Integration of Carbon, Nitrogen, and Oxygen Metabolism in *Escherichia coli*

Applicant/Institution: Princeton University
4 New South Building
Princeton, NJ 08544

Principal Investigator: Joshua D. Rabinowitz
Lewis-Sigler Institute for Integrative Genomics and Department of Chemistry
Princeton University
241 Carl Icahn Laboratory, Washington Road
Princeton, NJ 08544
Telephone Number: (609) 258-8985
Email: joshr@genomics.princeton.edu

Co-PI: Ned S. Wingreen
Lewis-Sigler Institute for Integrative Genomics and Department of Molecular Biology
Princeton University
347 Lewis Thomas Lab, Washington Road
Princeton, NJ 08544
Telephone Number: (609) 258-8476
Email: wingreen@princeton.edu

Co-PI: Herschel Rabitz
Department of Chemistry
Princeton University
253 Frick Laboratory
Princeton, NJ 08544
Telephone Number: 609-258-3917
Email: hrabitz@princeton.edu

DOE award number: DE-SC0002077
Period: September 1, 2009 to Aug 31, 2012

Personnel supported:

Rabinowitz group: Joshua D. Rabinowitz, Yifan Xu (graduate student), Daniel Amador-Noguez (post-doc), Junyoung O. Park (graduate student), Meytal B. Higgins (post-doc)

Wingreen group: Ned S. Wingreen, David Schwab (post-doc), Edvin Memet (undergraduate student), Robert M. Cooper (grad student)

Rabitz group: Herschel Rabitz, Xiao-Jiang Feng (post-doc), Genyuan Li (research chemist)

Proposal Submitted via NSF Interagency Program in Metabolic Engineering
DOE/Office of Science Program Office: Office of Biological and Environmental Research
DOE/Office of Science Program Office Technical Contact: Dr. Joseph Graber
Funded jointly with AFOSR Bioenergy Program, Grant Number FA9550-09-1-0580

Abstract

A key challenge for living systems is balancing utilization of multiple elemental nutrients, such as carbon, nitrogen, and oxygen, whose availability is subject to environmental fluctuations. As growth can be limited by the scarcity of any one nutrient, the rate at which each nutrient is assimilated must be sensitive not only to its own availability, but also to that of other nutrients. Remarkably, across diverse nutrient conditions, *E. coli* grows nearly optimally, balancing effectively the conversion of carbon into energy versus biomass. To investigate the link between the metabolism of different nutrients, we quantified metabolic responses to nutrient perturbations using LC-MS based metabolomics and built differential equation models that bridge multiple nutrient systems. We discovered that the carbonaceous substrate of nitrogen assimilation, α-ketoglutarate, directly inhibits glucose uptake and that the upstream glycolytic metabolite, fructose-1,6-bisphosphate, ultrasensitively regulates anaplerosis to allow rapid adaptation to changing carbon availability. We also showed that NADH controls the metabolic response to changing oxygen levels. Our findings support a general mechanism for nutrient integration: limitation for a nutrient other than carbon leads to build-up of the most closely related product of carbon metabolism, which in turn feedback inhibits further carbon uptake.

Aims summary

Aim 1: Reveal the interplay between different nutrient utilization systems

We proposed to measure the response of the *E. coli* metabolome to perturbation of the availability of one nutrient as a function of the other nutrients present. These experiments have been successfully conducted, and have led to important insights into nutrient coordination in *E. coli*. By combining LC-MS-based metabolomics with genetics (enzyme knockouts and point mutants), we have been able to link molecular-level regulatory events to their physiological consequences. For example, we identified the key regulator linking nitrogen availability and carbon utilization, α-ketoglutarate. This regulation involves low nitrogen resulting in build-up of α-ketoglutarate, which in turns inhibits Enzyme I of the phosphotransferase system (PTS) responsible for glucose uptake (Doucette et al., 2011). Mutants of the PTS system lack the ability to match glucose uptake to nitrogen availability. Similarly, investigated point mutants of the anaplerotic enzyme phosphoenolpyruvate carboxylase. These point mutants which lack positive regulation of anaplerosis by fructose-1,6-bisphosphate, a key metabolite indicating high sugar availability. Such mutants grow normally in steady glucose but are strikingly deficient in an oscillating glucose environment (Xu et al., 2012a; Xu et al., 2012b). Related work investigates additional nutrients and stresses, and the robustness of biological responses to them (Goodarzi et al., 2010; Hart et al., 2011; Peterson et al., 2012; Reaves and Rabinowitz, 2011a, b).

Aim 2: Advance quantitative understanding of nutrient integration

The initial aim was to develop an ordinary differential equation (ODE) model of central metabolism in *E. coli* that bridges multiple nutrient systems. This model aims to include most central metabolites of *E. coli* and their enzymatic connections, with kinetic parameters taken from the literature or computationally inferred based on the data from Aim 1. As proposed initially in the application, we are assembling the model in a modular fashion. We first built an ODE model on nitrogen assimilation system (Yuan et al., 2009). We successfully combined the simplified nitrogen assimilation model with simplified models of glycolysis and the TCA cycle to explain the linkage between nitrogen availability and carbon utilization (Doucette et al., 2011). We are now working with larger scale modular models, which include all enzyme connections. These models have been assembled in a manner that will soon allow their integration. We have coded the complete model and shown that it can obtain various physiologically relevant steady-states. We will begin testing its dynamical properties shortly.

Aim 3: Gain insight into the key principles underlying the coordination of metabolic activity across pathways

We proposed to develop simplified models that capture the fundamental processes that enable effective integrated metabolic regulation. The reductionist models were proposed to be assembled in light of typical pathway architectures, known cellular objectives, and physical constraints. To this end, we have successfully gained insights from our nitrogen-carbon network and glycolysis-anapleurosis pathways. We initially found that feedback inhibition alone is sufficient to enable both intracellular metabolite concentration homeostasis and optimal steady-state fluxes (Goyal et al., 2010). More recently, we demonstrated that feed forward regulation is important in oscillating conditions (Xu et al., 2012a; Xu et al., 2012b). In general, our experimental and computational results support a primary paradigm for nutrient integration: Limitation for a nutrient other than carbon leads to build-up of the most closely related product of carbon metabolism, which in turn feedback inhibits further carbon utilization. This principle was tested and confirmed by oxygen-carbon network: NADH built up and regulated central carbon metabolism(Amador-Noguez et.al. in preparation).

Research Achievements

Our research in this project has resulted in the development of novel experimental and computational approaches for the quantitative study of nutrient integration in *E. coli*. Our efforts concentrated on the coordination of nitrogen and carbon utilization, including the key interface between glycolysis (the primary glucose catabolic pathway) and the TCA cycle (where carbon and nitrogen metabolism directly intersect). Our investigation yielded significant advances in the understanding of *E. coli* nutrient integration, and more generally of microbial central metabolic regulation.

α-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition (Doucette et al., 2011)

This single seminal paper achieves a majority of each of the three initial grant aims. Microbes survive in a variety of nutrient environments by modulating their intracellular metabolism. Balanced growth requires coordinated uptake of carbon and nitrogen, the primary substrates for biomass production. Yet the mechanisms that balance carbon and nitrogen uptake were previously poorly understood. Here we report in Escherichia coli that a sudden increase in nitrogen availability results in an almost immediate increase in glucose uptake. The concentrations of glycolytic intermediates and known regulators, however, remain homeostatic. Instead, we find that α-ketoglutarate, which accumulates in nitrogen limitation, directly blocks glucose uptake by inhibiting enzyme I, the first step of the sugar-phosphoenolpyruvate phosphotransferase system (PTS). This inhibition enables rapid modulation of glycolytic flux without marked changes in the concentrations of glycolytic intermediates by simultaneously altering import of glucose and consumption of the terminal glycolytic intermediate phosphoenolpyruvate. Quantitative modeling shows that this previously unidentified regulatory connection is, in principle, sufficient to coordinate carbon and nitrogen utilization.

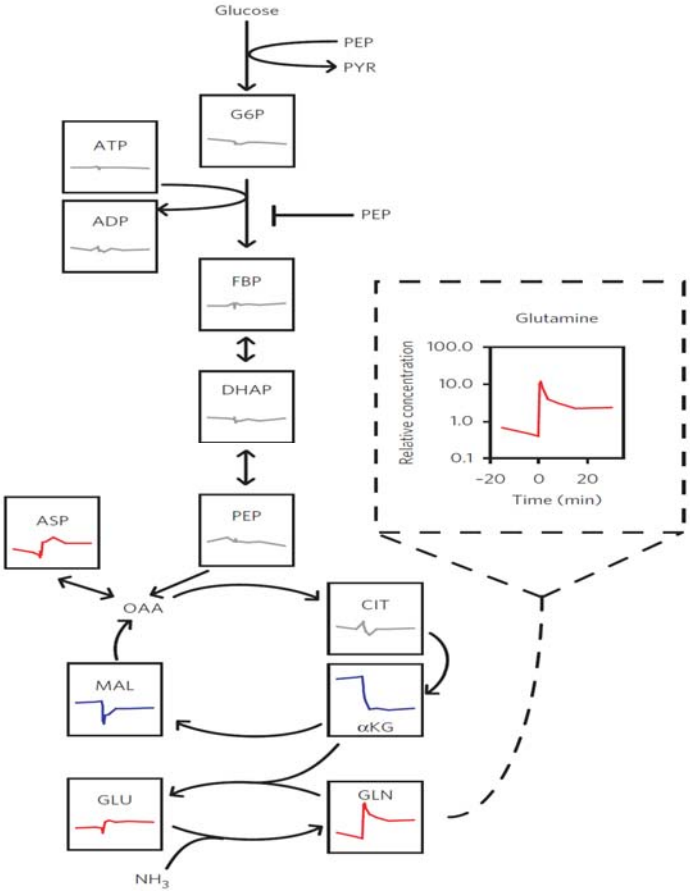

Figure 1. Glycolytic intermediates are homeostatic during nitrogen upshift. Wild-type *E. coli* cultures were subjected to a sudden 13-fold increase in extracellular ammonia at t = 0 that induced a 2.5-fold increase in growth rate. Each box shows the time-dependent concentration change of the indicated metabolite during the perturbation. Concentrations are relative to the pool size in cells grown on ample nitrogen. G6P, glucose-6-phosphate and its isomer fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; DHAP, dihydroxyacetone-phosphate; PEP, phosphoenolpyruvate; PYR, pyruvate; ADP, adenosine diphosphate; ATP, adenosine triphosphate; OAA, oxaloacetate; CIT, citrate; αKG, α-ketoglutarate; MAL, malate; GLU, glutamate; GLN, glutamine; ASP, aspartate.

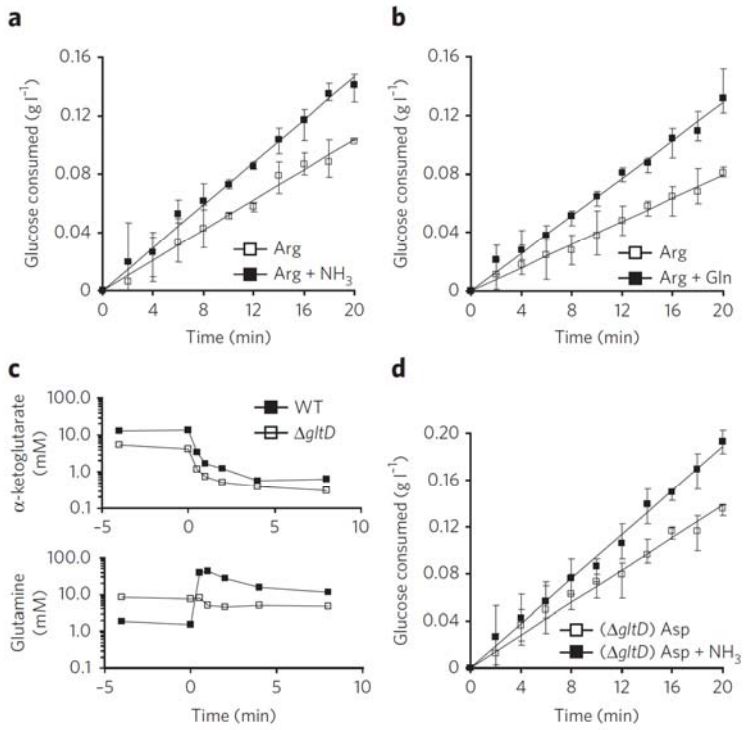

Figure 2. Glucose uptake increases rapidly upon nitrogen upshift, independent of glutamine concentration (**a**,**b**) Batch cultures of wild-type *E. coli* were grown to mid-logarithmic phase in medium containing 2 g per liter glucose and 2.5 mM arginine (Arg). At t = 0, the cultures were perturbed by addition of 10 mM ammonia in **a** or 2 mM glutamine (Gln) in **b**. (**c**) Intracellular α-ketoglutarate and glutamine concentrations during a nitrogen upshift of wild-type NCM 3722 and Δ*gltD* cultures. (**d**) As in **a** except that *E. coli* batch cultures were glutamate synthase deficient (Δ*gltD*) and the medium contained 10 mM aspartate (Asp) instead of arginine.

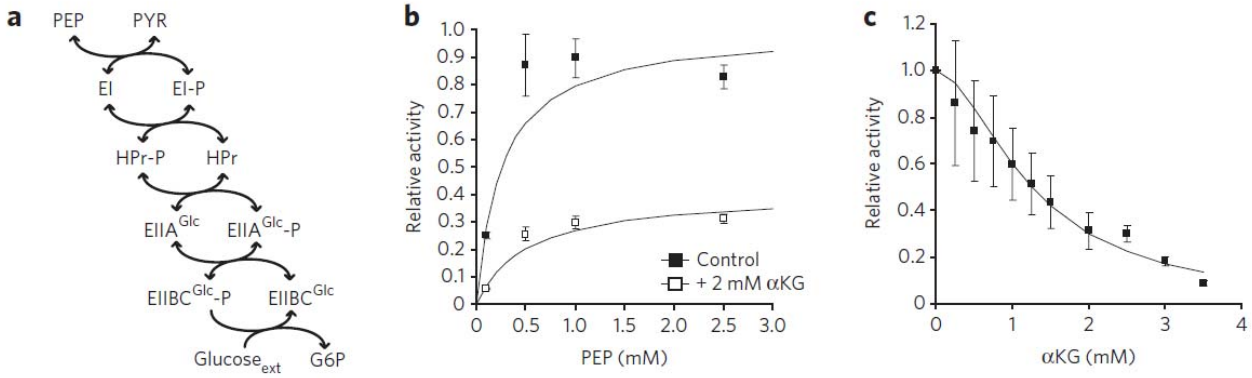

Figure 3. α-ketoglutarate inhibits enzyme I (**a**) Phosphoenolpyruvate (PEP) donates phosphate, which is transferred in turn to the cytosolic proteins enzyme I, HPr, and $EIIA^{Glc}$, the membrane transporter $EIIBC^{Glc}$ and finally glucose, which is simultaneously transported and phosphorylated. (**b**) The activity of purified enzyme I was determined in the presence (white squares) or absence (black squares) of 2 mM α-ketoglutarate. (**c**) Enzyme I activity was determined in the presence of 1 mM PEP and varying concentrations of α-ketoglutarate.

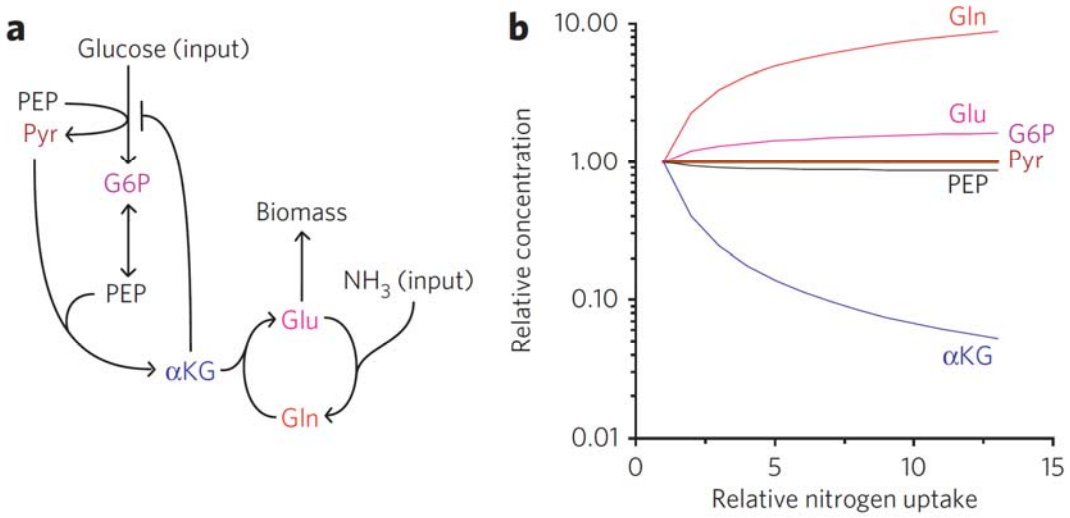

Figure 4. The differential equation model consists of five metabolic reactions: glucose transport and phosphorylation with concomitant conversion of PEP to pyruvate, reversible glycolytic conversion of glucose-6-phosphate to two molecules of PEP, condensation of PEP and pyruvate to α-ketoglutarate, nitrogen transfer from glutamine to α-ketoglutarate producing two molecules of glutamate, and glutamine synthesis from glutamate and ammonia. Glutamate is consumed in the outflux to biomass, which is controlled by glutamine and α-ketoglutarate concentrations. Glucose and ammonia availability are the inputs to the model. α-ketoglutarate inhibits its own production as well as glucose transport, and glutamine inhibits glutamine synthesis. (**b**) The model was used to simulate the steady-state pool sizes of each metabolite as external nitrogen availability (parameter S_N in the model) was varied over a 13-fold range with carbon availability kept constant. The equations were simulated with all variables in the same unitless concentration scale; here, the fold changes in pool size, relative to the condition where nitrogen is most scarce (S_N = 0.35), are plotted on a logarithmic scale.

Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase (Xu et al., 2012a)

Our investigation of nitrogen-carbon coordination hit a roadblock when we realized that existing information regarding integration of glycolysis and the TCA cycle was deficient. This paper rectifies this deficiency by revealing a predominant role for activation of anaplerosis by fructose-1,6-bisphosphate. The paper is also a landmark in investigation of allosteric regulation: it is the first to make allostery-altering directed mutations in central carbon enzymes and show both their metabolic and fitness consequences.

Anapleurosis is the filling of the tricarboxylic acid cycle with four-carbon units. The common substrate for both anapleurosis and glucose phosphorylation in bacteria is the terminal glycolytic metabolite phosphoenolpyruvate (PEP). Here we show that Escherichia coli quickly and almost completely turns off PEP consumption upon glucose removal. The resulting buildup of PEP is used to quickly import glucose if it becomes available again. The switch-like termination of anapleurosis results from depletion of fructose-1,6-bisphosphate (FBP), an ultrasensitive allosteric activator of PEP carboxylase. *E. coli* expressing an FBP-

insensitive point mutant of PEP carboxylase grow normally when glucose is steadily available. However, they fail to build up PEP upon glucose removal, grow poorly when glucose availability oscillates and suffer from futile cycling at the PEP node on gluconeogenic substrates. Thus, bacterial central carbon metabolism is intrinsically programmed with ultrasensitive allosteric regulation to enable rapid adaptation to changing environmental conditions.

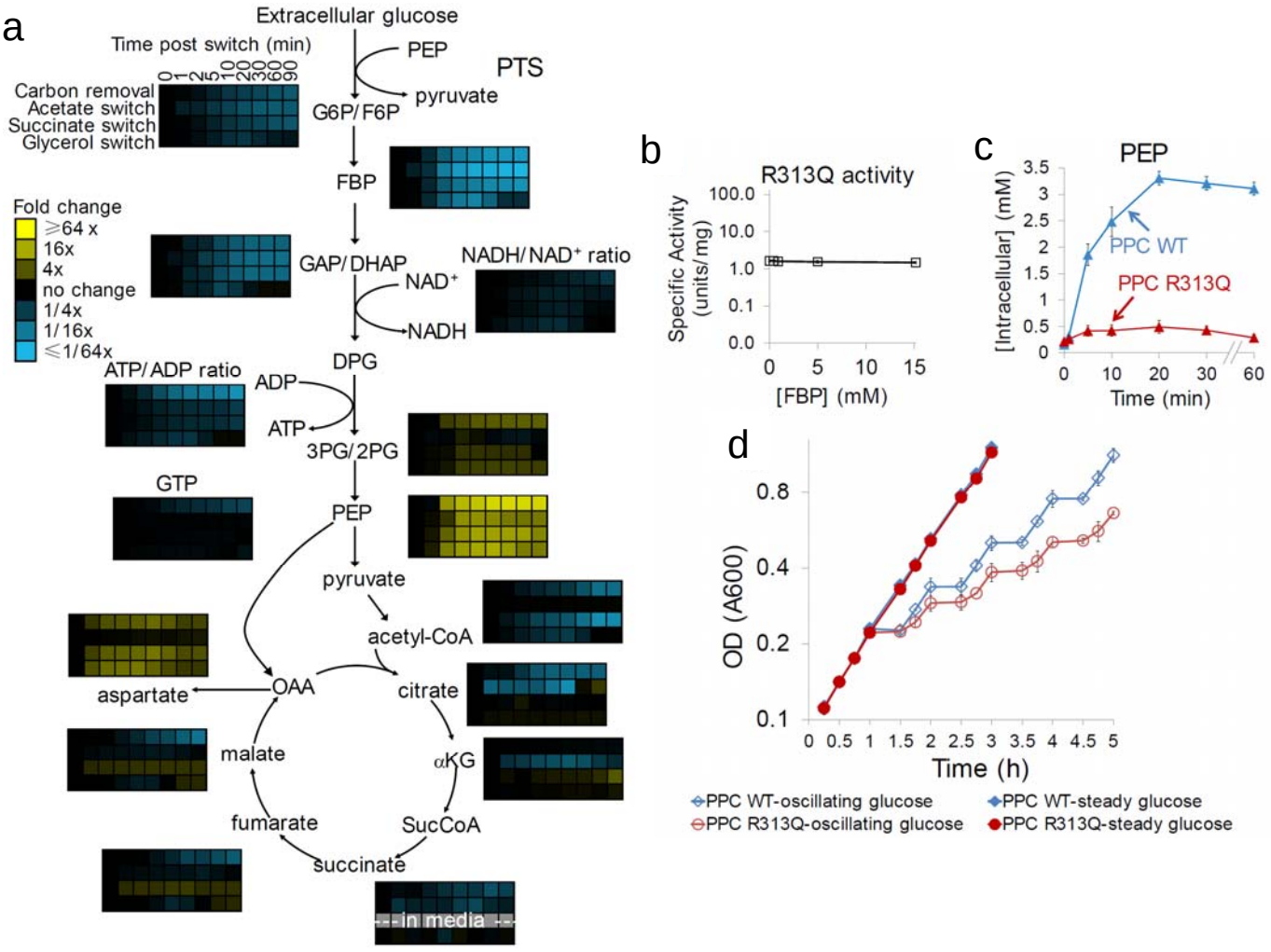

Figure 5. PEP carboxylase allosteric regulation causes PEP accumulation upon glucose removal and thereby enhances growth on oscillating glucose. (**a**) Glucose removal results in PEP buildup. *E. coli* cells growing freely in glucose minimal medium were switched to no carbon, acetate, succinate or glycerol as indicated. At the indicated times after glucose removal, the metabolome was quantified by LC-MS. Data are shown in heat-map format, with each line reflecting the dynamics of a particular compound in a particular culture condition. (**b**) PPC^{R313Q} has higher activity than the wild type in the absence of FBP and is desensitized to FBP activation. (**c**) Expression of PPC^{R313Q} ablates the PEP spike upon carbon starvation (**d**) Growth curves of Δ*ppc* pCA24N-*ppc* and Δ*ppc* pCA24N-ppc^{R313Q} cells. Cells were grown in either steady glucose (closed symbols) or were alternated between glucose minimal medium and no-carbon minimal medium every 30 min (open symbols).

Regulation of yeast pyruvate kinase by ultrasensitive allostery independent of phosphorylation (Xu et al., 2012b)

Our advances in understanding allosteric regulation of the glycolysis-TCA interface in *E. coli* motivated us to conduct similar studies in yeast, where allostery and covalent modification are major means of fast-acting metabolic regulation. Through this investigation, we not only illuminated control of the glycolysis-TCA interface, we also examined the relative contribution of allostery and covalent modification to an important *in vivo* regulatory event: the rapid decrease in pyruvate kinase flux in yeast upon glucose removal.

The main pyruvate kinase isozyme (Cdc19) is phosphorylated in response to environmental cues. It also exhibits positively cooperative (ultrasensitive) allosteric activation by fructose-1,6-bisphosphate (FBP). Glucose removal causes accumulation of Cdc19's substrate, phosphoenolpyruvate. Here we showed that this response is retained in strains with altered protein-kinase-A or AMP-activated-protein-kinase activity or with *CDC19* carrying mutated phosphorylation sites. In contrast, yeast engineered with a *CDC19* point mutation that ablates FBP-based regulation fail to accumulate phosphoenolpyruvate. They also fail to grow on ethanol and slowly resume growth upon glucose upshift. Thus, while yeast pyruvate kinase is covalently modified in response to glucose availability, its activity is controlled almost exclusively by ultrasensitive allostery.

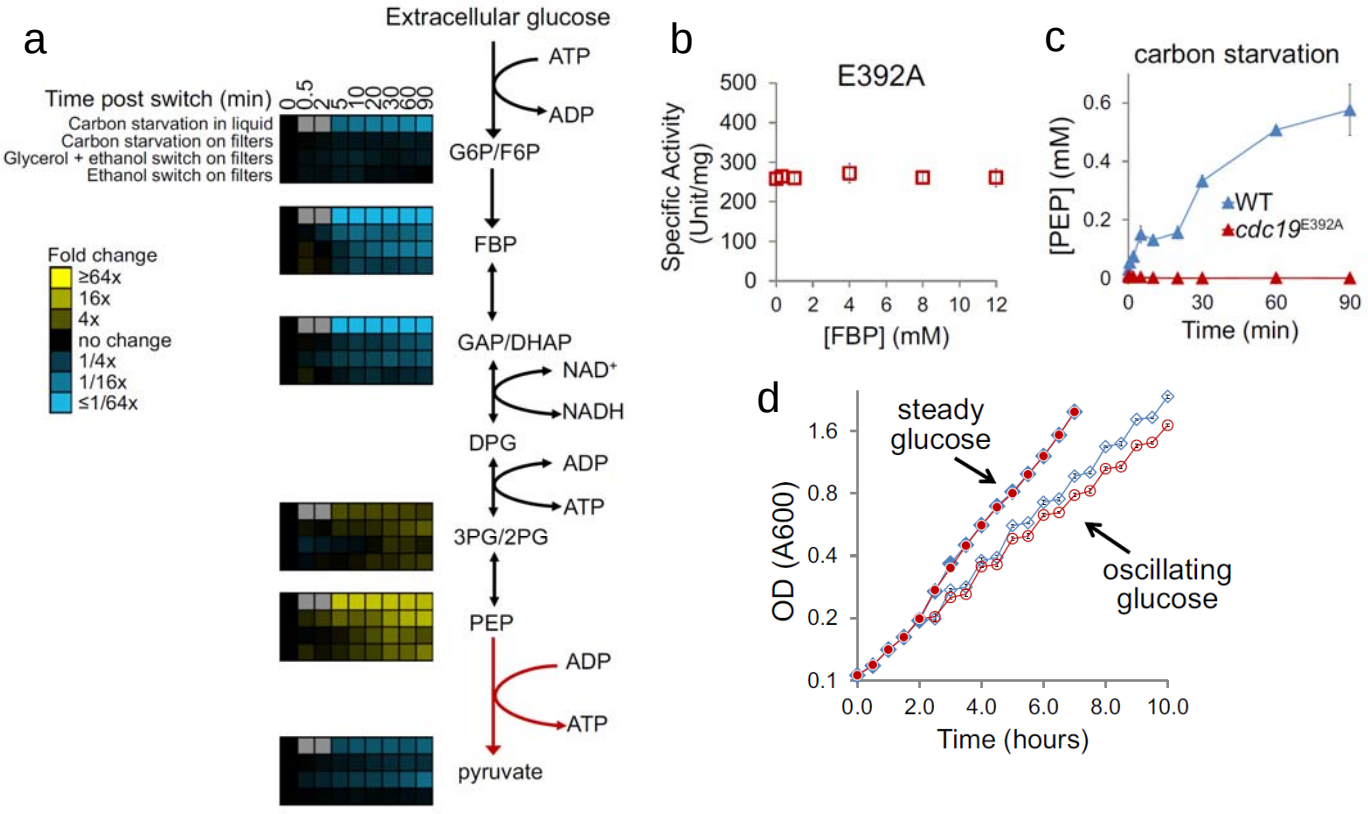

Figure 6. Pyruvate kinase allosteric regulation in yeast results in PEP accumulation upon glucose removal and enhanced growth on oscillating glucose.(**a**) Glucose removal results in PEP accumulation. Yeast in minimal media containing 2% glucose were switched to minimal media containing no carbon, 2% glycerol + 2% ethanol, or 2% ethanol. (**b**) The pyruvuate kinase E392A variant is active in the absence of FBP. (**c**) Genomic substitution of wild-type pyruvate kinase with E392A (*cdc19*^{E392A}) eliminates PEP accumulation upon glucose removal. (**d**) Growth of wild-type and *cdc19*^{E392A} yeast on steady versus oscillating glucose. Closed symbols, steady glucose; open symbols, oscillating; blue, wild-type; red, *cdc19*^{E392A}.

Characteristic Phenotypes Associated with *ptsN*-Null Mutants in Escherichia coli K-12 Are Absent in Strains with Functional *ilvG* (Reaves and Rabinowitz, 2011a)

The phosphotransferase system (PTS), encompassing EI, HPr, and assorted EII proteins, uses phosphoenolpyruvate to import and phosphorylate sugars. A paralog of EIIA of the sugar PTS system known as *ptsN* has been purported to regulate organic nitrogen source utilization in Escherichia coli K-12, and thus to play a role in nitrogen-carbon coordination. Its known biochemical function, however, relates to potassium homeostasis. The evidence for regulation of organic nitrogen source utilization by *ptsN* is based primarily on the defective growth of Δ*ptsN* mutants on amino acid nitrogen sources and other nutrient combinations. These observations were made with *E. coli* strains MG1655 and W3110, which carry a nonfunctional version of *ilvG*. There are three isozymes that effectively catalyze the first committed step of branched-chain amino acid biosynthesis, but *ilvG* is unique for doing so effectively across a range of potassium concentrations. Here we show that all of the nutrient utilization phenotypes attributed to *ptsN* are manifested selectively in strains lacking functional *ilvG*. We conclude that the *ptsN* gene product does not regulate organic nitrogen source utilization as previously proposed.

Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli* (Goodarzi et al., 2010)

The present paper diverges somewhat from the grant's focus on Understanding the genetic basis of adaptation is a central problem in biology. However, revealing the underlying molecular mechanisms has been challenging as changes in fitness may result from perturbations to many pathways, any of which may contribute relatively little. We have developed a combined experimental/computational framework to address this problem and used it to understand the genetic basis of ethanol tolerance in Escherichia coli. We used fitness profiling to measure the consequences of single-locus perturbations in the context of ethanol exposure. A module-level computational analysis was then used to reveal the organization of the contributing loci into cellular processes and regulatory pathways (e.g. osmoregulation and cell-wall biogenesis) whose modifications significantly affect ethanol tolerance. Strikingly, we discovered that a dominant component of adaptation involves metabolic rewiring that boosts intracellular ethanol degradation and assimilation. Through phenotypic and metabolomic analysis of laboratory-evolved ethanol-tolerant strains, we investigated naturally accessible pathways of ethanol tolerance. Remarkably, these laboratory-evolved strains, by and large, follow the same adaptive paths as inferred from our coarse-grained search of the fitness landscape.

Metabolomics in systems microbiology (Reaves and Rabinowitz, 2011b)

In this review article, we describe the role of metabolomics in dissecting microbial physiology, with a focus on examples of integrative analysis of microbial metabolism. Because of the importance of microbes as model organisms, biotechnology tools, and contributors to mammalian and ecosystem metabolism, there has been longstanding interest in measuring their metabolite levels. Current metabolomic methods, involving mass spectrometry-based measurement of cell extracts, enable routine quantitation of most central metabolites. Metabolomics alone, however, is inadequate to understand cellular metabolic activity: Flux measurement

and proteomic, genetic, and biochemical approaches with a metabolomics bent are all needed. Here we highlight examples where these integrated methods have contributed to discovery of metabolic pathways, regulatory interactions, and homeostasis mechanisms. We also indicate enduring challenges concerning unstable and low abundance compounds, subcellular compartmentalization, and quantitative amalgamation of different data types.

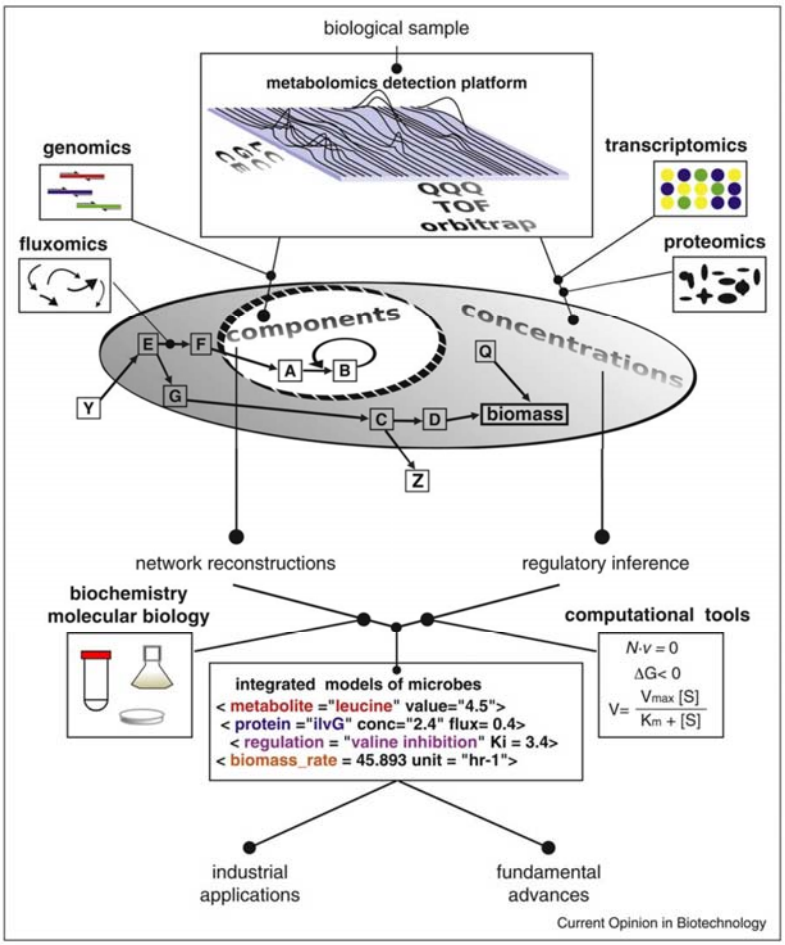

Figure 7. Overview of metabolomics in systems microbiology. Metabolomic and genomics provide complementary information for identifying an organism's metabolic capabilities. Concentration data for metabolites, proteins, and transcripts can be used for regulatory inference. Computational integration of such data aims to enable the development of mechanistically accurate, predictive metabolic models.

RpoS proteolysis is controlled directly by ATP levels in *Escherichia coli* (Peterson et al., 2012)

An important aspect of nutrient integration is global proteome remodeling in response to nutrient cues. The master regulator of stationary phase in *Escherichia coli*, RpoS, responds to carbon availability through changes in stability, but the individual steps in the pathway are unknown. Here we systematically blocked key steps of glycolysis and the citric acid cycle and monitored the effect on RpoS degradation in vivo. Nutrient upshifts trigger RpoS degradation independently of protein synthesis by activating metabolic pathways that generate small energy molecules. Using metabolic mutants and inhibitors, we showed that ATP, but not GTP or NADH, is necessary for RpoS degradation. In vitro reconstitution assays directly demonstrated that ClpXP fails to degrade RpoS, but not other proteins, at low ATP hydrolysis rates. These data suggest that cellular ATP levels directly control RpoS stability.

Robust control of nitrogen assimilation by a bifunctional enzyme in *E. coli* (Hart et al., 2011)

Bacteria regulate the assimilation of multiple nutrients to enable growth. How is balanced utilization achieved, despite fluctuations in the concentrations of the enzymes that make up the regulatory circuitry? Here we address this question by studying the nitrogen system of *E. coli*. A mechanism based on the avidity of a bifunctional enzyme, adenylyltransferase (AT/AR), to its multimeric substrate, glutamine synthetase, is proposed to maintain a robust ratio between two key metabolites, glutamine and α-ketoglutarate. This ratio is predicted to be insensitive to variations in protein levels of the core circuit and to the rate of nitrogen utilization. We find using mass spectrometry that the metabolite ratio is robust to variations in protein levels and that this robustness depends on the bifunctional enzyme. Moreover, robustness carries through to the bacteria growth rate. Interrupting avidity by adding a monofunctional AT/AR mutant to the native system abolishes robustness, as predicted by the proposed mechanism.

Achieving optimal growth through product feedback inhibition in metabolism (Goyal et al., 2010)

Recent evidence suggests that the metabolism of some organisms, such as Escherichia coli, is remarkably efficient, producing close to the maximum amount of biomass per unit of nutrient consumed. This observation raises the question of what regulatory mechanisms enable such efficiency. Here, we propose that simple product-feedback inhibition by itself is capable of leading to such optimality. We analyze several representative metabolic modules--starting from a linear pathway and advancing to a bidirectional pathway and metabolic cycle, and finally to integration of two different nutrient inputs. In each case, our mathematical analysis shows that product-feedback inhibition is not only homeostatic but also, with appropriate feedback connections, can minimize futile cycling and optimize fluxes. However, the effectiveness of simple product-feedback inhibition comes at the cost of high levels of some metabolite pools, potentially associated with toxicity and osmotic imbalance. These large metabolite pool sizes can be restricted if feedback inhibition is ultrasensitive. Indeed, the multi-layer regulation of metabolism by control of enzyme expression, enzyme covalent modification, and allostery is expected to result in such ultrasensitive feedbacks. To experimentally test whether the qualitative predictions from our analysis of feedback inhibition apply to metabolic modules beyond linear pathways, we examine the case of nitrogen assimilation in *E. coli*, which involves both nutrient integration and a metabolic cycle. We find that the feedback regulation scheme suggested by our mathematical analysis closely aligns with the actual regulation of the network and is sufficient to explain much of the dynamical behavior of relevant metabolite pool sizes in nutrient-switching experiments.

Identifying biological network structure, predicting network behavior, and classifying network state with high dimensional model representation (HDMR) (Miller et al., 2012)

This work presents an adapted Random Sampling - High Dimensional Model Representation (RS-HDMR) algorithm for synergistically addressing three key problems in network biology: (1) identifying the structure of biological networks from multivariate data, (2) predicting network response under previously unsampled

conditions, and (3) inferring experimental perturbations based on the observed network state. RS-HDMR is a multivariate regression method that decomposes network interactions into a hierarchy of non-linear component functions. Sensitivity analysis based on these functions provides a clear physical and statistical interpretation of the underlying network structure. The advantages of RS-HDMR include efficient extraction of nonlinear and cooperative network relationships without resorting to discretization, prediction of network behavior without mechanistic modeling, robustness to data noise, and favorable scalability of the sampling requirement with respect to network size. As a proof-of-principle study, RS-HDMR was applied to experimental data measuring the single-cell response of a protein-protein signaling network to various experimental perturbations. A comparison to network structure identified in the literature and through other inference methods, including Bayesian and mutual-information based algorithms, suggests that RS-HDMR can successfully reveal a network structure with a low false positive rate while still capturing non-linear and cooperative interactions. RS-HDMR identified several higher-order network interactions that correspond to known feedback regulations among multiple network species and that were unidentified by other network inference methods. Furthermore, RS-HDMR has a better ability to predict network response under unsampled conditions in this application than the best statistical inference algorithm presented in the recent DREAM3 signaling-prediction competition. RS-HDMR can discern and predict differences in network state that arise from sources ranging from intrinsic cell-cell variability to altered experimental conditions, such as when drug perturbations are introduced. This ability ultimately allows RS-HDMR to accurately classify the experimental conditions of a given sample based on its observed network state.

This work contributes new computational approaches for network identifications and control based on the increasing availability of high-throughput biological data, including metabolomics data.

Statistical mechanics of transcription-factor binding site discovery using Hidden Markov Models (Mehta et al., 2011)

Hidden Markov Models (HMMs) are a commonly used tool for inference of transcription factor (TF) binding sites from DNA sequence data. We exploit the mathematical equivalence between HMMs for TF binding and the "inverse" statistical mechanics of hard rods in a one-dimensional disordered potential to investigate learning in HMMs. We derive analytic expressions for the Fisher information, a commonly employed measure of confidence in learned parameters, in the biologically relevant limit where the density of binding sites is low. We then use techniques from statistical mechanics to derive a scaling principle relating the specificity (binding energy) of a TF to the minimum amount of training data necessary to learn it.

While this paper does not directly investigate nutrient integration in *E. coli*, it used statistical mechanics to study transcription-factor binding, which contributes to better insights about large data sets being generated in biology. By better understanding transcription factor binding, it lays the groundwork for future efforts to make longer time-scale dynamic metabolic models that include transcriptional regulation. By better understanding basic issues regarding applicability of HMM models to biochemical data, it lays the groundwork for their application to study also metabolic network regulation.

Other collaborators

In addition to collaborations among the three PIs, this grant benefited from valuable collaborations with Uri Alon (Weizmann Institute), Tania Baker (MIT), Jim Broach (Princeton, now Penn State), Saeed Tavazoie (Princeton, now Columbia), and Thomas Silhavy (Princeton).

Publications

Doucette, C.D., Schwab, D.J., Wingreen, N.S., and Rabinowitz, J.D. (2011). alpha-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition. Nat Chem Biol *7*, 894-901.

Goodarzi, H., Bennett, B.D., Amini, S., Reaves, M.L., Hottes, A.K., Rabinowitz, J.D., and Tavazoie, S. (2010). Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in E. coli. Mol Syst Biol *6*, 378.

Goyal, S., Yuan, J., Chen, T., Rabinowitz, J.D., and Wingreen, N.S. (2010). Achieving optimal growth through product feedback inhibition in metabolism. PLoS computational biology *6*, e1000802.

Hart, Y., Madar, D., Yuan, J., Bren, A., Mayo, A.E., Rabinowitz, J.D., and Alon, U. (2011). Robust control of nitrogen assimilation by a bifunctional enzyme in E. coli. Molecular cell *41*, 117-127.

Mehta, P., Schwab, D.J., and Sengupta, A.M. (2011). Statistical Mechanics of Transcription-Factor Binding Site Discovery Using Hidden Markov Models. Journal of statistical physics *142*, 1187-1205.

Miller, M.A., Feng, X.J., Li, G., and Rabitz, H.A. (2012). Identifying biological network structure, predicting network behavior, and classifying network state with High Dimensional Model Representation (HDMR). PloS one *7*, e37664.

Peterson, C.N., Levchenko, I., Rabinowitz, J.D., Baker, T.A., and Silhavy, T.J. (2012). RpoS proteolysis is controlled directly by ATP levels in Escherichia coli. Genes Dev *26*, 548-553.

Reaves, M.L., and Rabinowitz, J.D. (2011a). Characteristic phenotypes associated with ptsN-null mutants in Escherichia coli K-12 are absent in strains with functional ilvG. J Bacteriol *193*, 4576-4581.

Reaves, M.L., and Rabinowitz, J.D. (2011b). Metabolomics in systems microbiology. Current opinion in biotechnology *22*, 17-25.

Xu, Y.F., Amador-Noguez, D., Reaves, M.L., Feng, X.J., and Rabinowitz, J.D. (2012a). Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase. Nat Chem Biol *8*, 562-568.

Xu, Y.F., Zhao, X., Glass, D.S., Absalan, F., Perlman, D.H., Broach, J.R., and Rabinowitz, J.D. (2012b). Regulation of yeast pyruvate kinase by ultrasensitive allostery independent of phosphorylation. Molecular cell *48*, 52-62.

Yuan, J., Doucette, C.D., Fowler, W.U., Feng, X.J., Piazza, M., Rabitz, H.A., Wingreen, N.S., and Rabinowitz, J.D. (2009). Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli. Mol Syst Biol *5*, 302.