APPLICATION OF A GENERALIZED MWC MODEL FOR THE MATHEMATICAL SIMULATION OF METABOLIC PATHWAYS REGULATED BY ALLOSTERIC ENZYMES

TAREK S. NAJDI*

Institute for Genomics and Bioinformatics, Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, Irvine, California 92697
tnajdi@uci.edu

CHIN-RANG YANG*

Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center at Dallas Dallas, Texas 75390

BRUCE E. SHAPIRO

Jet Propulsion Laboratory, California Institute of Technology 4800 Oak Grove Drive, Pasadena, California 91109
bshapiro@caltech.edu

G. WESLEY HATFIELD

Institute for Genomics and Bioinformatics, Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, Irvine, California 92697
ghatfie@uci.edu

ERIC D. MJOLSNESS

Institute for Genomics and Bioinformatics, Donald Bren School of Information and Computer Science, University of California, Irvine, Irvine, California 92697
emj@uci.edu

In our effort to elucidate the systems biology of the model organism, *Escherichia coli*, we have developed a mathematical model that simulates the allosteric regulation for threonine biosynthesis pathway starting from aspartate. To achieve this goal, we used kMech, a Cellerator language extension that describes enzyme mechanisms for the mathematical modeling of metabolic pathways. These mechanisms are converted by Cellerator into ordinary differential equations (ODEs) solvable by Mathematica™. In this paper, we describe a more flexible model in Cellerator, which generalizes the Monod, Wyman, Changeux (MWC) model for enzyme allosteric regulation to allow for multiple substrate, activator and inhibitor binding sites. Furthermore, we have developed a model that describes the behavior of the bifunctional allosteric enzyme aspartate kinase I-homoserine dehydrogenase I (AKI-HDHI). This model predicts the partition of enzyme activities in the steady state which paves the way for a more generalized prediction of the behavior of bifunctional enzymes.

Keywords: Systems biology, Mathematical modeling, Allosteric regulation, MWC model, Threonine biosynthesis.

---

*These authors contributed equally to this work
Biology correspondence should be addressed to G.W.H. (gwhatfie@uci.edu).
Computation correspondence should be addressed to E.D.M. (emj@uci.edu).

1. Introduction

Systems biology uses the mathematical modeling of biological networks to allow scientists to understand and observe complex biological behaviors and predict the outcomes of metabolic and genetic perturbations. The major biological networks that are undergoing active modeling include: transcriptional regulation, metabolic networks, signal transduction and mechanical networks. To integrate the large amount of data produced by these networks, it is essential to develop mathematical models to simulate such complex biological systems.

Allosteric regulation has been applied to signaling molecules as diverse as regulatory enzymes, nuclear and membrane receptors and even spliceosome activation. In this type of regulation, regulatory effectors bind at sites distinct form substrates and affect the catalytic activity and substrate binding through a reversible conformational change which also allows for cooperative binding of regulators. In metabolic pathways, allosteric enzymes are located at key points to regulate metabolite flow. If the end product accumulates in excess of the cell's needs, it usually acts as an inhibitor of the first enzyme specific to its synthesis and keeps its own concentration and the concentration of other intermediates at steady state levels.4 This is where allosteric regulation of enzyme activity and its role in feedback regulation become important. In addition, the activity of allosteric enzymes can be positively or negatively regulated by the concentration of the substrate itself, or by intermediate or final products of other pathways. This type of regulation helps metabolic networks maintain a steady state, and achieve efficiency and coordination with one another.

We have previously described kMech,[1] a mathematical model tool for complex enzyme mechanisms, and applied it to model the biosynthesis of the branched chain amino acids, L-isoleucine, L-valine, and L-leucine in *Escherichia coli*.[2] As a rule of thumb, metabolic pathways are regulated tightly by the feedback inhibition of the end-products. In kMech, we provide three basic types of inhibition (competitive, non-competitive and uncompetitive). Here, we develop a Generalized Monod, Wyman, Changeux (GMWC) model for the more complex, allosteric feedback regulation for enzymes regulated by different binding sites of substrates, activators and inhibitors. The GMWC model is an extension of the original MWC concerted allosteric transition model published by Monod *et al.*[3] The major enhancement is that the GMWC model can accommodate multiple substrates with activator and inhibitor binding sites, whereas the original MWC can accommodate only single substrate, activator and inhibitor binding sites. The GMWC model is especially useful for modeling threonine biosynthesis in *E. coli*, since there are three allosteric enzymes, aspartate kinase I (AKI), aspartate kinase III (AKIII), and homoserine dehydrogenase (HDHI), of total six enzymes in this pathway (Fig. 1).

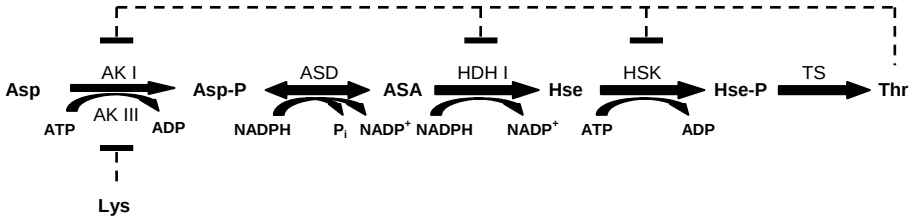

Fig. 1. The metabolic pathway for threonine biosynthesis from aspartate in *E. coli*. The abbreviations of metabolites are: Asp, aspartate; AspP, Aspartyl phosphate; ASA, aspartate semialdehyde; Hse, homoserine; HseP, homoserine phosphate; Thr, threonine. The abbreviations of enzymes are: AKI, aspartate kinase I (EC 2.7.2.4); AKIII, aspartate kinase III (EC 2.7.2.4); HDHI, homoserine dehydrogenase I (EC 1.1.1.3); ASD, semialdehyde dehydrogenase (EC 1.2.1.11); HSK, homoserine kinase (EC 2.7.1.39); TS, threonine synthase (EC 4.2.3.1). Feedback inhibition patterns are indicated by dashed lines.

Threonine synthesis, starting from aspartate, is a five-step metabolic pathway (Fig. 1). The first step of the pathway is a two-substrate, two-product (BiBi) reaction with the two substrates aspartate (Asp) and ATP, and the two products Aspartyl phosphate (AspP) and ADP. This step is catalyzed by three isozymes, AKI, aspartate kinase II (AKII) and AKIII. AKI is a bifunctional enzyme (AKI-HDHI) carrying both aspartate kinase and homoserine dehydrogenase activities (Step three). It is an allosteric enzyme made up of four subunits and inhibited by threonine.[5] The kinetic behavior of this enzyme can be simulated by the GMWC model. This model states that the enzyme exists in equilibrium between an active (R) state and an inactive (T) state, whereby binding of substrate or activator will shift the equilibrium towards the R state while binding of inhibitor will shift it towards the T state. The fractional saturation (Y_f) of enzyme bound by substrate is described as a function of the substrates and effectors.[3] AKII is also a bifunctional enzyme carrying both aspartate kinase and homoserine dehydrogenase (HDHII) activities. This enzyme has no effectors and exists at very low levels in *E. coli* K12.[6] Therefore; we did not include it in our simulation. AKIII is an allosteric enzyme made up of two subunits and inhibited by lysine.[7] Its kinetic behavior is also described by the GMWC model.[7]

In the second step of the pathway, aspartate semialdehyde dehydrogenase (ASD) catalyzes a reversible two substrates/three products (BiTri) reaction with AspP and NADPH as substrates and aspartate semialdehyde (ASA), NADP and inorganic phosphate as products.[8] The third step is a BiBi reaction with the two substrates ASA and NADPH and the two products homoserine (Hse) and NADP. This step is catalyzed by HDHI activity of the bi-functional enzymes (AKI-HDHI). HDHI is an allosteric enzyme made up of four subunits and is inhibited by threonine.[3] Its kinetic behavior also is described by the GMWC model.[3]

The forth step of the pathway is a BiBi reaction with the two substrates Hse and ATP and two products homoserine phosphate (HseP) and ADP. This step is catalyzed by homoserine kinase (HSK) and competitively inhibited by threonine.[9] The fifth and final step of the pathway is a simple one substrate/one product reaction that results in the conversion of HseP into threonine. This step is catalyzed by threonine synthase (TS) and has no effectors.[10]

The enzyme kinetics of individual enzymes of this pathway have been subjected to extensive study over the past 40 years. However, a detailed mathematical model of the pathway with allosteric regulation has not been available.

2. Methods and Tools

To generate a simulation of threonine biosynthesis, we used Mathematica™, Cellerator, and kMech. Mathematica™ is software developed by Wolfram Research and used in various applications including mathematical calculations, solving equations, and programming. Cellerator is a Mathematica™ package designed for the generation of chemical reaction networks that describe complex cascades as well as differential equation models that are derived from such networks.[11] These differential equations are solvable by Mathematica™. kMech is a Cellerator language extension that describes enzyme mechanisms for the mathematical modeling of metabolic pathways. Using the law of mass action, the more complex enzyme mechanisms supplied in kMech are interpreted by Cellerator to generate ordinary differential equations (ODEs) for each reactant. Allosteric enzymes get specialized rate laws as discussed later in this paper and the parameters in the ODEs were obtained as described in the section below. The differential equations and variable definitions are then passed to Mathematica™ where they are solved by its numeric solver (NDSolve) function and times vs. concentration plots are generated. This method was described in detail previously.[1]

2.1. Parameter estimation and optimization

The mathematical model for the pathway includes the entire forward and reverse single and multiple substrate enzyme kinetic reactions and the regulatory feedback inhibition mechanisms of the pathway (allosteric, competitive, and non-competitive). Enzyme kinetic constants for substrate (K_m), inhibitor (K_i), and activator (K_a) were obtained from the literature. Forward and reverse rate constants (k_f, k_r, k_{fi}, k_{ri},) were approximated from kinetic measurements (K_m, k_{cat}). The development of such approximation methods for estimating unavailable model parameters were previously described.[1] Intracellular enzyme concentrations are difficult to measure and usually are not available. For this reason, we previously described a method for approximating enzyme concentrations from DNA microarray data.[1, 2]

A Mathematica™ notebook of the simulation with the detailed kMech and GMWC models, corresponding ODEs, kinetic rate constants, and initial conditions for solving the ODEs for the pathway, a Mathematica™ executable kMech.m file, and a list of reported and optimized enzyme kinetic and physical parameters used to solve differential equations in the simulation and their literature sources are available at the University of California, Irvine (UCI), Institute for Genomics and Bioinformatics (IGB) website, http://www.igb.uci.edu/servers/sb.html. Cellerator, available at the same site, is free of charge to academic, U.S. government, and other nonprofit organizations.

2.2. Generalized Monod, Wyman, Changeux concerted allosteric transition model (GMWC)

According to the original MWC model,[3] an allosteric enzyme can exist in an active (R) state or an inactive (T) state. The fraction of active enzyme in the R or T states is determined by the concentrations and relative affinities of substrate (Asp for AKI and AKIII, and ASA for HDHI), inhibitor (Thr for AKI and HDHI, and Lys for AKIII), and activator (none present in this model) for the R and T states. The original model is described by Eq. (1) and Eq. (2).

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

$$Y_f = \frac{v_o}{V_{\max}} = \frac{L'c\alpha(1+c\alpha)^{n-1} + \alpha(1+\alpha)^{n-1}}{L'(1+c\alpha)^n + (1+\alpha)^n} \quad (2)$$

where $L' = L\frac{(1+\beta)^n}{(1+\gamma)^n}, \alpha = \frac{S}{K_m}, \beta = \frac{I}{K_i}, \gamma = \frac{A}{K_a}$; *S*, *I* and *A* are substrate, inhibitor and activator concentrations, respectively; K_m, K_i and K_a are their respective dissociation constants; *n* is the number of substrate and effector ligand binding sites; *c* is the ratio of the affinity of the substrate for the catalytically active R state and the inhibited T state; *L* is the equilibrium constant (allosteric constant) for the R and T states in the absence of ligands; v_o is the initial reaction velocity (initial rate of product generation, $d[P]/dt$); and V_{max} is the maximal reaction velocity.

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 ($Y_f = v_o/V_{max}$) of the enzyme occupied by substrate as a function of substrate and effector concentrations.[12]

Because the MWC model is an equilibrium model, it may be derived from its "partition function" *Z*,[13] where $Z = \sum_i e^{-\Delta Gi/\tau}$, where *i* is the index of all possible enzyme states and the probability of any given enzyme state, P_i, is $e^{-\Delta Gi/\tau} / Z$

Here, ΔG is Gibbs free energy difference of enzyme state and $\tau = kT$, where *T* is the temperature and *k* is Boltzmann's constant. For example, a single binding site for ligand *l* with concentration [*l*] and equilibrium constant, K_l, could have a partition function

$$Z_{site} = e^{-\Delta G(unbound)/\tau} + e^{-\Delta G(bound)/\tau} = 1 + [l]/K_l$$

If we defined the free energy differences with respect to the unbound state, then $\Delta G(unbound)$ is zero and the partition function of the ligand-bound state is the fugacity of the interaction between ligand and enzyme in solution, $[l]/K_l$. In the case of non-interacting systems Z_1 and Z_2, probabilities and partition functions both multiply: $Z = Z_1 Z_2$. If we imagined all binding sites, *j*, of an enzyme to be completely independent, the partition function would just be a product over sites:

$$Z_{bound} = \prod_j (1 + [l_j]/K_{lj})$$

However, the effect of binding sites on the enzyme activity is not considered in this simplified model and it has little biological function.

The MWC model is a minimal complication of the non-interacting model: it has two global states (an active R state and an inactive T state) in addition to the local binding site states, occupied and unoccupied. Although all binding sites act independently in each global state, each binding site couples to the specific global state. Therefore *Z* is a sum of two products over binding sites. In the R state, the independent binding sites consist of one substrate-binding site per subunit and one activator-binding site per subunit; all inhibitor binding sites have zero probability of occupancy. If the enzyme has *n* identical subunits, then the partition function, Z_R, for the R state is

$$Z_R = (1+\alpha)^n (1+\gamma)^n$$

Likewise the T state has independent binding for each substrate and each inhibitor, but not for any activators:

$$Z_T = L(1+c\alpha)^n (1+\beta)^n$$

The MWC partition function is just the sum of these terms:

$$Z_{MWC} = (1+\alpha)^n (1+\gamma)^n + L(1+c\alpha)^n (1+\beta)^n$$

Generally in equilibrium statistical mechanics, thermal averages may be computed in terms of suitable derivatives of the logarithm of the partition function *Z*. Since α is the fugacity variable pertaining to the substrate, the average number of binding sites occupied by the substrate is in general[13]

$$\frac{\partial \log Z}{\partial \log \alpha} = \alpha \frac{\partial \log Z}{\partial \alpha}$$

For example, if we had just the R state alone with no cooperativity whatsoever, this average occupancy would be

$$\frac{\partial \log Z_R}{\partial \log \alpha} = n \frac{\alpha}{1+\alpha}$$

exactly as one would expect for *n* independent binding sites.

For the MWC model, then, the average number of binding sites occupied by the substrate is

$$\frac{\partial \log Z_{MWC}}{\partial \log \alpha} = \frac{\partial \log Z_{MWC}}{\partial \alpha} \frac{d\alpha}{d \log \alpha} = \alpha \frac{\partial \log Z_{MWC}}{\partial \alpha}$$

$$= \alpha [n(1+\alpha)^{n-1}(1+\gamma)^n + Lnc(1+L\alpha)^{n-1}(1+\beta)^n] / Z_{MWC}$$

$$= n \frac{(1+\alpha)^{n-1}\alpha(1+\gamma)^n + Lc\alpha(1+c\alpha)^{n-1}(1+\beta)^n}{(1+\alpha)^n(1+\gamma)^n + L(1+c\alpha)^n(1+\beta)^n} = n \frac{\alpha(1+\alpha)^{n-1} + L' c\alpha(1+c\alpha)^{n-1}}{(1+\alpha)^n + L'(1+c\alpha)^n}$$

Therefore, $Y_f = \frac{1}{n} \frac{\partial \log Z_{MWC}}{\partial \log \alpha}$, which demonstrates Eq. (2).

However, the original MWC model can only accommodate one substrate, one activator and one inhibitor. In order to model multiple-ligand allosteric enzymes in *E. coli*, we developed a Generalized MWC (GMWC). In this model, each substrate will have its own activator and inhibitor and each triple of a substrate, its activator, and its inhibitor will have a corresponding triple of binding sites on each subunit.

The GMWC model can be derived in a similar way to the MWC derivation above, by elaborating the picture of the enzyme and its binding sites. We assume each subunit has *Q* different substrate binding sites, each of which binds a different but necessary substrate S_q. For simplicity, assume also each substrate has one corresponding activator a_q and one inhibitor i_q.

To find the partition function and therefore the kinetics for this GMWC model, we generalize the foregoing equations by reasoning as follows. If there were only one global state for the enzyme, all binding events would be independent of one another. The partition function would be a product of independent partition functions, one per binding site, as follows:

$$Z_{1-state} = \prod_n \prod_q (1+s_q)(1+a_q)(1+i_q) = \prod_q (1+s_q)^n (1+a_q)^n (1+i_q)^n$$

Here q indexes the Q substrates; s_q=[S_q]/K_{Mq} is the fugacity of the substrate at any of its (identical) sites, [S_q] is the concentration of substrate q, and K_{Mq} sets the scale for [S_q] and is a generalized version of the dissociation constant of the enzyme-substrate complex in steady state; similarly, a_q=[A_q]/K_{aq} where [A_q] is the concentration of activator for substrate S_q; i_q=[I_q]/K_{iq} where [I_q] is the concentration of inhibitor for S_q.

However, the actual situation is different: there are two global states of the n-subunit enzyme, the "R" and "T" states. In the R state, activators can bind but not inhibitors. In the T state, inhibitors can bind, but not activators. In either state, production occurs in proportion to the number of subunits containing all required substrates. Thus the partition function is a sum of two terms, one for the R state omitting inhibitors from the hypothetical single-state partition function shown above, and one for the T state omitting activators, with an extra multiplicative constant L due to the free energy difference between the two global states when all binding sites are empty, and also an extra multiplicative constant c for each substrate owing to the change in free energy when that substrate is bound to a site within the T state rather than the R state. The partition function is again a sum of non-interacting binding site products:

$$Z_{GMWC} = \prod_q (1+s_q)^n (1+a_q)^n + L\prod_q (1+cs_q)^n (1+i_q)^n$$

From this form, we can see that the rate of product generation (v_0) is proportional to the fraction of subunits at which all required substrates and no inhibitors are present, which is:

$$Y_f = [\prod_q (\frac{1}{n}\frac{\partial}{\partial \log s_q})] \log Z_{GMWC}$$

$$Y_f = \frac{\prod_q [(1+s_q)^{n-1} s_q (1+a_q)^n] + L\prod_q [(1+cs_q)^{n-1} cs_q (1+i_q)^n]}{\prod_q [(1+s_q)^n (1+a_q)^n] + L\prod_q [(1+cs_q)^n (1+i_q)^n]}, \text{ and}$$

$$\frac{d[p_i]}{dt} = v_0 = k_{cat}[E]Y_f$$

where $V_{max} = k_{cat}[E]$ and [E] is the total enzyme concentration. This is the differential equation used for the GMWC model. An illustration of Y_f function with q = 2 is shown in Fig. 2. A more detailed version of the model can be derived, using the same logic as above, to incorporate the effects of competitive inhibition at the substrate and activator binding sites:

$$Y_f = \frac{\prod_q [(1+s_q+\bar{s}_q)^{n-1} s_q (1+a_q+\bar{a}_q)^n] + L\prod_q [(1+\bar{a}_q)^n (1+cs_q+\bar{s}_q)^{n-1} cs_q (1+i_q)^n]}{\prod_q [(1+s_q+\bar{s}_q)^n (1+a_q+\bar{a}_q)^n] + L\prod_q [(1+\bar{a}_q)^n (1+cs_q+\bar{s}_q)^n (1+i_q)^n]},$$

where $s_q = [S_q]/K_{Sq}$, $a_q = [A_q]/K_{Aq}$, $i_q = [I_q]/K_{Iq}$, $\bar{s}_q = \sum_j \frac{[\bar{S}_{qj}]}{K_{Sqj}}$, $\bar{a}_q = \sum_j \frac{[\bar{A}_{qj}]}{K_{Aqj}}$, $\bar{i}_q = \sum_j \frac{[\bar{I}_{qj}]}{K_{Iqj}}$,

$\bar{s}_q$ is the net competitive inhibition for the substrate s_q, and $\bar{a}_q$ is the net competitive inhibition for the activator a_q. Yet more general parameterizations of this model could provide new c factors: $c_{(s)}$ for $\bar{s}_q$ and $c_{(a)}$ for $\bar{a}_q$ in the T state, so that their binding affinity may be unequal in the R and T states, and could also allow any of the factors c: $c_{(s)q}$ and $c_{(a)q}$ to depend on q. The detailed documentation of the implementation of the GMWC model can be found in the Cellerator web site and the "Help" section in Cellerator.

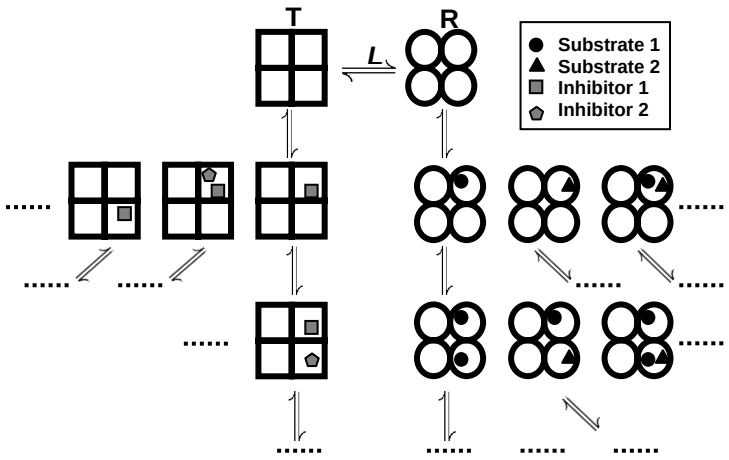

Fig 2. An illustration of the generalized MWC model. This is an example of an allosteric enzyme with four protomers, two substrates and two inhibitors. The black circles represent substrate 1; the black triangles represent substrate 2; the grey squares represent inhibitor 1; the grey ovals represent inhibitor 2. The diagram illustrates a part of possible enzyme-substrate and enzyme-inhibitor binding states. As described in the text, the Y_f function of the GMWC model is the fraction of the active enzyme bound by all required substrates for its function, which is the fraction of the active enzyme bound by both substrates 1 and 2 as illustrated in this diagram. The denominator of the Y_f function represents all possible enzyme-substrate and enzyme-inhibitor binding states, and the numerator of the Y_f function represents all possible enzyme-two substrates binding states.

3. Pathway Modeling

As discussed previously,[1, 2] traditional approaches to model enzyme kinetic pathways have mostly relied on the Michaelis-Menten kinetic equation for one substrate/one product reactions and the King-Altman method to derive equations for complex multiple reactant reactions. In these studies, non-linear differential equations are simplified into linear algebra equations.[14] On the other hand; kMech/Cellerator models include non-linear differential equations where complex enzyme mechanisms including single and multiple substrate enzyme kinetic reactions, and ligand activation and feedback inhibition mechanisms are taken into consideration.

3.1. Bifunctional Aspartate Kinase I-Homoserine Dehydrogenase I (AKI-HDHI)

AKI-HDHI is a bifunctional enzyme carrying both aspartate kinase (AKI) and homoserine dehydrogenase (HDHI) activities. It catalyzes two different steps of the same pathway. This bifunctional activity was modeled in a way that the binding of either substrate Asp or ASA shifts the equilibrium towards one of the activities, namely, the kinase or the dehydrogenase activity of the active (R) state of the enzyme, while binding of the inhibitor threonine shifts the equilibrium towards an inactive (T) state and binding of the substrates Asp and ASA shifts it back towards the active (R) state (Fig. 3).

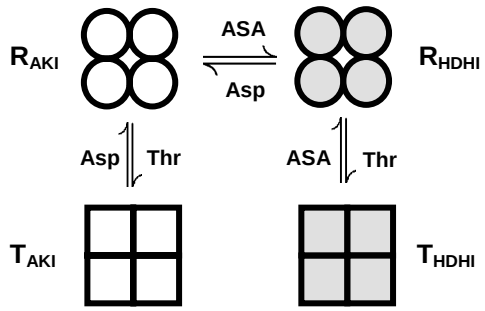

Fig. 3. A model for the bifunctional enzyme AKI-HDHI. The diagram shows that the substrates Asp and ASA will trigger the switch to the kinase and the dehydrogenase activities of the enzyme respectively, while the inhibitor Thr will shift the equilibrium from the active (R) state of the enzyme to the inactive (T) state. Abbreviations used here are: R_{AKI}, active state of Aspartate Kinase I; R_{HDHI}, active state of Homoserine Dehydrogenase I; T_{AKI}, inactive state of Aspartate Kinase I; T_{HDHI}, inactive state of Homoserine Dehydrogenase I; Asp, aspartate; ASA, aspartate semialdehyde; Thr, threonine.

The GMWC inputs for the interchange between AKI kinase and HDHI dehydrogenase activities are described by the two reactions below:

$$\{\{\{AKI \underset{\{\{\},\{\},\{\}\}}{\overset{ASA}{\Longrightarrow}} \{HDHI\},$$

GMWC [cGMWC ⟶ cASA, LGMWC ⟶ LASA, nGMWC ⟶ nASA, KGMWC ⟶ {KmHDHIASA}, kcatGMWC ⟶ kcatAKIHDHI]}}

$$\{\{\{HDHI \underset{\{\{\},\{\},\{\}\}}{\overset{Asp}{\Longrightarrow}} \{AKI\},$$

GMWC [cGMWC ⟶ cAsp, LGMWC ⟶ LAsp, nGMWC ⟶ nAsp, KGMWC ⟶ {KmAKIAsp}, kcatGMWC ⟶ kcat$HDHI$AKI]}}

Here, the interchanging enzymes are on either side of the reaction arrow and the substrates for these enzymes are on top of the arrow. The scheme is followed by the reaction type of enzyme mechanism and the parameters required for such mechanism. On the other hand, the GMWC inputs for the interchange of AKI and HDHI between each of their active (R) and the inactive (T) state are described by the following:

$$\{\{\{Asp \underset{\{\{\},\{Thr\},\{\}\}}{\overset{AKI}{\Longrightarrow}} \{AspP\},$$

GMWC [cGMWC ⟶ cAKI, LGMWC ⟶ LAKI, nGMWC ⟶ nAKI, KGMWC ⟶ {KmAKIAsp, KiAKIAspThr}, kcatGMWC ⟶ kcatAKIAsp]}}

$$\{\{\{ASA \underset{\{\{\},\{Thr\},\{\}\}}{\overset{HDHI}{\Longrightarrow}} \{AKI\},$$

GMWC [cGMWC ⟶ cHDHI, LGMWC ⟶ LHDHI, nGMWC ⟶ nHDHI, KGMWC ⟶ {KmHDHIASA, kiHDHIASAThr}, kcatGMWC ⟶ kcat$HDHI$ASA]}}

Here, and for the rest of the reaction schemes, the substrates are listed on the left hand side of the reaction arrow, the products on the right hand side of the arrow, the enzyme on top of the

arrow and the effectors, if present, below the arrow. This scheme is followed by the reaction type of enzyme mechanism and the parameters required for such mechanism. Cellerator translates the above GMWC models into the fraction of substrate saturated enzyme (Yf) as a function of the substrate (Asp for AKI and ASA for HDHI respectively) and the inhibitor (Thr) as seen in Eq. (1) and Eq. (2).

$$Yf = \frac{\left(\frac{Asp\left(1+\frac{Asp}{KmAKIAsp}\right)^{-1+nAKI}}{KmAKIAsp} + \frac{c\,L\,Asp\left(1+\frac{c\,Asp}{KmAKIAsp}\right)^{-1+nAKI}\left(1+\frac{Lys}{KiAKIAspLys}\right)^{nAKI}}{KmAKIIIAsp} \right)}{\left(1+\frac{Asp}{KmAKIAsp}\right)^{nAKI} + L\left(1+\frac{c\,Asp}{KmAKIAsp}\right)^{nAKI}\left(1+\frac{Lys}{KiAKIAspLys}\right)^{nAKI}} \quad (1)$$

$$Yf = \frac{\left(\frac{ASA\left(1+\frac{ASA}{KmHDHIASA}\right)^{-1+nHDHI}}{KmHDHIASA} + \frac{c\,L\,ASA\left(1+\frac{c\,ASA}{KmHDHIASA}\right)^{-1+nHDHI}\left(1+\frac{Thr}{KiHDHIASAThr}\right)^{nHDHI}}{KmHDHIASA} \right)}{\left(1+\frac{ASA}{KmHDHIASA}\right)^{nHDHI} + L\left(1+\frac{c\,ASA}{KmHDHIASA}\right)^{nHDHI}\left(1+\frac{Thr}{KiHDHIASAThr}\right)^{nHDHI}} \quad (2)$$

3.2. Aspartate kinase III (AKIII)

AKIII is an allosteric enzyme inhibited by lysine and is also described by the GMWC model:

$$\{\{\{Asp \underset{\{\{\},\{Lys\},\{\}\}}{\overset{AKIII}{\Rightarrow}} \{AspP\},$$

GMWC [cGMWC → cAKIII, LGMWC → LAKIII, nGMWC → nAKIII, KGMWC → {KmAKIIIAsp, KiAKIIIAspLys}, kcatGMWC → kcat$AKIII$Asp]}}

Cellerator translates this model into the fraction of substrate saturated enzyme (Y_f) as a function of the substrate (Asp) and the inhibitor (Lys) as seen in Eq. (3).

$$Yf = \frac{\left(\frac{ASA\left(1+\frac{ASA}{KmHDHIASA}\right)^{-1+nHDHI}}{KmHDHIASA} + \frac{c\,L\,ASA\left(1+\frac{c\,ASA}{KmHDHIASA}\right)^{-1+nHDHI}\left(1+\frac{Thr}{KiHDHIASAThr}\right)^{nHDHI}}{KmHDHIASA} \right)}{\left(1+\frac{ASA}{KmHDHIASA}\right)^{nHDHI} + L\left(1+\frac{c\,ASA}{KmHDHIASA}\right)^{nHDHI}\left(1+\frac{Thr}{KiHDHIASAThr}\right)^{nHDHI}} \quad (3)$$

3.3. Aspartate semialdehyde dehydrogenase (ASD)

This step catalyzed by ASD is a reversible two substrate/three products reaction (BiTri). Its kMech inputs are:

$$Enz\ [\{AspP, NADPH\} \overset{ASD}{\Leftrightarrow} \{ASA, NADP, Phosphate\},$$
BiTri [kfASDAspP, krASDAspP, KcatASDAspP]]

$$Enz\ [\{ASA, NADP, Phosphate\} \overset{ASD}{\Leftrightarrow} \{AspP, NADPH\},$$
TriBi [kfASDASA, krASDASA, KcatASDASA]]

The first kMech input is for the forward reaction, and the second kMech input is for the reverse reaction. The metabolites in the bracket on the left side of arrows are substrates, and on the right are products. Enzyme names are above the arrows. BiTri and TriBi indicate

enzyme mechanisms. Variable names with k_f - prefixes are rate constants of the enzyme-substrate associations; variable names with k_r - prefixes are rate constants of the enzyme substrate dissociations; variable names with k_{cat} - prefixes are catalytic rate constants for the formation of products. The detail description of how kMech parsing the enzyme models into association-dissociation reactions in Cellerator syntax, then translated into ODEs can be found in our previous publication[1] and at the UCI, IGB web site.

3.4. Homoserine kinase (HSK)

The step catalyzed by HSK is a two substrate/two products reaction (BiBi) reaction competitively inhibited by threonine (Thr). Its kMech input is:

```
                  HSK
Enz [{Hse, ATP} ⟺ {HseP, ADP},
BiBi [ kfHSKHse, krHSKHse, kcat$HSK$Hse], CI [ Thr, kfiHSKHseThr, kriHSKHseThr]]
```

CI indicates competitive inhibition. Thr is the competitive inhibitor that competes with substrate for the same binding site. Variable names with a k_{fi} - prefix are rate constants of the enzyme-inhibitor associations; variable names with a k_{ri} - prefix are rate constants of the enzyme-inhibitor dissociations.

3.5. Threonine synthase (TS)

The last step is a simple one substrate/one product reaction catalyzed by TS. This basic enzyme model is provided by Cellerator:

```
{{ASA ⟶ Ø, kfASA}}
{{Hse ⟶ Ø, kfHse}}
{{Thr ⟶ Ø, kfThr}}
```

4. Results

4.1. Data fitting for the GMWC model

In order to apply the GMWC model to AKI, HDHI and AKIII, several parameters are required as described above. *S*, *A*, *I*, *n*, K_m, K_a, and K_i, are usually available in literature. However, the values of *c* and *L* are often not available. Such values can be calculated by fitting substrate saturation curves in the presence and absence of various inhibitor concentrations.[1, 15] Here, the *c* and *L* values for AKI, HDHI and AKIII were calculated by finding the minimum sum of squared differences between theoretical data and experimental data with the non-linear programming Mathematica™ function, FindMinimum and fitting data from inhibition curves in the presence of the substrate available in the literature.[16, 17, 18] For example, the fractional saturation of AKIII in the presence of several concentrations of the inhibitor, Lys and how this fits with data obtained from the literature, is shown in Fig. 4.

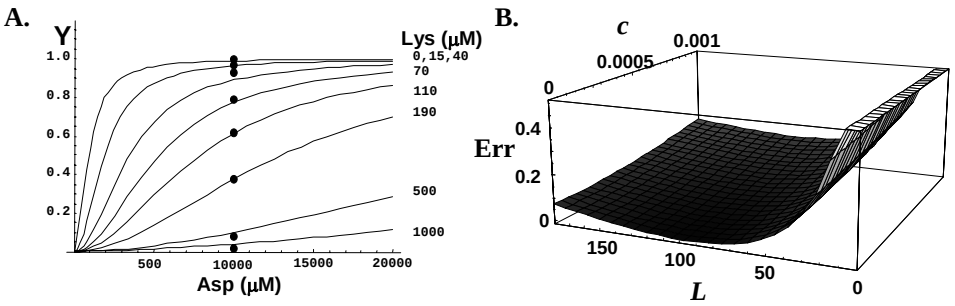

Fig. 4. Optimization of physical constants, c and L, for the concerted transition allosteric Monod, Wyman, and Changeux (MWC) model. (A) Black dots represent experimental measurement of Y_f values for *E. coli* Aspartate Kinase III with substrate (Aspartate, Asp) and various inhibitor (Lysine, Lys) concentrations. The solid curves represent theoretical Y_f values of calculated with c and L values determined from Panel 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 from 0 to 1800. The pair of c and L values that produce minimum errors are c=0.00005, L=1000.

In this case, and due to the lack of experimental data, we were limited to constructing our fitting curve using only one concentration of the substrate Asp and several concentrations of the inhibitor Lys. This fitting was sufficient for the calculation of the c and L values of AKIII. Mathematica™ notebooks fitting the experimental data of AKI, HDHI and AKIII, and computing the c and L values, respectively, are available online at our web site listed above. Our computations show that for AKI, c = 0.0001 and L = 10000, for AKIII, c = 0.00005 and L = 1000 and for HDHI, c = 0.008 and L = 3000. So in the case of these three enzymes, $c << 1$ and $L >> 1$ which indicates that the substrates will only bind to the enzymes in the active state, and the majority of enzymes will be in the inactive state when no substrate, activator or inhibitor are present. The values of c and L can be varied between 10-fold and one tenth of the estimated values above without changing the concentration of threonine predicted in our simulation significantly.

4.2. Threonine biosynthesis is feedback regulated by threonine

To simulate constant flux, the first derivatives of substrates such as Asp, ATP and NADPH were set to zero. Relative enzyme concentrations were inferred from DNA Microarrays data. Initial concentrations of threonine and the intermediates Asp-P, ASA, Hse and Hse-P were set to zero. Substrates of the pathway were set to values of intracellular concentrations reported in the literature, such as Asp whose initial concentration was set to 3600 μM.[19] Forward and reverse rate constants were approximated as described previously.[1] The differential equations were solved by Mathematica™ and plots for the formation of intermediates and products versus time were generated. As shown in figure 4, the concentration of Asp was kept at steady state conditions as reported in literature.[19] The concentrations of the other intermediates start at zero and reach a steady state level after a certain time has passed. At the beginning, a sharp rise in the concentrations of Asp-P, ASA, Hse and Hse-P is observed followed by a sharp rise in the concentration of threonine. This initial rise in the concentrations of Asp-P, ASA, Hse and Hse-P is followed by a sharp drop as soon as the concentration of threonine rises which demonstrates feedback inhibition mechanisms on AKI, HDHI and HSK by threonine. As a consequence, threonine levels start to drop and partially release inhibition in the concentrations of intermediates (e.g. ASA) (Fig. 5). At the end, all metabolites reach their

respective concentrations at steady state conditions. The intracellular concentration of threonine reported in literature is within the range of 290 to 520 μM which closely matches that seen in the simulation. [20, 21]

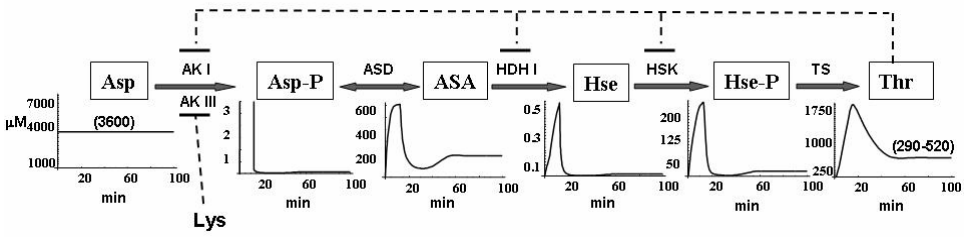

Fig. 5. Simulation of carbon flow through the threonine biosynthesis pathway. The graphical insets show the approach (minutes) to steady state (μM) synthesis and utilization of the substrates, intermediates, and end-products of the pathway. Abbreviations used here are: Asp, aspartate; AspP, Aspartyl phosphate; ASA, aspartate semialdehyde; Hse, homoserine; HseP, homoserine phosphate; Thr, threonine; AKI, aspartate kinase I; AKIII, aspartate kinase III; HDHI, homoserine dehydrogenase I; ASD, semialdehyde dehydrogenase; HSK, homoserine kinase; TS, threonine synthase. Feedback inhibition is indicated by dashed lines. Starting concentration of Aspartate is kept constant at 3600 μM.

4.3. The mathematical model predicts the partition of enzyme activities of a bifunctional enzyme

The allosteric bifunctional enzyme AKI-HDHI was modeled as described earlier in Fig. 3. The enzyme has the kinase and dehydrogenase activities and the switch to either activity depends on the amounts of substrate Asp or ASA available. The intracellular concentration of AKI-HDHI protein inferred from DNA microarrays is 16 μM and in the model, each activity is given an initial concentration of 8 μM when both substrates are not present. As seen in Fig. 6, constant Asp (3600 μM) leads to a higher fraction of the enzyme having an AKI kinase activity initially. This fraction drops when the concentration of ASA (substrate for HDHI) increases. At the same time, an increase in the fraction of enzymes having the HDHI dehydrogenase activity is observed. The opposite effect is then observed when threonine feedback inhibition occurs. The partition of both enzyme activities stabilizes when the concentrations of both substrates reach steady states. The model predicts the fraction of enzyme having the kinase activity is around 15/16 while that of the fraction having the dehydrogenase activity is around 1/16. This kind of prediction can be generally applied to other bifunctional enzymes important for the regulation of metabolic flux channeling.

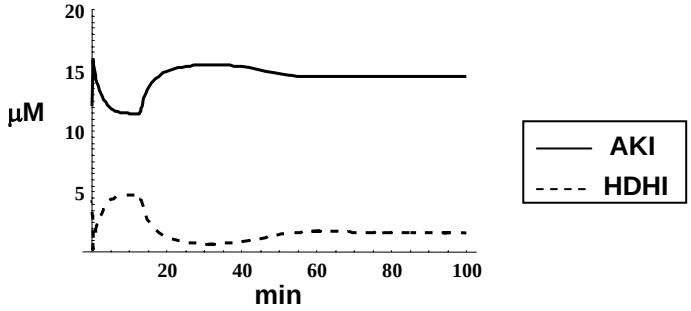

Fig. 6. Simulation of the fractional activities of the allosteric bifunctional enzyme aspartate kinase I-homoserine dehydrogenase I. The total concentration of the enzyme used here is 16 μM and the concentration of aspartate is kept constant at 3600 μM. The data show the concentrations (μM) of both kinase and dehydrogenase activities as steady state is approached (minutes), where the fraction of enzyme having the kinase activity is around 15/16 while that of the fraction having the dehydrogenase activity is around 1/16.

4.4. Simulating the metabolic engineering of a threonine over-producing E. coli K-12 strain

One of the goals of systems biology is to facilitate the metabolic engineering of microorganisms that overproduce certain chemicals such as amino acids. One important approach utilizes the production of strains with feedback resistant mutations in the enzymes that regulate the metabolic pathway for the production of the desired product. To test the ability of our model to simulate such an effect, the production of threonine was determined in the presence of various concentrations of lysine in both wild type and lysine feedback resistant AKIII mutant. As seen in Fig. 7, increasing concentrations of lysine result in a stronger inhibition and a lower threonine production at steady state in the wild type. On the other hand, an overproduction of threonine is observed in the lysine feedback resistant AKIII mutant even in the presence of a high concentration of lysine. This phenomenon is observed even in the presence of wild type AKI which is normally feedback inhibited by threonine. These results are consistent with the data published by Ogawa-Miyata *et al.*[21] where lysine feedback resistant AKIII mutants were shown to boost threonine production following the addition of lysine.

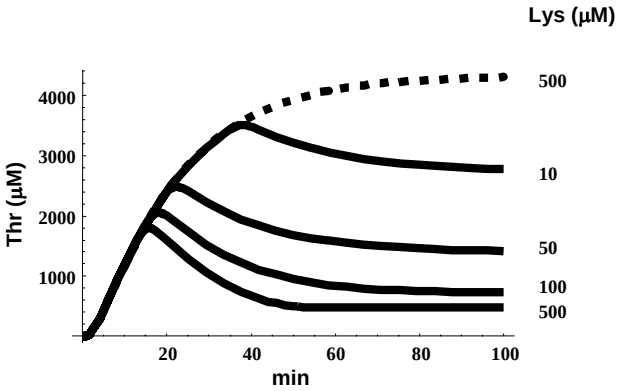

Fig. 7. Simulated effect of lysine feedback resistant aspartate kinase III on threonine production. The data show the concentrations (μM) of threonine as steady state is approached (minutes) in the presence of different concentration

(μM) of lysine. The solid line represents wild type AKIII while the dashed line represents lysine feedback resistant AKIII mutant.

5. Conclusion

To model the behavior of complex biological systems, we have elected a "bottom-up" approach that incorporates detailed enzyme kinetic and pathway-specific regulatory mechanisms from the literature into our model. Using kMech/Cellerator, models for enzyme mechanisms and their patterns of regulation are converted automatically into association-dissociation reactions, and then into differential equations. These equations are solved by Mathematica™ to simulate the model and generate the graphical output. In addition to simplifying the underlining mathematics of writing down differential equations, this approach allows us to examine the biochemical behavior of metabolites and enzyme states in the pathway with greater detail. Furthermore, the GMWC model expands our ability to simulate the behavior of allosteric enzymes and their feedback inhibition mechanisms through binding of different substrates, activators and inhibitors. As demonstrated here in two different cases, our simulation closely matches data in the literature. In addition, the bifunctional enzyme model allows us to predict the partition of enzyme activities in the steady state, and new hypotheses and experimental designs can be generated through this type of prediction.

To these ends, we have developed a technique for incorporating realistic enzyme mechanisms such as multiply regulated allosteric enzymes within metabolic models. The resulting rate laws take a novel algebraic form, and it cannot be promised that yet more realistic models of enzymes and regulation by molecular complexes will remain within this expanded vocabulary of algebraic forms for rate laws. Indeed consideration of "molecular machines" in the cell leads us to speculate that detailed models of molecular mechanism may become increasingly elaborate in their algebraic form. The question then arises as to whether the realistic models that reflect diverse mechanisms can be homogenized into approximating dynamics that are simpler than GMWC from the point of view of its traditional algebraic expression (Of course GMWC is very simply invoked in kMech.) We do not think that enough groundwork has been done yet in developing detailed enzyme-centric models of a variety of metabolic networks, including their regulation, to conclude with confidence whether they can be uniformly approximated within a modeling framework of simpler and more uniform algebraic form. However, we discuss here some of the future possibilities in this direction.

The most widely used homogeneous modeling framework in the field of metabolic engineering is metabolic control analysis (MCA). In contrast to our enzyme-centric approach, one major limitation of the traditional MCA is its assumption that a system is always in a steady state. Therefore, the traditional MCA approach is not suitable for modeling the transient phenomena from perturbations of metabolic parameters such as enzyme expression levels and the concentrations of metabolites and ligands that regulate enzyme activities by feedback.[22] Another major limitation of traditional MCA is its assumption of linearity, which is violated by allosteric enzyme mechanisms. These limitations are removed by several other modeling techniques such as the Power-law approximation (S-system),[23] the (Log)Linear refinement of the MCA model[22] and lin-log kinetics[24] which have been developed to consider enzyme concentrations, feedback regulators and reversible enzyme reactions in the models. Other homogeneous-form but nonlinear, dynamical network frameworks for

phenomenological modeling have been developed for regulatory networks outside of the field of metabolism.[25]

Whether any such models can closely approximate an enzyme-centric model such as ours is a potentially fruitful line of future investigation. Unfortunately, there is reason to believe that biology is more complicated than what those models can realistically reflect without the introduction of numerous extra dynamical variables. In the threonine biosynthetic pathway presented here as an example, in order to model the aspartate kinase isozymes (AKI, AKII and AKIII) that are controlled by different modes of regulation, the MCA models must allow multiple fluxes for the conversion of Asp to AspP and to model the bi-functional enzyme (AKI-HDHI) carrying both aspartate kinase and homoserine dehydrogenase activities, the fluxes of Asp to AspP and ASA to Hse have to be dependent on each other. Also in our model for the branched chain amino acid biosynthetic pathway[2], several enzymes are shared by two pathways and their fluxes vary and shift between their partner pathways. These facts point out the fundamental drawback of the MCA approach in considering metabolic networks as a collection of independent chemical conversions. In contrast, our non-linear, enzyme-centric modeling approach relies on simulating metabolic networks as a collection of enzymes depending on their individual catalytic and regulatory mechanisms. Our experience suggests this can be done systematically for models of successively larger and more inclusive metabolic systems. For these reasons, modeling key allosteric regulators using the GMWC model presented here is especially important to facilitate the genetic engineering of metabolic pathways.

Acknowledgments

This work was supported in part by the Biomedical Information Science and Technology Initiative (BISTI) grant (number 4 R33 GM069013) from the National Institute of General Medical Sciences, USA to Eric D. Mjolsness and NIH grants GM55073 and GM68903 awarded to G. Wesley Hatfield. Tarek Najdi and Chin-Rang Yang are trainees of the Biomedical Informatics Training (BIT) Program of the Univ. of California-Irvine Institute for Genomics and Bioinformatics, and the recipients of pre-doctoral and postdoctoral fellowships, respectively, from National Institutes of Health, National Research Service Award 5 T15 LM007443 from the National Library of Medicine, USA.

References

1. C.R. Yang, B.E. Shapiro, E.D. Mjolsness and G.W. Hatfield, “An enzyme mechanism language for the mathematical modeling of biochemical pathways,” *Bioinformatics*, **21**, 774-780 (2005).
2. C.R. Yang, B.E. Shapiro, S.P. Hung, E.D. Mjolsness and G.W. Hatfield, “A Mathematical Model of the Branched Chain Amino Acid Biosynthetic Pathways of *Escherichia coli*,” *J Biol Chem*, **280**(12), 11224-11232 (2005).
3. J. Monod, J. Wyman and J.P. Changeux, “On the Nature of Allosteric Transitions: A Plausible Model,” *J Mol Bio*, **12**, 88-118 (1965).
4. H. E. Umbarger, “Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine,” *Science*, **123,** 848 (1956).
5. J. Janin and G.N. Cohen, “The threonine-sensitive homoserine dehydrogenase and aspartokinase activities of Escherichia coli K 12. A study of the allosteric equilibrium,” *Eur J Biochem*, **11**(3), 520-529 (1969).

6. C. Chassagnole, B. Rais, E. Quentin, D.A. Fell and J.P. Mazat, "An integrated study of threonine-pathway enzyme kinetics in Escherichia coli," *Biochem J*, **356**(2), 415-423 (2001).
7. J.P. Mazat and J.C. Patte, "Lysine-sensitive aspartokinase of Escherichia coli K12. Synergy and autosynergy in an allosteric V system," Biochemistry, **15**(18), 4053-4058 (1976).
8. W.E. Karsten and R.E. Viola, "Chemical and kinetic mechanisms of aspartate-beta-semialdehyde dehydrogenase from Escherichia coli," *Biochim Biophys Acta*, **1077**(2), 209-219 (1991).
9. S.L. Shames and F.C. Wedler, "Homoserine kinase of Escherichia coli: kinetic mechanism and inhibition by L-aspartate semialdehyde," *Arch Biochem Biophys*, **235**(2), 359-370 (1984).
10. M. Szczesiul and D.E. Wampler, "Regulation of a metabolic system in vitro: synthesis of threonine from aspartic acid," *Biochemistry*, **15**(10), 2236-2244 (1976).
11. B.E. Shapiro, A. Levchenko, E.M. Meyerowitz, B.J. Wold and E.D. Mjolsness, "Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations," *Bioinformatics*, **19**, 677-678 (2003).
12. G.W. Hatfield and H.E. Umbarger, "Threonine deaminase from *Bacillus subtilis. II*. The steady state kinetic properties," *J Biol Chem*, **245**(7), 1742-1747 (1970).
13. T.L. Hill (Ed.), *Cooperativity Theory in Biochemistry, Steady-State and Equilibrium Systems.*, Springer-Verlag, New York, 1985.
14. S. Falk, A. Guay, C. Chenu, S. D. Patil and A. Berteloot, "Reduction of an eight-state mechanism of cotransport to a six-state model using a new computer program," *Biophys J*, **74**, 816-830 (1998).
15. I. H. Segel (Wiley Classics Library Ed.), *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems, p427, equation (VII-66)*, Wiley, New York, 1993.
16. J.C. Patte, P.T. Bachi and G. Cohen, "The threonine-sensitive homoserine dehydrogenase and aspartokinase activities of *Escherichia coli*. Evidence that the two activities are carried by a single protein," *Biochem Biophys Acta*, **128**, 426-439 (1966).
17. J.P. Mazat and J.C. Patte, "Lysine-sensitive aspartokinase of *Escherichia coli* K12. Synergy and autosynergy in an allosteric V system," *Biochemistry*, **15**(18), 4053-4058 (1976).
18. G. W. Hatfield, "The Regulation of L-threonine deaminase in Bacillus subtilis be Repression and Endproduct Inhibition," Ph.D. Thesis. Purdue University (1970).
19. S.C. Quay, T.E. Dick and D.L. Oxender, "Role of transport systems in amino acid metabolism: leucine toxicity and the branched-chain amino acid transport systems," *J Bacteriol*, **129**(3), 1257-1265 (1977).
20. S. Epelbaum, R.A. LaRossa, T.K. VanDyk, T. Elkayam, D.M. Chipman and Z. Barak, "Branched-chain amino acid biosynthesis in Salmonella typhimurium: a quantitative analysis," *J Bacteriol*, **180**(16), 4056-4067 (1998).
21. Y. Ogawa-Miyata, H. Kojima and K. Sano, "Mutation analysis of the feedback inhibition site of aspartokinase III of *Escherichia coli* K-12 and its use in L-threonine production," *Biosci Biotechnol Biochem*, **65**(5), 1149-11454 (2001).
22. V. Hatzimanikatis, J.E. Bailey, "MCA has more to say," *J Theor Biol*, **182**(3), 233-242 (1996).
23. R. Alves, M.A. Savageau, "Comparing systemic properties of ensembles of biological networks by graphical and statistical methods," *Bioinformatics*, **16**(6), 527-533 (2000).
24. L. Wu, W. Wang, W.A. van Winden, W.M. van Gulik, J.J. Heijnen, "A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics," *Eur J Biochem*, **271**(16), 3348-3359 (2004).
25. E. Mjolsness, D.H. Sharp, J. Reinitz, "A connectionist model of development," *J Theor Biol*, **152**(4), 429-453 (1991).