Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.0811091106">http://dx.doi.org/10.1073/pnas.0811091106</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Apr 15 08:07:22 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/50252">http://hdl.handle.net/1721.1/50252</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p

Joel F. Moxleya,1, Michael C. Jewettb,1, Maciek R. Antoniewiczc,1,2, Silas G. Villas-Boasb,1,3, Hal Alpera,c, Robert T. Wheelerd, Lily Tonga, Alan G. Hinnebuscha, Trey Idekerd, Jens Nielsenb, and Gregory Stephanopoulusa,4

aDepartment of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; cCenter for Microbial Biotechnology, Technical University of Denmark, BioCentrum-DTU, Building 223, DK-2800 Kongens Lyngby, Denmark; bWhitehead Institute for Biomedical Research, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02139; dLaboratory of Gene Regulation and Development, National Institute of Child Health and Human Development, Bethesda, MD 20892; and eDepartment of Bioengineering, University of California, San Diego, CA 92093

Edited by John Ross, Stanford University, Stanford, CA, and approved February 12, 2009 (received for review November 3, 2008).

Genome sequencing dramatically increased our ability to understand cellular response to perturbation. Integrating system-wide measurements such as gene expression with networks of protein–protein interactions and transcription factor binding revealed critical insights into cellular behavior. However, the potential of systems biology approaches is limited by difficulties in integrating metabolic measurements across the functional levels of the cell despite their being most closely linked to cellular phenotype. To address this limitation, we developed a model-based approach to correlate mRNA and metabolic flux data that combines information from both interaction network models and flux determination models. We started by quantifying 5,764 mRNAs, 54 metabolites, and 83 experimental 13C-based reaction fluxes in continuous cultures of yeast under stress in the absence or presence of global regulator Gcn4p. Although mRNA expression alone did not directly predict metabolic response, this correlation improved through incorporating a network-based model of amino acid biosynthesis (from \( r = 0.07 \) to 0.80 for mRNA-flux agreement). The model provides evidence of general biological principles: rewiring of metabolic flux (i.e., use of different reaction pathways) by transcriptional regulation and metabolite interaction density (i.e., level of pairwise metabolite-protein interactions) as a key biosynthetic control determinant. Furthermore, this model predicted flux rewiring in studies of follow-on transcriptional regulators that were experimentally validated with additional 13C-based flux measurements. As a first step in linking metabolic control and genetic regulatory networks, this model underscores the importance of integrating diverse data types in large-scale cellular models. We anticipate that an integrated approach focusing on metabolic measurements will facilitate construction of more realistic models of cellular regulation for understanding diseases and constructing strains for industrial applications.

Metabolic fluxes are informative indicators of cellular physiology and in vivo homeostasis. Fluxes are the model-independent rates of metabolite interconversion that emerge through the interplay of genes, proteins, and metabolites at multiple regulatory levels. Protein–protein interactions (1) and transcription factor binding (2, 3) can be used to predict a cellular behavior at the gene and protein level (4, 5). However, complex metabolite-enzyme interactions (6), translational regulation (7), and posttranscriptional mechanisms prevent direct linkage of transcriptional state and metabolic phenotype. Previous in silico attempts to address this gap included metabolic network and regulatory on/off switch models for growth and viability prediction (8, 9). In addition, experimental studies of posttranscriptional control have measured mRNA expression in conjunction with either 13C-based flux (10) or metabolite level measurements (11). Such studies have proven invaluable for emphasizing the importance of metabolic phenotypes; however, new approaches are needed which, by integrating these diverse data types in biological models, can identify specific mechanisms by which genetic regulatory circuits mediate metabolic flux phenotype.

Toward this goal, we created a pathway model of amino acid biosynthesis that includes genetic regulatory circuits and metabolite-enzyme interactions to simultaneously integrate flux, metabolite, and mRNA data into a functional biological model. Specifically, our investigation focused on native transcriptional control of amino acid biosynthetic enzymes via the Gcn4p-mediated stress response. This collection of pathways has been studied extensively by both classical genetics and genomic approaches (3, 12), and human counterparts of Gcn4p-regulated pathways have been associated with 209 genetic disorders including Phenylketonuria (PKU). Our study aimed to use this model system to demonstrate that integrating metabolic flux phenotypes would yield new and specific insights to a well-studied set of pathways. Here, we present the results of our measurements, a mechanistic model of amino acid biosynthesis, a predictor of metabolic flux changes, and a discussion of what we learned about the metabolic control structure for this model system.

Results
To study the response enabled by Gcn4p, we compared wild-type yeast to a gcn4Δ-knockout strain (i.e., stress response present versus absent) with both strains cultivated in a chemostat under constant pH, controlled growth rate, and starvation-induced stress (Fig. 1A). Chemostat cultivations, which enable growth under a tightly regulated steady-state environment, eliminate variability that is inherent in dynamically changing batch cultures. In shake-flask studies, for example, it is difficult, and in some cases impossible, to control several key cultivation parameters (e.g., growth rate, media composition, dissolved oxygen concentration, pH, etc.), which often obscures the “real” pa-


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.


2Present address: Department of Chemical Engineering, University of Delaware, Newark, DE 19716.

3Present address: AgResearch Limited, Grasslands Research Centre, Tennent Drive, Private bag 11008, Palmerston North, New Zealand.

4To whom correspondence should be addressed. E-mail: gregstep@mit.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811091106/DCSupplemental.
same specific growth rate (0.10 h
confidences to abundance ratios of 54 observed metabolites in separate, characterize, and quantify metabolites from whole-cell chromatography coupled to mass spectrometry (GC-MS) to
(3AT) concentration, both wild-type and
ment. Wild-type S288C and an isogenic
rays, GC-MS, and HPLC to measure mRNAs, fluxes, and metabolites (see text such as metabolite interaction densities.
view of the Gcn4p stress response and help to characterize network effects,
Gcn4p translational activation and transcriptional activation of hundreds of unsupplemented YNB minimal media regulated to 5
6478/H20841
metabolic pathways (15), and reaction fluxes—defined as the
grams observed in the wild-type and
macroscopic growth under the divergent transcriptional pro-
cell compensated at the level of metabolism to maintain similar
gcn4

cysteine, methionine, and lysine pathways.
increased at least 2-fold for the arginine, asparagine, tryptophan,
shown in Fig. 2
A
genes within amino acid biosynthesis and other pathways (3, 14),
wild-type population induces transcription of hundreds of target
measurements (Dataset S1) confirmed that active Gcn4p in the
differed markedly at the molecular level. Our gene expression
exploring the transcriptional response, which is known to change
rameter under investigation. This is particularly true when exploring the transcriptional response, which is known to change with specific growth rate (13).

Despite similar growth rate conditions and overall macroscopic growth characteristics, the responses of the 2 strains differed markedly at the molecular level. Our gene expression measurements (Dataset S1) confirmed that active Gcn4p in the wild-type population induces transcription of hundreds of target genes within amino acid biosynthesis and other pathways (3, 14), as shown in Fig. 2A. The median enzyme mRNA expression level increased at least 2-fold for the arginine, asparagine, tryptophan, phenylalanine, tyrosine, histidine, isoleucine, leucine, and valine pathways, and at least 1 transcript reached this threshold for the cysteine, methionine, and lysine pathways.

Despite these drastic changes at the molecular level, the wild type and gen4Δ chemostat cultures maintain a macroscopic physiological similarity. To explain this, we hypothesized that the cell compensated at the level of metabolism to maintain similar macroscopic growth under the divergent transcriptional programs observed in the wild-type and gen4Δ cultures. Metabolic compensation can be assessed by monitoring 2 features: metabolite levels—defined as the relative abundances of non-genetically-encoded substrates, intermediates, or products of metabolic pathways (15), and reaction fluxes—defined as the rates of conversion by each reaction in the metabolic network (16) (Fig. 1B). To measure metabolite levels, we used gas chromatography coupled to mass spectrometry (GC-MS) to separate, characterize, and quantify metabolites from whole-cell extracts. Additionally, we used high performance liquid chromatography (HPLC) to obtain more precise measurements of 17 of the 20 free amino acids. In total, we assigned statistical confidences to abundance ratios of 54 observed metabolites in the gen4Δ versus wild-type cells (Dataset S2).

Reaction fluxes provide a direct measure of metabolite interconversion rates. In the absence of experimental data, values of reaction fluxes are often estimated in silico under the constraints of maximized biomass production, an assumed cellular stoichiometry, a steady state metabolic network and fixed protein composition (17). However, these simulations result in an underdetermined system (more fluxes than measurements), and it is still an open question whether such methods adequately capture regulatory-induced large network perturbations. For this reason, we cultivated cells on 1-13C-glucose and experimentally measured condition-specific biomass amino acid content with HPLC and isotopic labeling patterns with GC-MS to overdetermine the reaction system (more measurements than fluxes) and thus more robustly estimate fluxes (18) that can be found in Dataset S3. Stable isotopic labeling is an established technique for flux determination whereby metabolic conversion of 13C-enriched substrates generates specifically labeled metabolites, i.e., isotope isomers, whose labeling patterns are direct functions of the flux configuration (16). As such, flux can be estimated from measurements (by GC-MS) of isotopic enrichment of metabolites and solving the inverse problem. Overall, our measurements yielded 17 amino acid fluxes into biomass and 250 13C-labeling abundances, thus providing a very overdetermined system (see below) for accurate and robust flux estimation.

To convert these 13C-labeling abundances to intracellular flux values and statistically verify the results (18), we developed a new method based on the yeast metabolic reaction network refined and expanded from Gomberg et al. (19). Compared with the 2 existing methods for flux determination (19, 20), we took greater advantage of redundant flux information (i.e., more measurements than fluxes) by (i) using the complete labeling distributions instead of lumping abundances into a summed fractional labeling [relative to the GC-MS method (19)], and (ii) simultaneously fitting the entire measurement set [relative to the NMR method (20)]. We used a network of 83 reactions (with corresponding carbon transitions) distributed among the 3 assigned compartments (cytosolic, mitochondrial, extracellular), using 75 metabolites (1 substrate, 51 balanced intracellular, and 23 products). Combining our measurement set with this network resulted in 105 redundant measurements for our now overdetermined system (SI Appendix, C13-Based Reaction Flux Determination). By adding experimentally obtained amino acid fluxes into biomass measurements (previously assumed unchanging), we expanded our resolution beyond that of previous methods that only use networks of central carbon metabolism. We also note the exceptionally high quality of fit in Fig. S2 that gives us high confidence in our experimentally measured fluxes. Because metabolite levels and fluxes are independent measurements (21), together they give a more complete view of the metabolic regulatory system.

Consistent with the previous observation that mRNA expression insufficiently predicts protein level (22), biosynthetic mRNA changes demonstrated poor correlation with flux changes (r = 0.02 for log ratios or differences in Fig. 2B, r = 0.07 for deltas or linear differences in Fig. 4B). Indeed, we found little correlation for any pairwise combination of mRNA, flux, and end-product metabolite changes (Fig. 2 B–D). Furthermore, Gcn4p DNA-binding strength did not always translate to transcriptional activation (Fig. 2A).

We next proposed that a mechanistic model of amino acid biosynthesis capturing regulatory knowledge would better correlate the mRNA, intracellular flux, and metabolite level datasets. Accordingly, we created a biosynthetic network consisting of gene/protein, reaction, and metabolite nodes connected by condition-specific transcription factor binding interactions (3), protein–protein binding interactions (23), and enzyme-reaction and reaction-metabolite interactions. This network contained at its core the metabolic reaction network used previously for (and statistically verified by) flux determination. Additionally, we curated all known metabolite–enzyme interactions (24) as an explicit avenue of enzyme-level regulation. The key feature of our updated model is that we no longer rely on stoichiometry
alone. Now, regulatory information (given by interactions within the network) is incorporated that might impact the correlation between mRNA levels and metabolic flux. This expanded network model linking all measurement types was visualized in Cytoscape (25) as shown in Fig. 3 (with a log2 color bar for measurement ratios of the wild type strain relative to gen-Δ).

To improve the poor mRNA-flux agreement seen in Fig. 2B, we used this expanded network model to construct a predictor of metabolic flux changes as presented in SI Appendix, section 3. A novelty of the model is the introduction of a parameter termed “metabolite interaction density” for each reaction that captures the degree to which the reaction’s enzymes are negatively or positively regulated by metabolites as represented in the above pathway model. Specifically, for a pathway this quantity is defined as the ratio of the number of metabolite-enzyme interactions to that of the total metabolic network stoichiometry and satisfy the steady state condition (7).

Using this concept, network flux changes were predicted from changes in mRNA levels and metabolite interaction densities, using the following equation: Δflux = \exp(-p1 \cdot d_{interaction})

\( \Delta mRNA = \frac{P2}{P1} \). The rationale for this equation is to allow metabolite interaction density \( (d_{interaction}) \) to modulate the extent to which changes in mRNA levels are allowed to propagate to changes in flux, with \( p1 \) and \( p2 \) determined to minimize the difference between actual and predicted fluxes (SI Appendix, mRNA-Flux Model). According to this equation, the effect of mRNA changes \( (\Delta mRNA) \) is attenuated by metabolite interaction density in causing changes in the pathway flux \( (\Delta flux) \). The last feature of our model is the additional constraint that fluxes predicted by the above equation must also be consistent with the overall metabolic network stoichiometry and satisfy the steady state reaction network, i.e., the equation \( S \cdot \Delta v = 0 \).

This model represents a new, hybrid approach to correlating mRNA and flux data that combines information from both interaction network models and flux determination models. A core element is that the model preserves metabolic network consistency both in flux determination and in flux prediction from mRNA data. In addition, the model introduces the new concept of metabolite interaction density that embodies a hypothesis that a simpler, less-connected regulatory network often has more correlated mRNA and flux changes.
Fig. 3. Integrated perspective of a large-scale network perturbation on amino acid biosynthesis. The central GCN4 node anchors a biomolecular gene/protein network expanded to include reactions and metabolites. The network contains condition-specific transcription factor binding, protein–protein binding, enzyme-reaction, reaction-metabolite, and metabolite-protein feedback edges (see node type and edge key). High metabolite interaction density is evident in A, B, and E. The dotted yellow edges emanating from the GCN4 node indicate an interaction found only in the wild type condition (i.e., not in gcn4Δ). Measurement ratios of wild type relative to gcn4Δ were visualized with a log2 colorbar (see node color key), and gray coloring indicates the lack of a measurement for that node. Findings for individual pathways are discussed in the text.

The model illustrates that flux control (i.e., invariance to perturbation) increases as metabolite—enzyme interaction density increases. In other words, greater interaction density lessens transcriptional control and gives metabolites an increasing role in regulating the flux phenotype. This model supports the body of literature on feedback inhibition in that increased metabolite interaction density

Fig. 4. Incorporating large-scale biomolecular networks improve flux phenotype understanding. (A) Metabolite interaction densities are illustrated for arginine and aromatics biosynthesis. (B) Changes in mRNA measurements were used to predict flux changes as described in the text and SI Text. By adding the statistically verified reaction network, biomass efflux bounds, and metabolite interaction density, the predicted fluxes from mRNA better matched measured fluxes with a correlation coefficient value of 0.80 relative to 0.07 corresponding to the initial assumption set (see SI Appendix, section 3). (C) Measured delta flux for individual reactions is labeled by end product metabolite and plotted versus the flux predicted from mRNA measurements and the 2 parameter model. Notably, adding metabolite interaction densities discriminates arginine and aromatic pathways to predict arginine’s up-regulated flux. Without the model incorporating metabolite interaction densities, aromatic pathways would have been wrongly predicted to increase in flux.
regulation of enzyme activity results in greater metabolic control. The high interaction density aromatics and isoleucine-leucine-valine pathways ($d_{interaction} = 13/14$ and $d_{interaction} = 7/11$ in Fig. 3 A and B, respectively) provide an illustrative foil to the low interaction density arginine and lysine pathways ($d_{interaction} = 1/10$ and $d_{interaction} = 2/7$ in Fig. 3 C and D, respectively) as depicted in Fig. 4A and discussed in SI Discussion. For the high metabolite interaction densities, we hypothesize that because end products are not being increasingly used in protein synthesis, the metabolite levels build up and cause feedback inhibition at the enzyme level, resulting in tightly regulated flux. In contrast, we hypothesize that low metabolite interaction density pathways use alternative strategies. These pathways rely on preferential vacuolar localization of products and activation of transcription factor regulators separate from Gcn4p to compensate for unnecessary metabolite build-up.

Despite high metabolite interaction densities of the aspartate and threonine pathways, $d_{interaction} = 6/8$, fluxes are transcriptionally regulated and increase significantly because of rewiring of glycine flux, as revealed by the $^{13}$C-labeling patterns (Fig. 3E). Glycine may be produced from 3 independently $^{13}$C-enriched precursors: glyoxylic acid (Arg1p), serine (Shm1p, Shm2p), and threonine (Gly1p). After discarding glyoxylic acid flux as 0 (within 3% noise) in all conditions, $\approx 99\%$ of glycine flux comes via serine in the reference condition; however, the presence of activated Gcn4p shifts 44% of that flux to a threonine precursor. We hypothesize that although free aspartate and threonine do not increase appreciably, the significant (FDR <2%) activation of $aat2$, $hom3$, and $hom2$ and near significant (FDR <5% or log2 ratio >1.0) activation of $thr1$, $thr4$, and $gly1$ likely provide an increased biosynthetic capacity. Thus, some glycine biosynthetic flux is rewired away from serine to a route depleting the now more available free aspartate and threonine and increasing fluxes (>50% change), i.e., the up-regulated green reaction nodes in Fig. 3E. Overall, these data suggest that global rewiring of flux through the network may be a mechanism to override high metabolite interaction density.

In total, the predictive network model highlights 5 general observations and 10 new system hypotheses not evident without high resolution metabolic flux data (Tables S1 and S2). Because transcriptional flux rewiring has not been studied extensively and, more specifically, glycine flux rewiring has not previously been reported as a potential conserved physiological for stress response, we chose to examine these observations and hypotheses in greater detail.

To validate the predictive network model for transcriptional changes that drive a rewiring of global network fluxes, we created additional perturbations of yeast transcriptional regulation in the vicinity of glycine biosynthesis (Fig. 3E) with $met28\Delta$, $cbf1\Delta$, $met31\Delta$, and $met23\Delta$ knockout strains. We measured mRNA levels for each strain in shake flask cultures and found that $met28\Delta$ and $cbf1\Delta$ displayed significant changes in amino acid biosynthetic pathways (Figs. S3–S6). Using the measured mRNA levels and the predictive network model from above without any further modification, we determined that both $met28\Delta$ and $cbf1\Delta$ should exhibit a glycine flux rewiring that favors a threonine precursor.

To experimentally validate the conclusions of the predictive network model, we used the $^{13}$C labeling analysis described above to measure glycine biosynthetic fluxes for the wild type, $met28\Delta$ and $cbf1\Delta$ strains in YNB minimal media supplemented with methionine to correct for the corresponding knockout auxotrophies. As before, we achieved an exceptionally high quality of fit in Fig. S7.

Notably, we observed that the predictive network model correctly identified this experimentally-determined glycine flux rewiring, using the mRNA data. We observed a shifting of glycine flux to a threonine precursor where serine-originating glycine flux decreased from $\approx 50\%$ in the wild type to $\approx 25\%$ in $met28\Delta$ to 0% in $cbf1\Delta$ as shown in Fig. 5. We hypothesize that the increasing shutdown of mRNA levels in the methionine pathways increased the availability of homoserine and thus threonine to yield greater conversion rates of threonine to glycine.

We were especially encouraged by 3 aspects of the glycine flux rewiring observations in our follow-on analysis. First, the $^{13}$C high resolution flux determination enabled the experimental observation of a previously unreported mechanism for which an interaction network allows mRNA to mediate a flux rewiring for a cellular response. Second, this glycine rewiring mechanism was experimentally observed for perturbations originating in different network locations indicating a convergent rewiring pathway activated by transcriptional regulation (Fig. 5B). Finally, we demonstrated that mRNA and network data may be used to correctly predict flux changes.

Discussion
Looking forward toward a full fusion of transcriptional and metabolic data, the suspected role of metabolite interaction density and metabolite-triggered activation of $AR09$ transcription (among others) demonstrate opportunities to substantially augment the currently available interactome. Despite modulating nearly all pathways observed, cell-wide data on interactions between metabolites and proteins are not yet available unlike protein–protein interac-
Materials and Methods

Chemostat Growth Conditions. Strains were cultivated in aerobic carbon-limited chemostat cultures in 2-L fermentors (Applikon) in standard YNB media obtained from Qbiogene with indicated levels of 3-amino-triazole (3AT). The working volume was 1.0 L and the dilution rate was 0.10 h⁻¹. Cultivations were carried out at 30 °C with an agitation speed of 600 rpm, pH of 5.0, and airflow rate of 1.0 standard L·min⁻¹ and dissolved oxygen >80% air saturation. The bioreactors were fitted with cooled condensers (4 °C), and the off-gas was directed to a gas analyzer (INNOVA) for measurement of CO₂ and O₂. Steady-state samples were taken after 5 volume changes. Biomass concentration, extracellular metabolite concentrations, and carbon dioxide evolution rate were constant over at least 3 measurements taken before sampling.

Biomass Amino Acid C13 Enrichment Analysis. C13 enrichment biomass was achieved by chemostat experiments performed in in-house-built reactors with a working volume of 200 mL. Cultivations were carried out at 30 °C with an agitation speed of 600 rpm, a dilution rate of 0.1 h⁻¹, a pH of 5.0, and airflow rate of 1.0 standard L·min⁻¹. Steady-state samples were taken after 5 volume changes. 100% of the glucose used was labeled in position 1 (1-13C glucose was from Omicron Biochemicals). The 13C-labeled biomass was harvested by centrifugation at 4,000 × g and 0 °C for 5 min. After centrifugation, the supernatant was poured off, and the cell pellet was frozen instantly in liquid nitrogen and stored at −80 °C. Hydrolysis, derivatization, and analysis were carried out as described by Antoniewicz et al. (29).

Metabolic Flux Determination. Metabolic fluxes were determined using the Metran software, using the elementary metabolic units (EMU) algorithm (30, 31). Metran accepts as input a user-defined metabolic network model consisting of biochemical reactions and atom transitions, and a set of 13C measurements and external fluxes; it produces as output metabolic fluxes for the entire network, confidence intervals for all fluxes, and statistical analysis of the goodness-of-fit (18).

Additional materials and methods on strains, shake flask growth conditions, whole genome expression analysis, biomass amino acid composition analysis, endometabolome analysis, and exometabolome analysis are in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Ryan Kelley, Tong Ihn Lee, Scott McCuine, Chris Workman, Jose Aleman, Richard Young, and Gerald Fink for advice on experimental design and analysis. This work was supported by the National Institutes of Health Grant 1R01 DK075650-01, the National Science Foundation National Research Fellowship Program, the Singapore–Massachusetts Institute of Technology Alliance, and the National Center for Research Resources award 018627.