A Mathematical Model of Allosteric Regulation for Threonine Biosynthesis in *Escherichia coli*

Tarek S. Najdi[1,3,*], Chin-Rang Yang[1,3,*], Bruce E. Shapiro[4],

G. Wesley Hatfield[1,3] and Eric D. Mjolsness[1,2]

[1]**Institute for Genomics and Bioinformatics, University of California, Irvine**
 Irvine, California 92697
[2]**Donald Bren School of Information and Computer Science,**
[3]**Department of Microbiology and Molecular Genetics, College of Medicine,**
[4]**Jet Propulsion Laboratory, California Institute of Technology**
 4800 Oak Grove Drive, Pasadena, California 91109 U.S.A.
* These authors contributed equally to this work.

Contact: Biology correspondence should be addressed to G.W.H. (gwhatfie@uci.edu).

Computation correspondence should be addressed to E.D.M. (emj@uci.edu).

ABSTRACT

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 <u>M</u>onod, <u>W</u>yman, <u>C</u>hangeux (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 a way for a more generalized prediction of the behavior of bifunctional enzymes.

1. INTRODUCTION

Systems biology utilizes 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.

We have previously described a mathematical model tool for complex enzyme mechanism, kMech (1) 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 <u>M</u>onod, <u>W</u>yman, <u>C</u>hangeux (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 take multiple substrate, activator and inhibitor binding sites, respectively verse the original MWC can only take one substrate, activator and inhibitor binding sites, respectively. The GMWC model is especially useful for modeling the threonine biosynthesis in *E. coli*, since there are three allosteric enzymes (out of total six enzymes) in this pathway. It is upstream of the pathways for the biosynthesis of L-isoleucine, L-valine, and L-leucine. This work is another step towards the elucidation of the systems biology of the model organism, *E. coli*, since most of allosteric enzymes play important roles in the regulation of metabolic flux channeling.

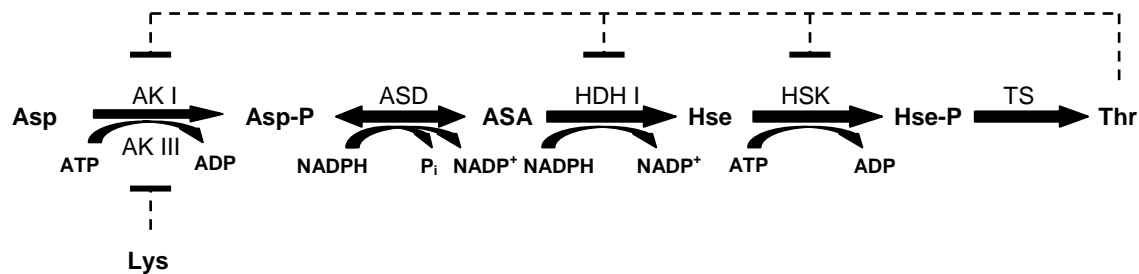Threonine synthesis starting from Aspartate is a five-step metabolic pathway (Fig. 1).

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; and 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.

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, aspartate kinase I (AKI), aspartate kinase II (AKII) and aspartate kinase III (AKIII). AKI is a bifunctional enzyme (AKI-HDHI) carrying both aspartate kinase and homoserine dehydrogenase (HDHI) activities (Step three). It is an allosteric enzyme made up of four subunits and inhibited by threonine (4). The kinetic behavior of this enzyme can be model by the MWC 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 (*Yf*) of enzyme also can be 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. The enzyme has no effectors and exists at very low levels in *E. coli* K12 (5). Therefore, we did not include it in our simulation. AKIII is an allosteric enzyme made up of two subunits and inhibited by lysine (6). Its kinetic behavior is also described by the MWC model (6).

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 (7). 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 MWC model (3).

The forth step of the pathway is a BiBi reaction with the two substrates HSE and ATP and the two products homoserine phosphate (HseP) and ADP. This step is catalyzed by homoserine kinase (HSK) and competitively inhibited by threonine (8). 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 (9).

The enzyme kinetics of this pathway has been subjected to extensive study over the past 40 years and many of the kinetic and physical parameters are reported in the literature. As a part of the effort to elucidate the systems biology of *E. coli*, we aim to model the metabolic pathway of synthesis of threonine from aspartate and integrate the data reported in literature to produce a simulation that describes carbon flow through this pathway, and predict the partition of enzyme activities within the bi-functional enzymes (AKI-HDHI).

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 reactions that describe complex cascades as well as the differential equations that are derived from this chemical network (10). These differential equations are solvable by Mathematica. kMech is a Cellerator language extension that describes enzyme mechanisms for the mathematical modeling of metabolic pathways. Such mechanisms are codified to generate a set of elementary reactions that can be translated by Cellerator into ordinary differential equations (ODEs) solvable by Mathematica (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 usually not available. For this reason, we previously described a method for approximating enzyme concentrations from DNA microarray data (1, 2). This can be accomplished based on the fact that purification tables in the literature (11) suggest that the intracellular concentration of threonine deaminase in *E. coli* is 4 µM and on recent experiments that show a positive correlation between mRNA levels measured with DNA microarrays and protein abundance in *E. coli* (12). Consequently, the intracellular levels of enzymes can be inferred and we have shown that this is a reasonable method and that simulations using such inferred values correlate well with experimental values (2). However, because of the many variables that can influence *in vitro* measurements, corrections are sometimes necessary for the kinetic measurements (K_m, k_{cat}) values to optimize the model to fit the experimental data.

A Mathematica notebook file 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 is available at the University of California, Irvine, Institute for Genomics and Bioinformatics 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), AKI, HDHI and AKIII 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 (Threonine 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 two equations:

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

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

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; 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 (13).

Instead of implementing the original MWC model, we developed a more flexible model in Cellerator, which generalizes the MWC model in the following ways. In MWC there is one substrate, and on each identical subunit there is one binding site for that substrate. There is also one binding site per subunit for an activator and another for an inhibitor. Here, we generalize to the case of multiple substrates. Each substrate will have its own activator and inhibitor. Each triple of a substrate, its activator, and its inhibitor will have a corresponding triple of binding sites on each subunit.

To find the partition function and therefore the dynamics for this Generalized MWC (GMWC) model, we generalize the above 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:

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

Where q is the number of substrates; $s_q = [S_q]/K_{Mq}$ is the fugacity of the substrate at any of its (identical) sites, where $[S_q]$ 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 (as in MWC itself). 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 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 total partition function is then

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

From this form, we can see that the full production rate is proportional to the fraction of subunits at which all substrates are present and no inhibitor is present, which is:

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

This is the form of the GMWC model we will use in our model here. A more detailed version of the model is also available to incorporate the effects of competitive inhibition at the substrate and activator binding sites:

$$\frac{d[P_i]}{dt} = k_{cat}[E] \frac{\prod_q \left(1+a_q+\bar{a}_q\right)^n \prod_q s_q \prod_q (1+s_q+\bar{s}_q)^{n-1} + L \prod_q (cs_q) \prod_q (1+cs_q+\bar{s}_q)^{n-1} \prod_q (1+i_q)^{n-1}}{\prod_q \left(1+a_q+\bar{a}_q\right)^n \prod_q (1+s_q+\bar{s}_q)^n + L \prod_q (1+cs_q+\bar{s}_q)^n \prod_q (1+i_q)^n}$$

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

The detailed documentation of the GMWC model can be found in the Cellerator web site and the "Help" section in Cellerator.

3. PATHWAY MODELING

As discussed previously (1, 2), traditional approaches to model enzyme kinetic pathways mostly relied on Michaelis-Menten kinetic equation for one substrate/one product reactions and the King-Altman method to derive equations for complex multiple reactant reactions. In those 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 to either one of the active (R) states 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. 2).
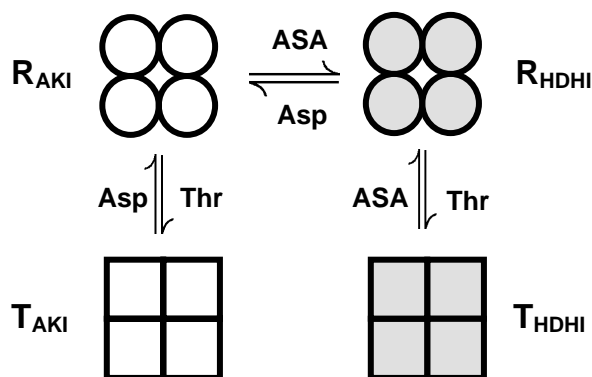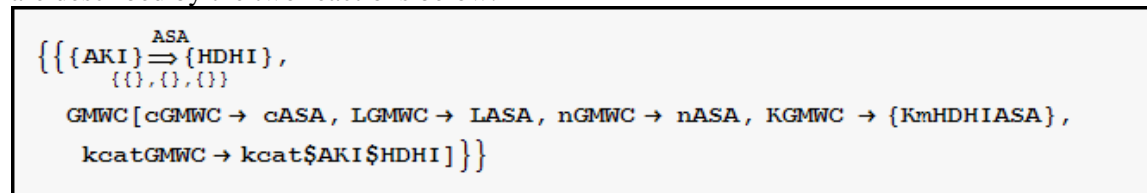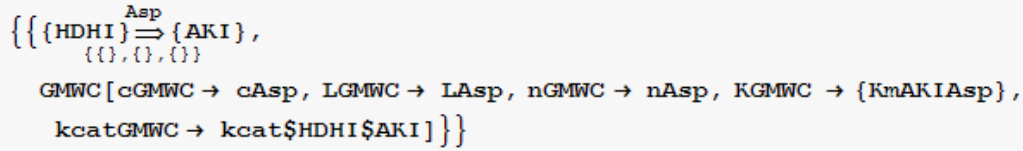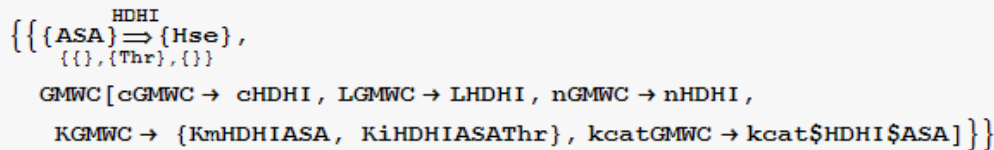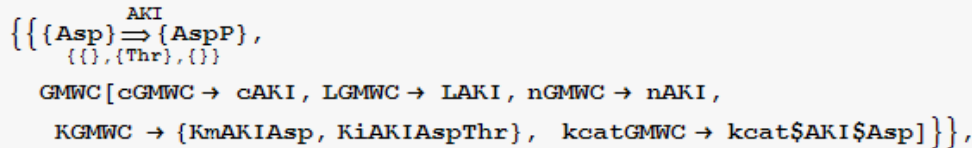

Fig.2 . A Model for the Bifunctional Enzyme AKI-HDHI. The diagram shows that the substrates Asp and ASA will trigger the switch to the kinase and the dehydrogenase activities of the enzyme respectively, while the inhibitor Thr will shift the equilibrium from the active (R) state of the enzyme to the 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:

```
         ASA
{{{AKI} ⟹ {HDHI},
      {{},{},{}}

  GMWC[cGMWC → cASA, LGMWC → LASA, nGMWC → nASA, KGMWC → {KmHDHIASA},
    kcatGMWC → kcat$AKI$HDHI]}}
```

```
{{{HDHI}⟹{AKI},
        Asp
     {{},{},{}}
   GMWC[cGMWC → cAsp, LGMWC → LAsp, nGMWC → nAsp, KGMWC → {KmAKIAsp},
    kcatGMWC → kcat$HDHI$AKI]}}
```

On the other hand, the GMWC inputs for the interchange of AKI and HDHI between the active (R) and the inactive (T) state are the following:
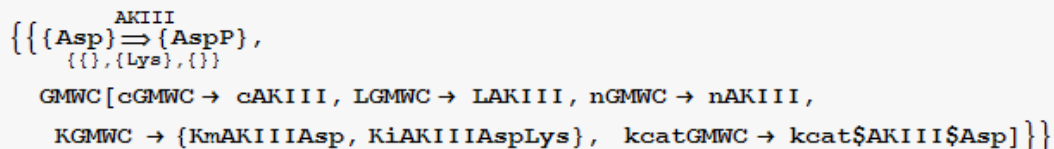
```
{{{Asp}⟹{AspP},
        AKI
     {{},{Thr},{}}
   GMWC[cGMWC → cAKI, LGMWC → LAKI, nGMWC → nAKI,
    KGMWC → {KmAKIAsp, KiAKIAspThr},  kcatGMWC → kcat$AKI$Asp]}},

{{{ASA}⟹{Hse},
        HDHI
     {{},{Thr},{}}
   GMWC[cGMWC → cHDHI, LGMWC → LHDHI, nGMWC → nHDHI,
    KGMWC → {KmHDHIASA, KiHDHIASAThr}, kcatGMWC → kcat$HDHI$ASA]}}
```

Cellerator translates the above GMWC models into the following equations that describe 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):

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

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

3.2 Aspartate Kinase III (AKIII)

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

```
{{{Asp}⟹{AspP},
        AKIII
     {{},{Lys},{}}
   GMWC[cGMWC → cAKIII, LGMWC → LAKIII, nGMWC → nAKIII,
    KGMWC → {KmAKIIIAsp, KiAKIIIAspLys},  kcatGMWC → kcat$AKIII$Asp]}}
```
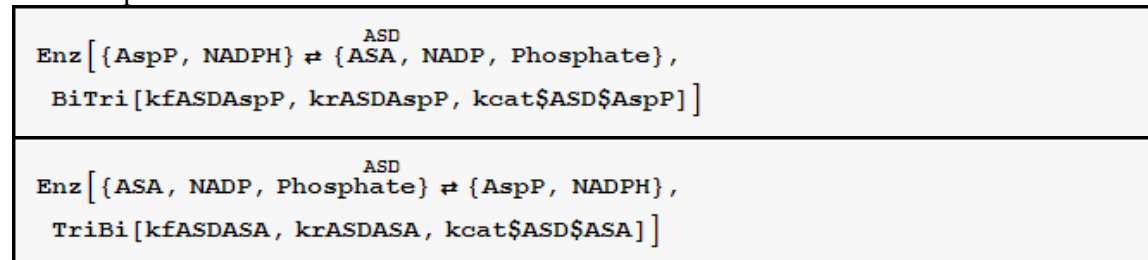
Cellerator translates this model into the fraction of substrate saturated enzyme (*Yf*) as a function of the substrate (Asp) and the inhibitor (Lys):

$$Yf = \frac{\left(\dfrac{Asp\left(1+\frac{Asp}{KmAKIIIAsp}\right)^{-1+nAKIII}}{KmAKIIIAsp} + \dfrac{c\,L\,Asp\left(1+\frac{c\,Asp}{KmAKIIIAsp}\right)^{-1+nAKIII}\left(1+\frac{Lys}{KiAKIIIAspLys}\right)^{nAKIII}}{KmAKIIIAsp} \right)}{\left(1+\frac{Asp}{KmAKIIIAsp}\right)^{nAKIII} + L\left(1+\frac{c\,Asp}{KmAKIIIAsp}\right)^{nAKIII}\left(1+\frac{Lys}{KiAKIIIAspLys}\right)^{nAKIII}}$$
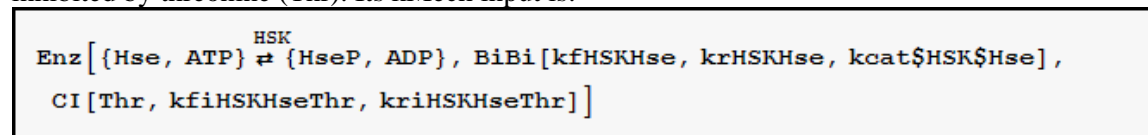
3.3 Aspartate Semialdehyde Dehydrogenase (ASD)

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

```
                    ASD
Enz[{AspP, NADPH} ⇌ {ASA, NADP, Phosphate},
  BiTri[kfASDAspP, krASDAspP, kcat$ASD$AspP]]
```

```
                            ASD
Enz[{ASA, NADP, Phosphate} ⇌ {AspP, NADPH},
  TriBi[kfASDASA, krASDASA, kcat$ASD$ASA]]
```

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 a k_f- prefix are rate constants of the enzyme-substrate associations; variable names with a k_r- prefix are rate constants of the enzyme substrate dissociations; variable names with a kcat- prefix 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 web site listed above.

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:

```
       TS
{{HseP ⇌ Thr, kfTSHseP, krTSHseP, kcat$TS$HseP}}
```
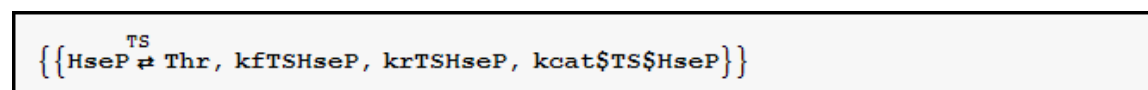
3.6 Consumption of Pathway Byproducts:

Some of the intermediates of the pathway like ASA, Hse and the end-product Thr are used as substrate in other amino acid synthesis pathways in *E. coli*. ASA is used for Lysine biosynthesis, Hse for methionine biosynthesis, and Thr for isoleucine biosynthesis. These consumptions were simulated with the Annihilation model in Cellerator as place holders for the future expansion:

```
{{ASA → ∅, kfASA}}
```

```
{{Hse → ∅, kfHse}}
```

```
{{Thr → ∅, 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 literature (16, 17). 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. 3. 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.
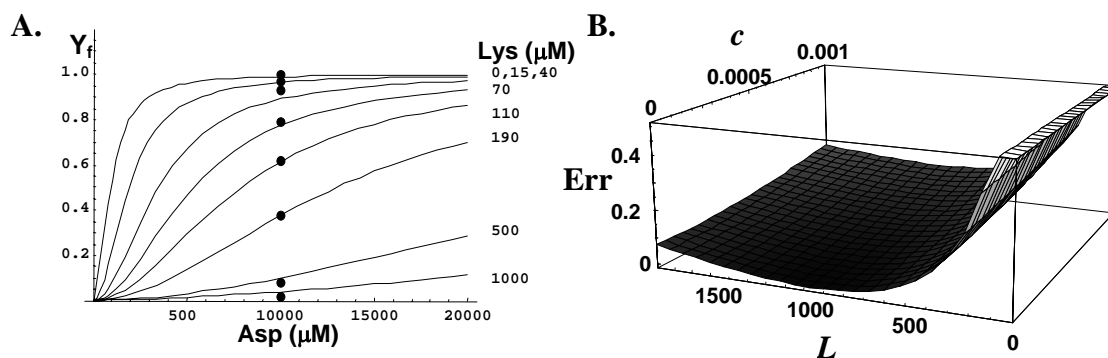

Fig.3 . Optimization of Physical Constants, c and L0, for the Concerted Transition Allosteric Monod, Wyman, and Changeux (MWC) Model. (A) Black dots represent experimental measurement of Y_f values for *E. coli* Aspartate Kinase III with substrate (Aspartate, Asp) and various inhibitor (Lysine, Lys) concentrations. The solid line 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.

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 literature, such as Asp whose initial concentration was set to 3600 µM (18). Forward and reverse rate constants are 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 (18). 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 release inhibition partially as observed by small rises in the concentrations of intermediates (e.g. ASA) (Fig. 4). 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 (18, 19) which closely matches that seen in the simulation.
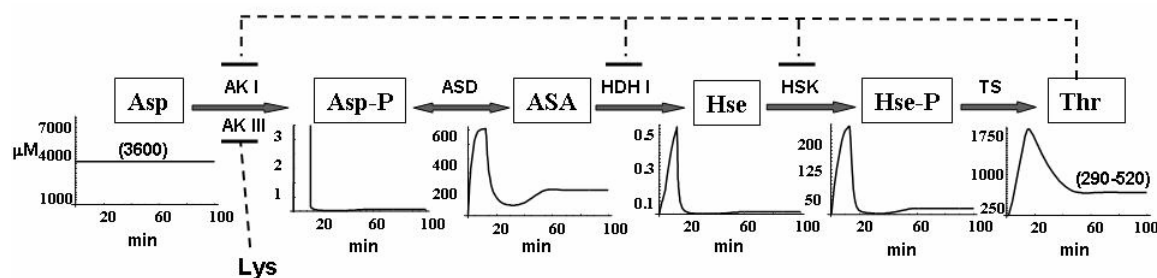

Fig.4 . Simulation of Carbon Flow through the Threonine Biosynthesis Pathway. The graphical insets show the approach (minutes) to steady state (µM) synthesis and utilization of the substrates, intermediates, and end-products of the pathway. Abbreviations used here are: Asp, aspartate; AspP, Aspartyl 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 A Mathematical Model that Predicts the Partition of Enzyme Activities of a Bifunctional Enzyme

The allosteric bifunctional enzyme AKI-HDHI was modeled as described earlier in figure 2. 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 figure 5, 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 which usually are important for the regulation of metabolic flux channeling. It would be very interesting and important if

experimental work can be carried out to confirm these theoretical partitions for these regulatory enzymes, and to better understand how metabolic pathways are regulated by this type of regulation.
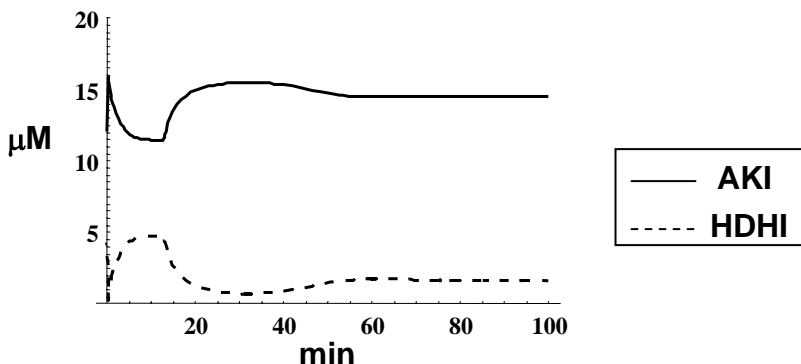

Fig.5 . Simulation of the Fractional Activities of the Allosteric Bifunctional Enzyme Aspartate Kinase I-Homoserine Dehydrogenase I. The total concentration of the enzyme used here is 16 µM and the concentration of aspartate is kept constant at 3600 µM. The data show the concentrations (µM) of both kinase 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.

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 bindings of different substrates, activators and inhibitors. As demonstrated here, our simulation closely matches data in 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. Our approach is especially useful for modeling the key player for the regulation of a given pathways.

6. REFRENCES

1. Yang CR, Shapiro BE, Mjolsness ED, Hatfield GW. An enzyme mechanism language for the mathematical modeling of biochemical pathways. *Bioinformatics* 2004 **in press** (Advance Access published on October 27, 2004; doi: 10.1093/bioinformatics/bti068)

2. Yang CR, Shapiro BE, Hung SP, Mjolsness ED, Hatfield GW. A Mathematical Model of the Branched Chain Amino Acid Biosynthetic Pathways of *Escherichia coli*. 2004 **submitted**

3. Monod J, Wyman J, Changeux JP. On the Nature Of Allosteric Transitions: A Plausible Model. *J Mol Biol.* 1965;12:88-118

4. Janin J, Cohen GN. The threonine-sensitive homoserine dehydrogenase and aspartokinase activities of Escherichia coli K 12. A study of the allosteric equilibrium. *Eur J Biochem.* 1969;11(3):520-9

5. Chassagnole C, Rais B, Quentin E, Fell DA, Mazat JP. An integrated study of threonine-pathway enzyme kinetics in Escherichia coli. Biochem J. 2001;356(Pt 2):415-23

6. Mazat JP, Patte JC. Lysine-sensitive aspartokinase of Escherichia coli K12. Synergy and autosynergy in an allosteric V system. *Biochemistry.* 1976;15(18):4053-8

7. Karsten WE, Viola RE. Chemical and kinetic mechanisms of aspartate-beta-semialdehyde dehydrogenase from Escherichia coli. *Biochim Biophys Acta*. 1991;1077(2):209-19

8. Shames SL, Wedler FC. Homoserine kinase of Escherichia coli: kinetic mechanism and inhibition by L-aspartate semialdehyde. *Arch Biochem Biophys*. 1984;235(2):359-70

9. Szczesiul M, Wampler DE. Regulation of a metabolic system in vitro: synthesis of threonine from aspartic acid. *Biochemistry*. 1976;15(10):2236-44

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

11. Calhoun, D. H., Rimerman, R. A., and Hatfield, G. W. *J Biol Chem* 1973: 248, 3511-3516

12. Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehle, M. M., Paegle, E. S., and Hatfield, G. W. *J Biol Chem* 2000; 275: 29672-29684

13. Hatfield, G. W., and Umbarger, H. E. *J Biol Chem* 1970: 245, 1742-1747

14. Falk, S., 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*. 1998; 74: 816-30

15. Segel, I. H. (1993) *Enzyme kinetics : behavior and analysis of rapid equilibrium and steady state enzyme systems, p427, equation(VII-66).* Wiley Classics Library Ed., Wiley, New York

16. Patte, JC, Bachi, PT, Cohen G. 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.* 1966: 128, 426-439

17. Mazat JP, Patte JC. Lysine-sensitive aspartokinase of Escherichia coli K12. Synergy and autosynergy in an allosteric V system. *Biochemistry*. 1976 Sep 7;15(18):4053-8

18. Quay SC, Dick TE, Oxender DL. Role of transport systems in amino acid metabolism: leucine toxicity and the branched-chain amino acid transport systems. *J Bacteriol*. 1977; 129(3):1257-65

19. Epelbaum S, LaRossa RA, VanDyk TK, Elkayam T, Chipman DM, Barak Z. Branched-chain amino acid biosynthesis in Salmonella typhimurium: a quantitative analysis. *J Bacteriol.* 1998; 180(16):4056-67