Consistency of Cybernetic Variables with Gene Expression Profiles: A More Rigorous Test

Frank DeVilbiss

School of Chemical Engineering, Purdue University, West Lafayette, IN 47907

Aravinda Mandli

School of Chemical Engineering, Purdue University, West Lafayette, IN 47907

Doraiswami Ramkrishna (iD)

School of Chemical Engineering, Purdue University, West Lafayette, IN 47907

Diauxic growth of Escherichia coli *is driven by a host of internal, complex regulatory actions. In this classic scenario of cellular control, the cell employs a rational algorithm to modulate its metabolism in a competitive fashion. Cybernetic models of metabolism, whose development now spans three decades, were first formulated to describe regulation of cells in complex, multi-substrate environments. They modeled this scenario using the hypothesis that the formation of the enzymatic machinery is regulated to maximize a return on investment. While this assumption is made on the basis of logical arguments rooted in evolutionary principles, little effort has been taken to validate if enzymes are truly synthesized in the same fashion that is predicted by cybernetic variables. This work revisits the original cybernetic models describing diauxic growth and compares their predictions of enzyme synthesis control with time series gene expression data in microarray and qRT-PCR formats. Three separate studies are made for two different strains of* E. coli. *The first is for the growth of* E. coli *BW25113 on a mixture of glucose and acetate, whose gene expression changes are metered by microarray. Another is also for the sequential consumption of glucose and acetate but involves strain MG1655 and employs qRT-PCR. The final is for* E. coli *MG1655 on glucose and lactose. By demonstrating how cybernetic variables for induced enzyme synthesis mimic the behavior of transcriptional data, a strong argument for using cybernetic models is made.* © 2018 American Institute of Chemical Engineers *Biotechnol. Prog.*, 34:858–867, 2018

Keywords: metabolic engineering, metabolic control, cybernetics, mathematical modeling

Introduction

Cybernetic models, which describe the dynamic regulation of metabolism as a goal-directed process, have been developed over the past three decades to describe a host of complex microbial growth scenarios. The first scenario cybernetic models sought to describe was that of diauxic growth of microorganisms.[1] These models assume that microorganisms are optimal control systems that actively modulate their metabolism through the cybernetic variables for enzyme synthesis u_i and activity v_i. These models further assume that microorganisms possess limited internal resources that they use in an optimal manner to achieve their goals. Using matching and proportional laws to describe enzyme synthesis and activity, respectively, these first cybernetic models were able to predict the distinct growth and lag phases of the diauxie using parameters taken from growth

data on individual substrates based on the notion that cells modulate their metabolisms to optimize growth rates.

From this initial formulation, cybernetic models have been extended to describe much more complex aspects of metabolism via incremental improvements over the years. Cybernetic models have been useful in not only describing complex substrate uptake patterns,[2] but have also yielded successful predictions of intracellular fluxes,[3] gene-knockout behaviors,[4] and a multiplicity of steady states in chemostats.[5] While cybernetic modeling's ability to predict complex cellular phenomena helps to realize the utility of this approach, no effort has taken place to verify the extent to which the cybernetic control mechanisms mimic cellular regulation. More specifically, the cybernetic variables for enzyme synthesis and activity, u_i and v_i, have not been directly compared with biological data that is representative of the internal regulation of metabolism in cells.

Experimental advances in the past decade have resulted in the collection of large amounts of cellular data which provide insights into regulation in cells.[6,7] Various techniques such as microarrays and RT-PCR are used to monitor the

---

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Correspondence concerning this article should be addressed to D. Ramkrishna at ramkrish@purdue.edu.

expression of mRNA from a large number of genes in cells under a variety of experimental conditions. These techniques have been used to characterize the gene expression changes that occur when *Escherichia coli* switches from growing on one substrate to growing on another.

To briefly explain the changes that occur, when *E. coli* switches from growing on glucose to growing on acetate, a number of regulatory mechanisms are activated that monitor the cell's nutritional status and the propensity for internal acidification.[8] These mechanisms then coordinate a host of actions on the gene expression level to induce the uptake of acetate predominantly through the ACS pathway, where it is converted to other growth precursors via increased flux through the TCA cycle.[9] Glycolysis ceases and carbon is channeled in a completely reversed direction via gluconeogenesis. Similarly, when *E. coli* switches from consuming glucose to consuming lactose, a more limited number of events occur. As glucose is depleted, cAMP levels are elevated and crp-dependent factors promote the expression of the lac operon which has already been derepressed by the presence of external lactose.[10] The lac operon is then transcribed and translated into enzymes relevant to the metabolism of lactose.

The present work focuses on comparing the predictions for cybernetic control variable for enzyme synthesis u_i with data for enzyme synthesis available in the form of mRNA measurements and measured using microarrays and RT-PCR. While cybernetic variables for enzyme synthesis do not directly describe mRNA concentrations, they describe the same process that mRNA synthesis drives. Although gene expression data do not perfectly correlate with the amounts of proteins in cells,[11] there are a number of models that have successfully recruited gene expression data for more accurate prediction of metabolic changes. These models include E-Flux,[12] tFBA,[13] and E-Fmin.[14]

The comparison made in this study seeks to investigate the similarity between the dynamics of mRNA creation and degradation, and the behavior of cybernetic variables for enzyme synthesis. Such a study involves a more detailed inspection of the prediction of regulated metabolism than evaluation of cybernetic models by comparing measurements of selected extracellular variables as in prior publications. Thus, this effort must be regarded as a more comprehensive assessment of cybernetic goals in explication of the observed dynamic variations in the activity of component pathways in metabolism. It is a verification of regulatory *trends* as the accuracy of proteomic data is not generally viewed to be sufficient for direct quantitative comparison with predicted enzyme levels. This difficulty arises from the significant differences in chemical complexity between proteins and mRNA. In this same vein, the cybernetic variable for enzyme activity, v_i, was not investigated because measurement of the dynamics of activity modulation in enzymes, stemming from conformational changes and phosphorylation, is not easily measured.

We have chosen for this consideration the early cybernetic models, which have been more recently denoted as LCM (lumped cybernetic models) by Song and Ramkrishna[15] that are based on grossly lumped pathways as against the hybrid cybernetic models (HCM) or lumped hybrid cybernetic models (L-HCM) that have a higher likelihood of relating to gene expression profiles because of being based on numerous elementary flux modes. Consequently, this investigation will also examine whether the pathway lumping in LCM obscures the regulatory features associated with metabolic reactions not explicitly incorporated in the model.

Toward this end, three separate studies wherein microorganisms grow on two substitutable substrate pairs are made. The first study is for the growth of *E. coli* BW25113 on a mixture of glucose and acetate whose gene expression changes are monitored using microarray. The second study is for the growth of *E. coli* MG1655 on glucose and acetate whose expression changes are monitored using RT-PCR. The final study is for the growth of *E. coli* MG1655 on glucose and lactose where gene expression changes are monitored using microarrays. By demonstrating that the cybernetic variables for induced enzyme synthesis mimic transcriptional data, we show that the cybernetic models capture not only the dynamic trends in the concentrations of extracellular substrates and products but accurately reflect changes in the regulatory phenomena that drive microorganisms.

Materials and Methods

To describe the diauxic growth phenomena that occur in the two previously described scenarios, models of similar structure were built. These models will project the cybernetic control of each lumped pathway's enzyme synthesis and are developed using prior formulations of cybernetic models.[1,16]

Most simply, a cybernetic model of metabolism is a set of coupled, first-order ODEs that describes the time-dependent rates of change in the concentrations of cells, metabolites, and enzymes. For each substrate, S_i, there is a lumped enzyme set E_i that is employed to digest each substrate and a yield describing the total amount of biomass, B, produced from each substrate according to the reaction:

$$B + S_i \xrightarrow{Ei} (1 + Y_i)B + \ldots$$
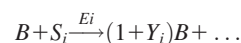
In diauxic growth, once the cell consumes the entirety of the preferred substrate, the production of lumped enzyme for the secondary substrate commences. In the cybernetic formulation, the rate at which each substrate is converted into biomass is governed by the concentration of the amount of pathway specific enzyme, e_i, present in the system. Also influencing this rate of reaction are the enzyme specific growth rate μ_i, substrate concentration, s_i, and the total concentration of cells c. This rate is formulated as

$$r_i = \frac{\mu_i e_i s_i c}{K_i + s_i}$$

where K_i is the substrate specific Michaelis–Menten constant. The formation of biomass is determined by the summation of the activity controlled biomass formation reactions and is written:

$$\frac{dc}{dt} = \sum_i r_i v_i$$

Each substrate's depletion is given by the negative substrate specific biomass formation rate normalized by the substrate's yield:

$$\frac{ds_i}{dt} = \frac{-1}{Y_i} r_i v_i$$

Finally, the production of the lumped enzymes is as follows:

$$\frac{de_i}{dt} = \alpha + u_i \frac{k_{E,i} s_i}{K'_i + s_i} - \left(\frac{d}{dt}\ln c + \beta\right) e_i$$

The above equation includes a number of terms for the phenomena related to changes in enzyme concentration. The first is for the constitutive enzyme synthesis, α.[16] The second is for induced enzyme synthesis which is controlled by the cybernetic variable, u_i. The rate of enzyme formation is constrained by a rate constant for induced enzyme synthesis, $k_{E,i}$, and is a function of the substrate level and a Michaelis–Menten constant, K'_i. Last, the enzyme quantity is reduced both by dilution due to growth and from enzymatic degradation β. Note that this enzyme control variable u_i appearing above will be the model variable that is compared to gene expression data.

Cybernetic Control Formulation

In the above model formulation, two control variables are presented. Control variable u_i specifies the regulation of enzyme synthesis for different pathways and v_i specifies the activity adjustment of the lumped pathways. To efficiently allocate cellular resources, single cells must make decisions regarding the regulation of competing metabolic pathways. From this model's perspective on diauxic growth, there are two competing metabolic options for which the cell can invest resources in to grow. At any given time, there is an unregulated growth rate for each substrate given by r_i which is a function of the amount of lumped enzyme present. This enzyme quantity, in turn, changes as a function of the instantaneous rate through the u_i control variable's effects. This control variable compares the different metabolic options and invests resources proportionally to a metabolic pathway's return on investment growth and is written as the following matching law:

$$u_i = \frac{r_i}{\sum_j r_j}$$

In this simple cybernetic model, return on investment for a lumped metabolic pathway is equal to its instantaneous rate of growth. This means that cells prioritize their limited capacity for enzyme synthesis for a pathway that has a higher payoff in terms of growth rate. This variable is normalized by the sum of returns on investment which represents the total pool of resources available to convert sugars into biomass. The set u_i control variables sum to 1. In this control variable, it is evident that there will be a higher investment of resources into the lumped enzyme for a pathway if there is a higher the return on investment or rate of growth. The models developed for this study make use of growth rate maximization as an objective function. Other objective functions such as carbon rate uptake maximization were not considered.

In addition to the efficient allocation of resources, pathways are also regulated by mechanisms that adjust their enzymatic activity. To model this, pathways that have lower rates of growth than the maximal pathway are turned down via the proportional law which is

$$v_i = \frac{r_i}{\max_j(r_j)}$$

This is justified in part by the fact that some metabolic pathways may require fluxes to route through different parts of

Table 1. Parameter Values for Diauxic Models of *E. coli*

Strain Parameter (units)	MG1655 Glucose	MG1655 Lactose	BW25113 Glucose	BW25113 Acetate
S_{i0} (g/L)	0.5	1.5	1.5	0.9
$\mu_{\max,j}$ (h^{-1})	0.642	0.618	0.442	0.426
Y_i (gDW/g)	0.400	0.440	0.339	0.129
K_i/K'_i (g/L)	0.005	0.50	0.0088	0.3317
α_i	1.67e-3	1.67e-3	1.67e-3	1.67e-3
$k_{E,i}$	1e-3	1e-3	1e-3	1e-3
β_i	0.05	0.05	0.05	0.05

Strains MG1655 and BW25113 growing on multiple substrate sources.

the metabolic network in opposite directions. For example, in the case of glycolytic and gluconeogenic substrates, glucose and acetate, pathways for one or the other should be turned down to reduce futile cycling. Therefore, the proportional cybernetic variables turn down pathways that do not represent the highest return on investment and fully express the pathway that does.

Estimation of Parameters

To develop two of the models in this work, sets of data including both batch growth data and time series microarray data taken during the switching of substrates were used. One data set is from Ref. 17 describing the growth of *E. coli* BW25113 on a mixture of glucose and acetate. The other set from Ref. 18 elucidates the changes in gene expression for *E. coli* MG1655 growing on a mixture of glucose and lactose. For both of these data sets, a growth curve and initial substrate data are available. Given that microarray data is inherently noisy, the findings for the glucose–acetate growth scenario were verified with RT-PCR data. An additional model was developed to describe the *E. coli* MG1655 strain using data from Ref. 19. RT-PCR data were not available in the literature to further validate the result from the glucose–lactose diauxic growth scenario.

Models were first parameterized for the microarray data. From the growth curves in the glucose–lactose condition, the parameters $\mu_i^{\max}$ and Y_i were estimated from the knowledge of the initial substrate levels, biomass concentrations, and slopes of the growth curves. Michaelis–Menten parameters were at first taken from Ref. 1. Michaelis–Menten constants, yields, and growth rates for the glucose–acetate data were determined using available single substrate data. Once these initial values were collected for both models, a genetic algorithm was used to refine the initial parameter values using a normalized least squares error function for the model's fit of the biomass data. The parameter values are listed in Table 1. Note that the Michaelis–Menten parameters K_i and K'_i take on the same value for both r_i and the induced enzyme formation terms for the same lumped metabolic pathway. For example, K_i and K'_i have the same value of 0.05 for the glucose growth pathway describing the glucose–lactose diauxie of *E. coli* MG1655.

The RT-PCR model was built in a similar fashion. The main departure in its formulation from the other models is the inclusion of an acetate production term. In this scenario, glucose is the only substrate initially. While consuming glucose, acetate is produced via overflow metabolism to adjust the intracellular redox ratio,[20] which is then consumed after glucose is exhausted. Note that acetate production is absent

from the microarray models as the dynamic changes in substrates and products are not given in the datasets and modeling this phenomena would require extraneous assumptions. For more details on the parameterization and development of this particular model, refer to Supporting Information 1.

Comparison of Dynamic Gene Expression Data and Model Variables

The general goal in the comparison between the u_i control variables and dynamic gene expression is the demonstration of qualitative similarity. These two entities cannot be compared in an absolute sense because they represent different phenomena. One represents an abstraction of control while the other represents the physical manifestation of cellular control. The matching law variables are constrained between values of 0 and 1 while gene expression levels vary over a greater range depending on how they are normalized. Moreover, given the nature of microarray, it would not be useful to make absolute comparisons as it is often referred to as a "semi-quantitative" method. Therefore, it is pertinent to normalize both the control variables and gene expression in a fashion that complements qualitative comparison. The method elected to do this is a standard normalization procedure where a data series is scaled by its standard deviation and centered around zero by subtracting the series' mean.[21] For some series, $C = \{c_1, \ldots c_n\}$ of either gene data or model variables, the normalization is

$$c'_i = \frac{c_i - \mu(C)}{\sigma(C)}$$

of which $\mu(C)$ is the mean of the data series and $\sigma(C)$ is the standard deviation of the series. The mean of a data series $\mu(C)$ is not to be confused with the enzyme specific growth rate for a given pathway μ_i. This normalization method aligns the data in such a way to make the comparison of the dynamic trends in the data more clear. To quantify similarity between the model variables and experimental data, the use of correlational statistics such as Pearson's correlation coefficient is still appropriate as this statistic is scale invariant.

The microarray and RT-PCR data were already provided as gene expression normalized sets in their original sources.[17–19] These data sets were not processed in any additional way. The significance of the observed expression changes are reported as fold changes in Table 2. The majority of fold changes observed in the data represented statistically significant changes in gene expression.

Results

Model capture of biomass formation

The models developed were able to accurately describe the growth of biomass. Both systems modeled capture the initial growth phase on glucose and the second growth phase on either lactose or acetate. They also correctly represent the timing of the lag phase as the batch culture transitions from the preferred substrate to the secondary one. The predictions of the glucose acetate models can be seen in Figure 1. Glucose–lactose substrate behavior is shown in Supporting Information, Figure S1.

From these model descriptions of the data, it can be gauged how the values of the control variables for each pathway change dynamically. As the glucose level depletes, the

Table 2. Fold Changes of Transcripts Compared to Cybernetic Variables

| Gene	Fold Change	Gene	Fold Change
Glucose–Acetate Microarray			
sdhA	21.677041	*gnd*	0.06591739
aceB	18.663797	*aceF*	0.07638358
fumC	15.381546	*aceE*	0.09660509
gltA	8.5310011	*pfkA*	0.11428783
sucA	7.5857758	*ackA*	0.24717241
sucB	7.4473197	*pykF*	0.25351286
yggF	2.2855988		
Glucose–Acetate qRT-PCR			
glgS	33.38495	*ptsG*	0.02199841
yfiA	19.902748	*ppc*	0.02351672
acs	8.548292	*cspA*	0.02529762
frd	5.5425171	*infA*	0.06812921
pck	4.6030265	*zwf*	0.10020879
ihfA	2.1322044	*pykF*	0.12090034
		cya	0.13806907
		edd	0.14985653
		maeA	0.16879325
		pfkA	0.20364643
		sdhD	0.27429994
		sucA	0.30113072
		mdh	0.38912586
		lacZ	0.40956123
		crp	0.52603482
Glucose–Lactose Microarray			
lacA	3459.3938	*crr*	0.10839269
lacY	215.77444	*ptsG*	0.12445146
lacZ	87.297137		
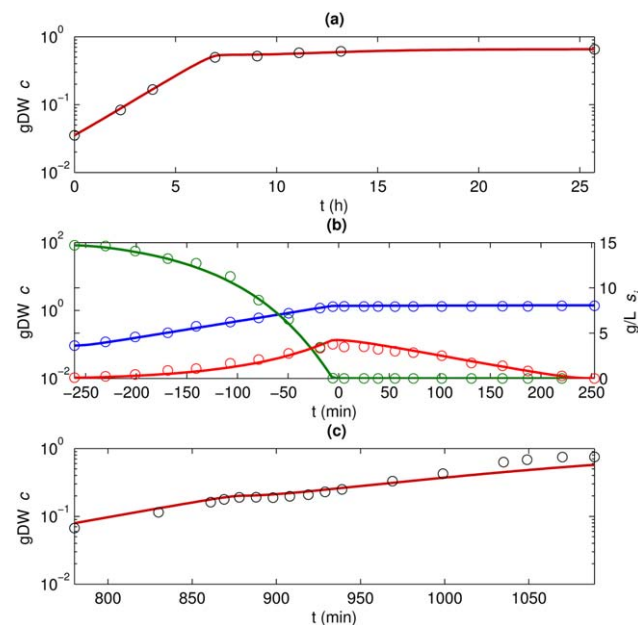
Figure 1. Predictions of external metabolite concentration data for the three diauxic systems.

(a) shows the biomass level for *E. coli* BW25113's diauxic growth on glucose and acetate. (b) exhibits *E. coli* MG1655's growth on glucose and acetate where RT-PCR data are available. The left axis shows biomass, c, in gDW and is depicted with a blue line. Substrate concentrations, s_i, are in g/L and are tracked on the right axis with the time profile of glucose in green and acetate in red. (c) shows diauxic growth of *E. coli* MG 1655 on glucose and acetate.

u_i variable for the glucose pathway decreases from one towards zero. This transition is made during the lag phase of the diauxic growth. At the same time, the u_i variable for the lactose or acetate pathways increases from a value close to
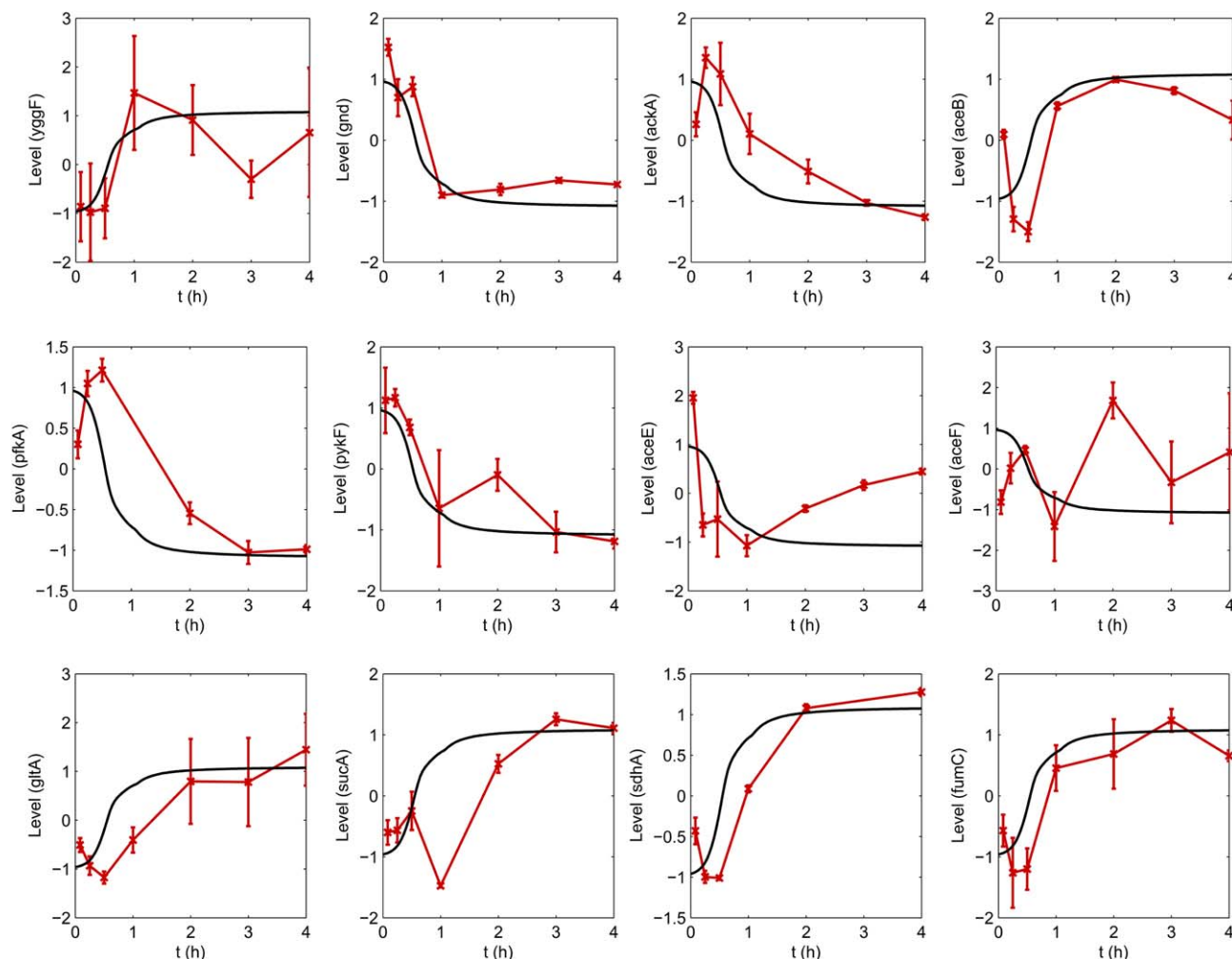
Figure 2.   Gene expression profile comparison with cybernetic variables for subsets of central carbon metabolism for strain MG1655's growth on glucose and acetate.

Gene profiles are plotted in red with error bars and u_i profiles are plotted in black. Subfigures with their respective pathways are labeled: (a) gluconeogenesis, (b) pentose phosphate pathway, (c) acetate excretion (d) glyoxylate pathway, (e) glycolysis, and (f) TCA cycle.

zero toward one. It is at this critical point that the comparisons between the control variable and gene expression will be made. All time scales refer to the time scales used in the original works from which this data was taken.

Study of glucose–acetate diauxie

Escherichia coli growing in a mixture of glucose and acetate can employ either glycolytic and gluconeogenic pathways to synthesize biomass and grow. In this mixture, the culture first consumes glucose and switches to acetate upon which it has a lower growth rate. When the culture shifts from the consumption of glucose to acetate, glycolysis is unneeded and gluconeogenic reactions must channel carbon to synthesize essential biomass precursors. A host of other changes happen in the central carbon metabolic pathways. Intuitively, acetate excretion goes down and uptake increases. Fluxes through the glyoxylate pathway and TCA cycle increase.[22] Also, the fluxes through the pentose phosphate pathway are significantly higher during growth on glucose compared to acetate.

These changes in the various metabolic pathways are observable on the gene expression level. The cybernetic variable profiles have been compared accordingly with the dynamics of these various central carbon metabolic pathway gene expression profiles. Genes associated with pathways whose

fluxes are upregulated have been compared with the acetate u_i cybernetic variable and genes associated with pathways whose fluxes go down have been compared with the glucose variable. An excerpt of this comparison is depicted in Figure 2. Comparing the changes in gene expression with the cybernetic variables for genes related to different metabolic modules demonstrates that cybernetic variables can be used to predict the general trends in gene expression for their respective pathways. Explicitly, the genes go from a sustained high state to a sustained low state. Before making the comparison, it must be noted that microarray data are inherently noisy. Because of this, many genes from these pathways were filtered out as their true behavior was hard to distinguish from the scatter in the data. The gene transcript measurements with significant fold changes were retained. These are listed in Table 2. For a complete picture of the behavior for these pathways, refer to Supporting Information, Figures S2–S8. It is evident in these figures that there is a significant amount of noise for some transcripts with the error bars, encompassing a range wider than the observed change in level.

Starting with excretion of acetate via the reaction of acetate kinase (*ackA*), the cybernetic variable for the glucose pathway was compared with the dynamic transcription data. It is evident that both show a downward shift from a higher

state to a lower one. The gene expression levels for acetate uptake via acetyl-CoA synthetase (*acs*) were also surveyed; however, an abundance of noise in the data made comparison difficult. *Acs* is compared with less noisy data using the RT-PCR data in the next section.

Next, the cybernetic variables for glycolysis and gluconeogenesis were compared with their respective model variables. Reactions for gluconeogenesis were first compared to the acetate control variable. Most transcriptomic data for the gluconeogenic pathway were also noisy with the exception of fructose-1,6-bisphosphatase (*yggF*), which is somewhat less noisy. This transcript profile matches the cybernetic variable's increase. However, after reaching its highest value, the trend does not stay at an elevated state and goes down to a value roughly between the initial low state and the maximum level.

Transcripts of the glycolytic pathway show a decreasing trend as predicted by the glucose pathway's cybernetic matching law variable. For example, mRNA levels for 6-phosphofructokinase-1 (*pfkA*) go down along with the model control variable. Also showing this behavior is pyruvate kinase I (*pykF*). Genes for the pyruvate dehydrogenase complex show varying behavior. Pyruvate dehydrogenase complex gene *aceE* transcription levels go down as the control variable does, but rebound after 2 h to a middle value. *aceF* shows significant fluctuations up and down and deviates significantly from the behavior predicted by the glucose pathway's cybernetic variable.

Finally, the TCA cycle and glyoxylate pathway were compared to the acetate control variable's behavior. Of the TCA cycle transcript data that did not have significant noise, all the genes show trends close to the model prediction. Succinate dehydrogenase (*sdhA*) and fumarase (*fumC*) show increasing levels very similar to the selected control variable. Citrate synthase (gltA), and 2-oxoglutarate dehydrogenase (*sucA* and *sucB*) also show general increases in expression over time. However, these expression level increases are delayed to varying degrees compared to the acetate pathway's u_i variable. They are also characterized by an initial decrease in the first 15–30 min of the data series.

The glyoxylate pathway also shows increased expression along with the model prediction. Isocitrate dehydrogenase gene expression, a key enzyme involved controlling flux into the glyoxylate pathway, shows little correlation with either variable. However, this TCA–glyoxylate branch point is well known for being controlled via phosphorylation.[23] When glucose is exhausted, the activity of isocitrate dehydrogenase goes down to increase the flux through the glyoxylate pathway. This change in enzyme activity is better modeled by comparing the v_i control variable for glucose activity regulation with the activity changes in isocitrate dehydrogenase. As shown in Supporting Information, Figure S9, qualitative comparison of data[24] representative of isocitrate dehydrogenase's enzyme activity and the v_i control variable for glucose reveals that both have similar behavior. Both show dramatic decline in value after glucose is exhausted.

Verification of microarray result using RT-PCR data from the literature

To cross-validate the findings for the glucose-acetate diauxie previously modeled above, a second model was developed to describe a similar scenario of acetate production
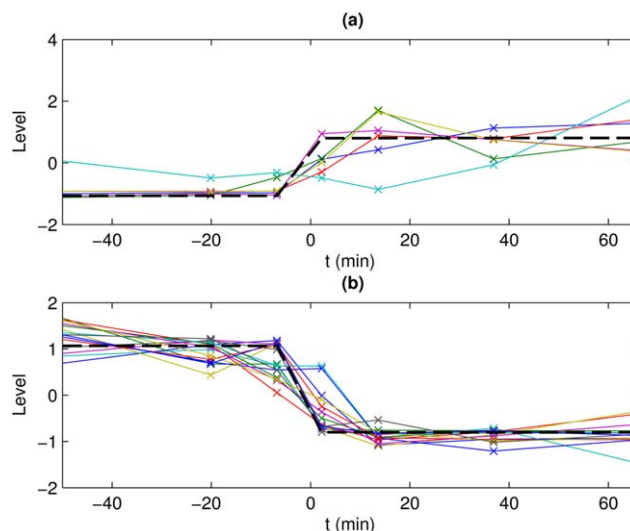
Figure 3. Gene profiles for all RT-PCR data. u_i variables are plotted in black dashed (- -) lines.

The various gene profiles are plotted with color markers (*x*) for each data point. In (a) is the comparison between the acetate model control variable and genes that match this behavior. (b) compares the glucose matching law variable and genes showing similar behavior.

during the consumption of glucose followed by acetate consumption after glucose exhaustion. RT-PCR data were collected for the transcripts with the largest changes in expression comparing steady-state growth on glucose and growth on acetate. For this analysis, the data have been partitioned into two sets. One set shows an increase in expression after glucose exhaustion corresponding to the acetate u_i variable and another represented by the glucose control variable shows a decreasing trend. A depiction of this comparison is in Figure 3.

In this system, the data for each transcript were not analyzed according to its pathway but its general behavior overall. The genes corresponding to each control variable are listed in Table 3. Looking at the Pearson correlation between the vector of model u_i values at different time points and the vector representing each gene's normalized expression level, there is a strong correlation between the model prediction of enzyme synthesis control and gene expression. Overall, these correlations are 0.9214 and 0.7814 for the glucose and acetate pathway with *p* values of 4.351e-044 and 1.017e-09, respectively. Correlational statistics and *p* values for all transcripts are listed in Supporting Information, Table sets S1–S3.

After measuring the general correlation between the gene expression data and the cybernetic variables, the grouping of genes into upregulated or downregulated categories was further analyzed. Starting with uptake reactions, RT-PCR data for acs for acetate uptake and glucose phosphotransferase (*ptsG*) for glucose uptake show appropriate grouping with their matching law variable. The glucose variable is also grouped appropriately with genes like *pfkA* and *pykF* from glycolysis. As expected, glucose-6-phosphate dehydrogenase (*zwf*) of the pentose phosphate pathway also shows similarity to the glucose control variable. Genes strongly related to the control of growth during glucose consumption, transcriptional regulators *crp* and *cya* also show a strong similarity to the glucose control variable.[25]

As for the acetate control variable, phosphoenolpyruvate carboxykinase (*pck*) representing the conversion of OAA to PEP is also upregulated similar to what is expected when the cells retool their metabolism for growth on acetate. This is a key entry point of carbon into gluconeogenesis. The genes of fumarate reductase (*frd*) of the TCA cycle are also upregulated stepwise along with the control variable for the acetate pathway. Other genes upregulated with this control variable are predicted glycogen synthase protein (*glgS*), transcriptional regulator, *ihfA*, and *yfiA* related to biofilm formation,[26] DNA regulation,[27] and translation inhibition.[28]

There were also some unexpected relationships between the glucose variable and various components of the central carbon metabolic pathways. TCA cycle transcripts for *sucA* and succinate dehydrogenase (*sdhD*) showed a downshift. Also indicating a downshift was malate dehydrogenase (*maeA*) of gluconeogenesis, which converts members of the TCA cycle to gluconeogenic products. This is further analyzed in the discussion.

Study of glucose–lactose diauxie

To verify the relationships of the enzyme synthesis cybernetic variables to gene expression in other diauxic growth systems, a glucose–lactose growth scenario for *E. coli* MG1655 was also studied. After glucose has been exhausted, genes associated with the consumption of lactose located of

the lac operon are upregulated to facilitate growth on the new substrate. From the comparison of cybernetic u_i variables with the expression data in Figure 4, it is clear that the lactose cybernetic variables and the regulation of genes related to the lactose pathway have a relationship. Both indicate an increase from a low to high value.

Given that lactose is a disaccharide composed of glucose and galactose, there is much overlap between the metabolic pathways for growth on lactose and growth on glucose. The main element not involved in lactose metabolism; the PTS glucose transporter complex consisting of *ptsG* and *crr* was surveyed.[29] The behavior of this gene shows some decrease as glucose is exhausted shortly before the glucose control variable goes down. However, afterward it rebounds upward and shows much fluctuation up and down with decreases in expression being related to decreases in growth rate indicated on the biomass curve.

Discussion

After making comparisons between the cybernetic variable for induced enzyme synthesis u_i and gene expression data for three separate growth scenarios, it is evident that the two entities are related. Overall, the similarities between gene expression dynamics and the predicted control are qualitatively similar and occur on the same time scale. The shifts in enzyme resource allocation predicted by the model occur within the same time period that shifts in gene expression data occur. Moreover, when looking at correlational statistics between the least noisy data set, the RT-PCR scenario, and the cybernetic control values, a strong statistical correlation is established. For the glucose variable showing a stepdown behavior, the Pearson correlation between the model and each gene related to that pathway is 0.9214 with a p value of 4.351e-044. For the acetate pathway, it is 0.7814 with a p value of 1.017e-09. These significant correlations indicate that there is a strong similarity between the dynamic

Table 3. Genes Grouped by Behavior After Glucose Exhaustion

Acetate Group (Up)	Glucose Group (Down)	
acs	crp	pfkA
frd	cspA	ppc
glgS	cya	ptsG
ihfA	edd	pykF
pck	infA	sdhD
yfiA	lacZ	sucA
	maeA	zwf
	mdh	
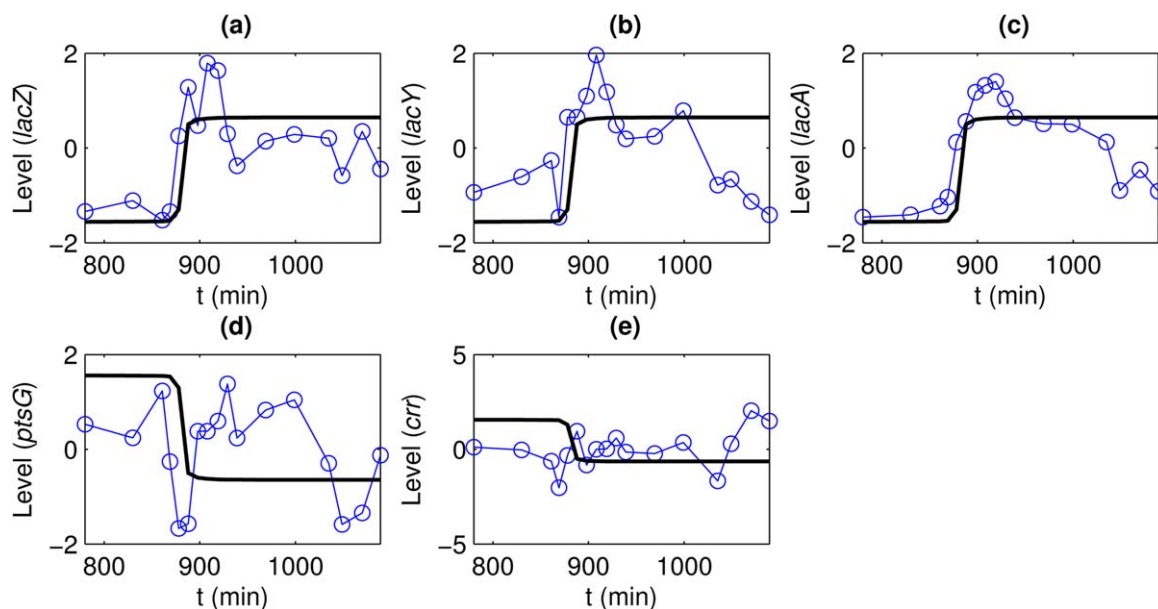
Figure 4. Glucose–lactose matching law variables with gene expression profiles for relevant genes.

(a–c) in the top row show the expression profiles of *lac* operon genes *lacZ, lacY,* and *lacA* along with the lactose pathway's u_i variables. The bottom row shows the profiles for genes relevant to glucose metabolism which are the phosphotransferase glucose transporter complex of (d) *ptsG* and (e), regulator *crr*.

behavior of the gene expression and the model's prediction of enzyme synthesis regulation.

It should be explicitly stated that the prediction of these cybernetic variable behaviors by the model is not made using any information from the gene expression level. The only input into the model's development and parameterization comes from time series data describing biomass and substrate levels. The ability to match gene expression is only made from the assumption that enzymes for substrate pathways are regulated in such a way to optimize the rate of formation of biomass. The fact that this assumption endows the model with an ability to generate predictions of control that mimic the behavior of gene expression in these scenarios helps to validate the idea that modeling *E. coli* cells from a goal oriented perspective is useful.

Modeling *E. coli's* growth using a cybernetic approach differs substantially from other modeling methods. Instead of exhaustively enumerating the minutiae of metabolic regulatory mechanisms as is done in a host of other, kinetically driven modeling frameworks for this scenario,[30,31] cybernetic models instead model the sum of these regulatory actions as cohesive, integrated machinery that attains some optimal behavior. By comparing the dynamics of *E. coli's* enzyme production machinery with the cybernetic model's prediction of enzyme synthesis regulation, the wealth information related to gene expression regulation is compressed down into a succinct description related to the organism's goal. Unlike FBA which also makes assumptions regarding optimal yields,[32] cybernetic models also retain consideration of the dynamic nature of optimality. Cells cannot be entirely prescient and, instead, react spontaneously to their changing environments.

The objective function of growth rate maximization even applies to scenarios such as glucose–acetate growth, where it has been observed that little to no anabolism occurs when *E. coli* consumes acetate.[33] Even if growth on acetate is infinitesimally small, the order of substrate consumption is still preserved by the growth rate maximization objective because the growth rate on glucose is still higher. Other objective functions, such as carbon uptake rate maximization, would still predict the same sequence of substrate consumption as well. Given that the order of substrate consumption does not change the cybernetic variables themselves should be roughly equivalent in either case. The predicted trends in gene expression still remain valid despite the fact that the growth rate on acetate approaches zero.

The present work only has surveyed the ability of lumped cybernetic models to predict changes in gene expression. More advanced cybernetic frameworks such as the hybrid cybernetic model (HCM)[34] and the lumped hybrid cybernetic model (L-HCM)[35] may be able to generate more detailed and accurate predictions of gene expression in these scenarios. This stems from the fact that these models may incorporate multiple pathways in the form of elementary modes to describe the consumption of a single substrate. The regulation of numerous metabolic pathways could provide a higher resolution prediction of gene expression. The development of such descriptions, however, would mandate more detailed information at the level of substrates to isolate the exact state of specific elementary modes. Without such information, these models would be undetermined and describing the control of any reaction would be impossible.

While more coarse-grained, the correlation between the trends of the present LCMs and gene expression is a vindication of the simplicity of this approach. Overall, the merit in LCM approximations lies in their simplicity and potential use in bioreactor modeling for the purposes of optimization and control. Other approaches that make use of metabolic network data, such as dynamic flux balance analysis,[36] are unable to capture the dynamics of the studied diauxic switch.

Despite an overall success of the forgoing comparison of cybernetic models and gene expression changes, some discrepancies when comparing these different systems did arise. For example, the glucose–acetate microarray data for BW25113 indicated increases in expression for most all of the TCA cycle genes including *frd*, *sucA*, *sdhA*, and *mdh*. In the RT-PCR data for the MG1655, the transcripts for TCA cycle genes have varying behaviors. Fumarate reductase (*frd*) shows an increase similar to the BW25113 strain data. However, *sucA*, *sdhD*, and *mdh* show declining expression. For these, it should be noted that the time periods over which the two data sets are taken varies. The MG1655 data are taken over only one hour after glucose exhaustion. On the other hand, the data analyzed in this scenario for the BW25113 strain go for 3 h longer than the MG1655 data. Both data sets do indicate an initial decrease. The data taken over a longer time period demonstrate that these genes ultimately reach higher levels of expression as indicated by fluxes taken at steady states for these growth scenarios. The similarity of these gene expression profiles to the u_i control variables is still notable. Depending on the time-frame selected for this comparison, there may be a varying but relatively small number of genes that change groupings among the different pathways. Considering the expression data from a broader angle, there still is a significant agreement between the control variables' behavior and gene expression dynamics for a majority of the genes surveyed.

The glucose–lactose system showed good agreement between the lactose pathway's cybernetic variable and the gene expression for genes dedicated to lactose metabolism. However, the glucose variable did not show a strong relationship with gene expression for genes related strictly to glucose metabolism. This is, in part, due to the fact that glucose and lactose metabolic pathways have significant overlap. Particularly, the entire pathway for glucose consumption is also used for lactose. Isolating particular chemical reactions for glucose's nonoverlapping metabolic reactions is a difficult task. The main component in the glucose pathway that is not in the lactose pathway is the transporter *ptsG*. The expression of *ptsG* does not behave the same way the glucose u_i variable behaves. It is expressed not only when glucose is present in culture, but also when lactose is being consumed. At first glance, the expression of this transporter appears to be linked more closely with growth rate and not glucose presence. When looking for explanatory information from the literature, however, *ptsG* expression is regulated by a number of intracellular transcriptional factors including *mlc*, *crp*-cAMP, and *Fis*[37–39] and such conclusions require more consideration to develop. To model this situation more accurately, a model that dedicates a larger portion of cellular resources toward nongrowth functions such as the production of *ptsG* could be used. Other formulations of cybernetic models have considered the idea of fractional resource investment into preadapted pathways based on other

regulatory mechanisms which may be another factor related to gene expression for the glucose pathway.[40]

By means of this study, it is apparent that a significant number of transcripts demonstrate a strong relationship with the model's projection of their dynamic behavior. This is especially true of genes related to the uptake of various substrates including *ptsG* and *acs* in the glucose–acetate scenario and for the lac operon of the glucose–lactose scenario. While not all genes conform to the trends established by the u_i cybernetic variables, future development of models that incorporate more information regarding the structures of metabolic networks may better predict regulatory behavior. The foregoing work helps to verify the assumption that cells dynamically regulate different metabolic pathways towards optimality. The assumed, cybernetic mechanism of enzyme synthesis control correlates strongly with how actual biological systems modulate enzyme synthesis. This indicates that cybernetic models have the potential to predict general trends in gene expression dynamics in other systems using information at the metabolite level and an assumption of the organism's goal for metabolic regulation.

Acknowledgment

Literature Cited

1. Kompala DS, Ramkrishna D, Jansen NB, et al. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. *Biotechnol Bioeng* 1986; 198628:1044–1055.

2. Ramakrishna R, Ramkrishna D, Konopka AE. Cybernetic modeling of growth in mixed, substitutable substrate environments: preferential and simultaneous utilization. *Biotechnol Bioeng* 1996; /10/05/199652:141–151.

3. Song HS, Ramkrishna D, Pinchuk GE, et al. Dynamic modeling of aerobic growth of *Shewanella oneidensis*. Predicting triauxic growth, flux distributions, and energy requirement for growth. *Metabol Eng* 2012; 2012.

4. Song H-S, Ramkrishna D. Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metabol Eng* 2012; /03//201214:69–80.

5. Kim JI, Song H-S, Sunkara SR, et al. Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. *Biotechnol Progr* 2012; /09/01/2012 28:1160–1166.

6. Park T, Yi S-G, Kang S-H, et al. Evaluation of normalization methods for microarray data. *BMC Bioinformat* 2003; 4:33. /09/02/2003

7. Overbergh L, Giulietti A, Valckx D, et al. The use of real-time reverse transcriptase PCR for the quantification of cytokine gene expression. *J Biomol Tech* 2003; /03//200314:33–43.

8. Wolfe AJ. The acetate switch. *Microbiol Mol Biol Rev* 2005; /03/01/200569:12–50.

9. El-Mansi M, Cozzone AJ, Shiloach J, et al. Control of carbon flux through enzymes of central and intermediary metabolism during growth of *Escherichia coli* on acetate. *Curr Opin Microbiol* 2006; 9:173–179. /04//2006

10. Cohen GN, Monod J. Bacterial permeases. *Bacteriol Rev* 1957; /09//195721:169–194.

11. Lee PS, Shaw LB, Choe LH, et al. Insights into the relation between mRNA and protein expression patterns: ii. Experimental observations in *Escherichia coli*1. *Biotechnol Bioeng* 2003; 84:834–841.

12. Colijn C, Brandes A, Zucker J, et al. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol* 2009; 5:e1000489.

13. Berlo RJPv, Ridder D, Daran JM, et al. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Trans Comput Biol Bioinformat* 2011; 8:206–216.

14. Song H-S, Reifman J, Wallqvist A. Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. *PLoS One* 2014; 9:e112524.

15. Song H-S, Ramkrishna D. Prediction of metabolic function from limited data: lumped hybrid cybernetic modeling (L-HCM). *Biotechnol Bioeng* 2010; 2010106:271–284.

16. Turner BG, Ramkrishna D. Revised enzyme synthesis rate expression in cybernetic models of bacterial growth. *Biotechnol Bioeng* 1988; /01/01/198831:41–43.

17. Kao KC, Tran LM, Liao JC. A global regulatory role of gluconeogenic genes in *Escherichia coli* revealed by transcriptome network analysis. *J Biol Chem* 2005; 280:36079–36087. /10/28/2005

18. Chang D-E, Smalley DJ, Conway T. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol Microbiol* 2002; 45:289–306. /07/01/2002

19. Enjalbert B, Letisse F, Portais J-C. Physiological and molecular timing of the glucose to acetate transition in *Escherichia coli*. *Metabolites* 2013; 3:820–837. /09/20/2013

20. Vemuri GN, Altman E, Sangurdekar DP, et al. Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol* 2006; 72:3653–3661. /05/01/2006

21. Ralanamahatana CA, Lin J, Gunopulos D, et al. Mining time series data. In: Maimon O, Rokach L, editors. *Data Mining and Knowledge Discovery Handbook*. New York, NY: Springer US; 2005:1069–1103.

22. Zhao J, Shimizu K. Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *J Biotechnol* 2003; 101: 101–117. /03/06/2003

23. Garnak M, Reeves H. Phosphorylation of isocitrate dehydrogenase of *Escherichia coli*. *Science* 1979; 203:1111–1112.

24. Edlin JD, Sundaram TK. Regulation of isocitrate dehydrogenase by phosphorylation in *Escherichia coli* K-12 and a simple method for determining the amount of inactive phosphoenzyme. *J Bacteriol* 1989; 171:2634–2638.

25. Perrenoud A, Sauer U. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *J Bacteriol* 2005; 187:3171–3179. /05/01/2005

26. Rahimpour M, Montero M, Almagro G, et al. GlgS, described previously as a glycogen synthesis control protein, negatively regulates motility and biofilm formation in *Escherichia coli*. *Biochem J* 2013; 452:559–573. /06/15/2013

27. Freundlich M, Ramani N, Mathew E, et al. The role of integration host factor in gene expression in *Escherichia coli*. *Mol Microbiol* 1992; 6:2557–2563. /09/01/1992

28. Maki Y, Yoshida H, Wada A. Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase *Escherichia coli*. *Genes Cells* 2000; 5:965–974. /12//2000

29. Stock JB, Waygood EB, Meadow ND, et al. Sugar transport by the bacterial phosphotransferase system. The glucose receptors of the Salmonella typhimurium phosphotransferase system. *J Biol Chem* 1982; 257:14543–14552. /12/10/1982

30. Peskov K, Mogilevskaya E, Demin O. Kinetic modelling of central carbon metabolism in *Escherichia coli*. *FEBS J* 2012; 279: 3374–3385.

31. Kotte O, Zaugg JB, Heinemann M. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol Syst Biol* 2010; 6.

32. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 1994; 60:3724–3731. /10//1994

33. Enjalbert B, Cocaign-Bousquet M, Portais J-C, et al. Acetate exposure determines the diauxic behavior of *Escherichia coli* during the glucose-acetate transition. *J Bacteriol* 2015; 197:3173–3181.

34. Kim JI. A hybrid cybernetic modeling for the growth of Escherichia coli in glucose-pyruvate mixtures. Theses and Dissertations Available from ProQuest 2008; 2008:1–185. /01/01/

35. Song H-S, Ramkrishna D. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnol Bioeng* 2011; 2011108:127–140.

36. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 2002; 83:1331–1340. /09//2002

37. Plumbridge J. Expression of ptsG, the gene for the major glucose PTS transporter in *Escherichia coli*, is repressed by Mlc and induced by growth on glucose. *Mol Microbiol* 1998; 29: 1053–1063. /08/01/1998

38. Seeto S, Notley-McRobb L, Ferenci T. The multifactorial influences of RpoS, Mlc and cAMP on ptsG expression under glucose-limited and anaerobic conditions. *Res Microbiol* 2004; / 04//2004155:211–215.

39. Shin D, Cho N, Heu S, et al. Selective regulation of ptsG expression by Fis. Formation of either activating or repressing nucleoprotein complex in response to glucose. *J Biol Chem* 2003; /04/25/2003278:14776–14781.

40. Mandli AR, Modak JM. Cybernetic modeling of adaptive prediction of environmental changes by microorganisms. *Mathemat Biosci* 2014; /02//2014248:40–45.