Title
The relationship between regulatory architecture and expression dynamics in transcriptional regulation of metabolic pathways

Author
Chubukov, Victor

Publication Date
2010

Peer reviewed|Thesis/dissertation

The relationship between regulatory architecture and expression dynamics in transcriptional regulation of metabolic pathways

by

Victor Chubukov

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Acknowledgments

I thank M. Guo, P. Kimmig, E. McCullagh, and J. Stewart-Ornstein for advice, technical assistance, and support, I. Zuleta for contributions to the data acquisition process, H. El-Samad, H. Madhani, and C. Tang for advice and editorial assistance, and H. Li for his role as primary advisor.

Abstract

A classic paradigm of gene regulation is metabolic enzyme expression in response to changes in metabolite levels. In this study we generate a quantitative picture of the response of a number of amino acid biosynthesis pathways in *S. cerevisiae* in response to nutrient depletion. We find that a striking pattern emerges that couples the architecture of the transcriptional regulatory network to the gene expression response. In particular, we find that networks controlled by the intermediate-activated architecture (IAA), in which an intermediate metabolite of the pathway activates transcription of pathway genes, exhibit the following response: the enzyme immediately downstream of the regulatory metabolite is under the strongest transcriptional control, while the enzymes upstream of the regulatory intermediate are only weakly induced. This pattern of separation of responses is absent in pathways not controlled by the IAA network and can be explained by a fundamental observation regarding the feedback structure of the network, which places downstream enzymes under a negative feedback loop, and upstream ones under a positive feedback loop. This general design principle for transcriptional control of a metabolic pathway can also be derived from a simple cost/benefit model of gene expression. Our results suggest that evolution of cis-regulation for metabolic pathway genes is strongly constrained by the overlying regulatory architecture.

List of Figures

Introduction

A classic paradigm of gene regulation is the regulation of metabolic enzyme expression in response to changes in external metabolite levels. By regulating enzyme levels, cells not only control the metabolic program, but also save resources and energy by not expressing enzymes which are not needed at a particular time. Countless studies have observed higher expression of enzymes in response to depletion of external nutrients. To achieve this regulation, a variety of strategies can be used, involving different interplay between metabolites and regulatory proteins. In many cases involving model organisms, the regulatory framework that controls this process is known, and research over past decades has indeed revealed a large number of different regulatory strategies (Jacob and Monod, 1961; Artz and Broach, 1975; Kohlhaw, 2003; Rébora et al., 2005) .

In this study we will mostly consider the example of a linear biosynthetic pathway in which an essential nutrient (hereafter called product) can be produced from an abundant precursor through a series of enzymatic steps, each catalyzed by a different gene product. A sample of possible regulatory architectures for such a pathway is shown in figure 1. It can be shown that any of the networks in figure 1b-1d can solve the problem of keeping pathway activity low when there is a large external product flux, but turning the pathway on and restoring product levels if the external flux disappears, as may happen if the cell found itself in a new environment with different nutrient sources. Aside from the network

structure, the parameters associated with the interactions – the numbers on the arrows (binding strengths, synthesis rates, etc) – also evolve and also contribute to the dynamics of the network (Ronen et al., 2002; Dekel and Alon, 2005; Hittinger and Carroll, 2007). The combination of architecture and parameters leads to the final phenotype, at the level of gene induction and nutrient recovery.

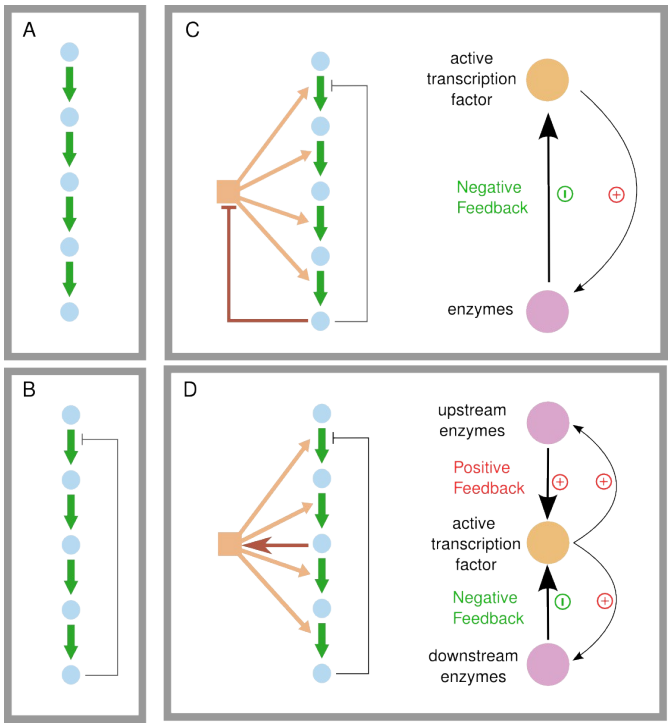

Figure 1. A sample of regulatory network architectures for a linear metabolic pathway. Metabolism can be regulated by simple mass action (a), by allosteric regulation of enzyme activity (b), or, typically, by a combination of allosteric regulation and

transcriptional regulation of enzyme levels (c-d). Two classic examples of transcriptional regulatory architectures are the end-product-inhibition (EPI) network (c) and the intermediate-activated architecture (IAA) network (d). The two networks create different feedback structures and place different constraints on gene expression. Examples of network (c) include arginine biosynthesis in *S. cerevisiae* and tryptophan biosynthesis in *E. coli.* Examples of network (d) include leucine and lysine biosynthesis in *S. cerevisiae* and methionine and lysine biosynthesis in *E. coli.*

Here we explore in detail the relationship between network architecture and gene expression. In particular, we consider several amino acid and nucleotide biosynthesis pathways in the yeast *S. cerevisiae* and measure their gene expression response to product depletion by using fluorescent protein constructs and single-cell measurements. We find a pattern that strongly suggests that the architecture of the regulatory network firmly constrains the dynamic profiles of gene expression. In particular we find that networks with intermediate-activated architecture (IAA, figure 1d), have a unique pattern of expression, involving strongest activation of the enzyme which catalyzes the step immediately after the regulatory intermediate, and often strong induction of other enzymes downstream of the intermediate. This pattern, which was observed earlier in the yeast leucine biosynthesis pathway (Chin et al., 2008), can be considered a universal feature of this network architecture, and is very likely to be based on the feedback structure imposed by the regulatory network, as shown in figure 1. In a pathway regulated by the end-product inhibition architecture (EPI, figure 1c), this feedback structure does

not exist and we do not observe similar gene expression dynamics. In the second half of the paper, we introduce a cost-benefit model for a generalized pathway. We show that the optimal solution of this very simple model predicts gene expression patterns that are consistent with what we observe *in vivo,* suggesting that such expression patterns may be a consequence of optimization through evolution.

Results

Gene expression profiles in amino acid and nucleotide biosynthesis pathways in S. cerevisiae

To look *in vivo* at the relationship between regulatory architecture and expression dynamics*,* we measured the transcriptional response to starvation in a number of amino acid and nucleotide biosynthesis pathways in *S. cerevisiae.* We choose amino acid and nucleotide biosynthesis pathways for several reasons. One is that they are a large and well-characterized class of pathways, in the sense that they come with a large body of knowledge about their regulation. A second is that the stimulus to turn on the pathway is easy to perform experimentally, by switching cells from a medium in which the amino acid is present to one where it is absent. Finally, since amino acids are fundamental cellular building blocks, we can estimate the amount of pathway flux that is necessary to sustain growth, and this aids us in constructing a quantitative model.

We used fluorescent reporter strains constructed by putting yeast-enhanced green fluorescent protein (yeGFP) (Heim et al., 1995) under the control of the natural promoter of each gene. The strains were grown in rich media, then moved quickly to media lacking one of the amino acids. Throughout the time course, fluorescence in single cells was monitored by flow cytometry, using an automated system for sample injection and data collection. We go through the results for each pathway sequentially.

Leucine biosynthesis

Leucine is synthesized from pyruvate via a pathway that also branches to valine biosynthesis, and involves some of the same enzymes as isoleucine biosynthesis. The linear leucine-specific branch of the pathway starts with alpha-ketoisovalerate, and the first of these steps is feedback-inhibited by leucine. The major transcriptional regulation involves the transcription factor Leu3, which is constitutively bound to DNA, but only activates transcription in the presence of the metabolic intermediate alpha-isopropyl-malate (αIPM), creating an example of the IAA network (Kohlhaw, 2003) (figure 3a). This pathway was explored in detail in Chin et.al (Chin et al., 2008), which was one of the first quantitative pictures of the dynamical transcriptional response in an IAA network. In that study, the authors observed a striking difference in the induction profiles of the enzymes which depended on the position of the enzyme within the pathway. While most genes were induced by about 2-fold, Leu1 and Leu2, the enzymes immediately

downstream of the regulatory intermediate αIPM, were induced by over 10-fold.

Here, we remeasured the transcriptional dynamics for the pathway in response to leucine depletion using promoter-GFP constructs instead of the fusion proteins used in the previous study (fig 3b). The results we obtained were consistent with the previous study. We confirmed that the strongest transcriptionaly regulated enzyme in the pathway was Leu1, with over 20-fold induction in response to leucine depletion. The next enzyme downstream of the control point αIPM, Leu2, was also relatively strongly induced, with a change of about 10-fold. In contrast, none of the enzymes upstream of the control point had more than about a 2-fold induction (ratios are for 6 hours after environmental shift, when most genes have reached steady state).

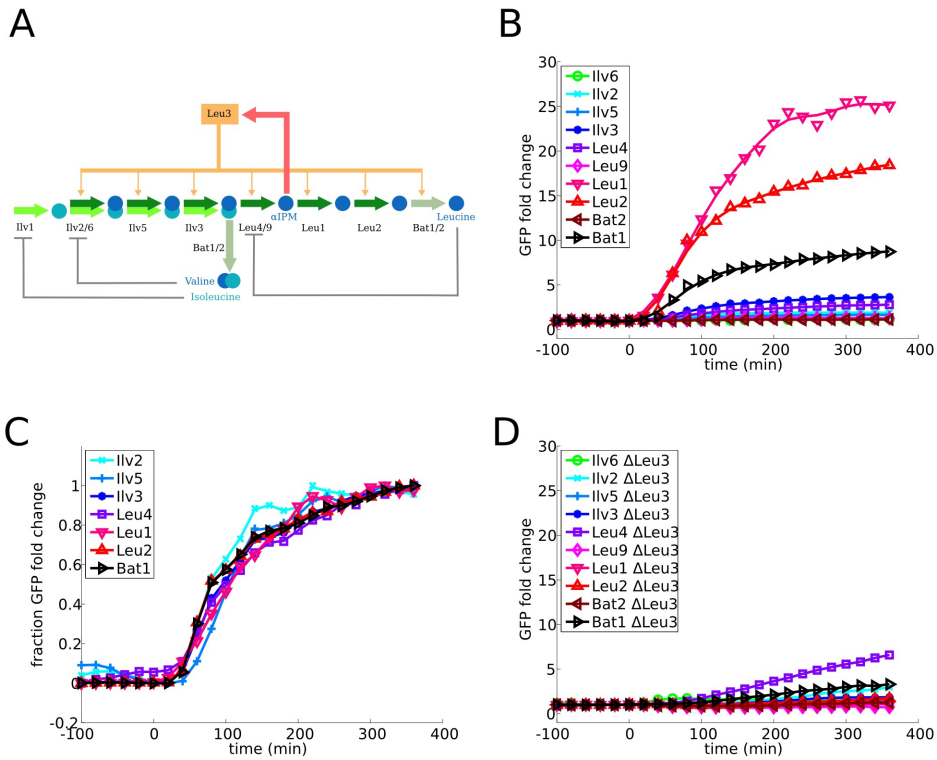

Figure 2: Dynamic profiles of leucine biosynthesis enzymes. (a) The leucine biosynthesis pathway in yeast is an IAA network regulated by the transcription factor Leu3, which senses the intermediate metabolite αIPM. (b) mean GFP fluorescence levels normalized to the level before environmental shift show that Leu1 is the highest induced gene in response to leucine depletion and Leu2 is also strongly affected. (c) GFP levels normalized to both initial and final levels shows no significant timing difference between pathway genes. Genes with negligible fold-changes are not shown. (d) deletion of *LEU3* abolishes Leu1 and Leu2 induction, but actually increases Leu4 induction. In figures 2-5, data shown is an average of two independent experiments done on the same day. Data from identical experiments on different days was also quantitatively consistent and is

shown in the supplementary information.

We did not observe significant timing differences between pathway enzymes when genes were normalized to both initial and final levels (figure 1c), suggesting that all genes in the pathway sense the stimulus at the same time. We also verified that the strong induction of Leu1 and Leu2 was dependent on the transcription factor Leu3. When *LEU3* was deleted, Leu1 and Leu2 induction fell drastically, whereas other enzymes were affected to a lesser degree, and Leu4 was actually induced more strongly (figure 1d). Evidence for Leu3 playing the role of transcriptional repressor instead of activator exists and has been discussed previously (Kohlhaw, 2003; Chin et al., 2008).

Lysine biosynthesis

Lysine biosynthesis in yeast occurs via a linear pathway from 2-oxoglutarate. Regulation is present by inhibition of the first committed pathway step (catalyzed by Lys20 and Lys21) by the end product lysine, and also by transcriptional regulation of the pathway enzymes. The major transcription factor involved is Lys14, which has been shown to regulate transcription of most enzymes in the pathway (Ramos et al., 1988). Lys14 can bind DNA constitutively but activates transcription only after a conformational change, induced by the binding of the small molecule α-aminoadipate-6-semialdehyde (α-AAS) (El Alami et al., 2002; Feller et al., 1999). α-AAS is an intermediate of the pathway,

which means that the regulatory network topology is another example of the IAA, similar to the leucine biosynthesis pathway.

We measured expression of all characterized genes in the pathway in response to lysine starvation (fig. 3). The most striking feature of the response is that Lys9, the enzyme immediately downstream of the regulatory intermediate α-AAS, has an induction ratio of over 40-fold, significantly higher than any other enzyme in the pathway. We verified that this strong induction was caused by regulation by the transcription factor Lys14. In a *lys14*Δ strain, the strong induction of Lys9 was completely abolished, while the other enzymes were not affected to nearly as large a degree. Lys1, the other enzyme downstream of the regulatory intermediate, had an intermediate level of induction, and interestingly, was largely unaffected by a *lys14*Δ perturbation, suggesting that it may be regulated by a completely separate mechanism. This observation is consistent with the previous data in Ramos et.al (Ramos et al., 1988). Similarly to our data for leucine biosynthesis, we did not observe significant timing differences among pathway enzymes (fig. 3c).

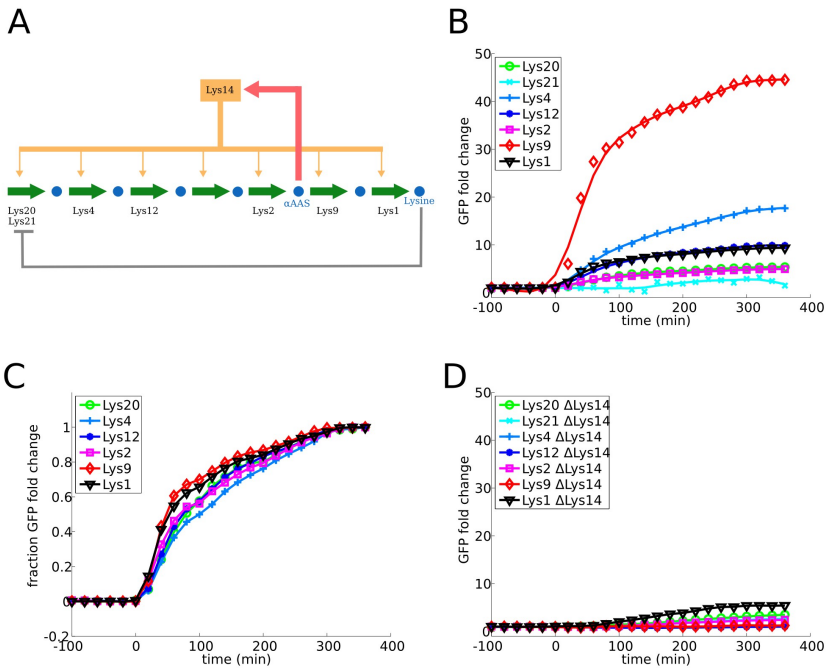

Figure 3: Dynamic profiles of lysine biosynthesis enzymes. (a) The lysine biosynthesis pathway in yeast is an IAA network regulated by the transcription factor Lys14, which senses the intermediate metabolite α-AAS. (b) mean GFP fluorescence levels normalized to the level before environmental shift show that Lys9 is the highest induced gene in response to lysine depletion. (c) GFP levels normalized to both initial and final levels shows no significant timing difference between pathway genes. Genes with negligible fold changes are not shown. (d) deletion of *LYS14* abolishes Lys9 induction but in general has a smaller effect on enzymes upstream of α-AAS.

Adenine biosynthesis

Similarly, adenine biosynthesis takes place via a nearly linear pathway starting with PRPP (phosphoribosyl pyrophosphate). The first committed step (catalyzed by Ade4) is inhibited by adenine, and transcriptional regulation of the pathway involves the factors Bas1 and Bas2. Only when Bas1 and Bas2 form a complex at the promoter do they activate transcription, and their interaction is modulated by either of the two small molecules AICAR and SAICAR, intermediates of the biosynthesis pathway (Rébora et al., 2001).This creates a third example of the IAA architecture. There is some interesting interplay between the histidine and adenine biosynthesis pathways as histidine biosynthesis can also produce AICAR as a side-product (Rébora et al., 2005), but we do not consider this aspect here.

The profiles for the all the genes in the pathway are shown in figure 4. Again, the striking feature of the results is that the Ade17 has by far the highest level of induction, with about 15-fold change in response to adenine depletion. Ade17 is one of the two isozymes which catalyze the conversion of the regulatory intermediate into the next metabolite. However, the other isozyme, Ade16, performs a smaller fraction of the synthesis, and its expression is known not to be regulated at all by adenine levels (Tibbetts and Appling, 2000) as observed also in our experiments. The other enzymes in the pathway exhibit about a 2-fold change, except for Ade1 and Ade2 which have about an 8-fold induction. When the gene encoding the transcription factor *BAS1* was deleted, Ade17 induction was completely abolished, showing that as in the previous two examples, the enzyme downstream of the intermediate has by far the strongest dependence on the transcriptional

regulator. While previous biochemical evidence suggests that both AICAR and SAICAR can affect transcription factor activity, we speculate that the major regulator is AICAR since Ade17, but not Ade13, was found to be very strongly induced by adenine depletion. We comment further on this in the discussion section.

After IMP synthesis, the step catalyzed by Ade17, the pathway branches into guanine (GMP) syntehsis and adenine (AMP) synthesis. Ade12 and Ade13, the enzymes on the AMP branch, were induced to only a moderate level and only depended weakly on Bas1. It is possible that the branching introduces other constraints on the network regulation due to different demands for flux on the two branches.

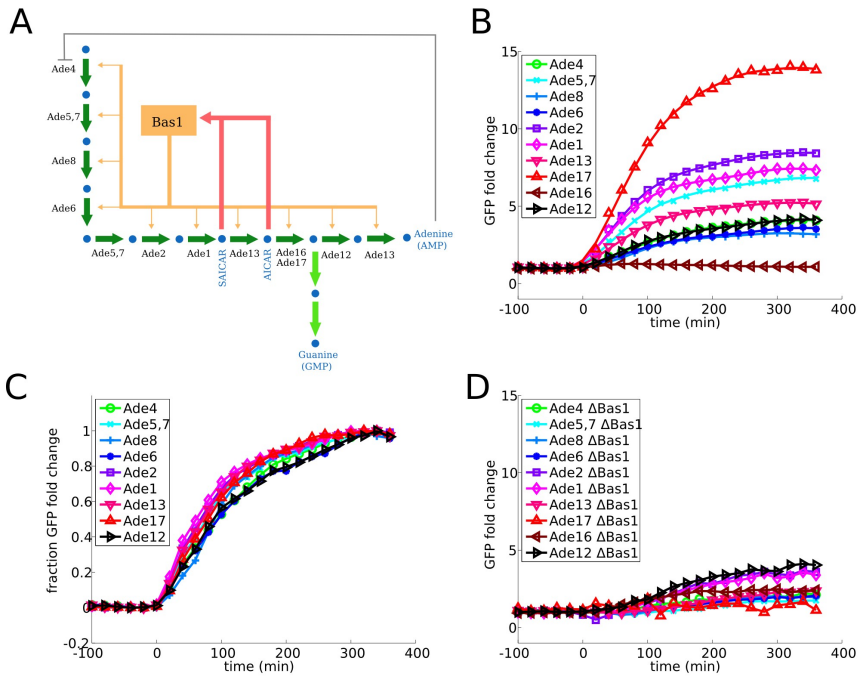

Figure 4: Dynamic profiles of adenine biosynthesis enzymes. (a) The adenine biosynthesis pathway in yeast is an IAA network regulated by the transcription factor Bas1, which senses the intermediate metabolites SAICAR and AICAR. (b) mean GFP fluorescence levels normalized to the level before environmental shift show that Ade17 is the highest induced gene in response to adenine depletion. (c) GFP levels normalized to both initial and final levels shows no significant timing difference between pathway genes. Genes with negligible fold changes are not shown. (d) deletion of *BAS1* abolishes Ade17 induction but in general has a smaller effect on enzymes upstream of AICAR.

Arginine biosynthesis

In all three pathways with IAA regulation, we observe a strikingly similar pattern of gene

expression. To contrast this with an example of the end-product inhibition architecture (fig. 5), we also measured the gene expression dynamics of the arginine biosynthesis pathway. Arginine is synthesized from glutamate via a nearly linear pathway. While there are still some unknowns regarding the allosteric regulation by arginine, it appears that Arg2 and Arg5,6, which catalyze the first two enzymatic steps, form a complex in which both steps are performed, and that both activities are inhibited by arginine (Abadjieva et al., 2001; Davis, 1986). The major mode of transcriptional regulation for this pathway involves a complex of several transcription factors which is generally called ArgR. In the presence of arginine this complex binds to DNA to repress transcription of the pathway genes, but is unbound in the absence of arginine (Amar et al., 2000), thus resulting the the end-product inhibition architecture as in figure 1c.

We found that all the enzymes in this pathway were induced by about 2-4 fold in response to arginine starvation, with similar timing (fig 5). This is consistent with the observation that in this regulatory network, all enzymes share the same effect on the level of active transcription factor and do not create different feedback structures.

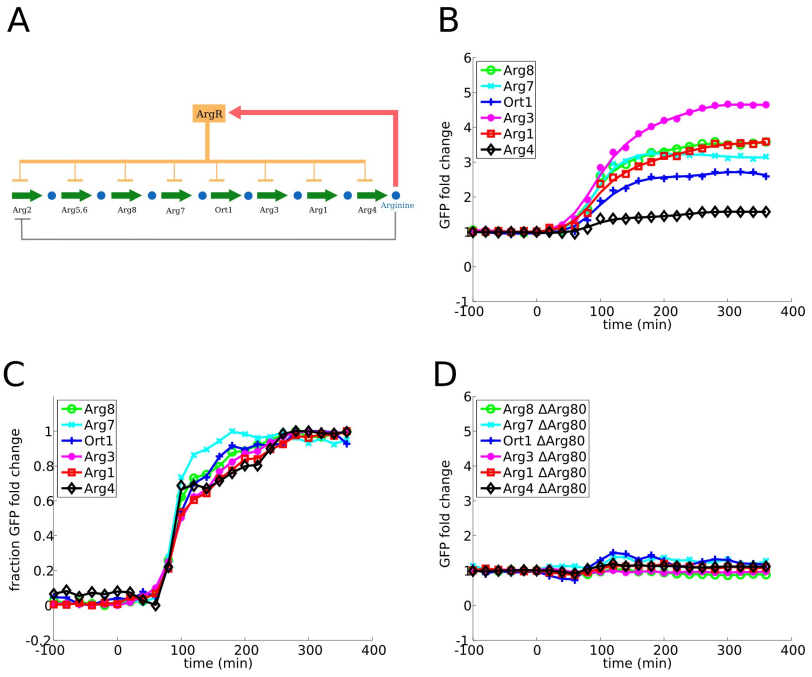

Figure 5: Dynamic profiles of arginine biosynthesis enzymes. (a) The arginine biosynthesis pathway in yeast is an EPI network regulated by the transcription factor complex ArgR, which senses arginine directly. Ort1, the ornithine transporter, is not a catalytic enzyme but rather transports the intermediate ornithine from the mitochondria to the cytosol for further conversion into arginine; it is shown in the appropriate position in the pathway. (b) mean GFP fluorescence levels normalized to the level before environmental shift show that all genes have approximately equal magnitudes of induction except for weaker induction of Arg4. We were unable to monitor expression of Arg2 or Arg5,6 due to very low basal levels of expression. (c) GFP levels normalized to both initial and final levels shows no significant timing difference between pathway genes. (d) deletion of *ARG80*, one of the components of ArgR, abolishes induction of all enzymes.

A theoretical cost-benefit model

Our observation of similar expression patterns across several metabolic pathways with IAA suggests that there may be a common design principle underlying their regulation. The contrast between IAA and EPI also indicates that the feedback structure of the regulatory network can severely constrain gene expression response. To explore whether optimization by natural evolution can create the observed pattern we developed a theoretical model to look quantitatively at the effect of different gene induction profiles on cellular growth. The key element of the model is the tradeoff between the costs of making a protein (energy and metabolic resources) and the benefits of making it (its cellular function), a fundamental idea in gene regulation. This tradeoff has been observed *in vivo* in a number of studies across different organisms in which higher fitness was observed for cells that did not express unnecessary genes (Suiter et al., 2003; Dekel and Alon, 2005; Lang et al., 2009; Gore et al., 2009).

This tradeoff has been the basis for a number of quantitative models for optimal gene expression in a metabolic pathway. Klipp et.al. modeled the metabolic cost by constraining the amount of available total enzyme concentration, and then asked if there was an optimal temporal profile for enzyme concentrations (Klipp et al., 2002). This resulted in a just-in-time (JIT) behavior, where enzymes were expressed in the order they

were needed in the pathway. However, this model assumed a black box regulatory network in which enzyme concentrations were assigned an arbitrary function of time. A followup on this model was published by Zaslaver et.al. in their study of amino acid biosynthesis pathways in *E. coli* (Zaslaver et al., 2004). Their model invoked a regulatory network with end-product inhibition architecture, but assumed an initial steady state in which all enzyme levels were at zero, and an infinitely fast drop in product concentration at time zero. In this study we use a more general model to explore different network architectures on top of the metabolic pathway, and allow for more realistic modeling of the starvation stimulus and the evolutionary environment.

We start by developing a model for cellular growth as a function of metabolic pathway response to starvation (fig. 6). We assume that two factors negatively impact cell growth: production of enzymes, which expends cellular reserves of energy and nutrients; and lack of flux through the metabolic pathway, i.e. lack of production of the nutrient which is absent in the environment and is required for growth (hereafter called "product"). While the functional form of the equation is inspired by the model in Zaslaver et.al, there are several key changes that make the model more realistic. The most important of these are the assumption that cells maintain a basal level of enzyme even in non-starvation environment, and that the product concentration does not drop from infinity to zero immediately upon starvation, but rather decays exponentially until the pathway is activated.

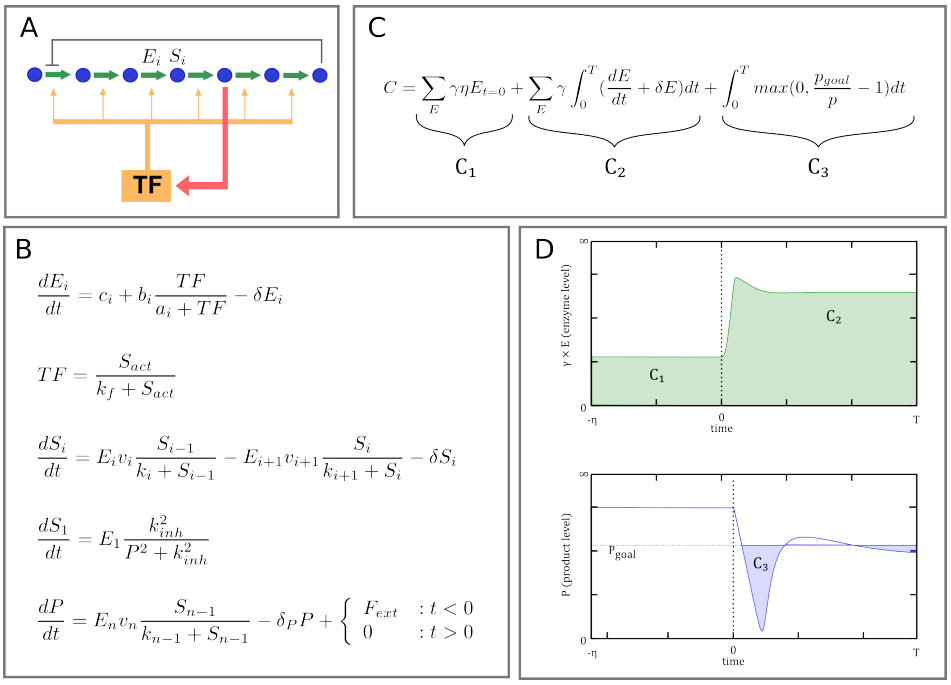

Figure 6: A cost-benefit model for gene expression in a metabolic pathway upon nutrient depletion. (a) a generalized linear pathway with one possible regulatory architecture. (b) a set of differential equations describes the dynamics of enzyme induction and product formation. E_1-E_6 are the concentrations of the enzymes. c_i, b_i and a_i are the regulatory parameters that define the effect of active transcription factor on enzyme synthesis. S_1-S_6 are the metabolites, with S_6=P being the product of the pathway and S_0 the upstream source. We assume S_0=∞. k_i and v_i are the Michaelis-Menten constant and the maximum rate of enzyme activity (catalytic constant) respectively for each enzyme. k_f is the concentration of regulatory metabolite necessary for half-maximal transcription factor activity, and k_{inh} is the concentration of product necessary for half-maximal inhibition of the first pathway step. δ is the rate of dilution of enzymes and substrates due to growth and δ_P is the rate of product usage. F_{ext} is the external product flux before starvation. (c) the cost function used for optimization of regulatory parameters is an estimate of the

growth penalty imposed by insufficient product flux or unneeded protein expression. The three terms are C1, which corresponds to the cost of the basal level of enzyme synthesis, C2, which corresponds to the cost of enzyme synthesis during starvation, and C3, which corresponds to the deficiency of product during starvation. The evolutionary parameters that define the relative magnitudes of these terms are **γ**, the unit cost of enzyme synthesis; **η**, the time spent in non-starvation conditions; and **T**, the time spent in starvation conditions. (d) a graphical illustration of the cost function components.

Thus, there are three terms in the cost function that correspond to reduced growth due to 1) basal level enzyme production 2) enzyme production upon starvation, and 3) lack of product during starvation. Assigning weights to the three terms results in three meta-parameters, which can also be thought of as environmental or evolutionary parameters, since they have clear interpretations in terms of the environment in which regulation of this pathway evolved. These are **γ**, which corresponds to the cost (growth reduction) from synthesizing one additional unit of enzyme; **η**, the amount of time spent in non-starvation conditions; and **T**, the amount of time spent in starvation conditions. By altering these parameters, our model can predict that different environments may cause evolution of different types of regulation.

We used our model to ask the question of whether different network architectures impose different constraints on gene expression profiles, under the assumption that the regulatory

parameters have been optimized by evolution. By fixing the biochemical properties of the enzymes, and allowing only the parameters related to transcriptional regulation to vary, we make the implicit assumption that more optimal cis-regulation can evolve much faster than more optimal enzyme properties, an assumption consistent with a number of studies that have suggested that cis-regulatory regions evolve more rapidly than protein coding regions (Kellis et al., 2003; Chin et al., 2005) and are under more efficient selection (Wray et al., 2003; Stern, 2000; Wray, 2007). We considered a linear pathway of six enzymatic steps, and considered six regulatory strategies, which corresponded to six different metabolites directly affecting transcription factor activity. For each one, we optimized the cost function using numerical methods. Interestingly, as shown in fig. 7, the different networks showed different gene expression strategies when allowed to evolve optimal regulation.

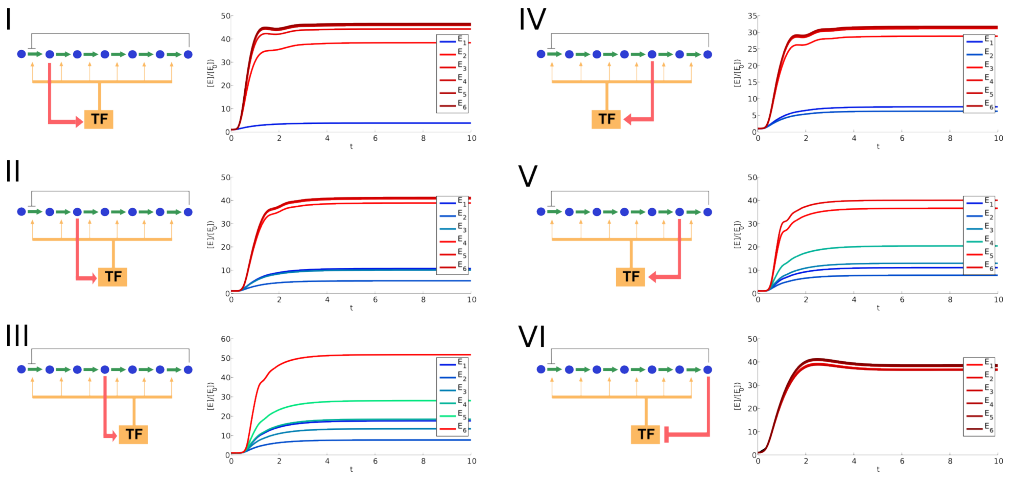

Figure 7. Predicted optimal dynamic profiles of gene induction under six different regulatory network architectures. Some curves which overlap perfectly have been

artificially separated by a small offset.

In particular, for networks I-V, in which an intermediate metabolite controls transcription factor activity (IAA networks), the optimal network response involves a separation of responses: strong induction for enzymes downstream of the controlling intermediate, and weaker induction for enzymes upstream of the controlling intermediate. We consider fold change as opposed to absolute level since the latter can be scaled by scaling k_{cat} for a particular enzyme, while the relative level is insensitive to the particular enzyme parameters.

This separation of responses is similar to our observed dynamics in the yeast leucine, lysine, and adenine biosynthesis pathways, and can be explained by a simple observation. Looking at the network topology from the point of view of the intermediate metabolite, upregulation of the upstream genes creates a positive feedback loop, since higher levels of upstream enzymes lead to higher levels of the intermediate metabolite. However, upregulation of downstream genes creates a negative feedback loop, since higher enzyme levels deplete the intermediate metabolite. Strong negative feedback has long been a well-known design principle for strong and fast upregulation of gene expression (Rosenfeld et al., 2002; Alon, 2006). While strong positive feedback is invaluable in switching or bistable systems, it can often have deleterious consequences for fast response in adaptive systems (Wilhelm, 2009).

The aforementioned results were obtained by choosing reasonable values for the evolutionary parameters $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, and $\mathbf{T}$. Since the true values are largely unknowable, we repeated the analysis using a wide range of values for these parameters, creating a phase diagram of network behaviors. The results for network III are shown in fig. 8. For a wide range of parameters, we see the behavior observed previously, with a separation of responses between upstream and downstream genes. The area in the lower left corner corresponds to very low enzyme synthesis cost or very frequent starvation. In these cases, the model predicts that it is actually optimal to have no regulation and to express all the enzymes constitutively. The area in the upper right represents the case when enzyme synthesis cost is very high. We find no particular pattern among these solutions, and for several reasons suggest that this area corresponds to solutions with little biological relevance. One of these is that the enzyme synthesis cost is so high that these solutions often involve induction ratios of well over 10^4, a situation rarely observed in real metabolic gene expression regulation, especially in yeast. Similar results were obtained for the other IAA networks (figure S1).

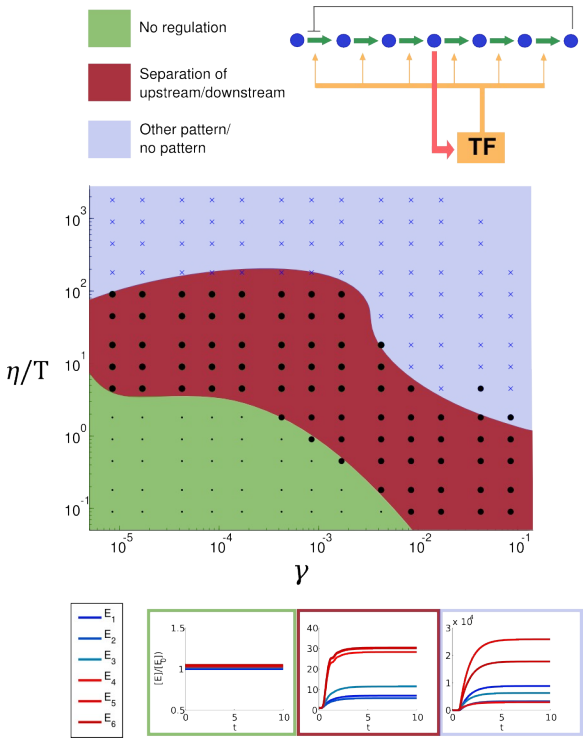

Figure 8. Phase diagram of optimal expression dynamics for an IAA network. The network is shown in the top right. Each point in the graph corresponds to a pair of values (**γ**, **η/T**) represents an optimization of the cost function over the regulatory parameters. Each solution was classified based on the resulting enzyme expression dynamics. The red region corresponding to intermediate values of the evolutionary parameters shows a pattern of separation of enzymes dynamics before and after the regulatory metabolite. The x-axis (**γ**) represents increasing cost of enzyme synthesis, relative to the cost of product starvation, while the y-axis (**η/T**) represents decreasing frequency of starvation. Behavior that depended on **T** independent of **η** was found only for very small **T**. At the bottom, gene expression profiles from sample solutions from the three regions are shown.

For network VI, in which the product of the pathway regulates the transcription factor, we found the optimal response to involve almost identical expression profiles for every enzyme in the pathway. This behavior was found to be robust for a very wide range of meta-parameters. In this network, there is no positive feedback loop, and thus no constraint on induction of the enzymes at the top of the pathway. This agrees with our observations in the yeast arginine biosynthesis pathway where all enzymes had similar induction dynamics.

Discussion

We have shown that in a number of biosynthetic pathways in yeast, gene expression dynamics depend on the underlying regulatory architecture. In particular, we have observed that under the IAA, the enzyme immediately downstream of the regulatory intermediate is induced much more strongly than any other enzyme in the pathway. These include Leu1 in leucine biosynthesis, Lys9 in lysine biosynthesis, and Ade17 in adenine biosynthesis. In some cases, other enzymes downstream of the intermediate also have a relatively strong induction level. In arginine biosynthesis, which is regulated by the EPI architecture, we did not observe significant differences in gene induction among pathway enzymes.

By using a theoretical model to balance the relative costs and benefits of gene expression

level and pathway flux in response to nutrient depletion, we have shown that organisms are likely to evolve different patterns of gene expression depending on the regulatory architecture used to control the pathway. In particular, for networks in which an intermediate metabolite interacts with the transcriptional regulator, the theoretical model predicts that the most favorable gene expression dynamics involve a much stronger induction for enzymes downstream of the intermediate than for those upstream of the intermediate.

While we have concentrated on the prediction of gene expression profiles for individual networks, our model also predicts a final cost value for each network, and we can compare the costs between the six networks in figure 7. We find that across virtually the entire phase space, network VI consistently has the lowest cost, followed by networks I, II, III, IV, V in that order. This is somewhat intuitive since this is the order in which the networks sense the starvation signal, from earliest (directly sensing lack of product) to latest (the last intermediate in the pathway). However, it does create the question of why not all biological pathways use the least costly regulatory topology. We hypothesize that this is due to the high evolutionary barrier of switching the regulatory program. A switch in regulatory program must involve the evolution of a new metabolite-TF interaction, as well as a concurrent evolution of each of the promoters involved, and is likely to create a very unfavorable intermediate state. Nevertheless, this question deserves attention, and multi-species studies are underway to investigate this topic.

Our theoretical model can make several predictions in the reverse direction – namely the prediction of network architecture based on gene expression data. One example of the first case is our discussion in the adenine pathway of the roles of AICAR and SAICAR. Our results suggest that AICAR is the major regulator of Bas1 activity (through Bas1/Bas2 interaction) based on the strong induction of Ade17 and weaker induction of Ade13. It may also be possible to use gene expression profiles to predict the regulatory architecture controlling a completely uncharacterized pathway.

Our model is undeniably a vast simplification of several biological features. One aspect of this is the accuracy of the cost function, which is meant to be inversely proportional to growth rate. Despite the many studies that have noted the effect of unnecessary gene expression on growth rate, no general quantitative model has been developed. In addition, it is naive to consider the flux through the linear pathway in isolation from the rest of cellular metabolism. For instance, turning the pathway on diverts flux from central metabolism that could be used for other biomass, or turning on the pathway could upset the balance of various cofactors. These effects are certainly complex and given these factors, we do not claim that our model is a quantitatively accurate model for cellular growth. Nevertheless, we feel that it captures the major components of the cost/benefit structure: a penalty for high gene expression, and a penalty for low pathway flux. Moreover, our use of meta-parameters to explore a large phase space in large part compensates for the uncertainty in the model.

We note also that the model predicts virtually identical expression profiles for all enzymes downstream of the control point in IAA networks, while our data only shows consistently a strong induction for the enzyme immediately after the control point. There are several reasons why this could be the case. One possibility is that we are observing some natural variation away from the optimal solution, which is lower for the enzyme immediately after the control point than for other downstream enzymes. Indeed we show (figure S2) that the sensitivity of the cost function is highest to parameters that affect the expression of the enzyme immediately downstream; that is, changing expression of the enzyme downstream of the intermediate is most detrimental to the behavior of the system. This is because this expression level tightly controls the level of regulator and will affect expression in the entire pathway. The second possibility is that our cost function has not captured fully the way by which fitness depends on expression level, as discussed in the previous paragraph. It is likely that cells are trying to optimize many things other than the cost of gene expression. One possibility is that strong regulation of the immediately downstream enzyme could enhance system stability or reduce noise. Nevertheless, both stability and noise reduction are aspects of strong negative feedback, which was the basis of differential induction in our model. We feel that the key finding of strong dependence of expression profiles on network architecture and in particular on the feedback structure will be robust as more interesting models for cellular fitness are considered.

We have noted that we do not observe significant timing differences between pathway

enzymes. Previous work had shown that in *E. coli* amino acid biosynthesis pathways, response to starvation involved sequential induction of pathway enzymes, in the order of their position in the pathway (Zaslaver et al., 2004). Furthermore, a theoretical model could produce a similar just-in-time type of behavior under some circumstances (Klipp et al., 2002; Zaslaver et al., 2004). Regarding the model, we found that we could reproduce just-in-time behavior, but only under two particular conditions. For the first, we considered a situation where instead of being constrained by a regulatory network, expression of each enzyme could be controlled arbitrarily. We reran our optimization algorithm under conditions where each enzyme could be controlled independently with an arbitrary sigmoid function, with the parameters of this sigmoid optimized according to the same cost function. We found (figure S3) that this produced the classic just-in-time behavior. A second scenario in which just-in-time regulation could be reproduced by our model was when steady state expression levels were much less important than transient levels. This occurred when the meta-parameter **T**, representing the duration of starvation, was very small. In that regime, the final concentration of the enzymes were irrelevant, and drastic overshooting is not punished by the cost function (supplementary figure). It can be shown that under the standard model of gene regulation that we use, involving one transcription factor, it is impossible to maintain identical basal and maximal levels for two different enzymes while maintaining different timing of responses. In order to create different timing when only one regulator is used, either the basal level or the induced level must also be different. Thus, given our model and cost function, in a regime where the duration of starvation is not insignificant, and where the basal level of the enzyme is

not zero, just-in-time is not optimal, since it will result in highly suboptimal steady state enzyme levels. It is possible that the reason just-in-time regulation was observed in *E. coli* amino acid biosynthesis pathway is precisely because *E. coli* evolved in such a regime where steady state levels were unimportant. However, we also note that this study only considered absolute enzyme abundance as opposed to fold change from basal level, since the basal levels are extremely difficult to measure accurately in *E. coli*. for most enzymes. It is possible that with a more careful measure of basal expression, the two results will become consistent. Another possibility is that the regulatory architectures for the pathways considered in that study are not yet fully understood. If a second factor is involved in regulating expression, a just-in-time phenotype is more feasible.

Another recent study suggests that the effect of separated induction levels under the IAA is present outside of yeast. We obtained time course data for expression of a large number of amino acid biosynthesis genes under depletion for six different amino acids in *E. coli* from Yamada et.al. (Yamada et al., 2010). While the same problem with measurement of basal level is present, these data suggest the same pattern that we have observed in yeast may also be present in *E. coli*. For instance, the lysine biosynthesis pathway in *E.coli* is regulated by the transcription factor LysR, which binds the metabolic intermediate diaminopimelate to become active (Stragier et al., 1983). Three other pathways: methionine biosynthesis, cysteine biosynthesis, and valine biosynthesis all are thought to be controlled by an IAA network, with intermediate coinducers homocysteine, acetylserine, and α-acetolactate respectively (Urbanowski and Stauffer, 1989; Ostrowski

et al., 1987; Wek and Hatfield, 1988) (review in (Schell, 1993)). In each case, the enzyme downstream of the intermediate (MetE, CysM/CysK, and IlvC, respectively) has a higher change in expression than any other enzyme in the pathway.

There are two design principles for the regulation of metabolic pathways which we have introduced in this work: 1) optimal dynamics of gene induction are strongly dependent on the underlying network architecture and 2) under the IAA, strong induction of enzymes downstream of the intermediate is highly favorable. The latter is one application of the more general principle of using strong negative feedback for fast response. Nevertheless, it is striking that these results can be deduced from an extremely simple mathematical model of evolutionary optimality. It is likely that the model captures the basic principles behind the biology, and it is possible that in the future, similar models could be used to tackle more complex problems.

Materials and Methods

Strains and plasmids

All yeast strains are derived from S288c MATα ura3-52. Transcription factor knockouts were constructed by replacement of the target gene with the NatMx casette by one-step homologous recombination. For promoter-GFP constructs, plasmid pVC02 was constructed by amplifying yeGFP from pFA6-GFP-KanMx with SacII and NgoMIV

flanking sequences, digesting, and subcloning into PRS306. In addition pVC02 contains a silent mutation in yeGFP creating a SacI digest site near the start codon, allowing integration of the promoter and 5' UTR before GFP with no restriction enzyme scar. The silent mutation had no effect on GFP expression. Plasmids with each promoter were then constructed by amplifying a 720bp fragment directly upstream of the start codon of each gene using primers with flanking digest sites E1 and E2 (E1={XbaI,NotI,BafmHI}, E2={SacI,SacII}). After digestion and ligation with appropriately digested pVC02, plasmids were verified by sequencing, and digested at a restriction site inside the URA3 gene for transformation into yeast via standard protocols (Gietz and Woods, 2002).

Media

Growth before starvation ($t<0$) was on SD-complete media. Amino acids and nucleotides were present at the following concentrations: adenine: 40mg/L arginine: 20mg/L; histidine: 20mg/L; isoleucine: 30mg/L; leucine: 100mg/L; lysine: 30mg/L; methionine: 20mg/L; phenylalanine: 50mg/L; threonine: 200mg/L; trypophan: 40mg/L; tyrosine: 30mg/L; uracil: 20mg/L; valine: 150mg/L. Starvation media was identical except for the absence of the appropriate amino acid.

Flow Cytometry

Cultures were grown in deep well 96-well plates, with a volume of 500ul per well. Overnight cultures were diluted and grown for 14-18 hours to OD 0.05-0.1 while fluorescence was monitored and observed to reach steady state. At that point media was removed by centrifugation, and cells were resuspended in dropout media. During both pre-starvation and post-starvation growth, a customized robotic liquid handler periodically diluted the cultures with fresh media and delivered samples to an LSRII flow cytometer (Beckton-Dickinson). GFP was excited with a 488nm laser and emission was collected with a 530/30 filter. Cell populations were filtered by gating on the forward and side scatter values, and total GFP fluorescence was normalized to side scatter to give an approximate measure of GFP concentration (Salzman, 2001).

Cost/benefit model and parameter optimization

All computation was done using software written by the authors. The SUNDIALS package (Hindmarsh et al., 2005) was used for numerical solution of the ordinary differential equations. For a given network architecture and values of the evolutionary parameters, optimization of the regulatory parameters was done by simulated annealing with the Metropolis Monte Carlo method (Metropolis et al., 1953; Kirkpatrick et al., 1983). Five independent simulations were done for each optimization problem, and variation in the final objective function was typically below 1%.

References

Abadjieva, A., Pauwels, K., Hilven, P., and Crabeel, M. (2001). A new yeast metabolon involving at least the two first enzymes of arginine biosynthesis: acetylglutamate synthase activity requires complex formation with acetylglutamate kinase. J. Biol. Chem *276*, 42869-42880.

Alon, U. (2006). An Introduction to Systems Biology: Design Principles of Biological Circuits 1st ed. (Chapman and Hall/CRC).

Amar, N., Messenguy, F., El Bakkoury, M., and Dubois, E. (2000). ArgRII, a component of the ArgR-Mcm1 complex involved in the control of arginine metabolism in Saccharomyces cerevisiae, is the sensor of arginine. Mol. Cell. Biol *20*, 2087-2097.

Artz, S. W., and Broach, J. R. (1975). Histidine regulation in Salmonella typhimurium: an activator attenuator model of gene regulation. Proceedings of the National Academy of Sciences of the United States of America *72*, 3453.

Chin, C., Chuang, J. H., and Li, H. (2005). Genome-wide regulatory complexity in yeast promoters: separation of functionally conserved and neutral sequence. Genome Res *15*, 205-213.

Chin, C., Chubukov, V., Jolly, E. R., DeRisi, J., and Li, H. (2008). Dynamics and Design Principles of a Basic Regulatory Architecture Controlling Metabolic Pathways. PLoS Biol *6*, e146.

Davis, R. H. (1986). Compartmental and regulatory mechanisms in the arginine pathways of Neurospora crassa and Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews *50*, 280.

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588-592.

El Alami, M., Feller, A., Piérard, A., and Dubois, E. (2002). The proper folding of a long C-terminal segment of the yeast Lys14p regulator is required for activation of LYS genes in response to the metabolic effector. Mol. Microbiol *43*, 1629-1639.

Feller, A., Ramos, F., Piérard, A., and Dubois, E. (1999). In Saccharomyces cerevisae, feedback inhibition of homocitrate synthase isoenzymes by lysine modulates the activation of LYS gene expression by Lys14p. Eur. J. Biochem *261*, 163-170.

Gietz, R. D., and Woods, R. A. (2002). Transformation of yeast by the LiAc/SS carrier DNA/PEG method. Methods Enzymol *350*, 87–96.

Gore, J., Youk, H., and van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. Nature *459*, 253-256.

Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995). Improved green fluorescence. Nature *373*, 663-664.

Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., and Woodward, C. S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Trans. Math. Softw. *31*, 363-396.

Hittinger, C. T., and Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. Nature *449*, 677-681.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol *3*, 318-356.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated

Annealing. Science *220*, 671-680.

Klipp, E., Heinrich, R., and Holzhutter, H. (2002). Prediction of temporal gene expression. Metabolic optimization by re-distribution of enzyme activities. Eur J Biochem *269*, 5406-5413.

Kohlhaw, G. B. (2003). Leucine biosynthesis in fungi: entering metabolism through the back door. Microbiol. Mol. Biol. Rev *67*, 1-15, table of contents.

Lang, G. I., Murray, A. W., and Botstein, D. (2009). The cost of gene expression underlies a fitness trade-off in yeast. Proc. Natl. Acad. Sci. U.S.A *106*, 5755-5760.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. 1953. J. Chem. Phys *21*, 108.

Ostrowski, J., Jagura-Burdzy, G., and Kredich, N. M. (1987). DNA sequences of the cysB regions of Salmonella typhimurium and Escherichia coli. J. Biol. Chem *262*, 5999-6005.

Ramos, F., Dubois, E., and Piérard, A. (1988). Control of enzyme synthesis in the lysine biosynthetic pathway of Saccharomyces cerevisiae. Evidence for a regulatory role

of gene LYS14. Eur. J. Biochem *171*, 171-176.

Rébora, K., Desmoucelles, C., Borne, F., Pinson, B., and Daignan-Fornier, B. (2001). Yeast AMP pathway genes respond to adenine through regulated synthesis of a metabolic intermediate. Mol. Cell. Biol *21*, 7901-7912.

Rébora, K., Laloo, B., and Daignan-Fornier, B. (2005). Revisiting purine-histidine cross-pathway regulation in Saccharomyces cerevisiae: a central role for a small molecule. Genetics *170*, 61-70.

Ronen, M., Rosenberg, R., Shraiman, B. I., and Alon, U. (2002). Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. Proc Natl Acad Sci U S A *99*, 10555-10560.

Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. J. Mol. Biol *323*, 785-793.

Salzman, G. C. (2001). Light scatter: detection and usage. Current protocols in cytometry/editorial board, J. Paul Robinson, managing editor...[et al.].

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol *47*, 597-626.

Stern, D. L. (2000). Perspective: Evolutionary Developmental Biology and the Problem of Variation. Evol *54*, 1079.

Stragier, P., Danos, O., and Patte, J. C. (1983). Regulation of diaminopimelate decarboxylase synthesis in Escherichia coli. II. Nucleotide sequence of the lysA gene and its regulatory region. J. Mol. Biol *168*, 321-331.

Suiter, A. M., Bänziger, O., and Dean, A. M. (2003). Fitness consequences of a regulatory polymorphism in a seasonal environment. Proc. Natl. Acad. Sci. U.S.A *100*, 12782-12786.

Tibbetts, A. S., and Appling, D. R. (2000). Characterization of Two 5-Aminoimidazole-4-carboxamide Ribonucleotide Transformylase/Inosine Monophosphate Cyclohydrolase Isozymes from Saccharomyces cerevisiae. Journal of Biological Chemistry *275*, 20920-20927.

Urbanowski, M. L., and Stauffer, G. V. (1989). Role of homocysteine in metR-mediated activation of the metE and metH genes in Salmonella typhimurium and Escherichia coli. J. Bacteriol *171*, 3277-3281.

Wek, R. C., and Hatfield, G. W. (1988). Transcriptional activation at adjacent operators in

the divergent-overlapping ilvY and ilvC promoters of Escherichia coli. J. Mol. Biol *203*, 643-663.

Wilhelm, T. (2009). The smallest chemical reaction system with bistability. BMC Systems Biology *3*, 90.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet *8*, 206-216.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003). The Evolution of Transcriptional Regulation in Eukaryotes. Mol Biol Evol *20*, 1377-1419.

Yamada, T., Ou, J., Furusawa, C., Hirasawa, T., Yomo, T., and Shimizu, H. (2010). Relationship between noise characteristics in protein expressions and regulatory structures of amino acid biosynthesis pathways. IET Syst Biol *4*, 82-89.

Zaslaver, A., Mayo, A. E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M. G., and Alon, U. (2004). Just-in-time transcription program in metabolic pathways. Nat. Genet *36*, 486-491.

Appendix

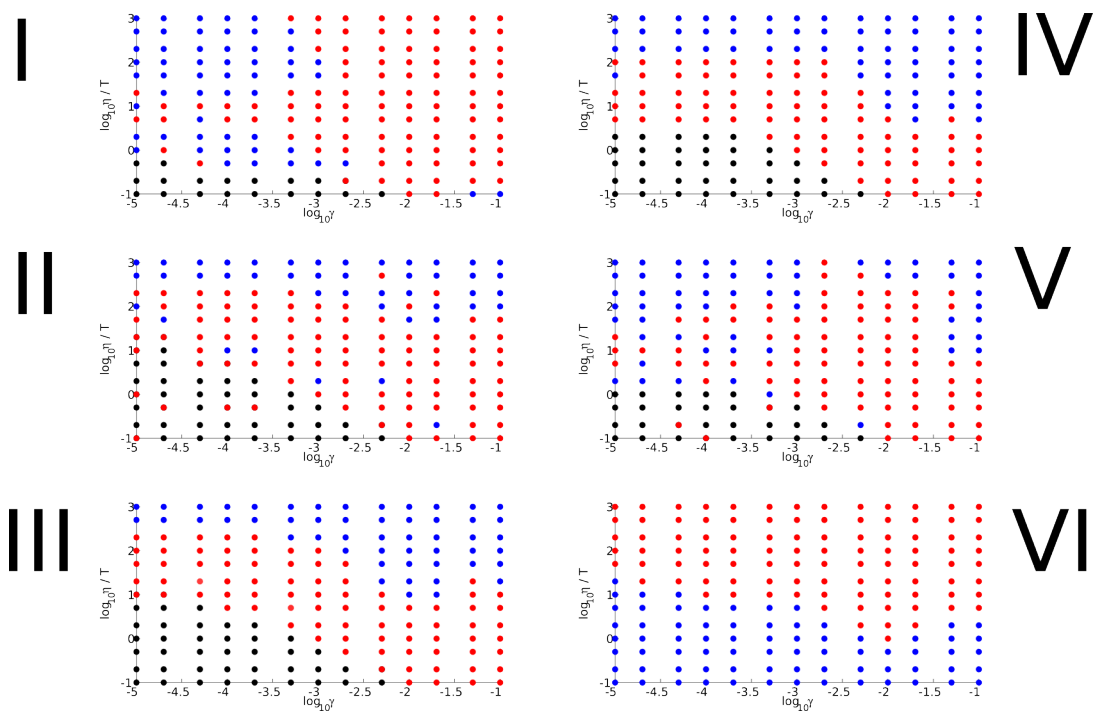

Figure S1. Phase diagrams for all networks. Networks I-VI correspond to the same networks in figure 7 in the text. For networks I-V, red points correspond to solutions with separation of regulation of downstream and upstream enzymes. For network VI, red points correspond to solutions with almost identical induction levels for all enzymes. Black points correspond to solutions with no appreciable induction of any enzymes. Blue points are solutions with no clear pattern.

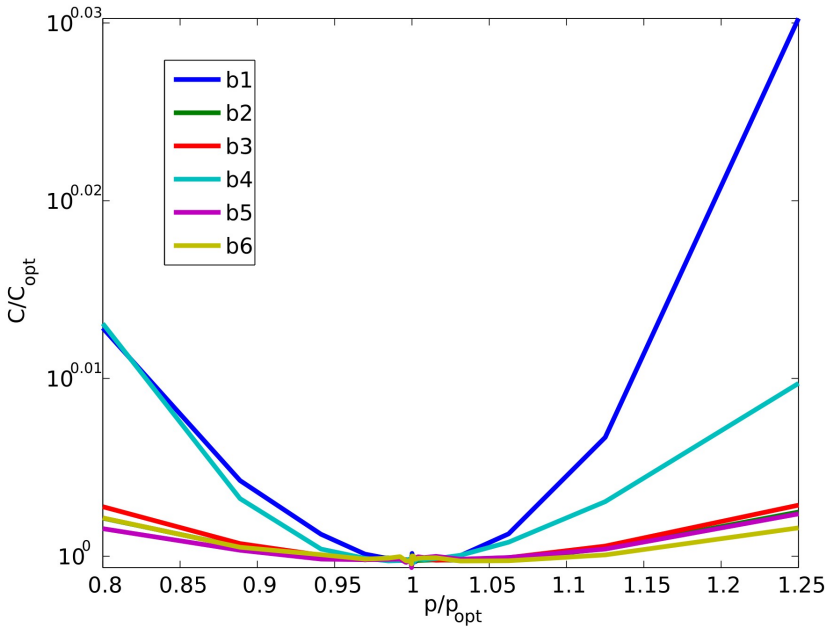

Figure S2. Sensitivity of the cost function to small changes in parameter values away from the found optimum. Parameters b1-b6 correspond to promoter dependence on transcription factor activity for enzymes 1-6 in the pathway. Enzyme 4 is the one immediately downstream of the regulatory intermediate. The cost function is most sensitive to b1 since enzyme 1 controls basal pathway flux, but also highly sensitive to b4 since enzyme 4 exerts the most control over the level of regulatory intermediate.

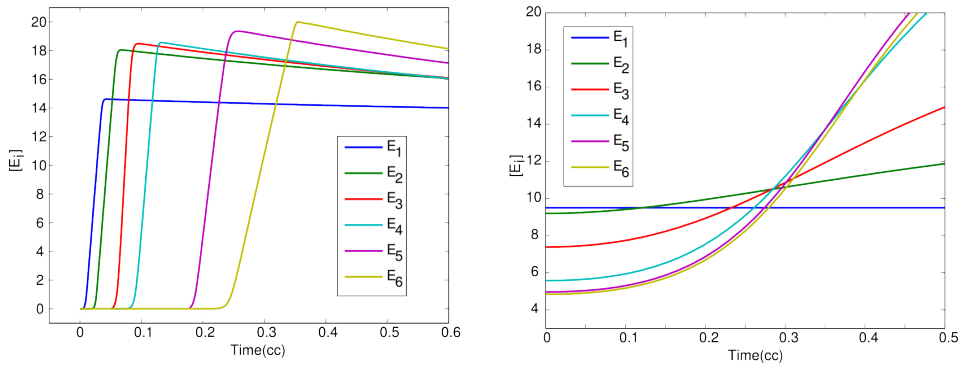

Figure S3. Examples of just-in-time behavior produced by our model. The left graph shows the profiles when the cost function is minimized with the similar meta-parameters to the rest of our analysis, but no regulatory network exists, and enzyme profiles are allowed to be a sigmoid function of three parameters. The right graph shows optimal profiles under network VI (end-product-inhibition) but with T=0.5, so steady state levels are largely irrelevant.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

6/9/2010

Date