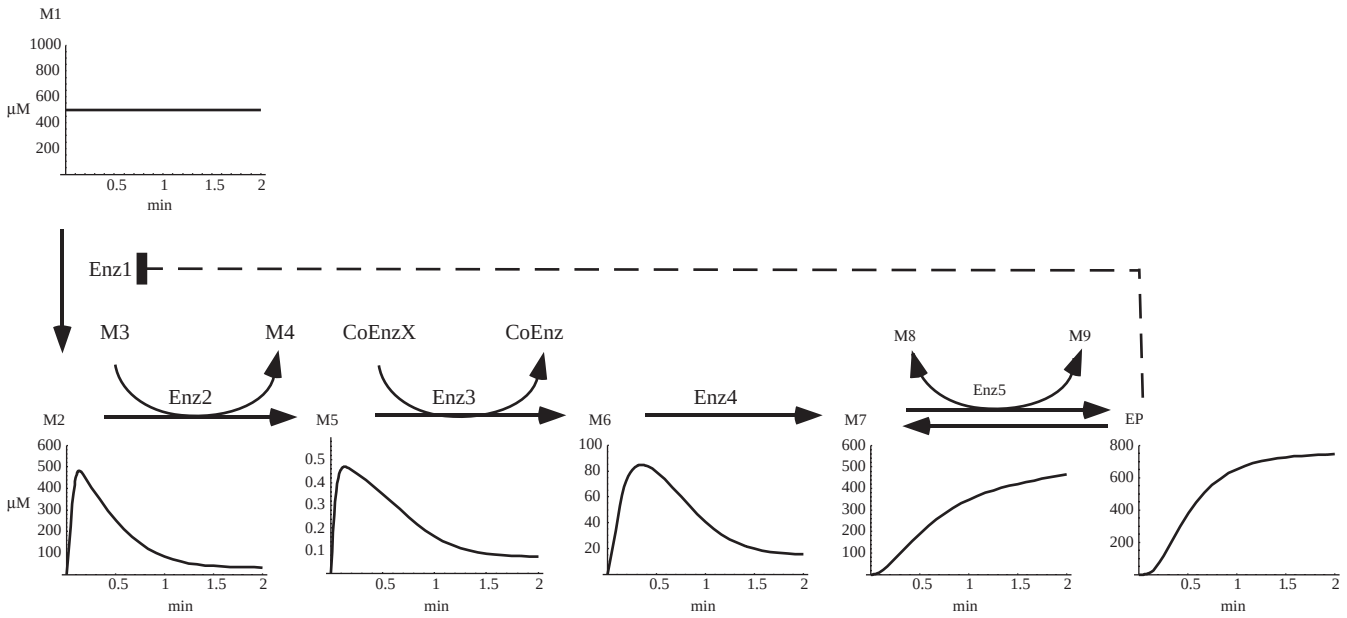


Fig. 1. Simulated flow of carbon through a metabolic Pathway. The graphical insets show the approach (min) to steady-state (μM) synthesis and utilization of the substrates, intermediates and end products of the pathway. The metabolites and enzymes are abbreviated as M1–M9 and Enz1–Enz5, respectively. EP stands for end product. CoEnz and CoEnzX represent states of the coenzyme. Feedback inhibition is indicated by dashed lines. The starting substrates M1 are supplied at rate to maintain constant levels of 500 μM.

of Ω is based on a rapid equilibrium assumption. The k_{fi} is approximated by

$$k_{\mathrm{fi}} = \Omega \cdot k_{\mathrm{f}} \tag{13}$$

and k_{ri} is approximated by

$$k_{\mathrm{ri}} = \Omega \cdot k_{\mathrm{f}} \cdot K_{\mathrm{i}}. \tag{14}$$

As shown in the detailed derivation provided as Supplementary Figure 2 online, the value of Ω can be an arbitrary number if free enzyme, enzyme–inhibitor complex, and enzyme–substrate complex are at equilibrium. Consequently, the value of Ω may be set to 1. Otherwise, the values of rate constants directly measured or empirically adjusted to fit experimental data can be used in the model.

For the purpose of demonstration, a simplified metabolic pathway (Fig. 1) consisting of various enzyme catalytic and regulatory mechanisms described above was constructed using kMech.

IMPLEMENTATION

kMech was written in the Mathematica™ language. It converts complex enzyme mechanisms into elementary Cellerator association–dissociation reactions. It was saved in a package format that can be read into Mathematica™ with Cellerator. kMech-modeled simulations can be executed by Mathematica™ software installed in a Microsoft Windows, MacOS or Linux operating system.

DISCUSSION

As interest in systems biology grows, a major challenge is to develop user-friendly software tools for biologists to model biochemical pathways without knowledge of the underlying mathematics. The kMech/Cellerator/Mathematica package described here addresses this need. To mathematically model a biochemical pathway, users need only to invoke kMech models for the enzyme mechanisms of a pathway without entering the differential equations that describe these mechanisms. Because of this simple user input and the integration of kMech, Cellerator and Mathematica™, human errors are greatly reduced.

Mathematica™ provides an excellent development environment that integrates symbolic user input (kMech) seamlessly with Cellerator, which converts these inputs into association–dissociation reactions and differential equations that are used by Mathematica™ to simulate the models and to supply graphic outputs. For demonstration purpose, a metabolic pathway that consists of five enzymes, ten metabolites, and one enzyme cofactor was constructed using kMech (Fig. 1). The mathematical model for this metabolic system consists of 27 ODEs, with 16 association and dissociation rate constants and 9 catalytic rate constants. The enzymes of these interacting pathways employ five distinct enzyme mechanisms (Enz1 is allosteric; Enz2 is Ping Pong Bi Bi; Enz3 is Bi Bi; Enz4 is simple catalytic and Enz5 is reversible Ping Pong Bi Bi). Enz1, feedback regulated by the pathway end product, was simulated with the MWC model employing physical parameters based on *E.coli* L-threonine deaminase (Calhoun *et al*., 1973). With the Michaelis–Menten pseudo steady-state assumption that substrate, enzyme and enzyme–substrate complex are under rapid steady-state equilibrium, the fractional saturation (Y_{f}) values of the enzyme occupied by substrate are equivalent to measured $v/V_{\max}$ values (Segel, 1993). These values of $v/V_{\max}(Y_{\mathrm{f}})$ are available from substrate (threonine) saturation data obtained in the presence and absence of inhibitor (isoleucine) reported in the literature (black dots) (Calhoun *et al*., 1973) (Fig. 2A). The error function (Err) in Figure 2B is the sum of squared differences of experimental substrate saturation data (black dots) that fit theoretical (Y_{f}) curves with values of c from 0 to 1, and

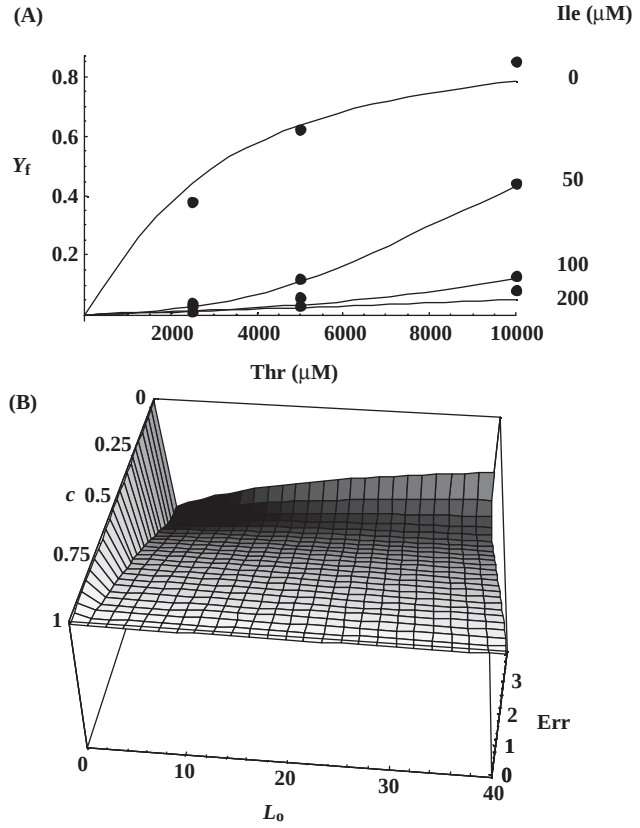


Fig. 2. Optimization of physical constants, c and L_0, for the concerted transition allosteric Monod, Wyman and Changeux (MWC) model. (**A**) Black dots represent experimental measurement of Y_f values for *E.coli* L-threonine deaminase with various substrate (threonine, Thr) and inhibitor (isoleucine, Ile) concentrations. The solid line curves represent theoretical Y_f values of calculated with c and L_0 values determined from (B). (**B**) The error function (Err) is the sum of squared differences between experimental data (black dots) and theoretical Y_f values calculated with values of c from 0 to 1 and L_0 from 0 to 40. The pair of c and L_0 values that produce minimum errors are $c = 0.013$, $L_0 = 1.05$.

L_0 from 0 to 40 as discussed in the Systems and methods section. The pair of c and L_0 values that produce the best-fit curves, determined with the non-linear programming Mathematica function, FindMinimum, are $c = 0.013$, $L_0 = 1.05$. The time-dependent approach to steady state for the six pathway intermediates and end products are shown in Figure 1. At $t = 0$, an initial burst of intermediate synthesis is observed that is inhibited as end product accumulates and feedback inhibits the first enzyme of the pathway. While the steady-state levels of the substrates for Enz3 and Enz4 (M5 and M6) are maintained at a low level, the level of the precursor of the end product, M7, reaches a high steady-state level comparable with that of the end product because of the reversible nature of this reaction catalyzed by Enz5. This pathway simulates the enzyme mechanisms of L-isoleucine biosynthesis in *E.coli* K12 (Umbarger, 1996). In this case, the last reversible reaction of the pathway is the transamination between L-isoleucine and L-glutamic acid. The detailed kMech inputs, corresponding ODEs, kinetic rate constants and initial conditions for solving the ODEs are presented in Supplementary Figure 3 online.

While deterministic, continuous, modeling of the type described here can provide valuable information such as predicted steady-state levels of metabolic substrates, intermediates and end products, and predict the outcomes of biochemical and genetic perturbations that require detailed enzyme kinetic and regulatory mechanisms, a major challenge of this approach is the measurement or estimation of rate constants required for the solution of rate equations. It nevertheless has a long history. For example, as early as 1943, Chance described methods to measure rate constants directly, and reports describing computer-based methods to fit rate constants to match experimental data date back to the 1960s (Barshop *et al*., 1983; Garfinkel *et al*., 1968; Schoeberl *et al*., 2002; Waas *et al*., 2001; Zimmerle and Frieden, 1989). However, since experimental results required for either the direct measurement or the indirect fitting of all required rate constants are not available for every enzyme, these methods have not been well-suited for large-scale simulations of metabolic networks. To address this problem we have described general methods for the approximation of some of the missing rate constants. These Lambda and Omega approximations, described in the systems and methods section, assume rapid equilibrium conditions between free enzyme and enzyme–substrate complexes, and between free enzyme and enzyme–inhibitor complexes, that allow us to estimate reaction rate constants from available kinetic data (K_m, k_{cat}). This facilitates the modeling of systems in which only these types of kinetic data are available.

Clearly, an ultimate goal of systems biology is the complete simulation of cellular systems (Tomita *et al*., 1999). However, given the uncertainty of rate constant measurements needed for solving differential equations, some investigators have turned their attention to more abstract methods. For example, Reed *et al*. (2003) have used metabolic flux balance analysis (FBA) methods to simulate steady-state metabolite flux through *E.coli* pathways representing more than 900 enzyme steps. An advantage of FBA is that it can make predictions about large metabolic networks without knowing anything about individual kinetic parameters. It assumes that for each node of the network the entrance and exit rates must equal one another. In this case, a simple set of equations can be generated stating that total flux into a node minus the total flux out of the node must equal zero. Thus, FBA can provide valuable information about biomass conversions (Edwards *et al*., 2001). However, since FBA does not consider pathway-specific regulation patterns it is less suited for the simulation of biochemical or genetic perturbations that require detailed enzyme kinetic and regulatory mechanisms. The kMech/Cellerator models described here represent the first step of a 'bottom-up' approach to model complex biological processes.

ACKNOWLEDGEMENTS

We are appreciative of support from the UCI Institute for Genomics and Bioinformatics and helpful advice from Donald F. Senear. This work was supported in part by National Institutes of Health grants GM55073 and GM68903 to G.W.H. C.-R.Y. is the recipient of a Biomedical Informatics Training (BIT) Program postdoctoral fellowship from the National Library of Medicine T15 LM-07443.

REFERENCES

Barshop,B.A., Wrenn,R.F. and Frieden,C. (1983) Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM—a flexible, portable system. *Anal. Biochem.*, **130**, 134–145.

Calhoun,D.H., Rimerman,R.A. and Hatfield,G.W. (1973) Threonine deaminase from *Escherichia coli*. I. Purification and properties. *J. Biol. Chem.*, **248**, 3511–3516.

Chance,B. (1943) The kinetics of the enzyme–substrate compound of peroxidase. *J. Biol. Chem.*, **151**, 553–577.

Edwards,J.S., Ibarra,R.U. and Palsson,B.O. (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, **19**, 125–130.

Falk,S., Guay,A., Chenu,C., Patil,S.D., and Berteloot,A. (1998) Reduction of an eight-state mechanism of cotransport to a six-state model using a new computer program. *Biophys. J.*, **74**, 816–830.

Garfinkel,D., Frenkel,R.A. and Garfinkel,L. (1968) Simulation of the detailed regulation of glycolysis in a heart supernatant preparation. *Comput. Biomed. Res.*, **2**, 68–91.

Hatfield,G.W. (1970) The regulation of L-threonine deaminase in *Bacillus subtilis* by repression and endproduct inhibition. PhD Thesis, *Department of Biological Sciences, Purdue University, Purdue*, pp. 103–112.

Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H., Arkin,A.P., Bornstein,B.J., Bray,D., Cornish-Bowden,A. *et al*. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Mendes,P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.*, **22**, 361–363.

Monod,J., Wyman,J. and Changeux,J.P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, **12**, 88–118.

Murray,J.D. (1993) Michaelis–Menten theory: detailed analysis and the pseudo-steady state hypothesis. *Math. Biol.*, **19**, 116–117.

Reed,J.L., Vo,T.D., Schilling,C.H. and Palsson,B.O. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR) *Genome Biol.*, **4**, R54.

Sauro,H.M., Hucka,M., Finney,A., Wellock,C., Bolouri,H., Doyle,J. and Kitano,H. (2003) Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS*, **7**, 355–372.

Schoeberl,B., Eichler-Jonsson,C., Gilles,E.D. and Muller,G. (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, **20**, 370–375.

Segel,I.H. (1993) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems*. Wiley, NY, p. 427, Equation (VII-66).

Shapiro,B.E., Levchenko,A., Meyerowitz,E.M., Wold,B.J. and Mjolsness,E.D. (2003) Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, **19**, 677–678.

Tomita,M., Hashimoto,K., Takahashi,K., Shimizu,T.S, Matsuzaki,Y., Miyoshi,F, Saito,K., Tanida,S., Yugi,K., Venter,J.C. and Hutchison,C.A.III (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84.

Umbarger,H.E. (1992) The origin of a useful concept–feedback inhibition. *Protein Sci.*, **1**, 1392–1395.

Umbarger,H.E. (1996) *Biosynthesis of the Branched-Chain Amino Acids.* Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*. ASM Press, Washington, DC.

Waas,W.F., Lo,H.H. and Dalby,K.N. (2001) The kinetic mechanism of the dual phosphorylation of the ATF2 transcription factor by p38 mitogen-activated protein (MAP) kinase alpha. Implications for signal/response profiles of MAP kinase pathways. *J. Biol. Chem.*, **276**, 5676–5684.

Zimmerle,C.T. and Frieden,C. (1989) Analysis of progress curves by simulations generated by numerical integration. *Biochem. J.*, **258**, 381–387.