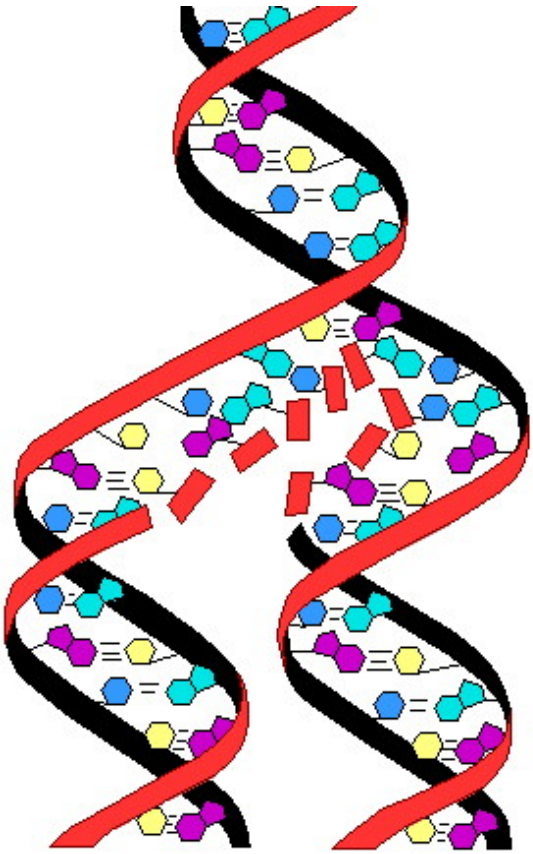
Proceedings of the

Inter-Agency Conference on Metabolic Engineering

2007

Metabolic Engineering Working Group

February 13, 2007



Workshop facilitation by the WORLD TECHNOLOGY EVALUATION CENTER, INC.

The World Technology Evaluation Center, Inc. (WTEC) provides assessments of foreign research and development in selected technologies under a cooperative agreement with the National Science Foundation (NSF). Michael M. Reischman, Deputy Assistant Director for Engineering at NSF's Engineering Directorate, is NSF Program Director for WTEC. Several other U.S. government agencies provide support for the program through NSF.

WTEC's mission is to inform U.S. scientists, engineers, and policymakers of global trends in science and technology in a manner that is timely, credible, relevant, efficient and useful. WTEC assessments cover basic research, advanced development, and applications.

WTEC is a private, nonprofit corporation. For other WTEC reports, see http://www.wtec.org.

Dr. R.D. Shelton
President

Dr. Michael DeHaemer
Executive Vice President

Mr. Geoff Holdridge
Vice President,
Government Services

Dr. David Nelson
Vice President
of Development

WTEC, Inc.
4800 Roland Avenue, Suite 201
Baltimore, Maryland 21210

INTER-AGENCY CONFERENCE ON METABOLIC ENGINEERING

2007

PROCEEDINGS

February 13, 2007

Metabolic Engineering Working Group

This document was sponsored by the National Science Foundation (NSF) under NSF Grant BES-0508384, awarded to the World Technology Evaluation Center, Inc. (WTEC, Inc.). The government has certain rights in this material. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States government, the authors' parent institutions, or WTEC, Inc.

NATIONAL SCIENCE FOUNDATION
4201 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22230

May 1, 2007

Dear Colleague:

We wish to present to you Proceedings from the recent Inter-Agency Conference on Metabolic Engineering, which was held at the Bethesda North Marriott Hotel & Conference Center, North Bethesda, Maryland on February 13, 2007 in conjunction with the Joint Genomics: GTL Contractor-Grantee Workshop V and the USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop 2007.

As part of the larger workshop, the session on Metabolic Engineering provided presentations which sought to identify opportunities for leveraging genomic technologies for Metabolic Engineering. Specific topics for discussion included:

- Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks
- The Challenge of Incorporating Regulatory Effect in Genome-scale Networks
- Biological Aspects of Deciphering and Engineering Regulatory Networks
- Perspectives in Metabolic Flux Mapping

Presentations were provided by Grantees of Awards resulting from the activities of the Inter-Agency Metabolic Engineering Working Group (MEWG). The electronic version of this report is available at the following url:

“http://www.metabolicengineering.gov/me2007/ReportTOC.html”.

This website has slide information from each of the presentations.

For current information on the activities of the MEWG, please refer to the MEWG web site located at this url:

“http://www.metabolicengineering.gov/”.

We thank you for your interest in Metabolic Engineering and invite your inquiries regarding the Inter-Agency activities.

With best regards, I am

Sincerely yours,

Fred G. Heineken
Chair

Metabolic Engineering Working Group (MEWG)

Department of Agriculture

Liang-Shiou Lin	(202) 401-5042	llin@csrees.usda.gov
Gail McLean	(202) 401-6060	gmclean@csrees.usda.gov
Chavonda Jacobs-Young	(202) 401-6188	cjacobs@csrees.usda.goc

Department of Commerce

Travis Gallagher	(301) 975-5726	travis.gallagher@nist.gov
Robert Goldberg	(301) 975-2584	robert.goldberg@nist.gov

Department of Defense

Jennifer Gresham	(703) 696-7787	jennifer.gresham@afosr.af.mil

Department of Energy

Valerie Sarisky-Reed	(202) 586-8014	valerie.sarisky-reed@ee.doe.gov
David Thomassen	(301) 903-9817	david.thomassen@science.doe.gov
Sharlene Weatherwax	(301) 903-6165	sharlene.weatherwax@science.doe.gov

Environmental Protection Agency

Michael Broder	(202) 564-3393	broder.michael@epa.gov
Mark Segal	(202) 564-7644	segal.mark@epa.gov
Barbara Karn	(202) 343-9704	karn.barbara@epa.gov
April Richards	(202) 564-2297	richards.april@epa.gov

National Aeronautics and Space Administration

Steve Davison	(202) 358-0647	sdavison@hq.nasa.gov
Jitendra Joshi	(703) 599-3674	jitendra.a.joshi@nasa.gov

National Institute of General Medical Sciences

Warren Jones	(301) 594-5938	jonesw@nigms.nih.gov

National Science Foundation

Fred Heineken	(703) 292-7944	fheineke@nsf.gov
Parag Chitnis	(703) 292-7132	pchitnis@nsf.gov
Lenore Clesceri	(703) 292-5313	lclesrer@nsf.gov

Table of Contents

1. Cover Letter i

2. Metabolic Engineering Working Group ii

3. Background

Metabolic Engineering 1
Scope 1
Introduction 1
Conference Theme 1

4. Agenda 3

5. Abstracts of Expert Presentations

James Liao 5
Bernhard Palsson 5
George Bennett 5
Jackie Shanks 6

6. Panel Session Summary 9

7. Presentations

Fred Heineken 21
James Liao 25
Bernhard Palsson 52
George Bennett 62
Jackie Shanks 83

8. Agency Activities in Metabolic Engineering

Department of Agriculture 103
Department of Commerce 103
Department of Defense 104
Department of Energy 105
Environmental Protection Agency 105
National Aeronautics and Space Administration 106
National Institutes of Health 107
National Science Foundation 107

Background

Metabolic Engineering

An emerging approach to the understanding and utilization of metabolic processes is Metabolic (or pathway) Engineering (ME). As the name implies, ME is the targeted and purposeful alteration of metabolic pathways found in an organism in order to better understand and utilize cellular pathways for chemical transformation, energy transduction, and supramolecular assembly. ME typically involves the redirection of cellular activities by the rearrangement of the enzymatic, transport, and regulatory functions of the cell through the use of recombinant DNA and other techniques. Much of this effort has focused on microbial organisms, but important work is being done in cell cultures derived from plants, insects, and animals. Since the success of ME hinges on the ability to change host metabolism, its continued development will depend critically on a far more sophisticated knowledge of metabolism than currently exists.

This knowledge includes conceptual and technical approaches necessary to understand the integration and control of genetic, catalytic, and transport processes. While this knowledge will be quite valuable as fundamental research, per se, it will also provide the underpinning for many applications of immediate value.

Scope

The Metabolic Engineering Working Group is concerned with increasing the science and engineering community's level of knowledge and understanding of ME. The Working Group strives to encourage and coordinate research in ME within academia, industry, and government in order to synergize the Federal investment in ME.

Introduction

In November 1995, Science Advisor John H. Gibbons of the Office of Science and Technology Policy (OSTP) released the report, "Biotechnology for the 21st Century: New Horizons." This report was a product of the Biotechnology Research Subcommittee (BRS) under OSTP, and identifies priorities for federal investment and specific research opportunities in biotechnology. These priorities include agriculture, the environment, manufacturing and bioprocessing, and marine biotechnology and aquaculture. The BRS formed several working groups to facilitate progress on some of these key priorities. The Metabolic Engineering Working Group (MEWG) was created to foster research in Metabolic Engineering, an endeavor that can contribute to all of the key priorities in the aforementioned report. The Working Group is composed of Federal scientists and engineers who participate as part of the activities of OSTP, and represent all of the major agencies involved in Metabolic Engineering research.

Conference Theme – Methods to Examine Biochemical & Metabolic Networks

The Metabolic Engineering Working Group (MEWG), in pursuit of its goals to promote the advancement of metabolic engineering and coordination of the Federal metabolic engineering research activities for maximum productivity, organized its seventh Inter-Agency Conference held on February 13, 2007.

The goals of any ME program comprise the conceptual and technical approaches necessary to understand integration and control of genetic, catalytic, and transport processes in cellular metabolism. The ability to modify biological pathways extends the fundamental knowledgebase of predictive systems biology towards driving practical applications and sustainable resources for bioenergy solutions.

This breakout session, conducted as part of the jointly-held DOE Genomics: GTL Contractor-Grantee Workshop V, the USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop, and the Metabolic Engineering Working Group Inter-Agency Conference on Metabolic Engineering 2007 Awardee Workshop 2007, seeks to identify opportunities for leveraging genomic technologies for metabolic engineering. Specific topics for discussion include:

- Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks
- The Challenge of Incorporating Regulatory Effect in Genome-scale Networks
- Biological Aspects of Deciphering and Engineering Regulatory Networks
- Perspectives in Metabolic Flux Mapping

Specific issues for each topic can be found in the Session Summary.

Agenda

Metabolic Engineering Breakout Session

Tuesday, February 13, 2006, 2:00 to 5:00 pm

Organizers: Gail McLean (USDA) & Sharlene Weatherwax (DOE)

Moderators: Mark Segal (EPA) & Fred Heineken (NSF)

2:00 pm	Opening Remarks – Fred Heineken (NSF)
2:10 pm	Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks – James Liao (University of California, Los Angeles)
2:40 pm	The Challenge of Incorporating Regulatory Effect in Genome-scale Networks Bernhard Palsson (University of California, San Diego)
3:10 pm	Break
3:40 pm	Biological Aspects of Deciphering and Engineering Regulatory Networks – George Bennett (Rice University)
4:10 pm	Perspectives in Metabolic Flux Mapping – Jackie Shanks (Iowa State University)
4:40 pm	General Discussion – Moderator: Mark Segal (EPA)
5:00 pm	Adjournment

Abstracts

Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks

James C. Liao
Chemical and Biomolecular Engineering
University of California, Los Angeles

The ultimate goal of systems biology and metabolic engineering is the prediction of cellular behavior, either for wild-type strains or the engineered strains. This ambitious task involves efforts at two levels: 1) prediction of regulatory networks, and 2) prediction of the behavior of the regulatory networks. We will discuss challenges in these problems and suggest some possible ways to tackle them. In particular, we will address how mathematical tools can interact with experimental data to maximize information output.

The Challenge of Incorporating Regulatory Effect in Genome-scale Networks

C.L. Barrett, M.J. Herrgard, B.K. Cho, E.M. Knight, J. Elkins, B.O. Palsson
Department of Bioengineering
University of California, San Diego

Genomic and bibliomic data has been used to reconstruct a number of genome-scale metabolic networks. The stoichiometric structure of these networks has enabled a series of basic and applied studies that address both proximal and distal causation in biology. One of the challenges going forward with computational models at the genome-scale is to account for regulatory effects. Regulation of metabolic enzymes occurs primarily at two levels; 1) the transcriptional level, and 2) the posttranscriptional (protein expression and activity) level. Significant progress is being made with characterizing, reconstructing and modeling transcriptional regulatory networks regulating metabolism. However, the development of the corresponding methods for incorporating posttranscriptional regulation into genome-scale models is still at an early stage. The current state and future challenges of incorporating both transcriptional and post-transcriptional regulation in genome-scale networks will be discussed.

Biological Aspects of Deciphering and Engineering Regulatory Networks

George N. Bennett
Department of Biochemistry & Cell Biology
Rice University

In efforts to detect and modify regulation of pathways for a particular goal there are strategies that can be implemented in two circumstances, one the synthetic approach where major components of the pathway are known and the appropriate regulation of the various enzymes can be adjusted by taking advantage of modeling and combinatorial assembly methods; and the other if the proteins to be altered are not obvious and thus the modification must take a more empirical approach with selection or screening methods being the issues.

In known expression systems, it is still needed to detect levels of regulation, for example determination of the level of functioning of a protein (enzyme for metabolic process) more than knowledge of the level of gene expression is needed to understand how the activity varies under the physiological condition contemplated for use. This is the situation if genes from various sources are placed together in a new way to form a nonendogenous pathway as often is the case in metabolic engineering. Detection of proteins, metabolites etc from systems biology approach of measurements in different genetically engineered cells under various conditions

can help in this endeavor. Due to the large number of possible combinations of mutations and conditions a way is needed to minimize or focus the experimental measurements on the most appropriate ones to examine. Such experiments should make an effort to take into account the effects of intercellular conditions produced by introduced changes on related protein activities (other enzymes of pathway, regulatory factors, functional state of activity of the enzyme or regulatory proteins) and models that include this interaction information would be more comprehensive.

In order to carry out appropriate modification of regulation there are relatively straightforward approaches in the case of desired changes in specific known expression controls such as through modification of transcriptional events and to a lesser extent, modification of more general aspects of cellular physiology (redox and air, enzyme stability, enzyme parameters, osmotic conditions). To target regulation to specific pathway, designed regulation of small units can be employed. These involve the use of known regulated promoters that are varied to adjust the level of constitutive expression, and can be combined with terminators or RNA structural elements to afford variation of level of expression. In the case of less known processes found in many industrial organisms that have less genetic and biochemical literature, efforts are more a matter of perturbing a somewhat more global functioning system and screening or selection for those altered cells that perform better, then analyzing and combining the most promising. The idea of eliminating complicating or undesired processes that may obscure the regulation you would like to enhance is an useful experimental strategy. For these wider scope effects, or regulation of unknown factors with less obvious connections to the metabolically engineered process, modification of transcription factors such as sigma factors, Zn-finger motif factors, general or global transcription factors may be used in combination with powerful selection or screening systems for the desired property.

Perspectives in Metabolic Flux Mapping

Jacqueline V. Shanks
Department of Chemical and Biological Engineering
Iowa State University

Metabolic flux maps provide a quantitative depiction of carbon flow through competing metabolic pathways, thus providing: analysis of substrate utilization and product formation; flexibility or rigidity of carbon flow at network nodes; the rate of a given enzymatic reaction in vivo; and inferred availability of NADPH or ATP. Thus, metabolic fluxes are an important physiological characteristic complementary to levels of transcripts, proteins, and metabolites. The system-wide quantification of intracellular fluxes in an organism is called metabolic flux analysis (MFA). The most basic approach to MFA is stoichiometric MFA, which involves writing balances for intracellular metabolites based on the stoichiometry of the biochemical reactions in the metabolic network. This results in a system of linear equations, which are solved by employing extracellular and biomass synthesis flux measurements to resolve some or all degrees of freedom. Genome-wide or *in silico* flux models provide the solution space of feasible fluxes resulting from optimization of the balances to a global cellular goal, such as maximum growth rate. Recently, constraints to the *in silico* models provided by data from ^{13}C labeling experiments, have narrowed the solution space.

^{13}C metabolic flux analysis (^{13}C MFA), with isotope detection via GC/MS or NMR of metabolites (e.g. amino acids from hydrolyzed protein), quantifies intracellular metabolic fluxes for smaller reaction networks, where the fluxes are completely determined (in contrast to the *in silico* models). ^{13}C MFA provides redundant measurements for flux quantification, as well as testing the consistency of the network topology for the physiological conditions. Isotopomers, which are

isomers of a metabolite that differ in the labeling state (^{13}C or ^{12}C) of their individual carbon atoms, are a central concept in the analysis and mathematical modeling of ^{13}C MFA.

Increasing levels of information can be obtained from ^{13}C labeling data when coupled with a stoichiometric model of the biochemical pathways and computational methods to solve for flux data in the smaller network. More rigorous analysis is indicated as one moves from analytical (a few flux ratios at metabolic branchpoints) to ^{13}C constrained flux analysis (stoichiometric model with a few flux ratios as constraints) to fully integrated determination of fluxes from all the experimental data and the stoichiometric and isotopomer balances. Iterative methods have been used to solve the full relationship of isotopomer balances and the NMR or GC measurements to provide consistency, and routines to minimize error from the overdetermined data sets are required.

^{13}C MFA studies of aerobic glycolysis in microorganisms have become “higher throughput” since simplifications to the metabolic network can be made, and ^{13}C constrained flux analysis can be used. For alternative physiological conditions, for example where anapleurotic pathways are active, reversibility of reactions are indicated, or substrates other than glucose are used, the development of a consistent network topology and the strategy for the choice of the label to obtain identifiable fluxes are not as straight-forward. Furthermore, due to compartmentation and the existence of parallel pathways in plants, more experimental measurements are needed than in microbial systems, and the number of isotopomer balances increases, further increasing the computational burden. As a note, to date *in silico* models for plants have not been developed. Thus, at this point, these more challenging systems are not yet ready for “high-throughput” measurements. However, a growing knowledge base in ^{13}C MFA in these systems should enable movement towards more genome-wide flux estimation. This presentation will summarize these points with approaches from our laboratory in determining ^{13}C-based metabolic flux maps in plants and microbes.

Summary of Metabolic Engineering Breakout Session at GTL-MEWG 2007

Fred Heineken (NSF) welcomed participants to the workshop and provided an overview of the interagency Metabolic Engineering Working Group (MEWG). The overview described the federal participants, goals, accomplishments, program solicitations, awards, and future plans. Proceedings from last year's meeting were provided at the session and access to them via the MEWG website was also noted. The session was moderated by Mark Segal who noted each of the speakers would make a presentation on a specific topic area and following the presentations, there would be a general discussion on how the MEWG can facilitate progress in metabolic engineering research. Ultimately MEWG would like to revise and update future calls for proposals to address the most pressing issues in metabolic engineering as discussed at this meeting.

The session addressed the topic area "Pathway Interactions and Regulatory Processes." While much research has focused on identifying and understanding components of metabolic pathways, the interactions between components and associated regulatory processes need to be better understood. This understanding will enable the metabolic engineering of functional networks that can be regulated as needed and desired. The Metabolic Engineering Breakout Session concentrated on four general areas related to identifying, analyzing, and manipulating pathway interactions and regulatory processes for metabolic engineering.

- **Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks - James Liao, University of California, Los Angeles**

  How do we perform large scale adaptive evolution?

- **The Challenge of Incorporating Regulatory Effect in Genome-scale Networks - Bernhard Palsson, University of California, San Diego**

  What are methods, tools, challenges to understanding interaction between components involved in metabolic pathways at the systems level? How can we make downstream 'omics,' such as metabolomics, more feasible and effective at the metabolite or enzyme scale as currently done for genomics and microarrays?

- **Biological Aspects of Deciphering and Engineering Regulatory Networks - George Bennett, Rice University**

  How do we detect and modify the regulation of a metabolic pathway to achieve a specific goal or product? How do we target regulation to a specific metabolic pathway?

- **Perspectives in Metabolic Flux Mapping - Jackie Shanks, Iowa State University**

  What are the methods, approaches, challenges for understanding and regulating flux? How do we elucidate genome-scale fluxes?

Summaries of the Presentations:

Challenges in Predictive Modeling for Engineering/Deciphering the Regulatory Networks - James Liao, University of California, Los Angeles

The first presentation was by James Liao and focused on strategies to model regulatory networks. Several questions were proposed for consideration regarding modeling:

1. What is the purpose of the model?
2. What data to use?
3. What are existing and alternative approaches?
4. What is the scope of the model?
5. What are the mathematical details?
6. What is the basis of prediction?

Dr. Liao used the modeling of active Nitric Oxide (NO) response networks in *E. coli* demonstrate how these questions would be addressed.

NO is used as a major mammalian signaling molecule and functions as a vasodilator, neurotransmitter, smooth muscle relaxant, and in the immune response. Vasodilation can lead to septic shock, the 10th leading cause of death in hospitals. To characterize the active nitric oxide (NO) response networks (question 1, the purpose of the model), the researchers utilized an NO donor compound (DeaNO) to release NO within 2-3 minutes. The DeaNO inhibited growth of *E. coli* longer than the lifetime of NO, leading to the question, "What is happening in the bacteria to produce this response and cause the growth inhibition?" For the second question (data to use), the researchers selected microarrays, specifically an *E. coli* whole genome array. There are a lot of different types of data and analyses available for the use of microarrays, for example by varying environmental conditions or examining transient expression (question 3, existing/alternative approaches). Typically biologists use an *ad hoc* analysis, listing all the genes that are up or down-regulated and then examining each one individually. While a researcher can generate a lot of insight with this type of analysis, it's not systematic and it's easy to lose key points. There is also cluster analysis, involving statistics, which uses the reasoning that genes that respond in a similar fashion can provide keys to common function.

The scope of the model (question 4) can include mRNA, transcription factors, degradations, etc. but needs to be kept manageable. The goal is to develop a model which predicts what the data are describing. For the example study, the scope was focused on transcription networks having the challenges of multiple levels of regulation. Transcription factors regulate transcription and are themselves subject to post-transcriptional regulation. The binding of active transcription factors to the transcription complex can be gene-dependent, condition-dependent, and variable in strength. The researchers chose a power-law model to address these variables and characteristics. Such a model is not perfect for any gene but roughly good for all genes, providing a needed balance. Network Component Analysis (NCA) was used to form the mathematical model (question 5, mathematical details). In its simplest form, the model is rate of formation minus rate of degradation. While transcription factor activity is also dependent on time, in this case, a quasi-steady state assumption is made and the time dependence is ignored. This leads to a relationship between microarray data and changes in transcription factor activity and strength/effect of binding. Included in this is a bi-linear optimization process involving connectivity constraints and regulatory constraints to derive a unique solution. NCA serves as the biological model and gives biological significance and limitation. It provides the

identifiability criteria, including yield mathematical constraints needing to be satisfied. It also provides connectivity theorems, generating insight into connectivity density.

Once all of the constraints are determined, the NCA schematics are devised. Connectivity data are collected, from literature such as a painstaking literature search or website queries where information has already been tabulated or from experiments such as ChIP-chip analysis. From the connectivity data, subnetworks are selected to satisfy NCA criteria, and then microarray data are collected, resulting in a potential predictive model. For question 6, in this example, the bases for prediction are: transcriptome data (mRNA abundance), transcription connectivity, and the power-law model. This leads to active transcription factors and active networks. Is there a statistical way to determine how active is active? What is statistically active? To approach this, one randomizes data and then constructs null distribution of transcription factor activity by random networks. Then one re-does analysis, plots the null distribution, gets a *p* value, 0.005, and sees if reported the network of interest falls significantly away from the mean.

Thus, using NCA, the NO response regulators were identified. Of the approximately 80 transcription factors found, DNA microarrays were used to measure network behavior, see which transcription factors mediate NO response to transcriptional perturbations, and find the cause for transcription factor activation. A chemogenomic approach was used to complement the transcriptome NCA and involved linking chemical analysis, “reactome,” and phenotype analysis to identify sensor, targets, and metabolic responses. Knockout experiments can also be added to verify predicted changes in transcription factor activity. Using these approaches, primary and secondary effects or direct and indirect effects can be distinguished. In the end, this approach deduced two parts of the *E. coli* network. One part actively participates in NO defense, utilizing the transcription factors IscR, NorR, and NsrR. The other part targets IlvD, LeuC, CyoB, involved in branched chain amino acid synthesis, a metabolic response. This is an indirect response, observed in microarray data. Without this type of modeling effort, a researcher would have a hard time separating a direct defense mechanism from an indirect metabolic response. This validates the methodology for solving this real biological problem.

Questions and Comments:

How robust is the method for missing connectivity links?

If you have a transcription factor that regulates many genes, missing a few is no big deal. If there are few targets, then missing anything is a real problem.

Another verification step is to see whether transcription factors are binding sites upstream of putative target genes. What about binding at mathematically predicted sites? Have you looked at this?

Yes, collaborating researchers are doing this, to verify via biochemical means our bioinformatics-predicted sites.

Mark Segal questioned if we need to increase the utility and availability of bioinformatics tools for biochemical validation. The conclusion was yes; such tools are needed for general scientific research as well as for metabolic engineering. It was also concluded that the scientific community needs a way to increase the availability of and data for bioinformatics tools for prediction in organisms less commonly used and lacking basic background data.

The Challenge of Incorporating Regulatory Effect in Genome-scale Networks - Bernhard Palsson, University of California, San Diego

The second speaker was Dr. Bernhard Palsson who brought up contemporary issues for incorporating regulatory effects in genome-scale networks. He first discussed the history of metabolic reconstruction. The history of the *E. coli* metabolic model, generated over 17 years ago, was summarized in a timeline starting with the 1990 Majewski and Domach *Biotechnology and Bioengineering* article up to a 2007 Feist *et al*. article currently under review. Reconstruction of the *E. coli* metabolism includes reactions, genes, and metabolites, with the timeline showing an increasing number of capabilities, from amino acid and nucleotide biosynthesis to compartmentalized reconstruction in the periplasm. Research now is moving towards gap filling. The latest version of the *E. coli* metabolic model will be fully incorporated into EcoCyc. Literally dozens of papers have used this type of reconstruction, including articles on metabolic engineering, microbial evolution studies, network analysis, phenotypic behavior analysis, and biological discovery.

Metabolism can now be described at a genome level. Pathways operate as elements of a network, not as independent entities. As a result, methods being used to interrogate genome scale modeling fall into constraint-based modeling approaches, as described in the 2004 Price *et al. Nature Reviews Microbiology*. Genome-scale constraint-based modeling is a field that has developed rapidly over the last 5 years. Using a top-down, constraint-based process view of regulation, regulation of expression can be described as a shrinking solution space. There is a need for reconstruction of transcription regulatory networks now to overlay them with metabolic pathways. This is now doable. The regulation of activity, or its location within a shrunken solution space, is a future task as it is the mathematical sampling of the solution space.

An example of state of the art in genome scale measurements is RNA polymerase (RNAP) and transcription factors. Measurables include RNAP binding, transcription factor binding, expression, and binding motifs and methods include high-density tiling microarrays and ChIP-chip analyses. Expression on tiling arrays provides transcriptome data. Location of binding factors and output can be assayed through genome-wide identification of active promoters. RNAP ChIP-chip analyses can indicate RNAP binding sites and, using rifampicin, the dynamics of RNA polymerase binding. Dynamic RNAP binding maps correlate with expression profiling so both the population of genes activated and the temporal activation by RNAP can be measured. Transcription factor ChIP-Chip can detect regulation by transcription factors. Also the genome-wide determination of noncoding RNAs in regulation using fine-resolution expression profiling shows that there are small RNAs regulated by transcription factors that bind in front of the small RNAs. Also in *E. coli*, drill-down studies on transcriptional sub-networks provide an entire, systematic program which fully elucidate all the transcription factors, from global regulators to major and minor regulators, to target genes. While these data are obtainable today, it can be expensive and time-consuming, up to $40,000 to do an experiment.

With today's technology, a researcher can collect the data needed for a genome scale view of transcription regulatory networks in bacteria. Now scientists can begin to integrate metabolism and transcription regulatory networks. *In silico* methods to assemble and analyze data are being developed. A 2005 article by Barrett *et al.* in *Proceedings of the National Academy of Sciences* computed the metabolic and transcriptional regulatory networks for *E. coli* and assessed the properties of the functional states in 15,580 distinct minimal growth media. The data showed that the electron acceptor has the biggest impact on distinguishing specific clusters in the integrated network. Such studies are helping identify the principle modalities in

the regulatory network. Mathematical formulations are now developing which integrate metabolic and transcriptional regulatory networks at the genome scale.

An impact of these types of studies on metabolic engineering occurs through increased understanding of causation in biology and is leading to new approaches to strain design. The current bacterial strains are unstable, for example due to batch processing and cell line banks. The hope is to reverse this and move towards synthetic biology with a self-optimizing strain performing continuous processing. The question is how many different strains can we actually engineer this way? A useful approach can be re-sequencing to study the genetic basis for transcriptional regulatory network adaptation at the genome scale.

To summarize, the presentation showed the establishment of metabolic reconstruction methods and a conceptual framework for transcriptional regulatory network function at genome scale. Technology is also now available for direct measurement of transcriptional regulatory network events on a genome scale. This has led to the identification of 180 putative transcription factors in *E. coli*. Analyses of function of each of these transcription factors are expensive, and *in silico* methods to assemble and analyze the data are being developed. Finally, the integration of transcriptional regulatory networks with metabolism is possible, and the genetic basis for adaptation of transcriptional regulatory networks can now be monitored.

Questions and Comments:

You thought that the electron acceptor has the highest impact on regulatory state—which ones are you talking about?

It turns out that the electron acceptor can be oxygen, fumarate, DMSO, TMAO, nitrates.

In regard to ChIP-chip and microarray -- do you see evidence for major differences for different RNAs? Will the stability of a message influence the correlation of the occupancy of the RNAP with a specific message?

We don't know yet—maybe transcripts known to be more labile (if you want to come to California and analyze the data, you're welcome to!)

Comments focused on the amount of data available and how to analyze it all. It takes 2 to 3 times longer to analyze data than to generate data, and lots of data are archived in warehouses and never fully analyzed. Funding agencies should know this. How do we make this widely available?

Peter Karp brought up the issue of meta-analysis noting there are lots of *E. coli* microarrays that can be pooled and predictions by other groups — when they asked the question of how much goes into EcoCyc, a reviewed database, they concluded that they, hopefully, are synthesizing more accepted information rather than being very complete with not-well-accepted data. It's hard for them to know which analyses to put in and which are not well accepted yet.

Maybe integrating across it computationally is better, although a challenge? Bernhard Palsson agrees that integrating all chip data from all labs would be difficult. Do you want to build one database or associations between databases? There is the challenge of integrating data in one place versus meta-analysis to tell you what the data really says; high throughput data are very hard to integrate. The end points of the genomic adaptation are actually hard to see from lots of low-resolution data.

Mark Segal noted that this sounds like another challenge for future consideration – do we want one large database or multiple databases that can communicate for these types of analyses?

The audience also discussed insights on mathematical representation of data, such as biochemical, genetic, and genomic-structured data. Rather than inference, researchers can now make direct measurements. This leads to questions of how to mathematically represent data and chemical conversions. Currently we have causation described by Boolean statements (for example, either there is oxygen or not) that are static but we can now use a matrix so that the description is dynamic rather than static. R-matrix formalism is a way to take the Boolean statements and put them into a matrix, quasi-stoichiometric formalism. Stoichiometric matrices have the advantage of one-step integration.

Biological Aspects of Deciphering and Engineering Regulatory Networks - George Bennett, Rice University

The third speaker was George Bennett who addressed the biological aspects of characterizing and modifying regulatory networks. For metabolic engineering, the cell should be consider as a production organism rather than a production line since the organism can effectively negate the engineered pathway through homeostatic control mechanisms. Thus, it is important to obtain appropriate balance in levels of expression of pathway genes. Such balance includes taking into account the regulation at enzyme level, including feedback connections, and at the global physiological level. Expression level includes active promoter selection and mRNA stability and translational efficiency. Tools for obtaining appropriate balance in pathway gene expression include a library of promoters with varying strengths, combinations of promoters and terminators, ribosomal binding sites for multicistronic mRNAs, RNA regulators (antisense, RNA binding proteins), and selection of plasmid or chromosomal location.

Interconnections among regulators and small molecules help modulate enzyme level and feedback connections. Researchers need to define which regulatory interactions are most important. We may know a great deal about individual components, but predicting and understanding the reasons why a specific metabolic condition prevails in a particular situation is not so easily done. There are a lot of situations where a number of regulators participate. Consequently, detailed experiments are needed to define the most important regulatory interactions in a specific situation or environments. Also, the function of a particular protein depends on more than gene expression. Using an example for aerobic/anaerobic *E. coli*, the ArcA that is activated under anaerobic conditions stimulates PFL and cyd. In fact, however, PFL fluxes are actually higher in the *arcA* mutant. ArcP activates *pfl* expression and PFL is reactivated by YfiD. In chemostat runs, ArcA- leads to lower cyd. Lower cyd cannot handle all NADH and redox equivalents build up, leading to expression of *yfiD*, which reactivates Pfl in the *arcA* mutant leading to higher flux. This sort of example gives the perspective that you need a variety of experiments to determine the important regulatory events.

Sometimes you can take advantage of the alternation of global cell physiology resulting from manipulation. The manipulation may alter levels of cofactors, sensors, and transporters, resulting in a global effect. For example, cells may be genetically engineered to produce more NADPH. This manipulation leads to rearrangements of metabolic fluxes in the system and pathway measurements show the pentose pathway that usually supplies most of the NADPH is much reduced. The higher availability of NADPH might allow it to be used in another pathway, and examination of a mutant shows a visible increase in the lycopene production pathway.

There are many challenges in engineering networks with unknown genes and regulators. A significant challenge is defining the network and identifying which genes to modify and control for a specific metabolic engineering goal. This can be accomplished using mathematical models, gene expression or proteomic measurement, tracing of phosphorylation networks, comparison with similar protein-protein interaction networks in other species. Genetic

exploration is another approach and involves use of individual and/or multiple gene knockouts or overexpression, novel regulatory factors to affect pathway gene expression, and high throughput screens, enrichments, or selections. For instance, one strategy is to use high throughput screens to gather data and then merge in model/computational and analytical approaches. Another strategy is to use a random approach based in genetics to find new regulatory tools.

Identification of important regulators is another challenge. These may include individual specific regulators for an operon, global regulators, 2-component regulators, or small RNA regulators. Factors to consider include that global regulators affect expression of a lot of genes and that some regulators respond to many environmental growth conditions. Thus, the phenotypes of regulatory genes should be tested under different conditions and used to construct a database of conditional regulators. Approaches to identify functions of regulators include ChIP-chip to determine binding regions in the DNA and two-hybrid or cross-linking to determine protein interaction networks. A combination of bioinformatics, microarrays, expression pattern analysis, and mutant production and analysis is useful for determining the use of small RNAs as regulators.

Even if we don't know what the regulatory network is, we can use genetics to help us find useful changes to make and influence our metabolic engineering strategies. Metabolic engineering of a pathway might simply be practically motivated and focused solely on improved production; the mechanism is not of immediate importance and can be researched later. In this case, random or artificial evolution approach can be used. A selection or screen for increased production is developed followed by analysis of the mutants. A genome sequence approach is possible where the genome of cells adapted to a higher growth-production state is analyzed to see what has changed in the genome. This approach is often used in industry to characterize the production strain as compared to the original parent strain.

Another option for altering regulation is the development of novel transcriptional regulators. Mutations can be introduced into known regulators to find those that generate an effect and result in the desired phenotype. For example, a mutation in the TATA binding factor was used to engineer the yeast transcription machinery for improved ethanol tolerance and production as described in the 2006 Alper *et al. Science* article. Bae and Kim in the 2006 *Molecules and Cells* describe a procedure for selecting active artificial zinc finger proteins from a zinc finger library.

Questions and Comments:

Are screening selections a complete replacement for rational design? Wouldn't it be better to do rational design? It would be hard to screen for random mutants and get the many simultaneous changes that have the effects of rational design. So it might be more effective to perform it sequentially.

Yes, these approaches might work best in concert with others. But we wouldn't want to rely completely on a single approach. Rational design can go quite far and especially in some organisms where there is a lot of background information and in combination with more advanced programs can take the process of metabolic engineering to an advanced point, but not all things are known that affect the system so it is in this area that the more random approach can bring advances.

Mark Segal brought up the question of practicality: obtaining strains for industrial development – how willing will people be to release such strains. Similarly, what about obtaining strains from industrial sources where there is a sequence of development. How can these be obtained for

re-sequencing? The general response was that it would be an internal issue for the company, but some of them might make them available upon publication.

In this section, I mentioned that industrially adapted strains are found over time that have gained productivity beyond that of the parental strains that have known key mutations. Sequence analysis of these can and has given companies insight into other factors that contribute to the goal of high production and stability. I do not think it is for NSF to analyze industrial production strains, this is more suitable for companies and their proprietary issues. But I was considering the further genomic and metabolic analysis of laboratory strains that have been adapted by selection under conditions similar to a production environment and to investigate those adapted strains because such analyses have potential to identify currently unknown factors (metabolic interactions) and genes that allow a more robust productive strain. Manipulation of such targets (factors) could then be used in a variety of practical ways, and also these factors could be included in predictive computer models once the mechanisms of their contributions to improved performance is understood. Analyzing how these specific genetic changes make the strain better under a useful condition is another interesting area where the tools and data described by other speakers could be employed to characterize the physiological consequences of newfound improved variants.

Perspectives in Metabolic Flux Mapping - Jackie Shanks, Iowa State University

The final presentation was by Jacqueline Shanks who concentrated on metabolic flux mapping. The charge was to describe the methods and challenges and extrapolate to genome scale. She used her work in soybean as a case study, which will illustrate natural challenges and things to think about. Several different pathways can be examined for metabolic flux. Her work will focus on steady state rates through central carbon metabolism.

Fluxes reflect integrations of genetic and metabolic regulation and are obtained using material balances. Metabolic flux analysis involves building a stoichiometric model to solve for unknown fluxes. For this model, intermediates experience rapid turnover and a quasi-steady state is assumed. How many measurements are needed? A stoichiometric model has m unknown fluxes and n internal metabolite balances. For the more complex networks, $m>n$; there are an insufficient number of measurements, yielding multiple solutions. Both $m=n$ and $m<n$ reflect simpler networks and sufficient (or greater) number of measurements; they also indicate unique flux distribution. Given this, how is the true solution chosen? First, the researcher should choose a solution that optimizes a logical cellular objective and/or add internal measurements. Labeling studies using GC-MS or NMR give measurements at internal branch points; this option is expensive but more accurate. Measurements via GC/MS use one metabolite and are more sensitive. Measurements via 2D NMR use two metabolites and are less sensitive. 2D NMR will separate out the carbon positions so you will know which position is labeled but with less sensitivity. With GC/MS the researcher can begin to get position information but more analysis is required. Shanks' lab uses 2D NMR. In the last year, Palsson and Maranas have provided genome wide scale models using some of these measurements given some constraints. The talk will focus on smaller scale networks.

An adapted method http://www.biotech.biol.ethz.ch/sauer/ using ^{13}C labeling of proteinogeneic(sic) amino acids or metabolites is used from comparative, analytical, and integrated information. The resulting labeling data are of increasing complexity from mutant/condition discrimination to flux ratios to absolute fluxes. These algebraic relationships assume some kind of labeling pattern and operation of your network. Physiological data, external fluxes, and biomass composition data can be solved comprehensively via iterative isotopomer balancing. Algebraic methods can be used to solve for absolute fluxes, using ^{13}C-

constrained flux analysis. This approach is less rigorous but also less computationally intensive. Sometimes the researcher has to determine what is available for their network. If analyzing several mutants, the researcher might be able to bootstrap from the simpler approaches and achieve a good result without having to go through the labor-intensive isotopomer balancing method.

Flux estimation is difficult in plants for several reasons. One reason is the plasticity of plants. The compartmentation results in parallel pathways and different fluxes. For instance, how can the glycolysis in the cytosol and the plastid be distinguished? The different tissue and different cell types within tissues are another source of difficulty for flux analysis in plants. Currently, the issues of the tissue and cell types cannot be addressed. However, the plasticity issue is being addressed in Shanks' research through the development of an *in vitro* culture for developing soybean embryos. Soybean seed is essentially a little reactor with sucrose, sugars, and other metabolites that makes protein, oil, and starch. Using glutamine as a nutrient (carbon and nitrogen) source, she adds 10% U-^{13}C sucrose for uniform label; an exponential growth rate allows the assumption of the pseudo steady state condition. The embryos are harvested and fractionated into traditional "agronomic" fractions, starch, protein, oil. The hydrolyzed protein and starch are then analyzed by NMR and a mathematical model used to obtain metabolic fluxes. Additional measurements including biomass accumulation, substrate and product rates, biomass composition, and HPLC protein analysis are also made. NMR only provides a ratio and these additional measurements connect the NMR data to absolute flux.

There is no soybean network to download from the web to generate a reconstructed network. Consequently, Shanks has used the literature, Singh (1999), Soybase, and KEGG encyclopedia to generate her own soybean metabolic network map. Upon hydrolyzing protein in the fractions, the resulting labeled amino acids reflect their precursors. For example, alanine and valine will reflect the labeling in pyruvate in the plastid fraction. Thus, hydrolyzed fractions are markers for metabolic flux information and a separate analysis of the plastid and cytosol with 2D NMR is not needed. Biochemical knowledge is used to determine the compartment. The relative abundance of doublets and singlets represents the relative concentrations of intact and biosynthetic bonds in the same metabolite molecule (in other words, different pathways). When looking at ^{13}C labeling, attached to ^{12}C, a singlet is most likely formed by a biosynthetic bond between a ^{13}C and a ^{12}C molecule. A ^{13}C attached to a ^{13}C doublet is most likely formed from an intact ^{13}C molecule. Thus the different metabolic histories are visible.

All the different isotopomers and all possible labeling patterns of a molecule have to be counted since there is some naturally occurring ^{13}C. Most of the amino acids are detected on HSQC with a separate one to detect aromatic regions. Peak fine structure shows multiplets, and multiplets are proportional to isotopomers. The flux evaluation methodology involved measuring156 isotopomers. Istopomer abundances are related to fluxes. This nonlinear relationship is not trivial to solve; consequently, a computational framework was developed to convert automatically raw data to fluxes and perform these "modules" in the program. Briefly the program modules are: (1) guess fluxes, (2) simulate isotopomer abundances, (3) compare simulated and experimental abundances, and (4) statistically analyze fluxes. Reiterations are performed between modules 2 and 3 to eliminate error. Approximately 3-500 simulations may be needed due to error in NMR measurements.

The strategy to identify segregation of pathways (cytosol vs. plastid) relies on analysis of isotopomer distribution. If isotopomer distribution doesn't match for an amino acid that should be in all compartments, then the data indicate that there are separate pathways. For example, hexose pools are isotopically segregated in the cytosol and the plasmid so distinct glycolytic and pentose phosphate pathways (PPP) are seen. The rapidly exchanging pools of triose are

also detected. Statistical analysis of evaluated fluxes takes into account experimental error of NMR measurements. Diagnostics also include the assumed network topology. For instance, a study of reductive PPP can point out errors in the network topology. These analyses show that reductive PPP fluxes are substantial in the plastid but negligible in the cytosol. This is consistent with genetic and enzymatic data from some other higher plants although it has been reported that reductive PPP may not be present in the cytosol in most plants. The model suggests that the plastid pathway has more flux and this is confirmed by the literature. Thus, Shanks is both performing flux measurements and learning about her system at the same time. Nodes in the network can be flexible or rigid, depending on conditions at which experiments are conducted, such as different temperatures and use of different genotypes.

Metabolic flux mapping has also been applied to *E. coli* anaerobic metabolism. *E. coli* is a well-studied system under aerobic conditions, even smaller under anaerobic conditions. Challenges of this system are smaller flux to biomass, many products, and less rearrangements (TCA cycle incomplete). Some of the measurements are used to determine flux while others are used to test how well the metabolites are being measured. Results were confirmed by HPLC.

The labeling chosen for use can help identify specific fluxes. Using two labeled compounds *in silico* can help determine which to use *in vivo* or *in vitro* but requires some knowledge of the network. As an example, icd is in the cytoplasm but this can't be distinguished in the carbon skeleton rearrangements in Shanks' model. In this case, then, she used a dual label experiment to identify a specific compound. This approach, however, requires that the researcher have some base value for fluxes.

To summarize, this research has led to development of both a carbon labeling experiment to evaluate compartmented fluxes and a computational framework to aid in flux evaluation including reversibilities and confidence intervals. Going to larger networks, how well can different network topologies be discerned? The ideal is a closed mass balance; difference will be due to experimental error. Extracellular, biomass, and composition measurements are important. High throughput methods such as metabolite profiling as well as retrobiosynthetic approaches can be used. How to design experiments to identify fluxes? This includes effective methods for analysis of isotopomer abundances. For larger networks, more labeling can be used to constrain and identify fluxes keeping in mind uncertainty or range in flux estimates.

Questions and Comments:

As you see where you are now, where are the major challenges faced by the field? Experiments? Computation? Instrumentation?

The methods are there. Dr. Shanks is limited by the experiments for her network. Her network had not been solved computationally yet. It's relatively small but nothing really in the literature. Solving the system and concentrating on a set of fluxes was a challenge. Theoretically one should be able to measure even more things so some of the computational methods and solutions are needed. Another issue is that with increasing numbers of measurements, there are more errors. How much better can we do? Some type of "propagation of error" analysis is needed. Given constraints of an experiment, one needs an error analysis. A defined system like *E. coli* where you know all the pathways, all the carbon skeletons, all the substrates, is just not available for the system she is working with.

Bernhard Palsson commented that the bottom line is that this field is computationally-limited, even for existing data, and data limited.

Regarding compartmentalization, if you get something like a cyanobacterial method or data, could you extrapolate to the plastid component? Would that help for specific parts of the network?

Yes, you could solve that subnetwork, but it doesn't help you with the overall plant network. In the soybean seed, we assume all the cells are homogeneous. But the next step requires a lot more computation, seeing the cells in different regions of the tissue and analyzing and building out. Also you need to consider cell to cell interactions, and thus computation for cell to cell, compartment to compartment. There is also the issue of differentiation.

Enzymes are known to have preferences for different isotopes. What is the impact of this?

In soybean, you have to consider the isotope effects and growth effects. Shanks compares growth rate for different levels of isotope to see no change in growth rate. With CO_2 there is a demonstrated isotope effect and one needs to know that for one's system. An option is adding more labeled substrate (such as Gln). The limitations of adding enough Gln to satisfy computational needs might be swamped out by isotope effects of Gln for growth. That's why Shanks likes to keep levels very low. She looks at growth rate, amino acid composition, lipid fraction, etc. One can't rule out there will be an effect on an enzyme which affects growth; that's why Shanks measures other factors besides growth such as biomass. But she still can't rule out that some enzymes might prefer ^{12}C vs. ^{13}C substrates.

Are there many experimental studies using both NMR and MS together on the same sample?

No, but she has thought about it.

General Discussion and Comments

At the conclusion of the presentations, Mark Segal opened up the discussion to identify important issues raised in the talks and to compile a follow-up list for future interagency Metabolic Engineering announcements. Items on the list are:

1. Increased data for and availability of bioinformatics tools for prediction, particularly important for organisms less commonly studied and where there isn't background data. Is there a way to use current data from other systems to further research in less commonly studied organisms?
2. Integration of metagenomic analysis and multiple transcriptional/translational datasets. Should there be one large database or should there be multiple databases that can communicate for meta-analyses?
3. Multi-institutional consortium to look fully at all the transcription factors in an organism under common growth conditions. Now that we have methods for genome-scale measurements of transcriptional regulatory networks, a multi-institutional, multidisciplinary approach will be useful and probably most effective. Palsson suggests the organism of choice be *E. coli* although he recognizes that DOE may not want to pay for it ($15-$20M annually for 3-4 years). The organism of choice will have to have a good genetic system.
4. Increased knowledge at the basic biochemistry level, to couple with models. This type of biochemical knowledge can also be connected to the biofuel issue to develop more efficient methods of biofuel production.
5. Encouragement of algorithmic development, more than MS and NMR. This type of project would be relatively inexpensive. Studies that combine methods such as MS and

NMR to see how they synergize would also be useful. Jackie Shanks mentioned she could use this type of knowledge.

6. Potential applications of research. There seems to be a lot of information on overall framework but scientists need to know what is important for a specific engineering goal. Thus a more pragmatic approach is needed. For example, on which network(s) should research be focused?
7. Strains for industrial development. Will these types of strains or data on them be easily accessed and available?
8. High throughput processes for estimating kinetic parameters, moving beyond looking at fluxes.
9. High throughput processes for examining metabolic processes in communities (metagenomics). Research needs to be broader instead of focused on a single pathway or organism. Community scale metabolism will have to be examined at a gross level rather than in a lot of detail.

Metabolic Engineering

Seventh
Interagency Conference
on
Metabolic Engineering

Fred Heineken
Chair
Interagency Metabolic Engineering Working Group

February 13, 2007

Metabolic Engineering

Participating Agencies/Departments

- Department of Agriculture
- Department of Commerce
- Department of Defense
- Department of Energy
- Environmental Protection Agency
- National Aeronautics and Space Administration
- National Institutes of Health
- National Science Foundation

Metabolic Engineering

Metabolic Engineering Working Group

Goals

- Promote the Advancement of Metabolic Engineering
- Coordinate Federal Metabolic Engineering Research
- Identify and Address Gaps in Metabolic Engineering Research

Metabolic Engineering

Accomplishments (12 Years)

- Held Three Interagency Workshops on Metabolic Engineering
- Held Seven Interagency Conferences on Metabolic Engineering
- Issued Seven Interagency Announcements of Opportunities in Metabolic Engineering
- Generated an Interagency Web Site (www.metabolicengineering.gov)
- Research Results have been Published in High Impact Journals

Metabolic Engineering

6 Interagency Announcements of Opportunities in Metabolic Engineering

Topic Areas of Interagency Interest

- Instrumentation, Tools, and Methods to facilitate the Study of Metabolic Pathways in Cells
- Quantitative and Conceptual Models
- Bioinformatics
- Engineering of Metabolic Pathways for a Desired Output

Metabolic Engineering

Responses to Announcements

Proposals	236
Awards	54
Award Total	$34M

Metabolic Engineering

Future Plans

- Issue New Interagency Announcements of Opportunities in Metabolic Engineering
- Sponsor Annual Interagency Conferences on Metabolic Engineering
- Maintain an Up-To-Date Interagency Web Site for Metabolic Engineering

Metabolic Engineering

Conference Theme

Metabolic Engineering

Pathway Interactions and Regulatory Processes

Challenges in Modeling for Engineering/Deciphering Regulatory Networks

James C. Liao
Department of Chemical and Biomolecular Engineering
University of California, Los Angeles

Modeling

- What is the purpose of the model?
- What data to use?
- What are the existing/alternative approaches?
- What is the scope of the model?
- Mathematical details?
- What is the basis of prediction?

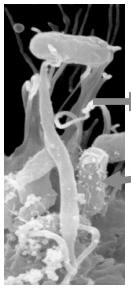
Modeling

- What is the purpose of the model?
  - *To elucidate active Nitric Oxide (NO) response networks in E. coli*
- What data to use?
- What are the existing/alternative approaches?
- What is the scope of the model?
- Mathematical details?
- What is the basis of prediction?

E. coli, Nitric Oxide, and Septic Shock

NO is a mammalian signaling molecule

- vasodilator
- neurotransmitter
- smooth muscle relaxant
- **immune response**



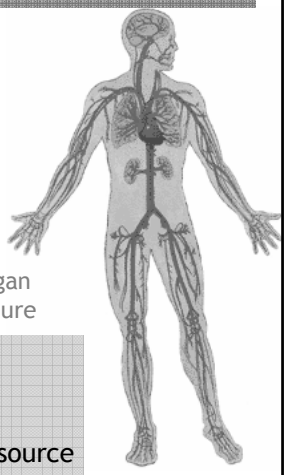
septicshock.com

Septic shock

NO production → vasodilation → hypotension → organ failure

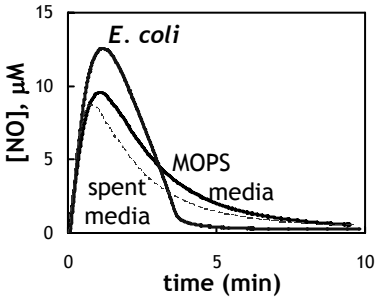
- average patient cost $50,000
- 10th leading cause of death (2000)

- urinary tract (UTI) major infection source
- *E. coli* accounts for ~90% of UTIs

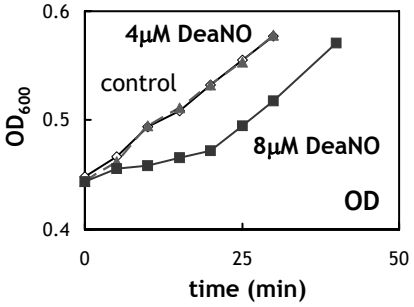


E. coli NO Response

NO Release from 10μM DeaNO



***E. coli* Growth Response**



- NO consumption increases in the presence of *E. coli*

DeaNO inhibits growth for a period longer than the NO lifetime

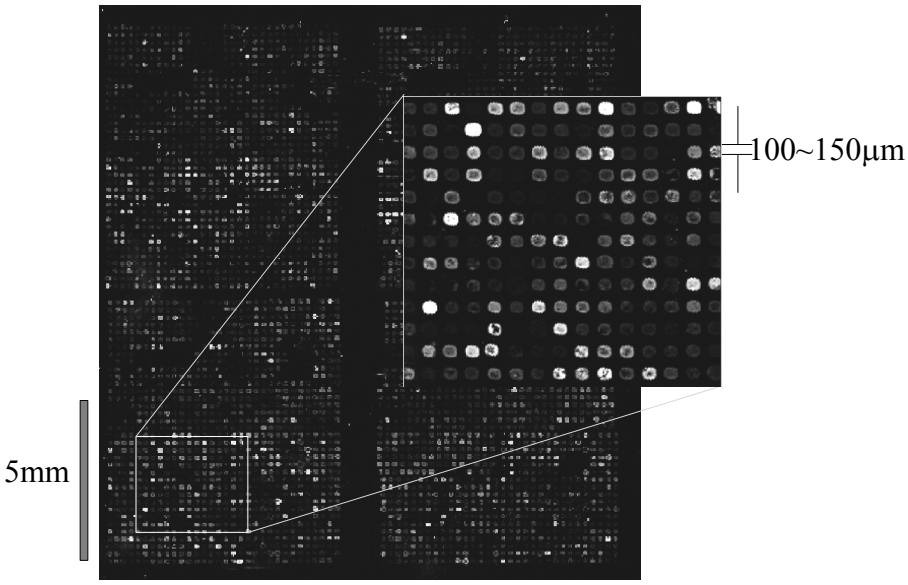
What causes the growth inhibition?

DeaNO = *diethylamine NONOate*
$t_{1/2}$ *(37^{0}C, pH 7.4) = 2 min*

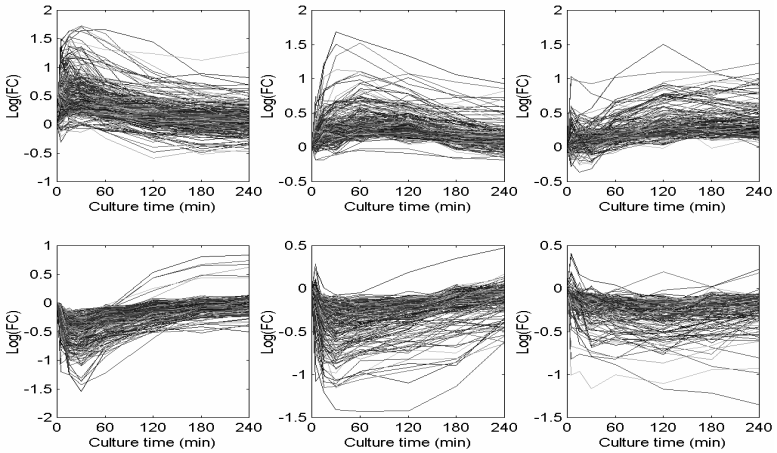
Modeling

- What is the purpose of the model?
  - To elucidate active Nitric Oxide (NO) response networks in E. coli
- What data to use?
  - *Microarray*
- What are the existing/alternative approaches?
- What is the scope of the model?
- Mathematical details?
- What is the basis of prediction?

E. coli whole-genome Microarray



Microarray Data



What do we learn from these data?

Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - *ad hoc analysis, clustering…*
- What is the scope of the model?
- What data to use?
- Mathematical details?
- What is the basis of prediction?

Traditional Microarray Analysis Looks at One Gene at a Time

in WT

Gene Name	Fold Change
carA	-4.76
sdhA	-4.72
sdhC	-4.72
sdhB	-3.91
ndk	-3.72
sdhD	-3.46
argC	-3.03
metF	-3.03
yeeF	-2.90
b0725	-2.89
iscU	3.13
iscS	3.15
iscR	3.93
cydB	4.34
b1797	4.95
cydA	5.48
b4209	16.88

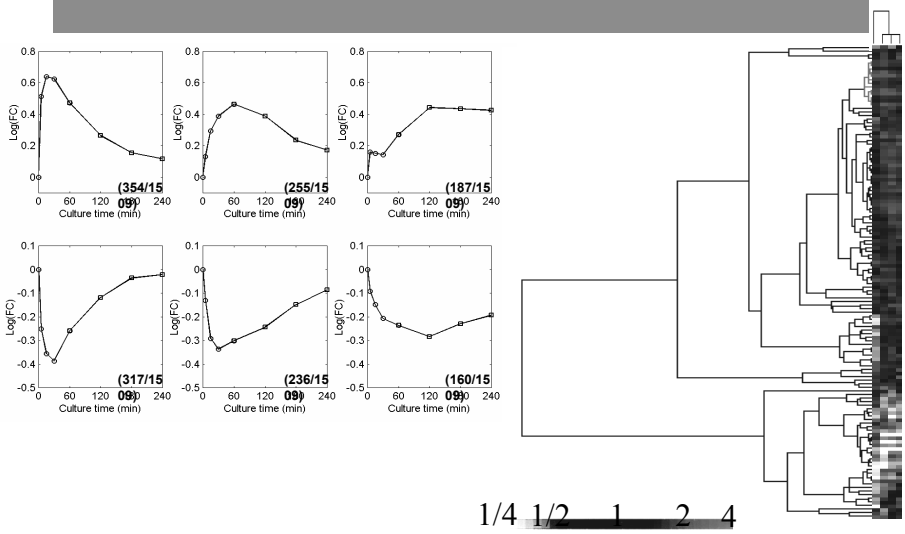
in ΔiscR

Gene Name	Fold Change
b0725	-3.16
sdhA	-3.04
sucB	-2.79
sdhC	-2.69
sdhD	-2.19
nadD	-2.17
sucA	-2.09
fdoI	-1.96
sdhB	-1.90
b2984	-1.86
yjgV	1.79
cydA	1.89
ompT	1.98
cydB	2.57
b1797	3.05
oppA	3.22
norW	5.67

in ΔarcA

Gene Name	Fold Change
carA	-10.22
cvpA	-3.76
metF	-3.28
yeeF	-3.19
purM	-2.96
rnpA	-2.81
ndk	-2.80
purH	-2.71
pyrD	-2.66
codB	-2.52
iscS	3.50
yhiO	3.95
iscR	4.08
ilvC	4.20
b0939	5.62
b1797	5.91
norV	15.69

Cluster Analysis



Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - *mRNA, transcription factors, degradation factors…*
- What data to use?
- Mathematical details?
- What is the basis of prediction?

Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - *mRNA, transcription factors, degradation factors…*
- What data to use?
  - *Microarray transcriptome data*
- Mathematical details?
- What is the basis of prediction?

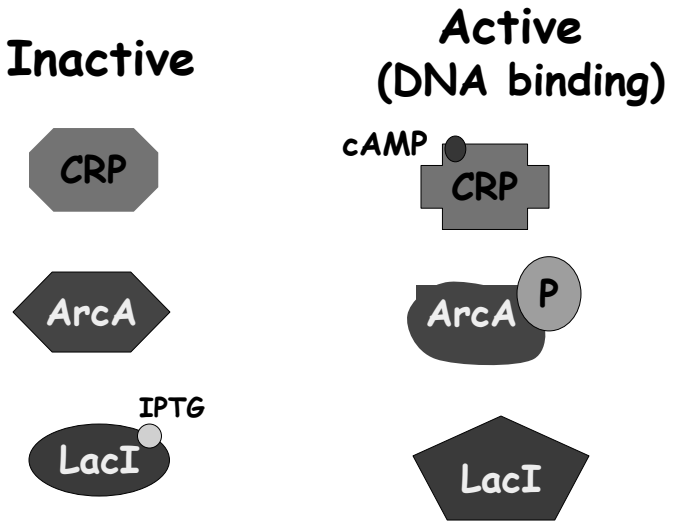
Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

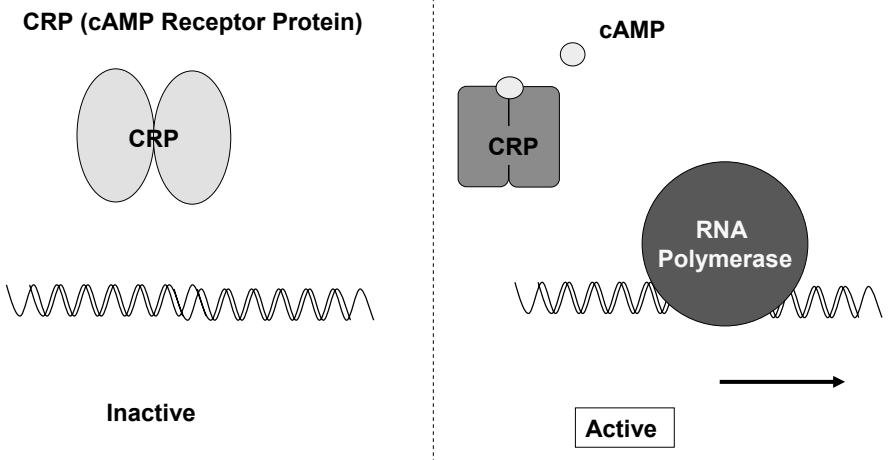
Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

Transcription Factor Activity (TFA)

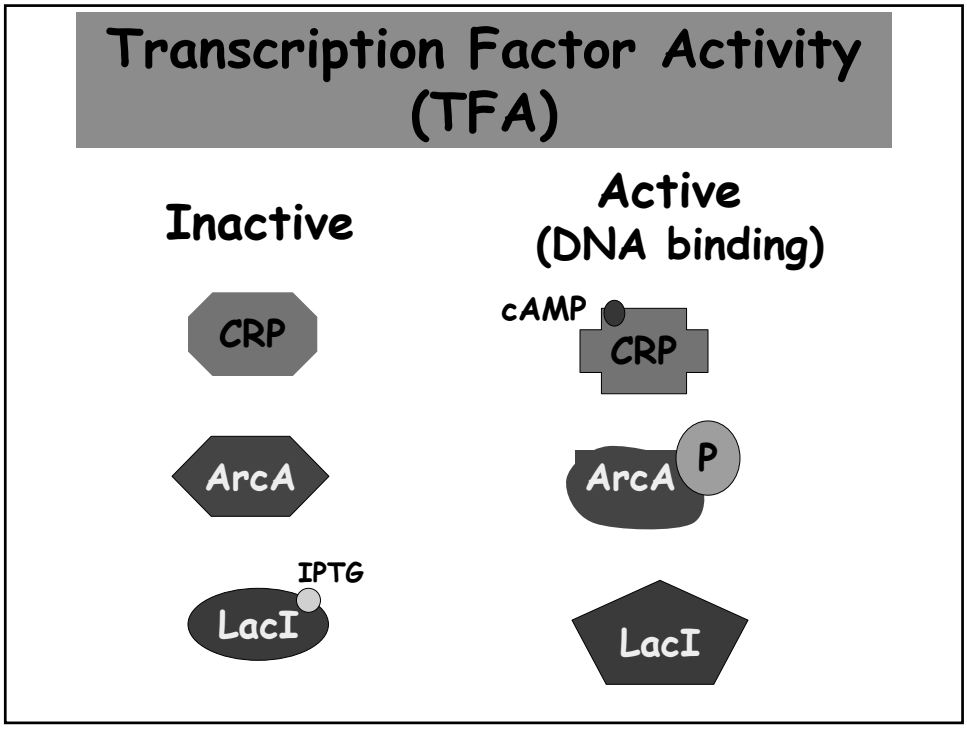


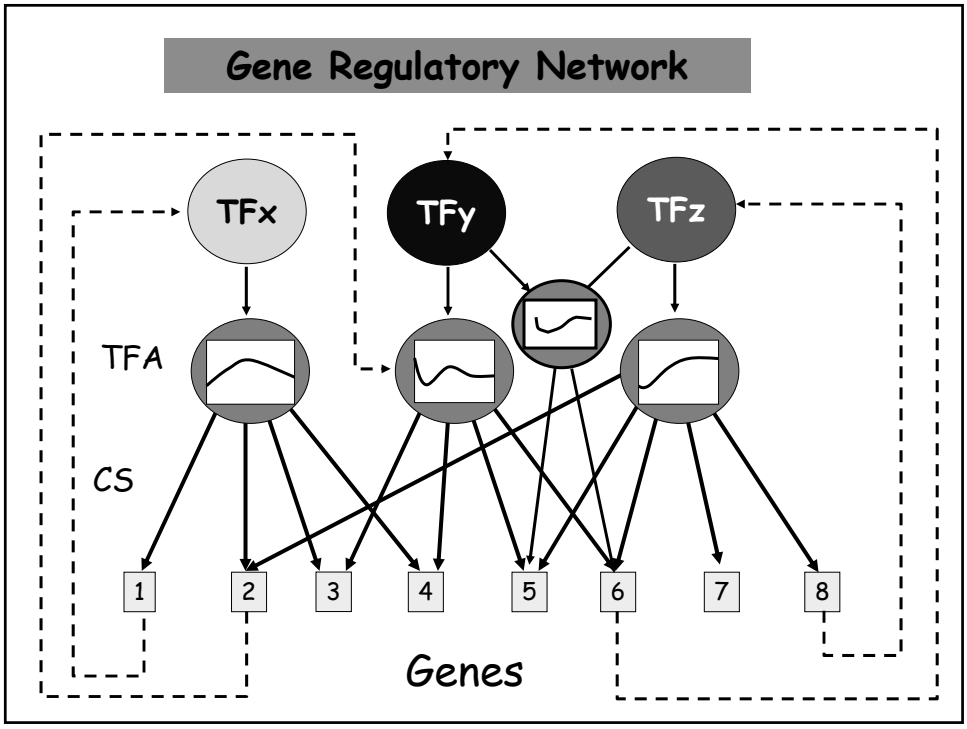
CRP induction mechanism



Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factors bind to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

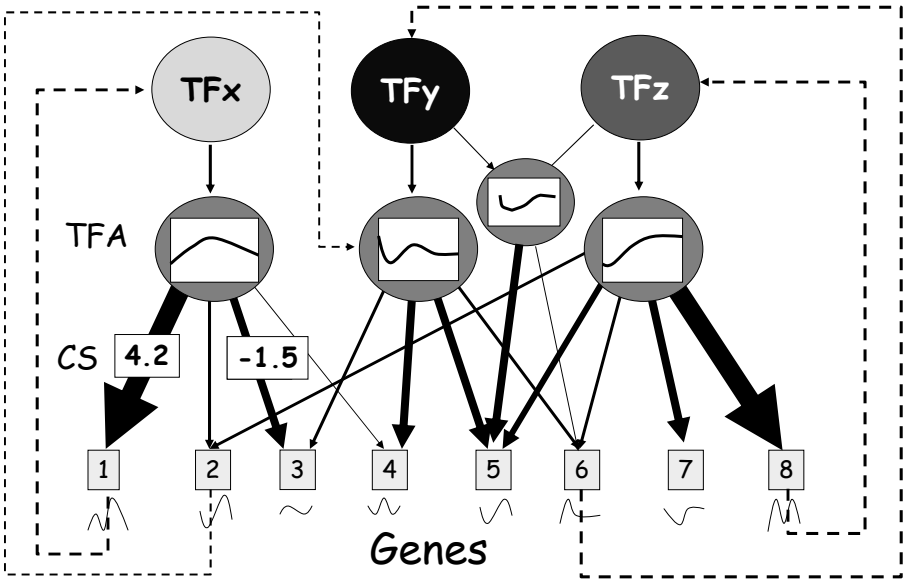

Transcription Factor Activity (TFA)
Inactive
Active (DNA binding)
CRP
cAMP
CRP
ArcA
ArcA
P
IPTG
LacI
LacI


Gene Regulatory Network
TFx
TFy
TFz
TFA
CS
1
2
3
4
5
6
7
8
Genes

Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

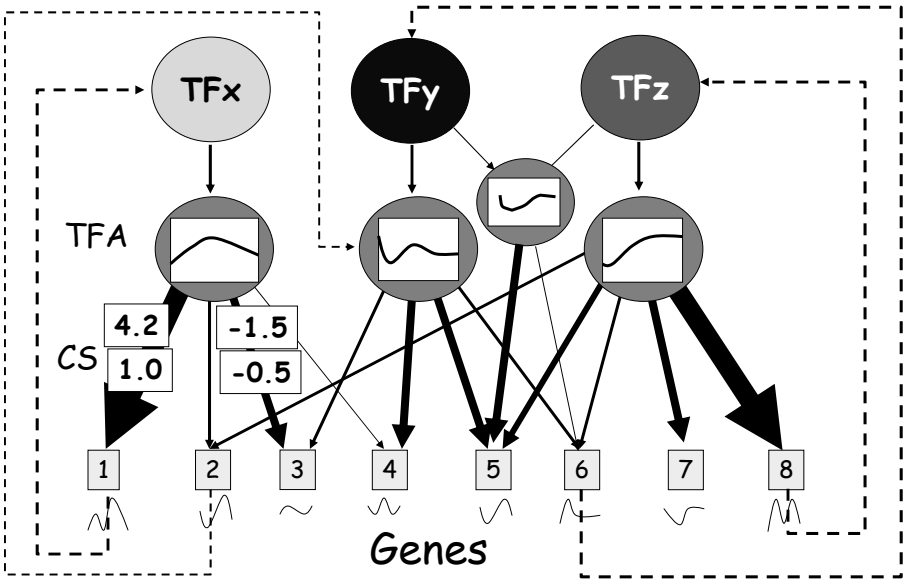
Gene Regulatory Network

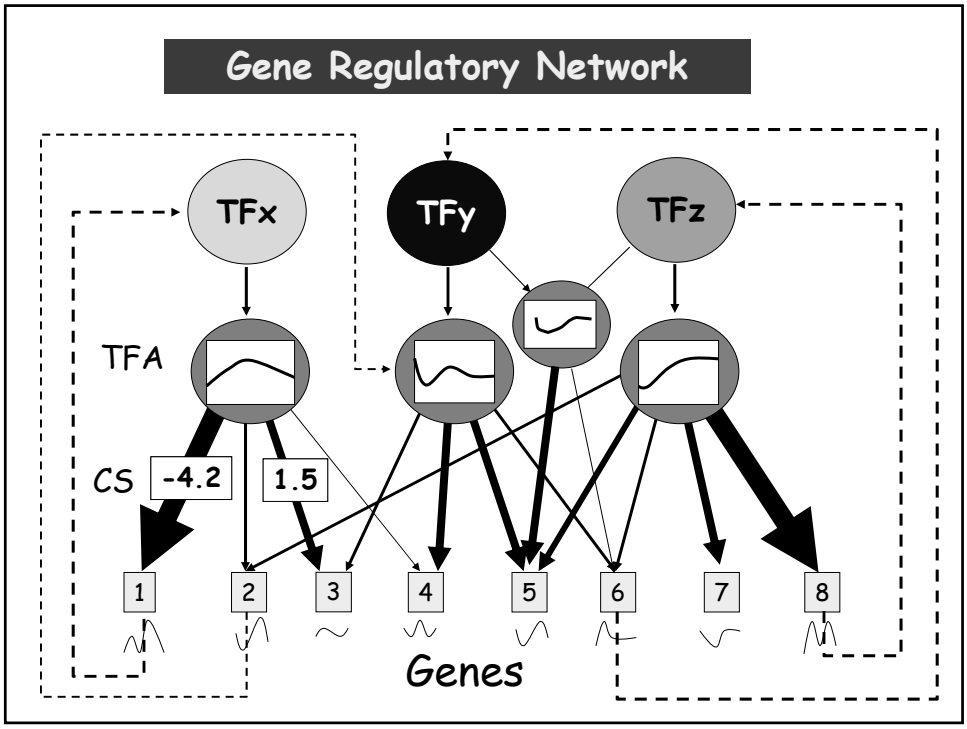


Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

Gene Regulatory Network





Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent
- A power-law model: not perfect for any gene but roughly good for all genes

Network Component Analysis (NCA)

- Formulation of a mathematical model

$$\frac{d[g_i]}{dt} = V_{synthesis} - V_{\deg radation}$$

$$= \prod TFA_j^{\alpha_{ij}} - k_{di}[g_i] \cong 0$$

Quasi-SS assumption

$$\frac{dTFA_j}{dt} = F_j(g(t), m(t), \theta)$$

Determination of Transcription Factor Activity

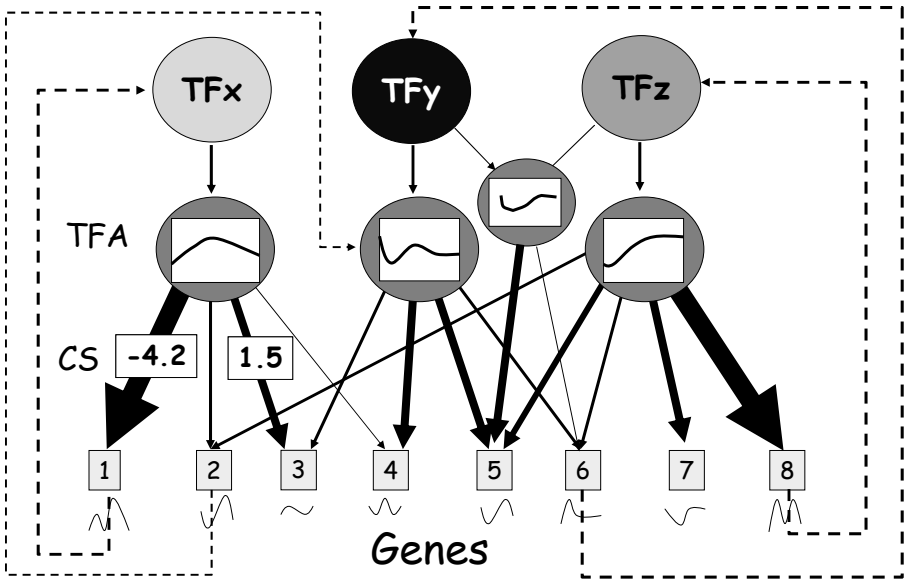
$$\frac{d[mRNA_i]}{dt} = \prod TFA_j^{\alpha_{ij}} - k_{di}[mRNA_i] \cong 0$$

$$\frac{[mRNA_i(t)]}{[mRNA_i(0)]} = \prod \left(\frac{TFA_j(t)}{TFA_j(0)} \right)^{\alpha_{ij}}$$

$$\underbrace{\log\left(\frac{[mRNA_i(t)]}{[mRNA_i(0)]} \right)}_{\text{Microarray data}} = \sum_j \alpha_{ij} \log\left(\frac{TFA_j(t)}{TFA_j(0)} \right) = \underbrace{\sum_j \alpha_{ij} \log\left(\Delta TFA_j(t) \right)}_{\Delta \text{TFA}}$$

[E] = [A] [P]+ Γ

Gene Regulatory Network



A Mathematical problem.

$$[E] = [A]\,[P] + \Gamma$$

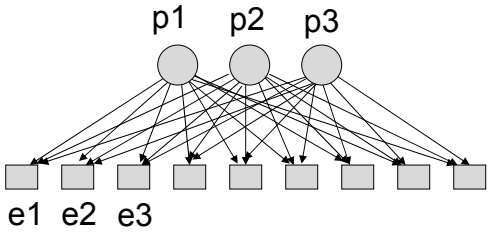
- The matrix decomposition is non-unique.

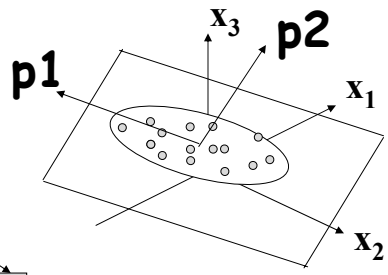
$$[E] = ([A]\,X)\;\;(X^{-1}\,[P]) + \Gamma$$

- Unless it is properly constrained.

Principal Component Analysis (PCA or SVD)

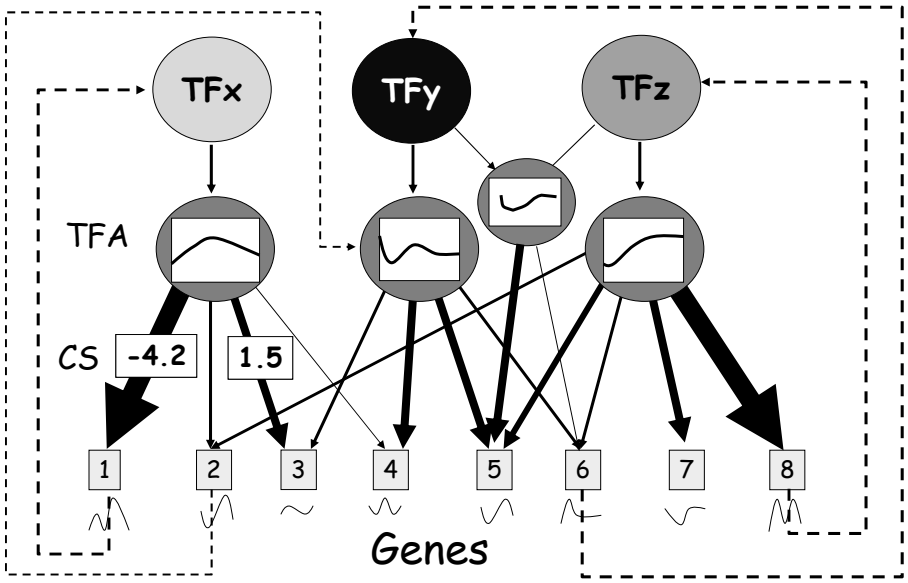
$$[E]_{NxM} = [A]_{NxL}[P]_{LxM}$$





- **Explains major variations in data**
- **Useful for visualization and classification**
- **No mechanistic insight**

Gene Regulatory Network



Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - mRNA, transcription factors, degradation factors…
- What data to use?
  - Microarray transcriptome data
- Mathematical details?
- What is the basis of prediction?

Network connectivity provides constraints

$$[E] = [A]\,[P] + \Gamma$$

Constraints: $a_{ij} = 0$

$$for \quad a_{ij} \in Z_A$$

$$\begin{bmatrix} * & 0 & * & 0 & * \\ 0 & * & 0 & * & 0 \\ 0 & 0 & * & * & 0 \\ * & * & 0 & 0 & * \\ * & 0 & * & 0 & 0 \end{bmatrix}$$

Network Component Analysis (NCA)

$$[E]_{NxM} = [A]_{NxL}[P]_{LxM} \quad \text{s.t.} \quad a_{ij} \in Z_A, \; p_{ij} \in Z_P$$

If

1. Each column of A has at least L-1 zeros
2. The non-zero members of a column should not be a subset of others.
3. The sub-matrix P involved in each gene has full row rank.

then, the decomposition is unique up to a scaling matrix.

NCA forms a bi-linear optimization problem

$$E = AP + \Gamma$$

$$\min_{A,P} \| E - AP \|^2$$

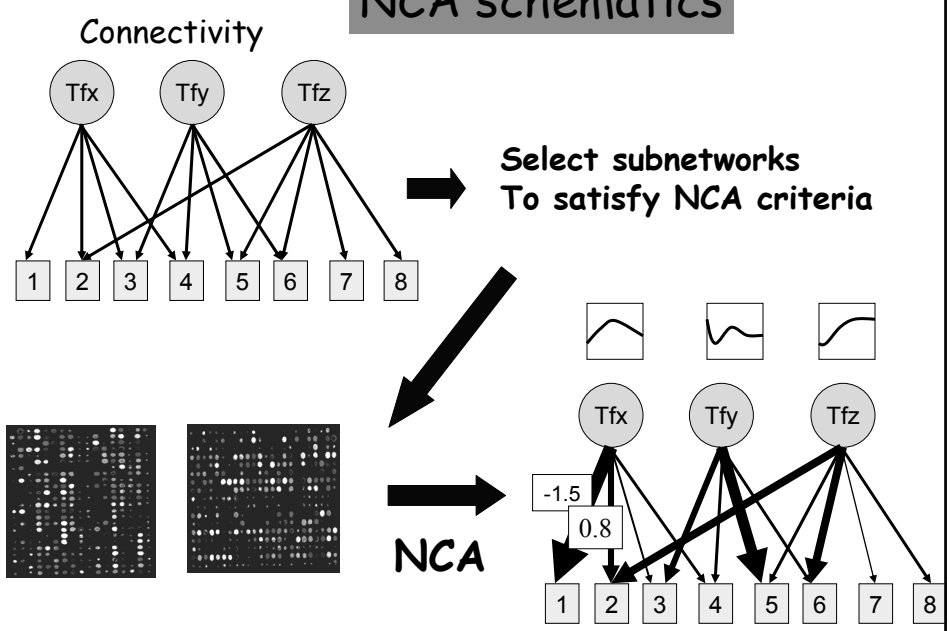
$$\text{s.t. } A \in Z_A \longleftarrow \text{Connectivity constraints}$$

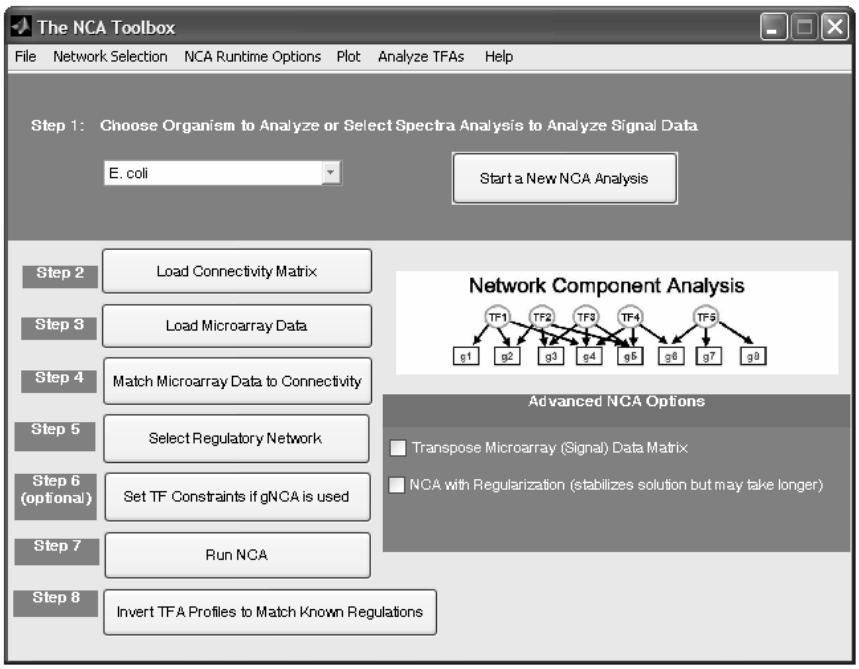
$$P \in Z_P \longleftarrow \text{Regulatory constraints}$$

Network Component Analysis (NCA)

- Biological model
  - Gives biological significance and limitation
- Identifiability criteria
  - Yield mathematical constraints need to be satisfied
- Connectivity theorems
  - Generate insight into connectivity density

NCA schematics



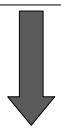


Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - mRNA, transcription factors, degradation factors…
- What data to use?
  - Microarray transcriptome data
- Mathematical details?
- What is the basis of prediction?

Basis for Prediction

- Transcriptome data (mRNA abundance)
- Transscription connectivity
- Power-law model

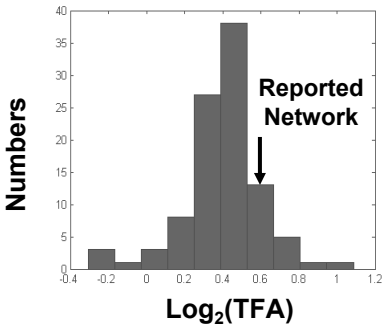


Active transcription factors
Active networks

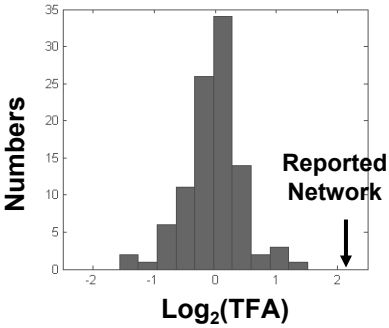
Statistical Significance of TFA

- Constructing null distribution of TFA by *random networks*
  - ✓ Expression data: randomly sampling from genenome
  - ✓ Connectivity: the same with reported network

Null distribution of E2F TFA in Akt-Tg after inducing 12hrs

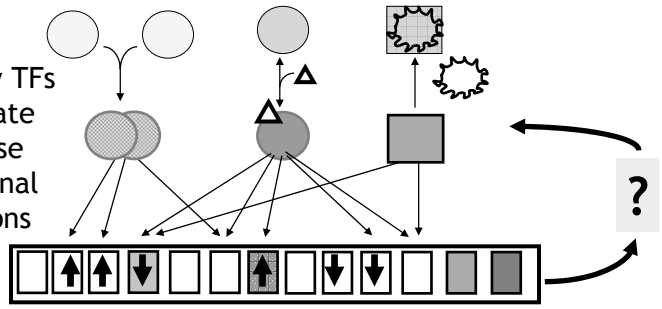


Null distribution of MYBL2 TFA in Pten-null prostate at PIN state



NCA Identifies NO Response Regulators

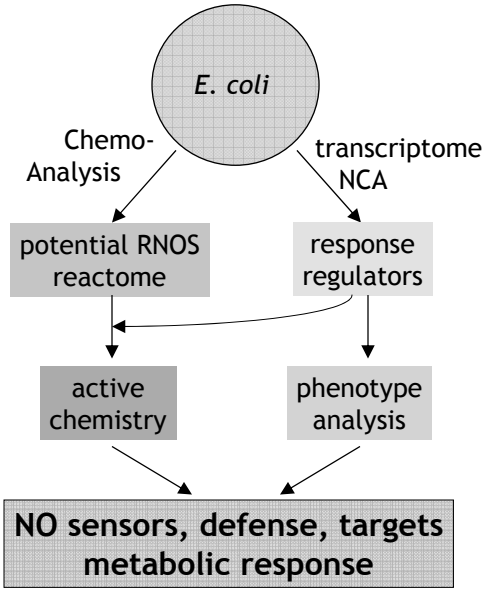
Goal: Identify TFs which mediate NO response transcriptional perturbations

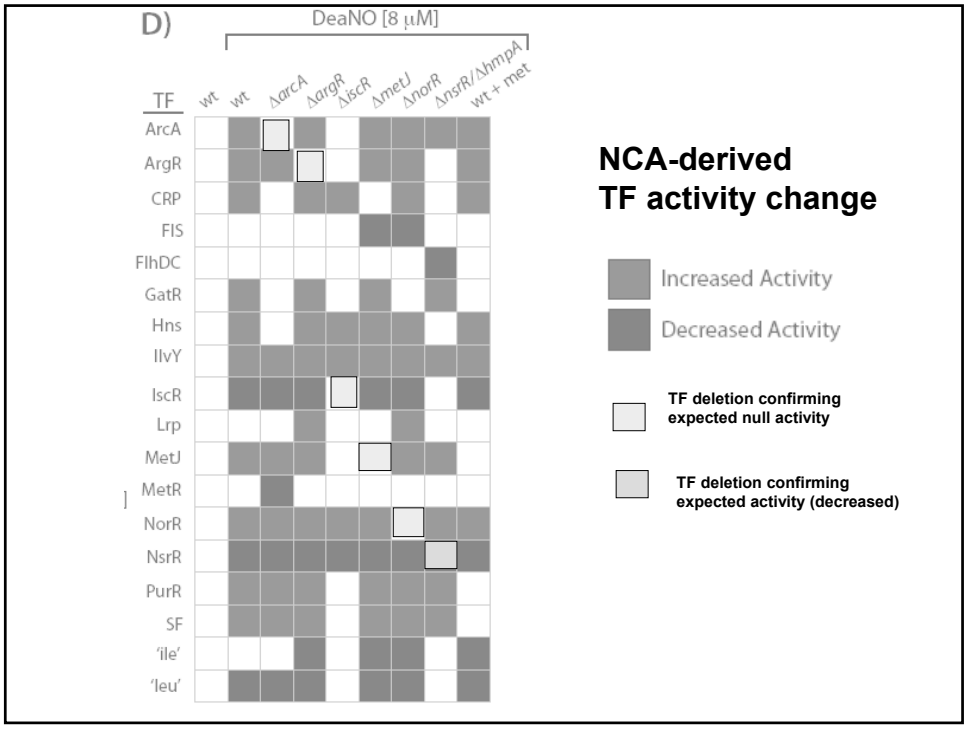


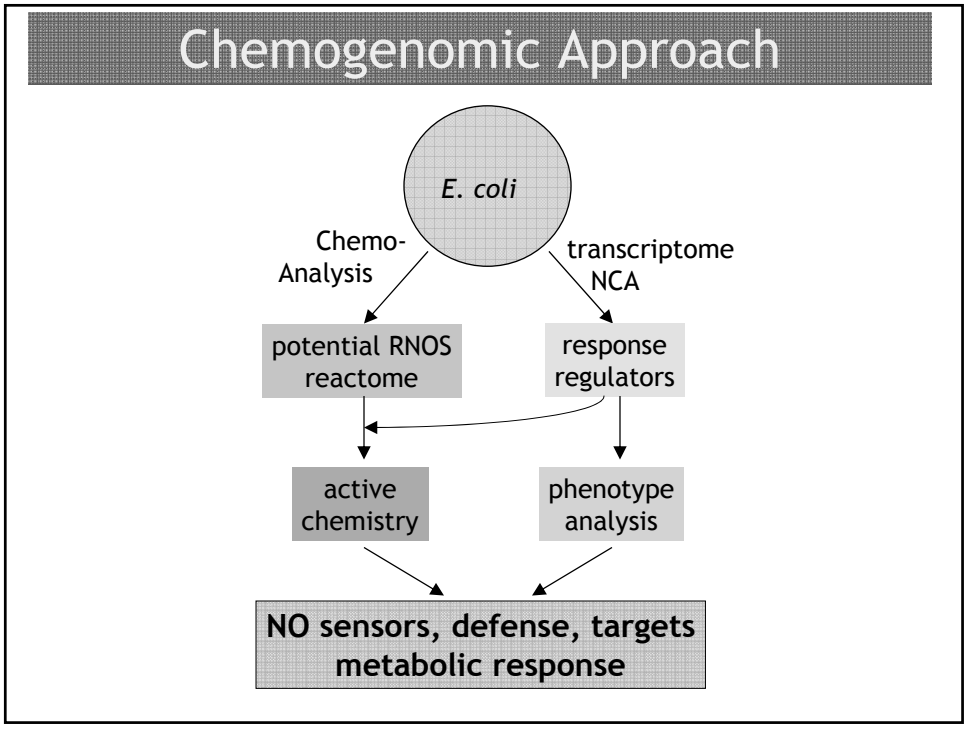
use DNA microarray to measure network behavior

AcrR	Ada	**ArcA**	**ArgR**	ArsR	AsnC	AtoC	BetI	BirA	CadC	Cbl
CpxR	**Crp**	CspA	CsrA	CynR	CysB	DcuR	EmrR	ExuR	FadR	FecI
FhlA	Fis	FlhDC	FliA	Fnr	FruR	Fur	**GatR**	HcaR	HipB	his
HNS	IclR	IdnR	IHF	ile	**IlvY**	**IscR**	KdpE	LexA	**leu**	Lrp
LysR	MalT	MaoB	**MetJ**	MetR	MhpR	Mlc	ModE	NadR	NagC	NarL
NorR	**NsrR**	OxyR	PhoB	phe	**PurR**	RcaSB	RhaSR	Rob	RpiR	RpoE
RpoH	RpoN	RpoS	RtcR	SoxR	trp	TyrR	XapR	XylR	YjbK	**SF**

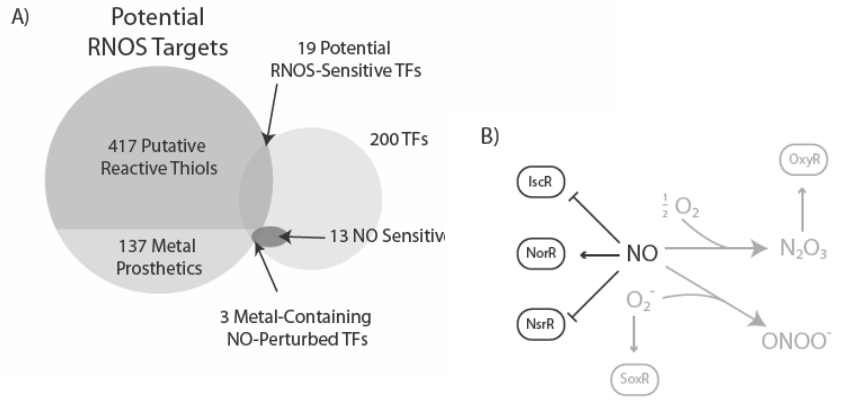
Chemogenomic Approach




D)
DeaNO [8 μM]
TF
wt
wt
ΔarcA
ΔargR
ΔiscR
ΔmetJ
ΔnorR
ΔnsrR/ΔhmpA
wt + met
ArcA
ArgR
CRP
FIS
FlhDC
GatR
Hns
IlvY
IscR
Lrp
MetJ
MetR
NorR
NsrR
PurR
SF
'ile'
'leu'
NCA-derived
TF activity change
Increased Activity
Decreased Activity
TF deletion confirming expected null activity
TF deletion confirming expected activity (decreased)

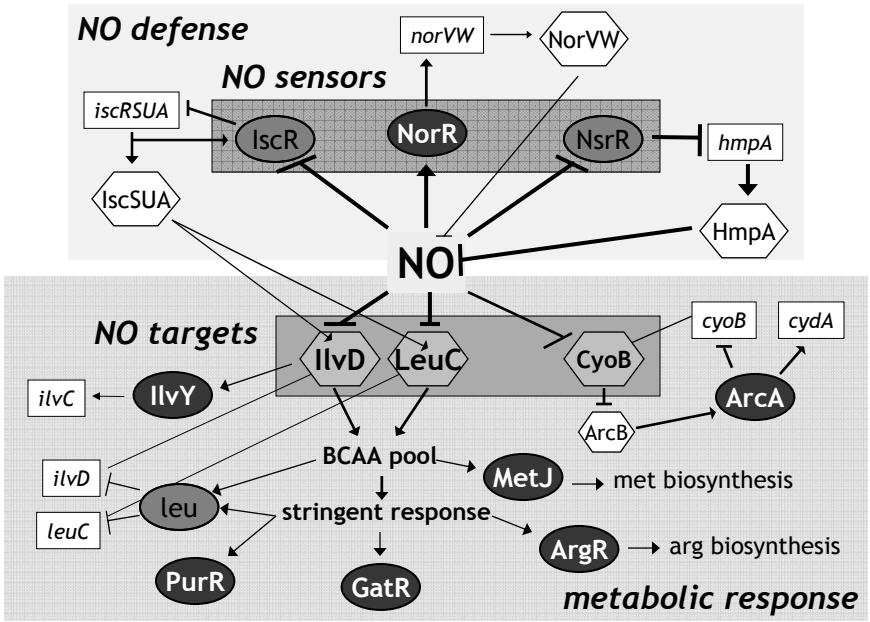

Chemogenomic Approach
E. coli
Chemo-Analysis
transcriptome NCA
potential RNOS reactome
response regulators
active chemistry
phenotype analysis
NO sensors, defense, targets metabolic response

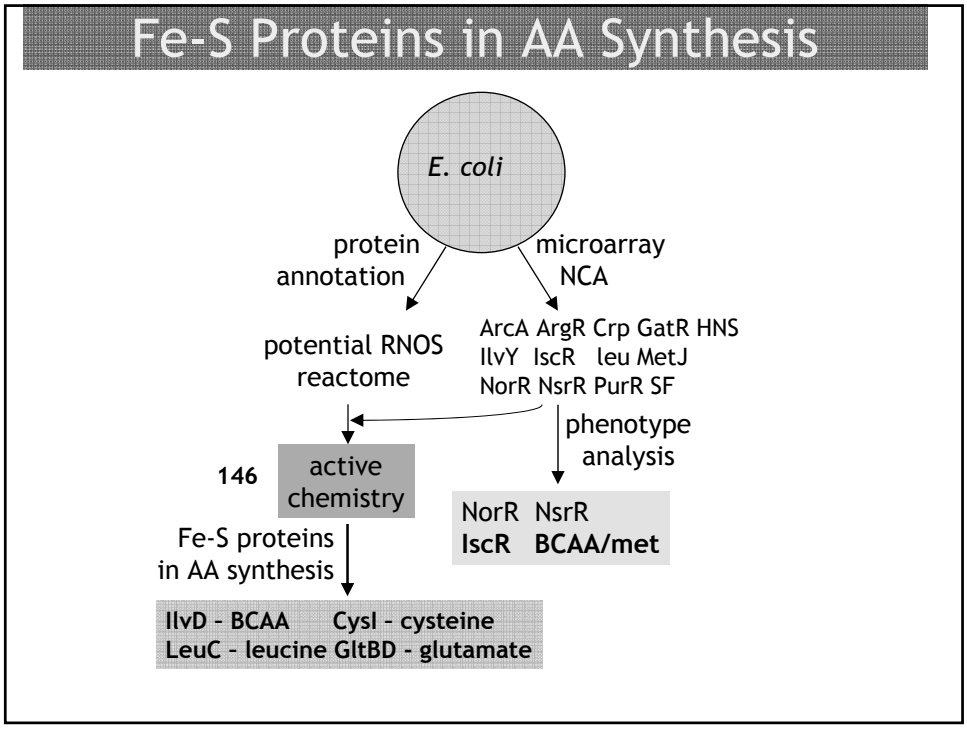
Combination of Active TFs and Reactome Identifies chemistry involved.

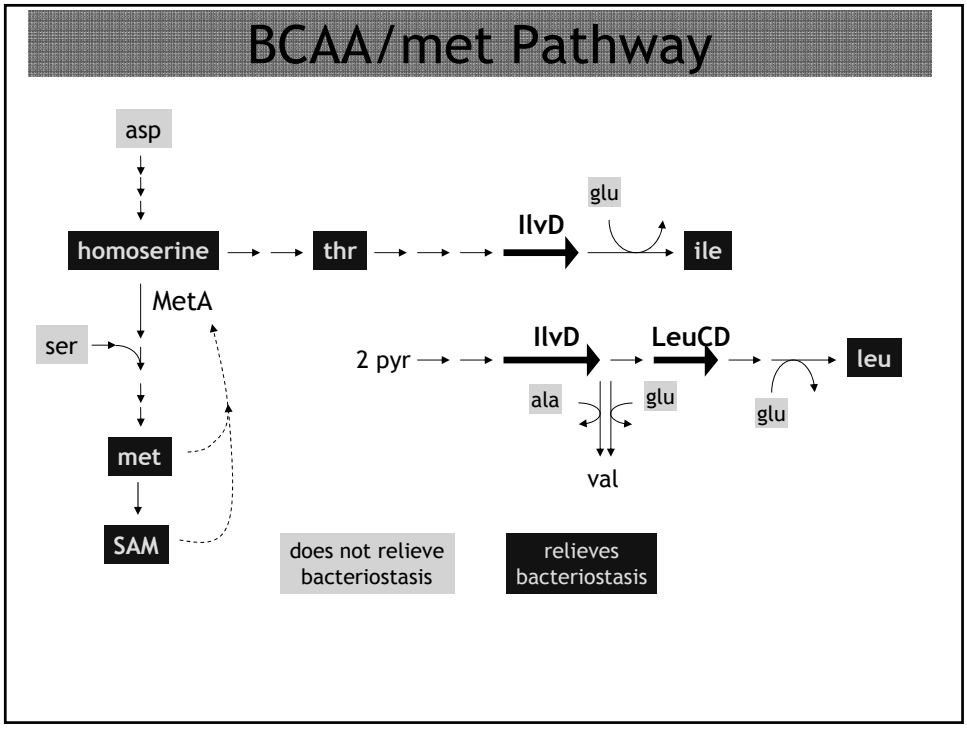


Direct NO-metal interaction is the chemistry involved,

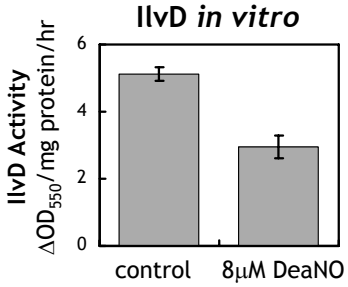
NO Response Network

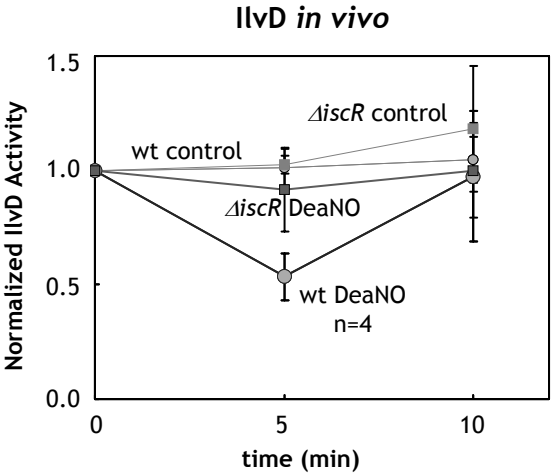


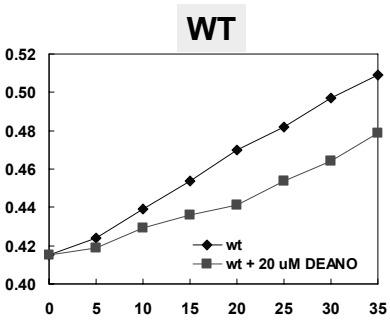

Fe-S Proteins in AA Synthesis
E. coli
protein annotation
microarray NCA
potential RNOS reactome
ArcA ArgR Crp GatR HNS
IlvY IscR leu MetJ
NorR NsrR PurR SF
phenotype analysis
146
active chemistry
NorR NsrR
IscR BCAA/met
Fe-S proteins in AA synthesis
IlvD - BCAA CysI - cysteine
LeuC - leucine GltBD - glutamate

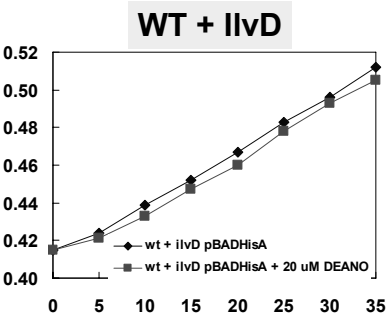

BCAA/met Pathway
asp
homoserine
thr
IlvD
glu
ile
MetA
ser
met
SAM
2 pyr
IlvD
LeuCD
leu
ala
glu
glu
val
does not relieve bacteriostasis
relieves bacteriostasis

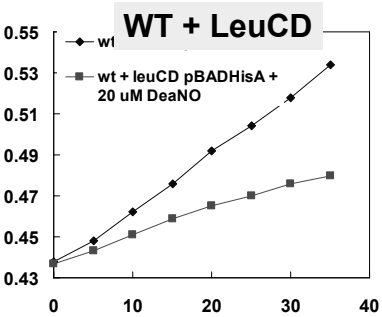
IlvD is NO Targets

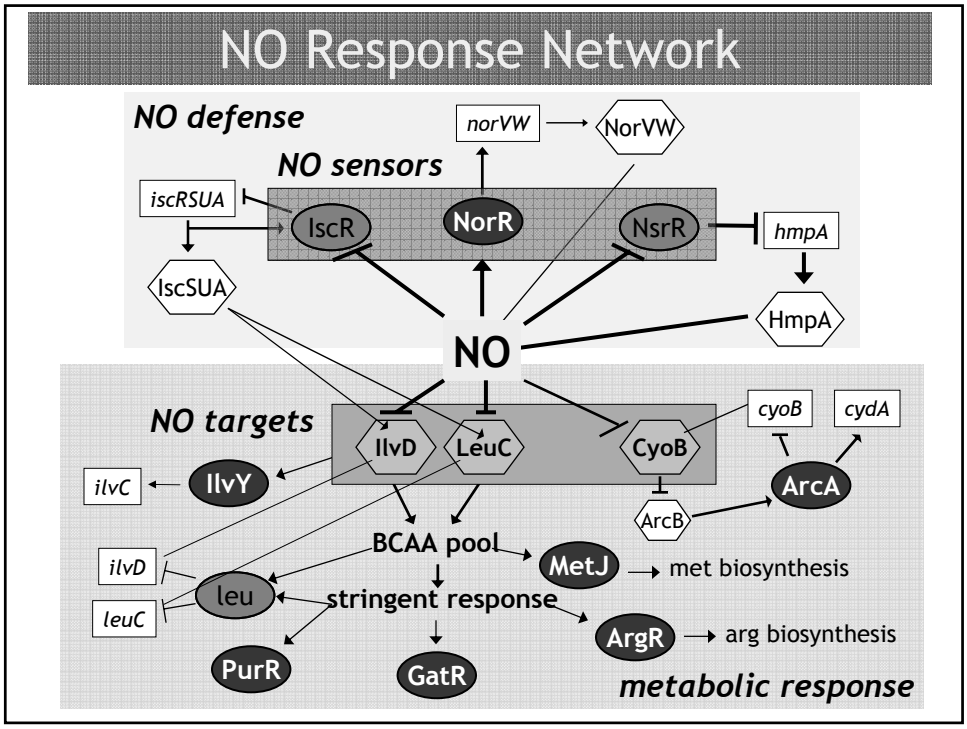











NO Response Network
NO defense
NO sensors
norVW
NorVW
iscRSUA
IscR
NorR
NsrR
hmpA
IscSUA
HmpA
NO
NO targets
IlvD
LeuC
CyoB
cyoB
cydA
ilvC
IlvY
ArcA
ArcB
BCAA pool
MetJ
met biosynthesis
ilvD
leu
leuC
stringent response
ArgR
arg biosynthesis
PurR
GatR
metabolic response

Acknowledgment
Young Lyeol Yang
Linh Tran
Katy Kao
Laura Jarboe
Matt Rizk
Mike Connor
Jeff Bernstein
Thomas Bulter
Anthony Cann
Ganesh Sriram
Katherine Chu
Mark Brynildsen
Simon Galbraith
Jason Suen
Wilson Wong
Eileen Fung
Jason Dean
NIH, DOE, NASA, NSF

The Challenge of Incorporating Regulatory Effect in Genome-scale Networks

Bernhard Palsson
Bioengineering, UCSD
GtL meeting, Feb 07

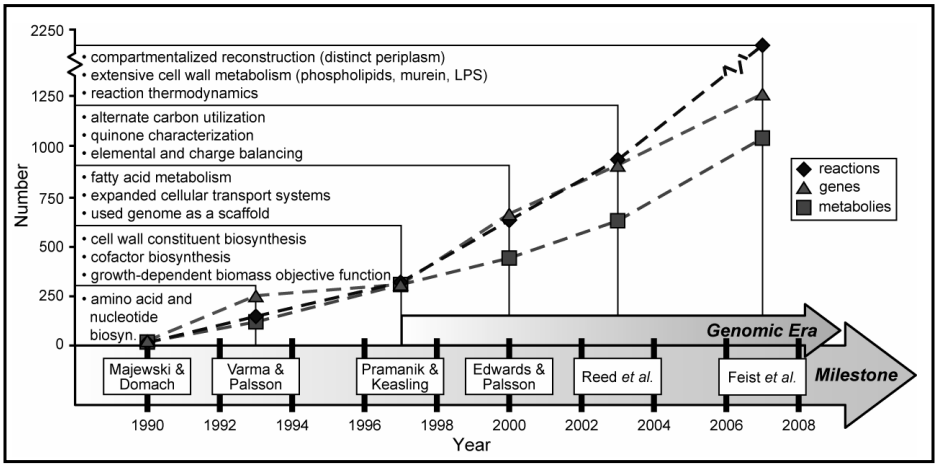
University of California, San Diego
Department of Bioengineering
Systems Biology Research Group
http://systemsbiology.ucsd.edu

Outline

- History of metabolic reconstruction
- Uses of metabolic reconstructions
- Regulation as a constraint-based process
- Measuring TRN components on a g/s
- TRN reconstruction
- Integrated analysis of regulation and metabolism

University of California, San Diego
Department of Bioengineering
Systems Biology Research Group
http://systemsbiology.ucsd.edu

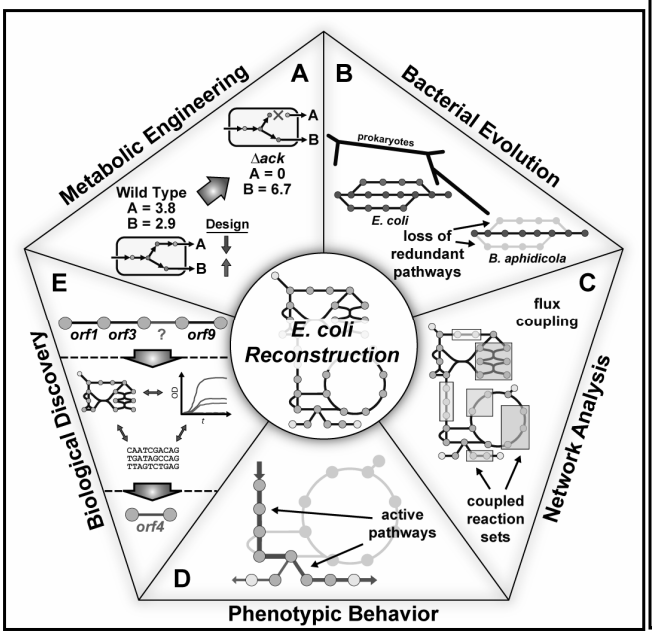
Reconstruction of *E. coli* Metabolism



References:

R.A. Majewski and M.M. Domach, *Biotechnol Bioeng* **35**, 732 (1990)
A. Varma, B.W. Boesch, and B.O. Palsson, *Appl Environ Microbiol* **59** (8), 2465 (1993) & *Biotechnol Bioeng* **42** (1), 59 (1993)
J. Pramanik and J.D. Keasling, *Biotechnol Bioeng* **56** (4), 398 (1997) & **60** (2), 230 (1998)
J.S. Edwards and B.O. Palsson, *Proc Natl Acad Sci U S A.* **97** (10), 5528 (2000)
J.L. Reed, T.D. Vo, C.H. Schilling et al., *Genome Biology* **4** (9), R54.1 (2003)
A.M. Feist, C.S. Henry, J.L. Reed et al., *Under Review* (2007)

Uses of the *E. coli* Reconstruction



Metabolic Engineering:
1. *Biotechnol Bioeng* **84**, 647 (2003)
2. *Biotechnol Bioeng* **84**, 887 (2003)
3. *Genome Res* **14**, 2367 (2004)
4. *Metab Eng* **7**, 155 (2005)
5. *Nat Biotechnol* **23**, 612 (2005)
6. *Appl Environ Microbiol* **71**, 7880 (2005)
7. *Metab Eng* **8**, 1 (2006)
8. *Appl Microbiol Biotechnol* **V73**, 887 (2006)
9. *Biotechnol Bioeng* **91**, 643 (2005)

Bacterial Evolution:
10. *Bioinformatics* **21 Suppl 2**, ii222 (2005)
11. *Nat Genet* **37**, 1372 (2005)
12. *Nature* **440**, 667 (2006)

Network Analysis:
13. *Bioinformatics* **19**, 1027 (2003)
14. *Genome Res* **14**, 301 (2004)
15. *Nature* **427**, 839 (2004)
16. *Biophys J* **88**, 37 (2005)
17. *PLoS Comput Biol* **1**, e68 (2005)
18. *BMC Bioinformatics* **7**, 111 (2006)
19. *Proc Natl Acad Sci U S A* 102, 19103 (2005)

Phenotypic Behavior:
20. *Mol Syst Biol*,Under Review (2007)
21. *Biophys J* 83, 79 (2002)
22. *Proc Natl Acad Sci U S A* **99**, 15112 (2002)
23. *Bioinformatics* **21**, 2008 (2005)
24. *Proc Natl Acad Sci U S A* **102**, 7695 (2005)
25. *J theor Biol* **237**, 401 (2005)
26. *Biophys J* **90**, 1453 (2006)
27. *BMC Bioinformatics* **7**, 118 (2006)
28. *Mol Syst Biol* **2**, 2006 0034 (2006)
29. *Biophys J* **91**, 2304 (2006)
30. *BMC Bioinformatics* **7**, 512 (2006)
31. *Biophys J* **92**, 1792 (2007)
32. *Nature Genetics* **36**, 1056 (2004)
33. *Nat Biotechnol* **19**, 125 (2001)
34. *Nature* **420**, 186 (2002)

Biological Discovery:
35. *Biophys J* **90**, 2659 (2006)
36. *Genome Biology* **7**, R17 (2006)
37. *BMC Bioinformatics* **7**, 177 (2006)
38. *Proc Natl Acad Sci U S A* **103**, 17480 (2006)
39. *Nature* **429**, 92 (2004)
40. *PLoS Computational Biology* **2**, e72 (2006)

#1 Pathways operate as elements of a network

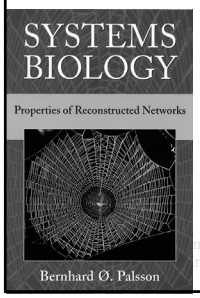
We can now describe metabolism at a genome level

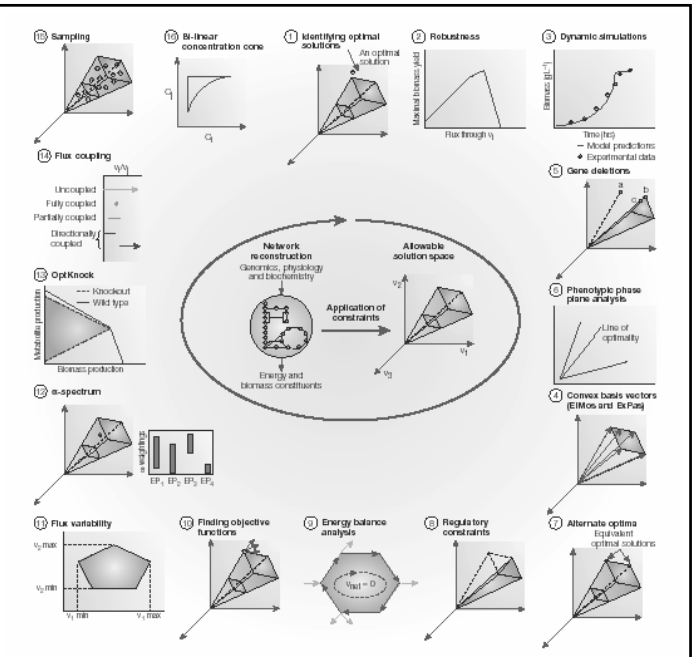
University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Genome-scale constraint-based modeling:

a rapidly growing





UCSD Extension

Price, et al Nat Rev Microbiol. Nov 2004

Top-down view of regulation

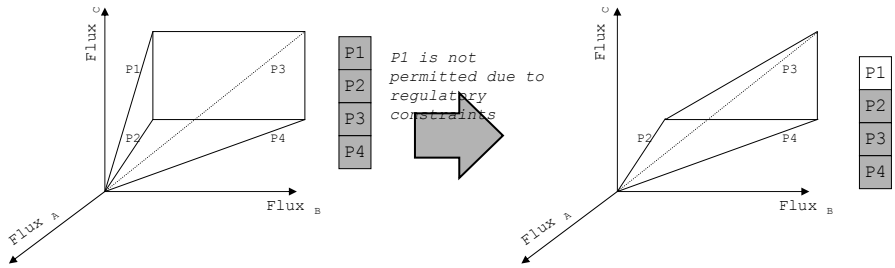
#2: Regulation of expression: shrinking solution space

- Need TRN reconstructed -- now

#3: Regulation of activity: location within a shrunken solution space

- Sampling -- future

Extreme Pathways and Regulatory Constraints



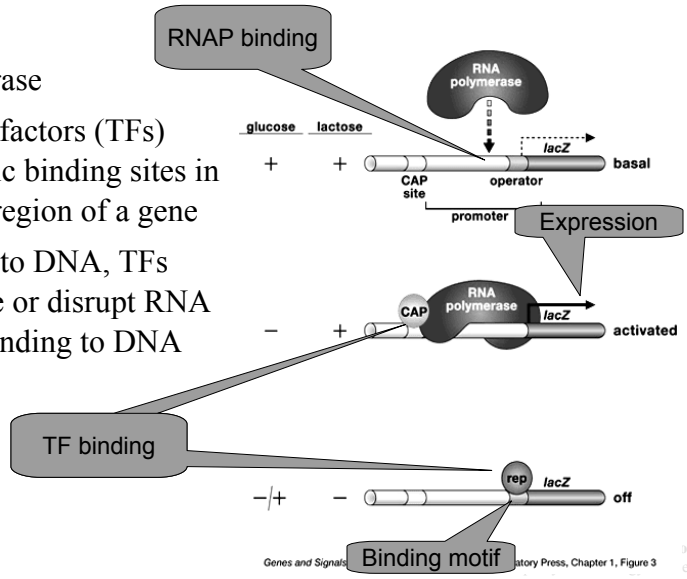
Consider the entire solution space of a metabolic network, bounded by extreme pathways P1-P4…

One or more of these pathways may not be feasible, depending on the environment and corresponding regulatory effects…

This leads to a reduced solution space bounded by fewer extreme pathways

Transcription: RNAP & TFs

- RNA polymerase
- Transcription factors (TFs) bind to specific binding sites in the promoter region of a gene
- After binding to DNA, TFs either enhance or disrupt RNA polymerase binding to DNA



Drill-down Studies on *E. coli* Transcriptional Sub-networks

Condition	Regulator(s)	Target Functionality	Target Locus
Nitrate	NarL/NarP	Glycolytic Pathways	tpi, pfkA, pflAB, focA, mdhA, aceAB, aceEF, pykF, eno, pgmA
Nitrate	NarL/NarP	Hydroxylamine Reductase	ybjW
Nitrate	NarL/NarP	Methylglyoxal Synthesis	mgsA
Nitrate	NarL/NarP	Molybdopterin Biosynthesis	moaABCDE, mogAB, moaAB
Nitrate	NarL/NarP	O6 and O4-methylguanine-DNA methyltransferase	ogt
Nitrate	NarL/NarP	Threonine Degradation	kbl, tdh
Oxygen	ArcA/Fnr	Acetate Metabolism	ackA, acnAB, acs, gltA, mdhA, pta
Oxygen	ArcA/Fnr	Fatty Acid Metabolism	fadL, fadD, fadE, fadBA, fadIJ, atoSBDA
Oxygen	ArcA	Galactose Transport	mglABC
Oxygen	ArcA/Fnr	Glycerol Metabolism	glcA, ugpABCEQ, glpABC, glpD
Oxygen	ArcA	Glycerol Transport	ugpBAECQ
Oxygen	Fnr	Glycolytic Pathways	pykA, pgmA
Oxygen	ArcA/Fnr	Methionine Degradation	yeaA
Oxygen	ArcA	Nucleotide Metabolism	nudE
Oxygen	Fnr	Peptide Degradation	pepE
Oxygen	ArcA	Potassium Transport	kefB
Oxygen	ArcA/Fnr	α-Ketoglutarate Transport	kgtP
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Acetohydroxybutanoate Synthesis	ilvN, ilvB
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Alternative DMSO Respiratory Pathway	ynfEFGH
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Aspartate-Asparagine Interconversion	aspA
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Betaine Biosynthesis	betABI, proP
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Curli (Amyloid) Synthesis	csgABDEFG, crl
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Folic Acid Biosynthesis	folE
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Glucose Degradation	gcd
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Nucleotide Metabolism	udhA
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Oligopeptide Transport	oppABCDF, cstA
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Stationary Phase/ Stress Response	rpoS, crl, yjeE
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Superoxide Response	soxS, rsxABCDGE
Oxygen/Nitrate	ArcA/Fnr, NarL/NarP	Thiamin Biosynthesis	thiC

#4: can get the necessary data for g/s reconstruction of TRN

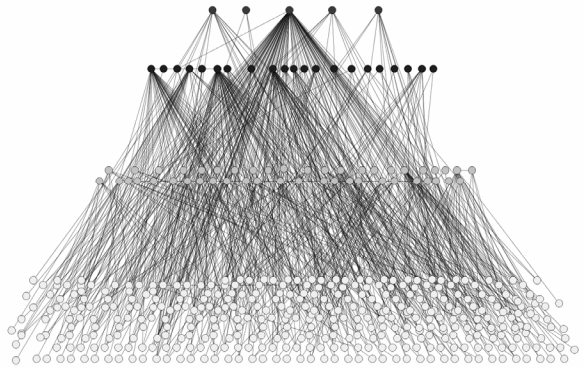
MC1010 – Hierarchy of regulator

global regulators

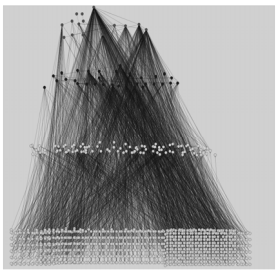
major regulators

minor regulators

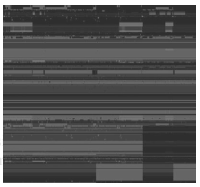
target genes



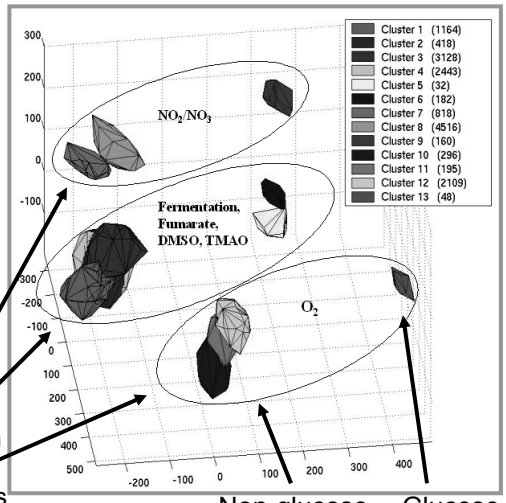
Integrating metabolism and TRN



- Computed Metabolic & Regulatory State in **15,580** distinct minimal growth media



The electron acceptor distinguishes clusters



Non-glucose media

Glucose media

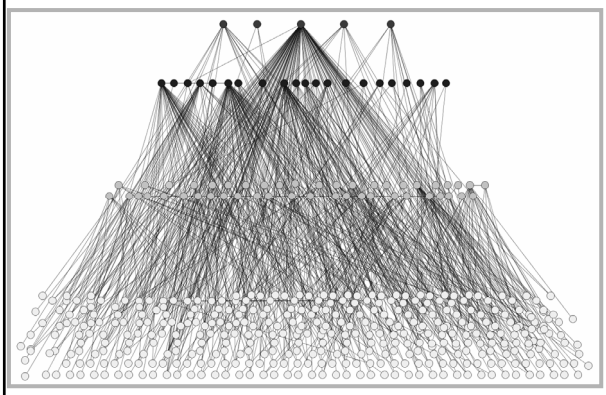
Barrett et al., *PNAS*, 102, 19103-8 (2005)

R-Matrix Formalism

- quasi-stoichiometric formalism
- a structured and self-contained representation of Transcriptional Regulatory Network (TRN)
- can be quantitatively interrogated relying on the principles of the constraint-based modeling approach

Where is this going (short-term)?

E. coli Regulatory Network

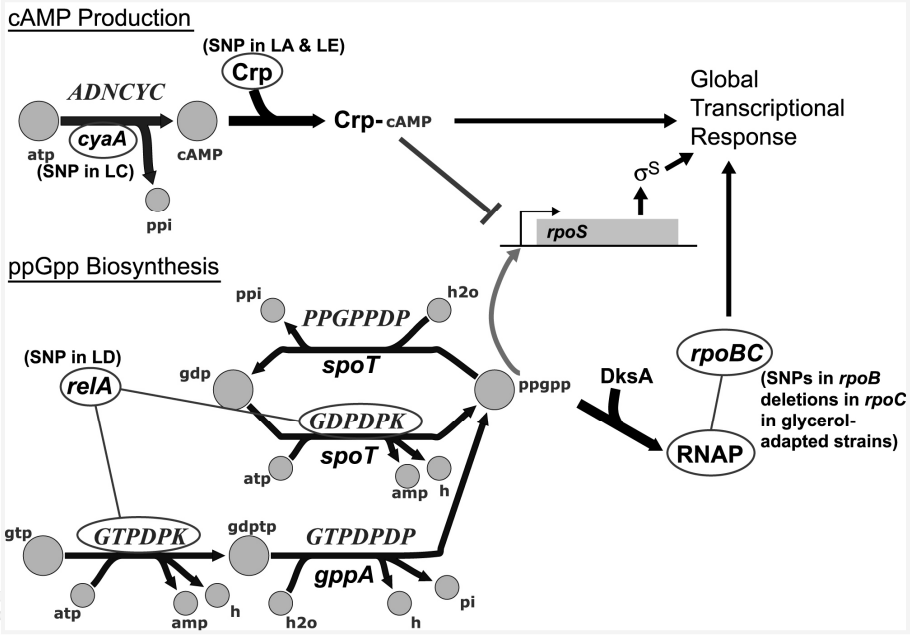


Reconstruction:

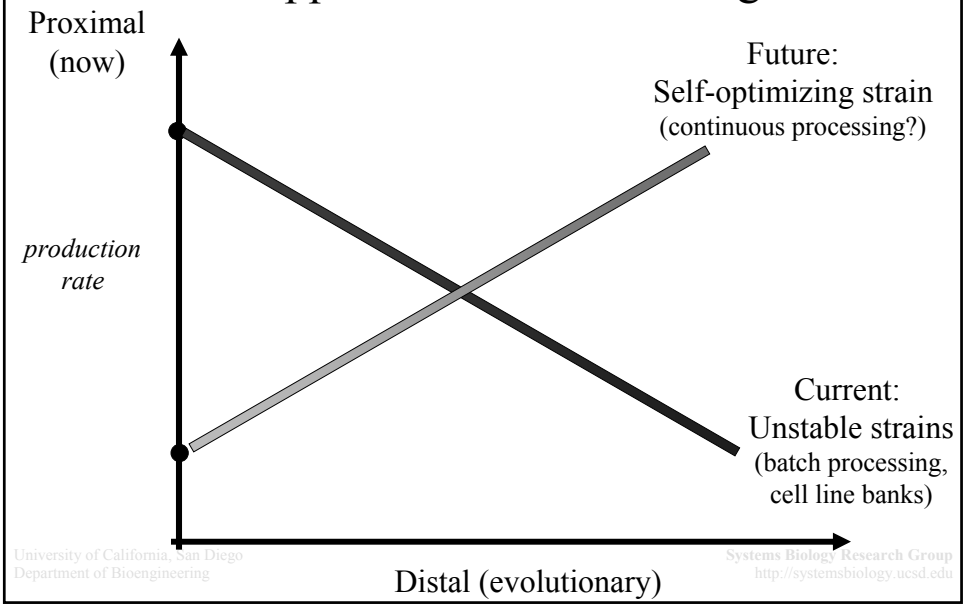
- 640 regulated genes
- 95 metabolites, 101 transcription factors, 26 reaction fluxes, 14 specific environmental factors
- 3 levels of transcription

#5: mathematical formalisms are developing that integrate metabolic and TRN at the g/s

Global Regulatory Response



Causation in Biology: new approaches to strain design



University of California, San Diego
Department of Bioengineering
Systems Biology Research Group
http://systemsbiology.ucsd.edu

#6: we can re-sequence to study the genetic basis for TRN adaptation at the g/s

Conclusions/issues

- Metabolic reconstruction methods established
- Conceptual framework fro TRNs function at g/s in place
- Technology now available for direct measurement of TRN events on g/s
- 180 putative TF in E. coli -- expensive
- In silico methods to assemble, analyze the data are developing
- Integration with metabolism possible
- The genetic basis for adaptation can now be monitored

Biological Aspects of Deciphering and Engineering Regulatory Networks

George N. Bennett

Department of Biochemistry and Cell Biology

Rice University
Houston, Texas

Metabolic Engineering Workshop
February 2007

Hans V. Westerhoff and Boris N. Kholodenko:
Metabolic Engineering in the Post Genomic Era

- **Consider the cell as a production organism rather than as a production line.**
- **The organism may largely do away with much of the engineering by invoking its homeostatic control mechanisms**
- **Metabolic engineering should be directed at optimizing both the production flux and the functioning of the organism itself**

Challenges in engineering networks of known genes & regulators

- **Obtaining high levels of expression of desired genes**
- **Obtaining appropriate balance in levels of expression of pathway genes**
- **Consider regulation at enzyme level and feedback connections**
- **Consider the alternation of global cell physiology from the manipulation**

Obtaining high levels of expression of desired genes

- Active promoters
- mRNA stability & translational efficiency
- Functionally active protein

Obtaining appropriate balance in levels of expression of pathway genes

- Library of promoters of varying strength
- Combinations of promoters & terminators, RBS for multicistronic mRNAs
- RNA regulators (antisense, RNA binding proteins)
- Plasmid vs chromosomal location

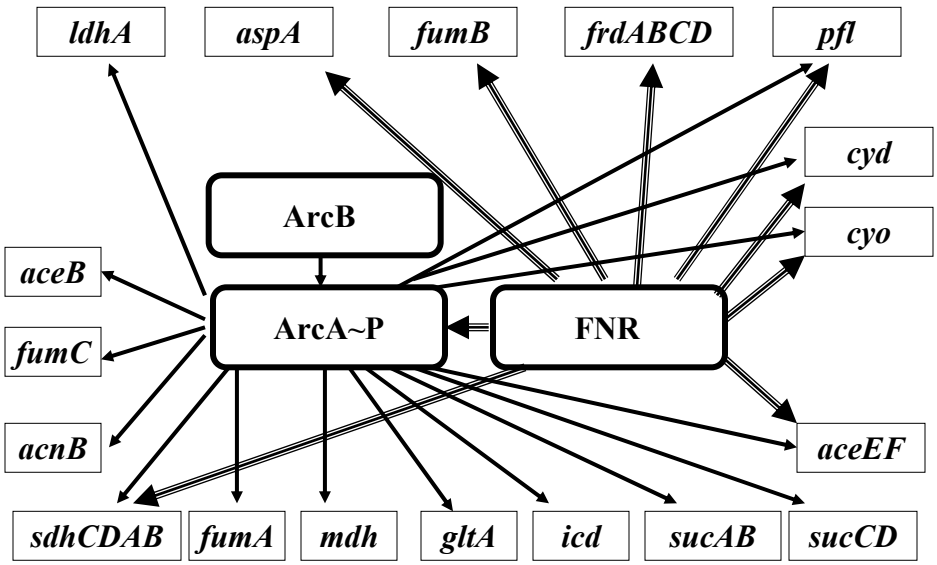
Taking into account the regulation at enzyme level and feedback connections

- **Interconnections among regulators and small molecules**
- **Picking the best enzyme from various sources or evolve such attributes**

We may know a great deal about individual components but the overall metabolic consequences of various alterations under defined conditions may not be so easily predicted

- Need detailed experiments to define ***which regulatory interactions are most important in a situation***
- Functional activity of a protein depends on more than gene expression

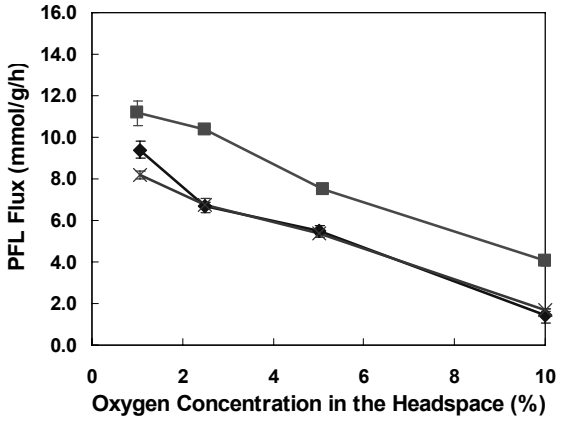
A quick example from aerobic/anaerobic E. coli



red denotes repress and **green** denotes activate.

Active ArcA~P (anaerobic situation) stimulates PFL and cyd

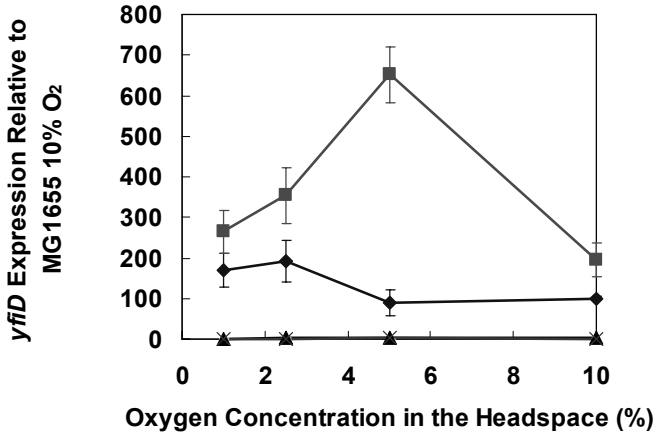
PFL flux higher in arcA mutant



diamond MG1655 ■ MG1655 [Δ*arcA*] × MG1655 [Δ*fnr*]

Sagit Shalel Levanon

Why are the PFL fluxes higher in the arcA mutant? (Since Arc~P activates pfl expression): PFL is reactivated by YfiD

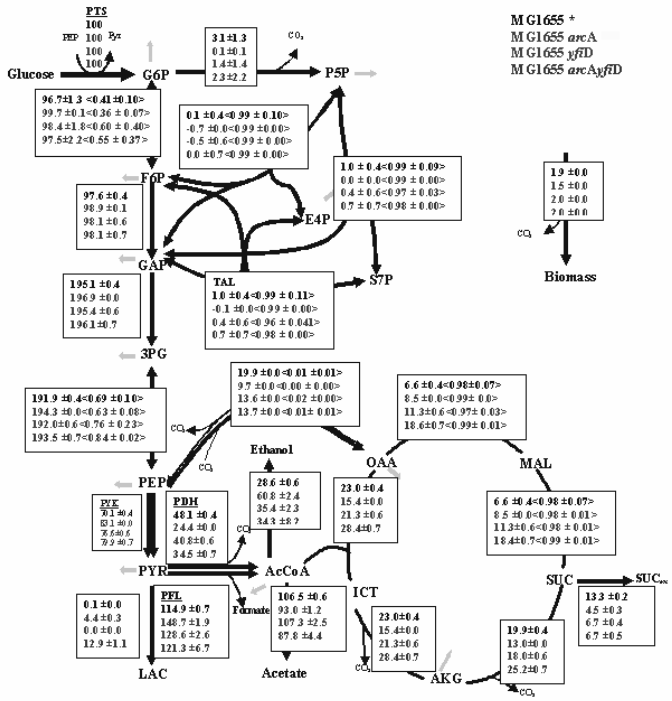


diamond MG1655 ■ MG1655 [Δ*arcA*] × MG1655 [Δ*fnr*] ▲ MG1655 [Δ*arcA* Δ*fnr*]

C-13 label flux analysis at 2.5% oxygen

YfiD contributes 18% of PFL flux in arcA mutant

PDH flux is ~25% of flux from pyruvate to acetyl-CoA in all strains



Overall series of actions (microaerobic)

arcA$^-$ leads to lower cyd

Low cyd can not handle all NADH and the level of reduced NADH rises

Higher NADH leads to expression of yfiD

Higher YfiD reactivates Pfl leading to higher flux through Pfl in arcA mutant in microaerobic conditions

Perspective

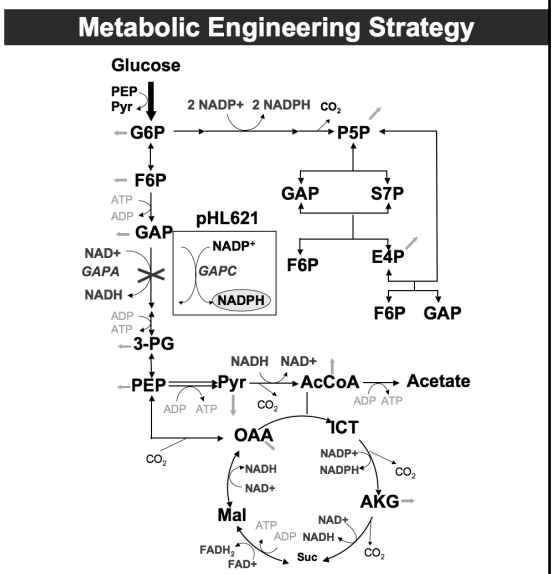
- **Need a variety of experiments to see what are the important regulatory events in the network under a particular circumstance**
- **Methods to analyze a number of parameters in defined sample**

Sometimes you can take advantage of the alternation of global cell physiology from a manipulation

- **Alter levels of certain cofactors, sensors, transporters**

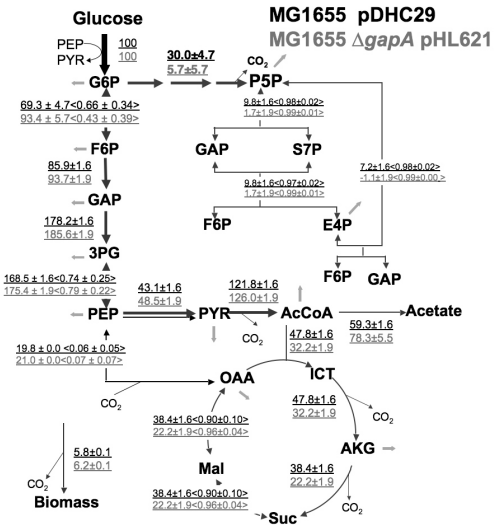
Redox reactions involving NADPH

- Engineer cells to produce more NADPH



Pathway measurements

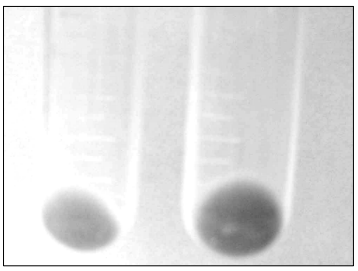
- The pentose pathway that usually supplies most of cells NADPH is much reduced in the modified stain
- Measurements made by GC-MS of C-13 labelled amino acids in steady state cultures



If there is higher NADPH availability can we use this in another pathway?

Lycopene Production Result (24 h)

WT Mutant



Challenges in engineering networks containing unknown genes & regulators

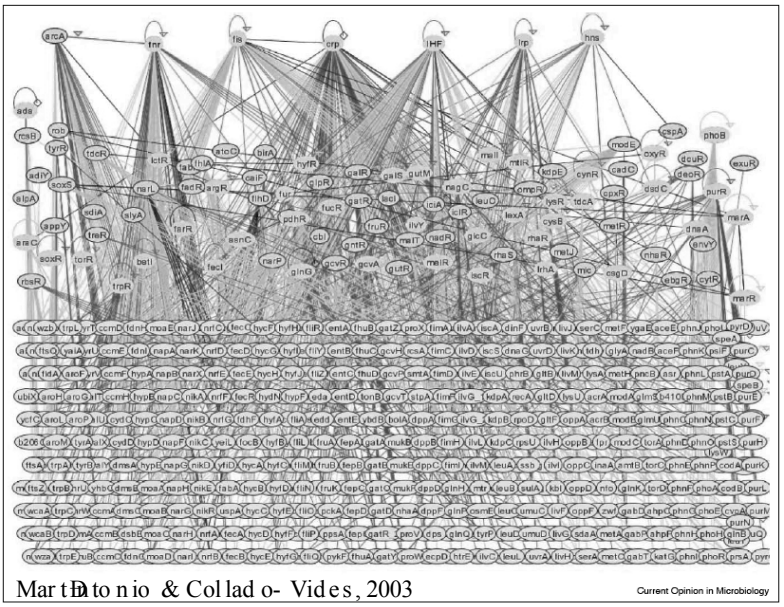
- **Identifying which genes to try to modify (define network)**
  - **Mathematical models**
  - **Gene expression microarray experiments or proteomic measurements**
  - **Tracing phosphorylation networks and connections**
  - **Similarities of protein-protein interaction networks to those of other species**

  Genetic exploration
- **Individual & multiple gene knockouts or overexpression**
- **Adaptation to higher production and genome analysis**
- **Employing novel regulatory factors to affect pathway genes**
- **Use of high throughput screens, enrichments or selections**

Considerations

- **How to identify and control genes useful for a specific metabolic engineering goal?**
- **One strategy- use high-throughput means to gather data, merge in model-computational and analytical approach**
- **Another strategy- use random approach genetic based-can find things you don't already know about**

Partial E. coli regulatory network



Many global regulators affect a large number of genes

Table 1

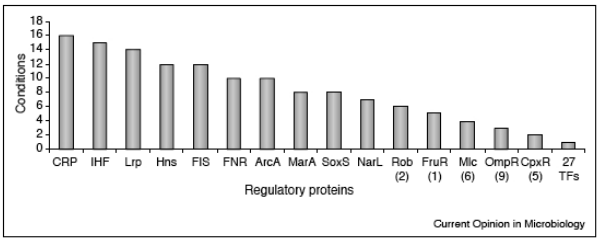
Summary of transcriptional interactions of major TFs, in the transcriptional regulatory network of *E. coli*.

Transcription factor	Genes regulated*	Co-regulators†	TFs regulated‡	Sigma factors§	Functional classes of genes regulated#	Family (members)¶
CRP	197	47	22	$\sigma^{70,38,32,24}$	48	CRP (2)
IHF	101	28	9	$\sigma^{70,54,38}$	26	HI-HNS (2)
FNR	111	20	5	$\sigma^{70,54,38}$	22	CRP (2)
FIS	76	15	4	$\sigma^{70,38,32}$	20	EBP (14)
ArcA	63	18	2	$\sigma^{70,38}$	17	OmpR (14)
Lrp	53	14	3	$\sigma^{70,38}$	15	AsnC (3)
Hns	26	14	5	$\sigma^{70,38,32}$	17	Histone-like (1)
NarL¥	65	10	1	$\sigma^{70,38,}$	14	LuxR/UhpA (17)
OmpR	10	9	3	$\sigma^{70,38}$	5	OmpR (14)
Fur¥	26	8	2	$\sigma^{70,19}$	9	Fur (2)
PhoB	26	1	3	σ^{70}	9	OmpR (14)
CpxR	9	2	1	$\sigma^{70,38,24}$	5	OmpR (14)
SoxRS	9	10	3	$\sigma^{70,38}$	10	AraC/XylS (24)
Mlc¥	5	3	1	$\sigma^{0,32}$	3	NagC/XylR (7)
CspA¥	2	2	1	σ^{70}	2	Cold (9)
Rob**	7	8	2	$\sigma^{70,38}$	6	AraC/XylS (27)
PurR**	28	7	1	σ^{70}	10	GalR/LacI (13)

*Total number of genes regulated directly. †Number of different TFs with which at least a gene or TU is jointly co-regulated. ‡Number of regulated genes that codify for TFs. §List of σ factors of the regulated promoters. #Number of functional classes of the gene products regulated [44]. ¶TF family and in parenthesis the number of members of the family. In addition to the seven global TFs considered here there are TFs suggested by ¥Babu and Teichmann, 2003, [42**] and **Shen-Orr *et al.*, 2002, [50**].

Martõntonio & Collado-Vides, 2003

Some regulators respond to many environmental growth conditions



Global environment growth conditions in which TFs are regulating. To see the detailed list of conditions see RegulonDB page: http://www.cifn.unam.mx/Computational_Genomics/regulondb/SupMat/conditions. Numbers in brackets indicate how many additional TFs participate in the same number of conditions.

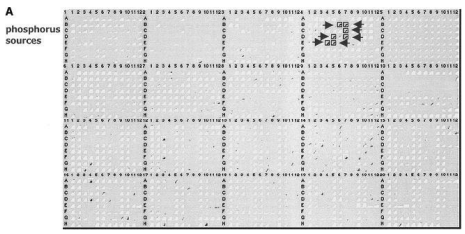
Martõntonio & Collado-Vides, 2003

Identifying important regulators in your network

- Individual specific regulators for operon
- Global regulators
- 2-component regulators
- Small RNA regulators

Test phenotypes of many regulatory genes under many different conditions-database.

- If a regulator has effect under conditions are important to production conditions; can manipulate those regulators and look for effects on the desired system?



•Zhou et al JBact 185, 4956, 2003 screened many mutants in hundreds of conditions

General Approaches to identifying functions of regulators

- **What proteins bind to what gene regulatory regions in the chromosome?**
- **Address by variations of ChIP-chip**
- **What regulatory proteins interact with each other?**
- **Address with two-hybrid system variations or crosslinking and chemical analysis, protein interaction networks**

Correlations to determine choices of genes to modify

- **Two-component regulators often increased in level of their gene expression with condition where they exert their activity**

Example

- **In Clostridium:**
- **Which proteins are involved in initial kinase reaction that starts solvent formation and sporulation**

If there were kinases that might phosphorylate Spo0A in *Cac*… how could we look for them?

Genome analysis: Compare genome characteristics of phosphorelay kinases with those of Bacillus subtilis

***Cac* has 36 His Kinases, 5 of them:**

LIKELY CANDIDATES?

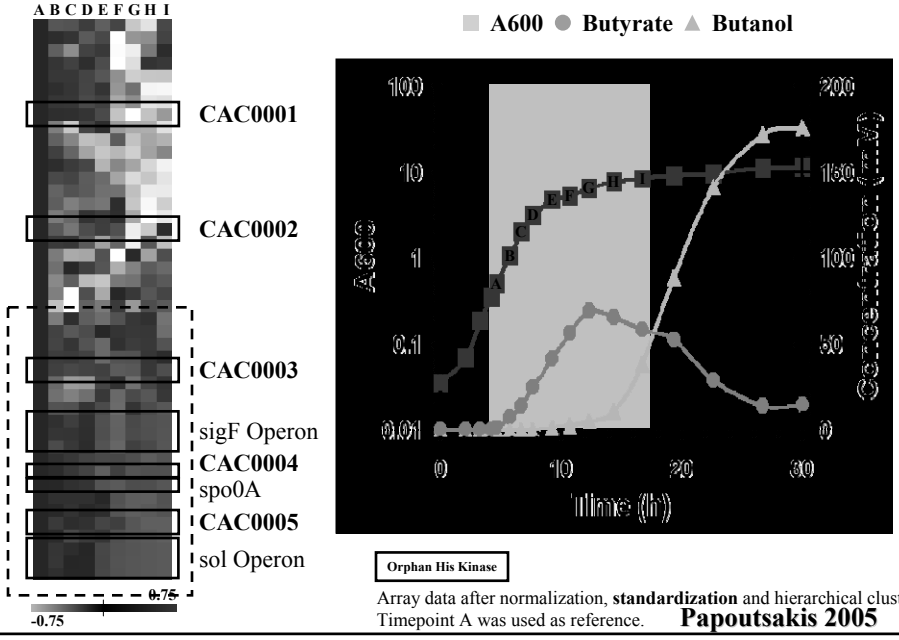
•Experimental analysis:

•Suitable temporal expression patterns?

•Structural similarity?

Papoutsakis 2005

All Cac His Kinase Expression Profiles



Orphan His Kinase

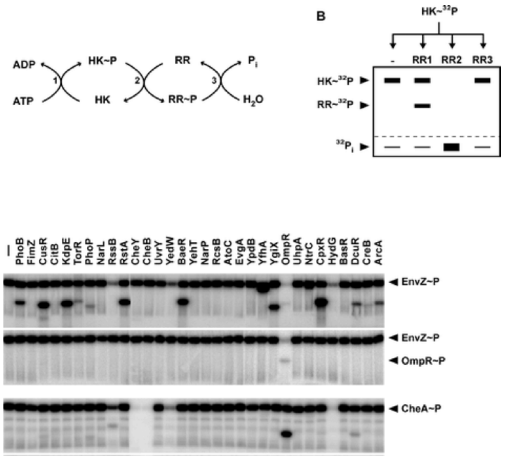
Array data after normalization, **standardization** and hierarchical clustering
Timepoint A was used as reference. **Papoutsakis 2005**

Another strategy is more biochemical

- **Use action of proteins in vitro to follow regulatory network**

Tracing phosphorylation networks

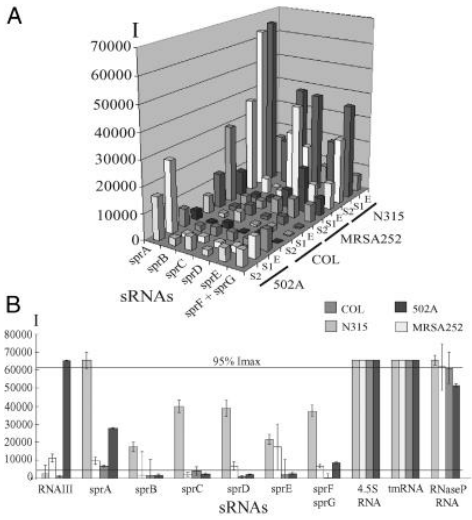
- **A regulator is autophosphorylated and then tested to see what other proteins it can transfer the labeled phosphate to in vitro**



•Skerker et al Plos Biology 3 (10) e334

Action of small RNAs in regulation

- **Identify possible RNAs by bioinformatics**
- **Do microarray with these intergenic regions**
- **Look at expression patterns of the RNAs**
- **Make overexpression and deletion mutants**
- **Explore these mutants by microarray to identify genes under their control**



Pichon & Felden PNAS 2005

Take advantage of the complexity of regulatory network

let genetics help find useful changes to make

just want improved production?

- **Find out mechanism later**

- **Random or artificial evolution approach**
- **Selection or screen then analyze**

Seeking effects of gene knockouts, gene overexpression, or regulatory sequence insertions

- **Large scale library of variants**
- **Screen by optical sorting system based on metabolite or key protein to find +s**
- **Grow under stress of desired production condition- look for those tagged mutants found in higher abundance or greatly diminished in population**

Genome sequence approach

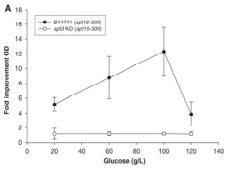
- **Adaptation of cells to higher growth-production state-analyze genome of selected variant to see what has changed**
- **Used in industry to characterize production strain vs original parent**

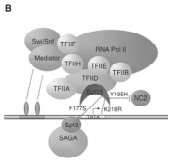
Examining the effect of novel transcriptional regulators

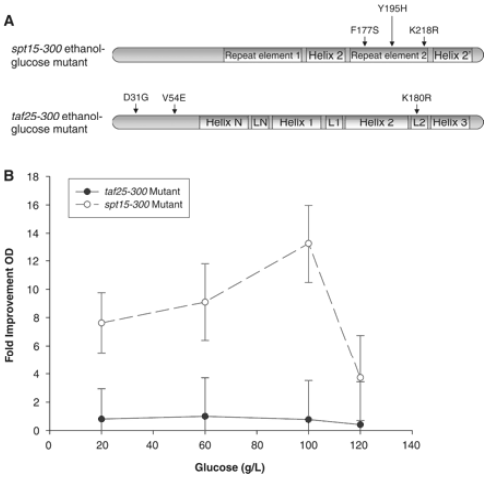
- **Introduce mutations in known regulators**
- **Artificial Zn finger protein variants**
- **Find those that generate desired phenotype**
- **Then can perhaps identify and manipulate the affected genes and see how the change produced its effect**

Mutations in known regulators

- **Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production**
- **Mutate TATA binding transcription factor**





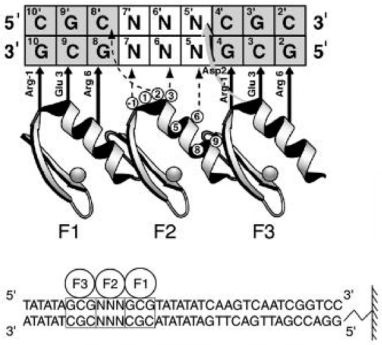


Alper et al Science 314, 1565, 2006

Zn fingers

- Directly select **artificial zinc finger** proteins from a zinc finger protein library- can join to other protein

Bae & Kim Mol Cell 376, 2006



Bulyk et al PNAS 7158, 2001

Acknowledgments

Collaborators:

Ka-Yiu San
Steve Cox
E. T. Papoutsakis

Ailen M. Sanchez
Jiangfeng Zhu

Henry Lin
Sagit Shalel Levanon

Our work was supported by grants from the National Science Foundation and USDA
(BES-0222691, BES-0000303, BES-0420840, and 2002-35505-11638)

Metabolic Flux Mapping

Jacqueline Vanni Shanks
Dept of Chemical and Biological Engineering
Iowa State University

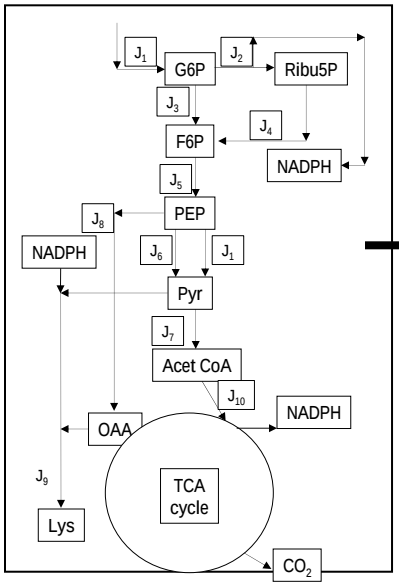
Flux

$$A \xrightarrow[J_1]{E_1} B \xrightarrow[J_2]{E_2} C$$

- Enzyme - E -- the biocatalyst
- Metabolic flux - J
  - Net rate of conversion of one metabolic precursor to a product
  - Based upon a material balance:
    $d[B]/dt = J_1 - J_2$
- Fluxes reflect integration of genetic and metabolic regulation

Metabolic Flux Analysis

Fluxes are obtained using material balances



$$\frac{d[\mathrm{G6P}]}{dt} = J_1 - J_2 - J_3$$

$$\frac{d[\mathrm{Ribu5P}]}{dt} = J_2 - J_4$$

$$\ldots$$

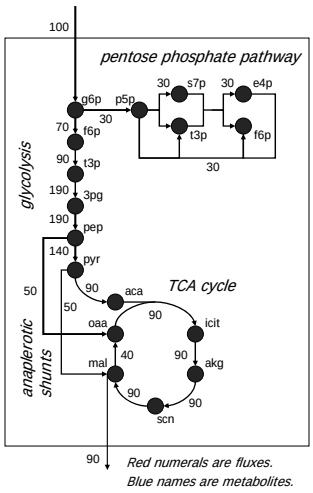
$$\frac{d[\mathrm{NADPH}]}{dt} = 2J_2 + J_{10} - 4J_9$$

$$\frac{d\overline{\mathbf{X}}}{dt} = \overline{\overline{\mathbf{A}}} \cdot \overline{\mathbf{J}}$$

assume
quasi -steady state

$$\overline{\overline{\mathbf{A}}} \cdot \overline{\mathbf{J}} = \mathbf{0}$$

Models and measurements



Hypothetical flux map, for purpose of illustration.

Stoichiometric model has *m* unknown fluxes and *n* internal metabolite balances

m>n	m=n	m<n
more complex networks	simpler networks	
Insufficient number of measurements. Multiple solutions.	Sufficient (or greater) number of measurements. Unique flux distribution.	

How do we choose the "true" solution?

Choose a solution that **optimizes** a logical cellular objective.

Add internal measurements (GC-MS, NMR). Expensive, but more accurate.

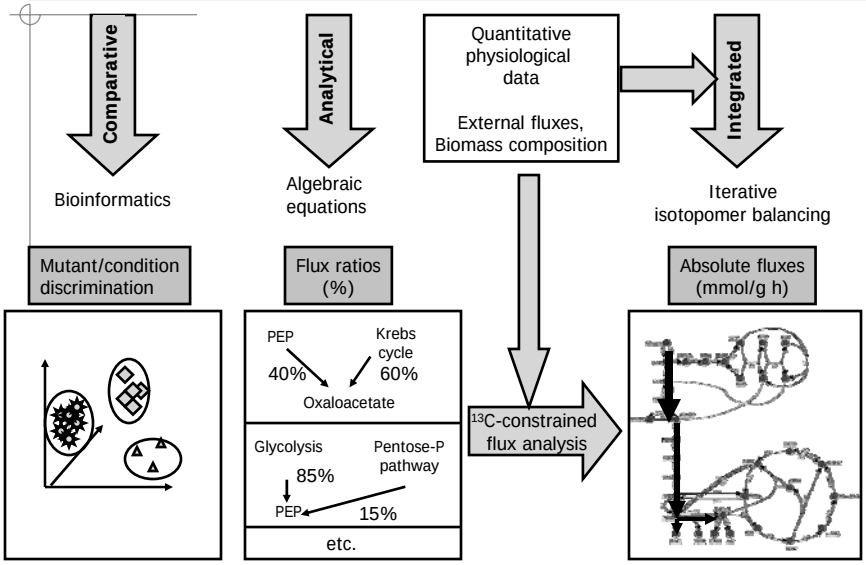
GC-MS vs 2D NMR

	GC-MS	2D NMR
^{13}C, ^{12}C		
	1 metabolite	2 metabolites
	more sensitive	less sensitive

http://www.biotech.biol.ethz.ch/sauer/

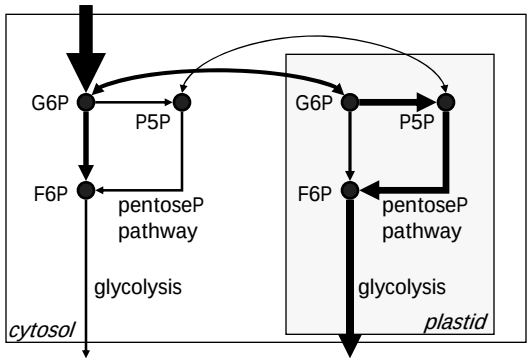
^{13}C Labeling Data

(from proteinogenic amino acids or metabolites)



Flux estimation is difficult in plants

- Plasticity
  - Compartmentation:
    - Parallel pathways, different fluxes
- Different tissues
- Cell types within tissues



In vitro culture for developing soybean embryos

carbon-labeling experiment

growth chamber

12 DAF

In vitro culture

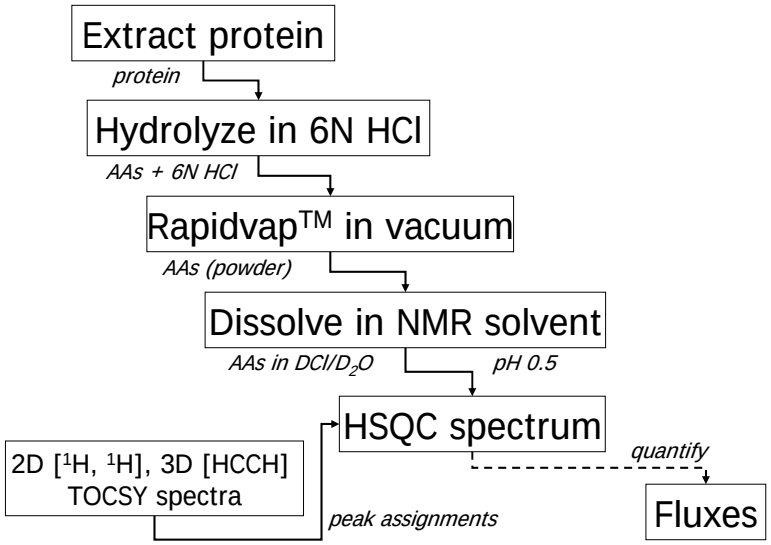
146 mM sucrose (**10% U-^{13}C sucrose**)
37 mM glutamine
vitamins, micronutrients, pH=5.5)

6 days

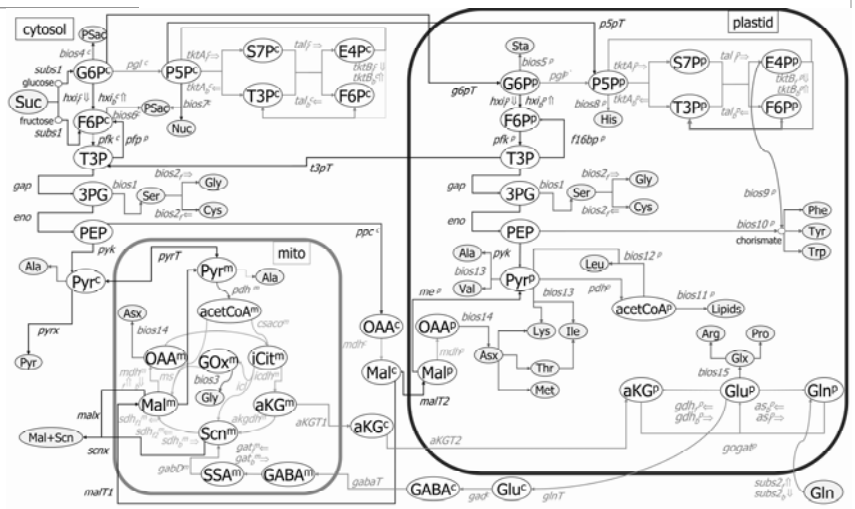
- Extract protein, starch, oil.
- Obtain NMR spectrum of hydrolyzed protein and starch
- Analyze spectrum, use mathematical model to obtain metabolic fluxes.

- Biomass accumulation
- Substrate and product rates
- Biomass composition
- HPLC analysis of protein

Sample Preparation



Metabolic network



(Recent literature; Singh, 1999; Soybase; KEGG encyclopedia)

^{13}C fine structures

^{13}C attached to ^{12}C singlet

Most likely formed by a biosynthetic bond between a ^{13}C and a ^{12}C molecule

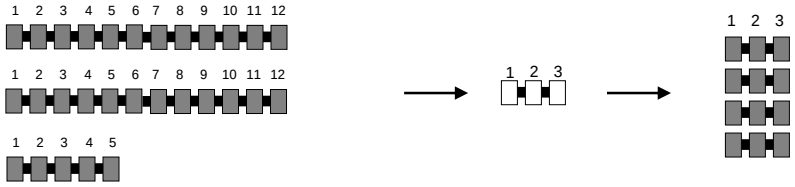
^{13}C attached to ^{13}C doublet

Most likely formed from an intact ^{13}C molecule

Different metabolic histories

Relative abundance of doublets and singlets represents the relative concentrations of intact and biosynthetic bonds in the same metabolite molecule (e.g. different pathways)

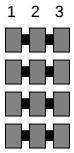
Isotopomer



substrate mixture
sucrose (^{13}C, ^{12}C) + glutamine (^{12}C)

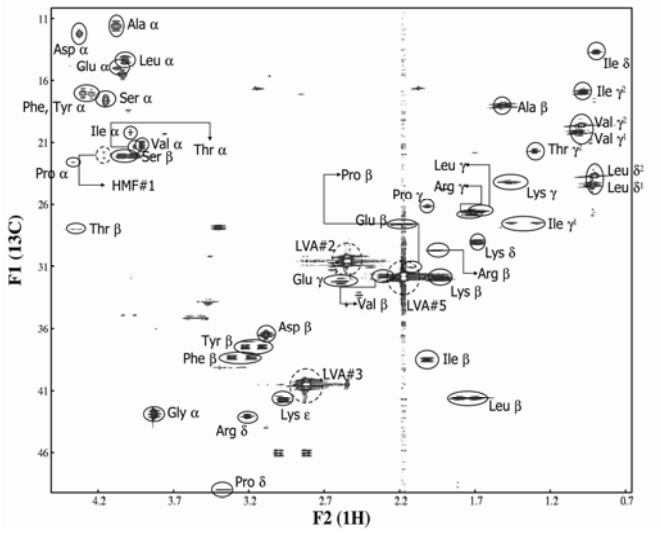
metabolite
e.g. T3P

Isotopomers of T3P



1-1-1, $T3P_7$
1-1-0, $T3P_6$
0-1-1, $T3P_3$
0-1-0, $T3P_2$

2-D [^{13}C,^{1}H] HSQC spectrum

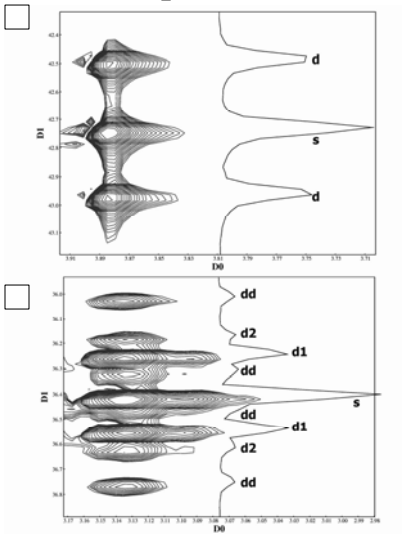


Multiplets and isotopomers

Peak fine structure shows multiplets

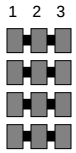
Multiplets are proportional to isotopomers

- singlet
  - ^{12}C-^{13}C-^{12}C
- doublet
  - ^{12}C-^{13}C-^{13}C
  - ^{13}C-^{13}C-^{12}C
- double doublet
  - ^{13}C-^{13}C-^{13}C



Flux evaluation methodology

- 156 isotopomers were measured
- Isotopomer abundances are related to fluxes
- Relationship is nonlinear and nontrivial to solve
- A computational framework was developed to automatically convert raw data to fluxes



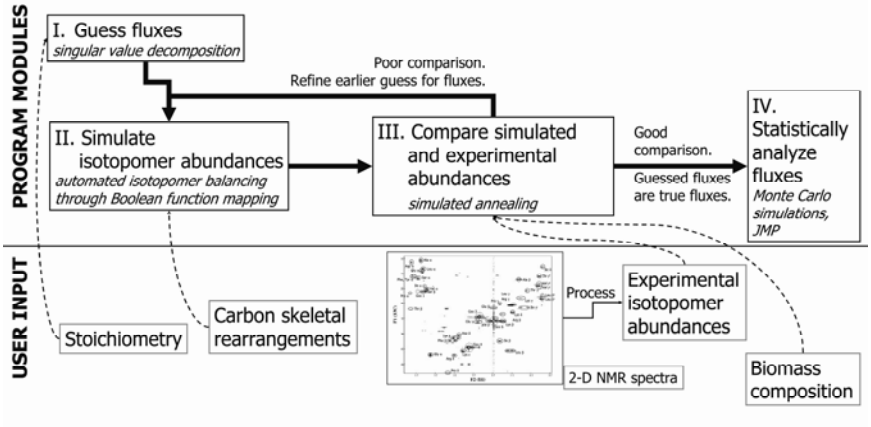
^{13}C-^{13}C coupling indicates intact bonds

^{13}C-^{12}C coupling indicates broken bonds

Natural abundance needs to be considered

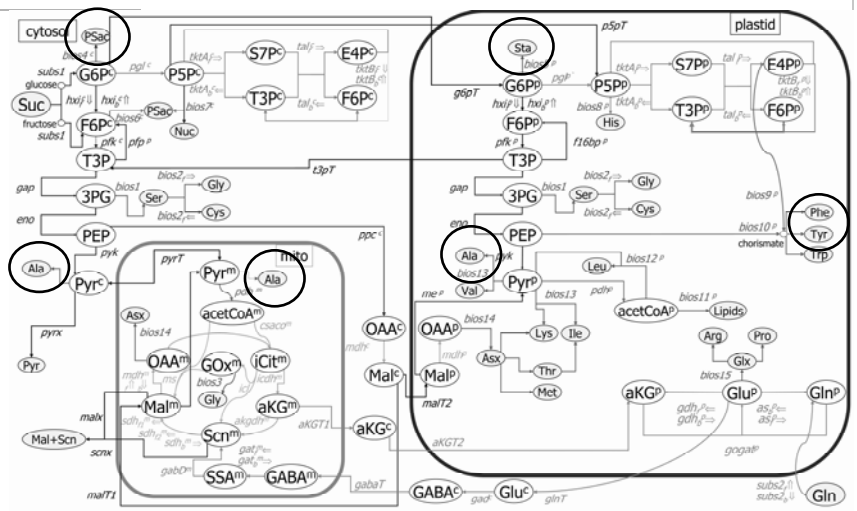
Exact models can be constructed

Flux evaluation methodology



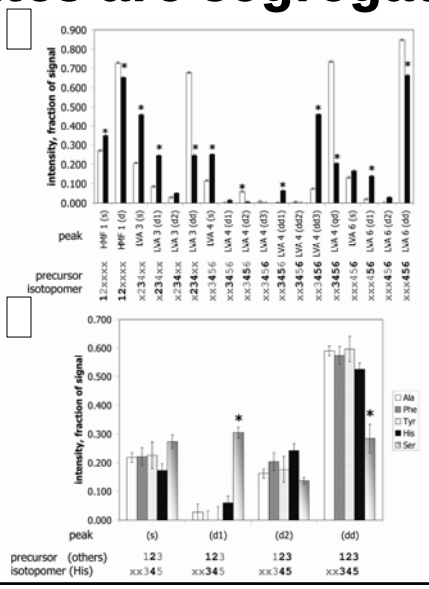
Computer program, NMR2flux (C, Red Hat Linux)
A posteriori model modification is possible

Strategy to identify segregation of pathways

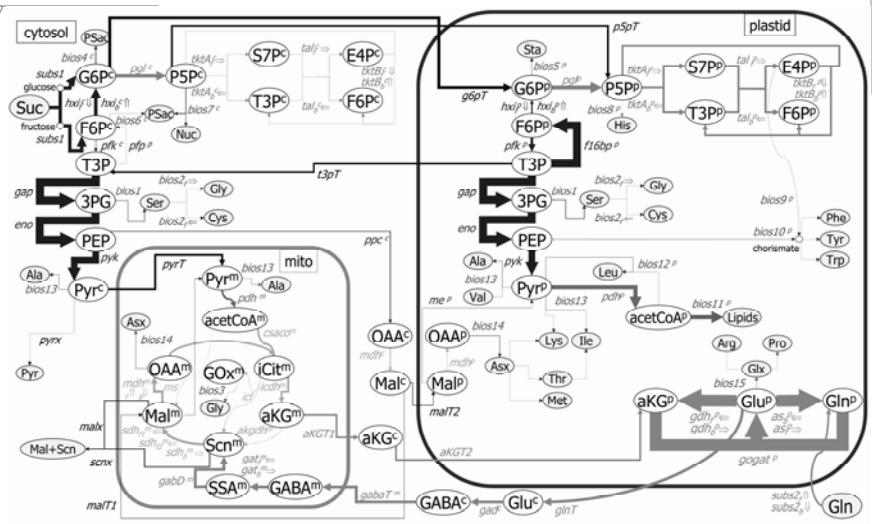


Hexose phosphates are segregated

- Between cytosol and plastid:
  - Hexose pools are isotopically segregated
  - Distinct glycolytic and pentose phosphate pathways exist
  - Triose pools are exchanging rapidly (except Ser)

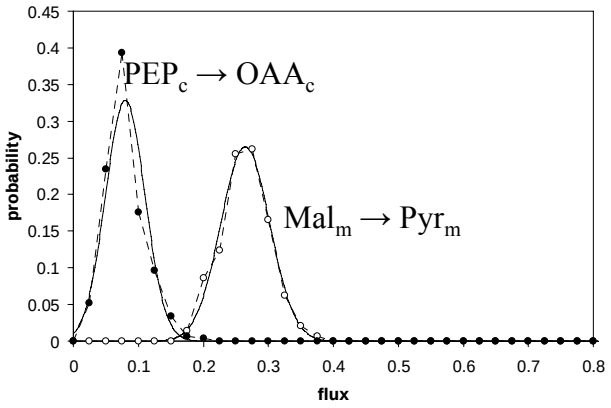


Flux map



Sriram et al.2004. Plant Physiology, 136, 3043-3057.

Statistical analysis of evaluated fluxes

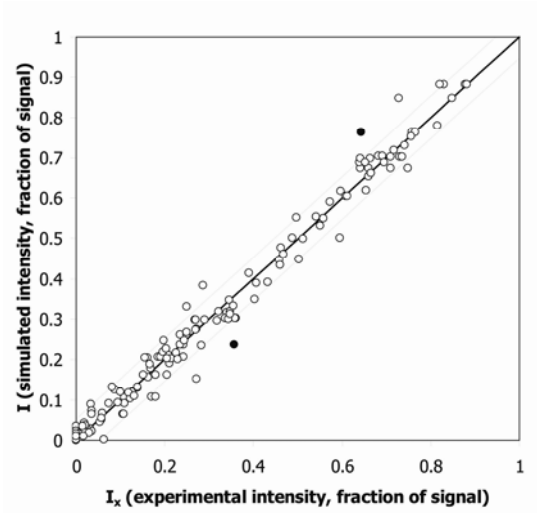


Takes into account experimental error of NMR measurements

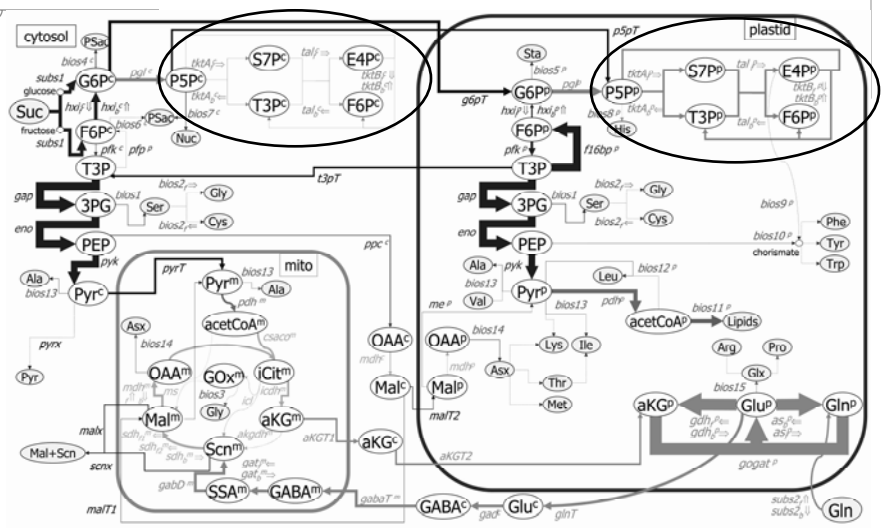
Fluxes, reversibilities, std. deviations

Reaction name	Stoichiometry	Net flux Mean	Net flux SD	Reversibility Mean	Reversibility SD	Comment
Glycolysis and oxPPP						
hxi$_f^c$	$G6P^c \rightarrow F6P^c$	-113.4	18.0	51.8	26.7	Flux is in the direction $F6P^c \rightarrow G6P^c$
hxi$_b^c$	$F6P^c \rightarrow G6P^c$					
hxi$_f^p$	$G6P^p \rightarrow F6P^p$	-43.4	50.2	97.0	3.2	Flux is in the direction $F6P^p \rightarrow G6P^p$
hxi$_b^p$	$F6P^p \rightarrow G6P^p$					
hxi$_f$	$G6P \rightarrow F6P$	-156.8	57.3	—	—	Combined *hxi* flux (cytosol+plastid)
pgl^c	$G6P^c \rightarrow P5P^c + CO_2$	107.9	45.7	irrev.	—	
pgl^p	$G6P^p \rightarrow P5P^p + CO_2$	129.3	58.9	irrev.	—	
pgl	$G6P^c \rightarrow P5P + CO_2$	237.0	52.0	irrev.	—	Combined *pgl* flux (cytosol+plastid)
tktA$_f^c$	$P5P^c + P5P^c \rightarrow S7P^c + T3P^c$	7.2	3.0	90.4	17.9	
tktA$_b^c$	$S7P^c + T3P^c \rightarrow P5P^c + P5P^c$					
tktA$_f^p$	$P5P^p + P5P^p \rightarrow S7P^p + T3P^p$	64.6	17.8	88.5	14.6	
tktA$_b^p$	$S7P^p + T3P^p \rightarrow P5P^p + P5P^p$					
tal$_f^c$	$S7P^c + T3P^c \rightarrow F6P^c + E4P^c$	7.2	3.0	46.7	25.1	
tal$_b^c$	$F6P^c + E4P^c \rightarrow S7P^c + T3P^c$					
tal$_f^p$	$S7P^p + T3P^p \rightarrow F6P^p + E4P^p$	64.6	17.8	76.2	16.1	
tal$_b^p$	$F6P^p + E4P^p \rightarrow S7P^p + T3P^p$					
tktB$_f^c$	$P5P^c + E4P^c \rightarrow F6P^c + T3P^c$	7.2	3.0	6.7	10.9	
tktB$_b^c$	$F6P^c + T3P^c \rightarrow P5P^c + E4P^c$					
tktB$_f^p$	$P5P^p + E4P^p \rightarrow F6P^p + T3P^p$	60.2	17.8	34.2	30.7	
tktB$_b^p$	$F6P^p + T3P^p \rightarrow P5P^p + E4P^p$					
pfk^c	$F6P^c \rightarrow T3P^c + T3P^c$	14.3	17.3	irrev.	—	
pfk^p	$F6P^p \rightarrow T3P^p + T3P^p$	81.4	20.6	irrev.	—	
pfk	$F6P \rightarrow T3P + T3P$	95.7	16.9	irrev.	—	Combined *pfk* flux (cytosol+plastid)
pfp^c	$T3P^c + T3P^c \rightarrow F6P^c$	3.7	4.3	irrev.	—	
f16bp^p	$T3P^p + T3P^p \rightarrow F6P^p$	257.5	59.3	irrev.	—	
gap	$T3P \rightarrow 3PG$	267.4	16.8	n. d.	—	For *gap*, *eno* and *pyk*, cytosolic and plastidic fluxes are indistinguishable. Therefore, only combined fluxes are reported.
eno	$3PG \rightarrow PEP$	253.5	16.8	n. d.	—	
pyk	$PEP \rightarrow Pyr$	219.4	13.4	n. d.	—	
pdh^p	$Pyr^p \rightarrow ACA^p + CO_2$	136.0	< 1.0	irrev.	—	
pdh^m	$Pyr^m \rightarrow ACA^m + CO_2$	66.3	16.0	irrev.	—	

Simulation and experiment



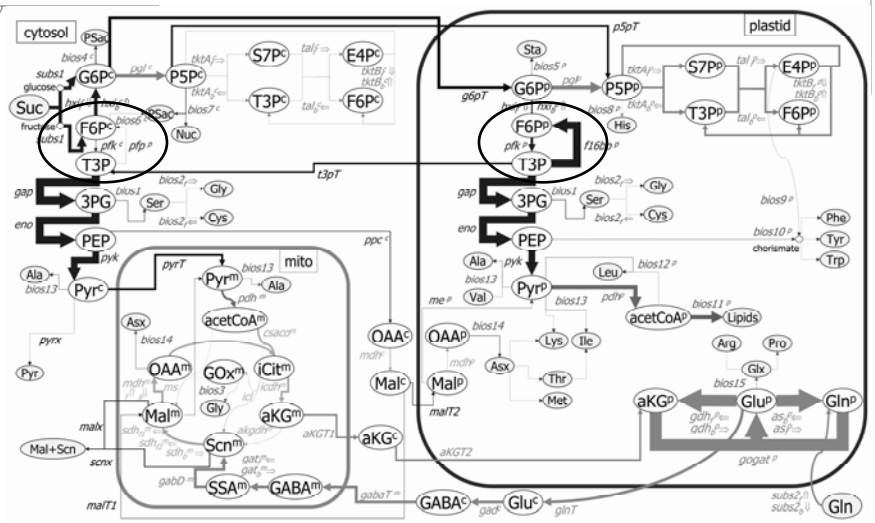
Reductive PPP



Reductive PPP compartmentation

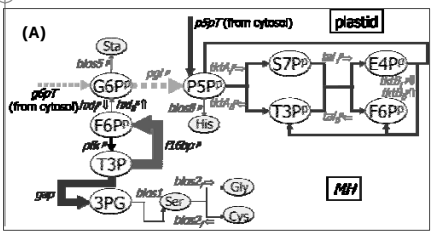
- Reductive PPP fluxes negligible in cytosol (7.4 ± 3.0 units), substantial in plastid (64.6 ± 17.8 units)
- Consistent with genetic and enzymatic data from several higher plants:
  - Arabidopsis genome lacks cytosolic genes for reductive PPP (Eicks *et al.*, 2002)
  - Spinach and pea did not contain cytosolic enzymes for reductive PPP (Debnam and Emes, 1999)
- Reductive PPP may not be present in the cytosol in most plants (Kruger and von Schaewen, 2003)

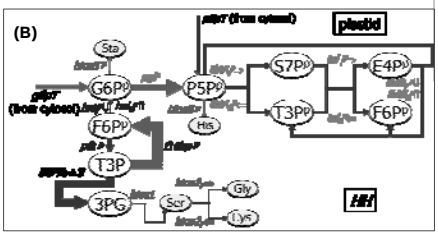
Network Addition of Fructose-1,6-bisphosphatase



A posteriori model modification

Flexibility and Rigidity of Nodes





oxPPP node for MH temp. condition (Case 1- MH, MM, ML)

oxPPP node for HH temp. condition (Case 2- HH, MM, LL)

RIGID

FLEXIBLE

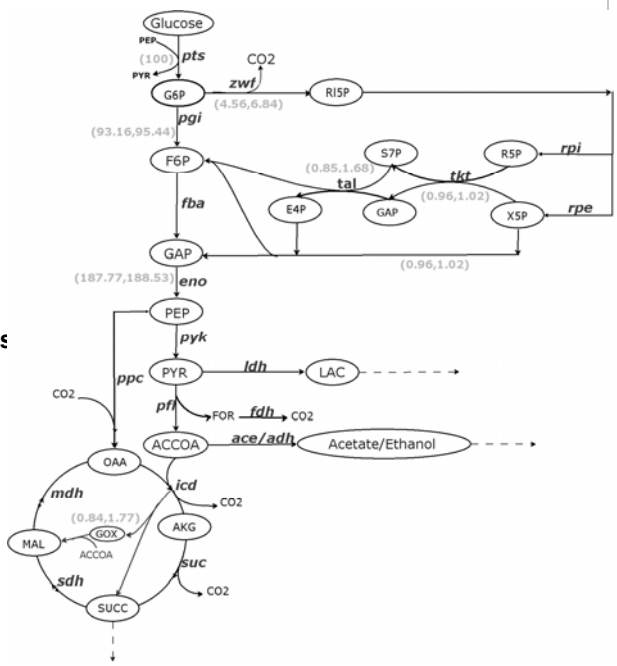
E. coli

Anaerobic Metabolism

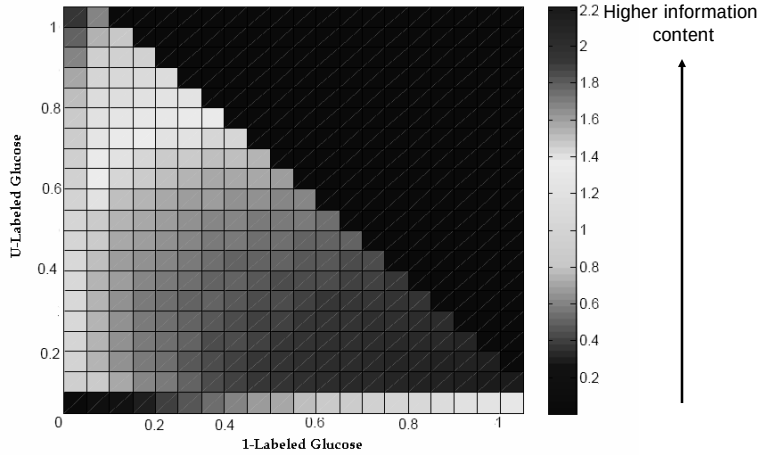
<u>Challenges:</u>
Smaller flux to biomass

Many products

Less rearrangement – TCA cycle incomplete.



Identifiability of fluxes with two labelled compounds



Summary

- Implemented a carbon labeling experiment to evaluate compartmented fluxes
- Developed a computational framework to aid in flux evaluation
  - Computer software (NMR2flux) to evaluate fluxes, reversibilities, and confidence intervals

Larger Networks

- Network Topology
- Extracellular, biomass, composition measurements
- Metabolite profiling/retrobiosynthetic approach
- Effective methods for analyzing isotopomer abundances
- Choice of label for identifiability of fluxes
- Uncertainty or range in flux estimates

Acknowledgments

Flux Map Creation in Plants

- Ganesh Sriram (ChE)
  - Omar González-Rivera (ChE)
- Vidya Iyer (ChE)
- Prof. Mark Westgate (Agronomy)
  - Joan Peterson
  - Ruilian Zhou
- Prof. Marty Spalding (GDCB)
- Dr. Louisa Tabatabai (Protein Facility)
- Dr. Bruce Fulton (NMR Facility)

Metabolic Engineering of E. coli Sugar-Utilization Regulatory Systems

- Prof Ramon Gonzalez, Rice
- Prof. Ka-Yiu San, Rice
- Madhuresh Choudhury (ChE)

Funding: Interagency Program in Metabolic Engineering (NSF,USDA,DOE,EPA)

Funding: NSF, Plant Sciences Institute, Iowa Soybean Board, Cargill

Isotopomer balance

$$0 = \left\{ \begin{array}{l} v_{tal_f^p} \cdot \mathrm{S7P^p}[\mathbf{12}3\mathrm{xxxx}] \cdot \mathrm{T3P^p}[\mathbf{12}3] + v_{hxi_f^p} \cdot \mathrm{G6P^p}[\mathbf{1}23\mathbf{45}6] \\ + v_{f16bp^p} \cdot \mathrm{T3P^p}[\mathbf{12}3] \cdot \mathrm{T3P^p}[1\mathbf{23}] + v_{tktB_f^p} \cdot \mathrm{P5P^p}[\mathbf{12}\mathrm{xxx}] \cdot \mathrm{E4P^p}[1\mathbf{23}4] \end{array} \right\}$$
$$- \left\{ v_{pfk_b^p} \cdot \mathrm{F6P^p}[\mathbf{1}23\mathbf{45}6] + v_{hxi_b^p} \cdot \mathrm{F6P^p}[\mathbf{1}23\mathbf{45}6] + v_{tal_b^p} \cdot \mathrm{F6P^p}[\mathbf{1}23\mathbf{45}6] + v_{tktB_b^p} \cdot \mathrm{F6P^p}[\mathbf{1}23\mathbf{45}6] \right\}$$

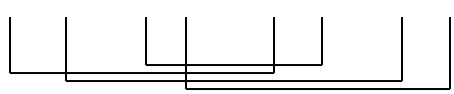
- Our network had 564 balances, and 33 unknowns.
- However, balances for all isotopomers (whether measured or not) in the metabolic network must be written and solved to calculate fluxes.

Boolean function mapping

- **What:** A method to convert reactant isotopomers to product isotopomers automatically (Sriram G., and J. V. Shanks, 2004, *Metab. Eng*).
- **How:** Metabolic reactions can be represented by the following schema:

  R1 + R2 → P1 + P2 (overall reaction)

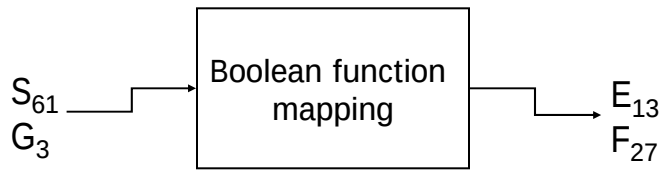
  $R1^L\text{-}R1^R + R2^L\text{-}R2^R \rightarrow P1^L\text{-}P1^R + P2^L\text{-}P2^R$ (typical carbon skeleton rearrangement)



- We recognize four moves in this schema, which are represented as Boolean functions:
  - Fragmentation — $R1 \rightarrow R1^L\text{-}R1^R$
  - Reversal — $R1^R \rightarrow \underline{R1^R}$ (occasional, e.g. F6P → T3P reaction, fumarase reaction)
  - Transposition — $R1^R \rightarrow P2^L$
  - Condensation — $P2^L\text{-}P2^R \rightarrow P2$

Example: transaldolase reaction

- Assume reactant isotopomers are S_{61} (0111101) and G_3 (011)
- Manual isotopomer enumeration shows that **E_{13} (1101)** and **F_{27} (011011)** should result.



- 67% time reduction compared to existing methods (Wiechert *et al.*, 2001).
- Time/iteration (soybean network) = 0.48 s on a 3 GHz pentium.

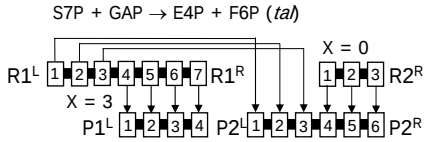
Sriram, G., and Shanks, J. V., 'Improvements in metabolic flux analysis using carbon labeling experiments: bondomer balancing and Boolean function mapping.' Metab. Eng., 2004.

return

Boolean function mapping

Sriram, G., and Shanks, J. V., 'Improvements in metabolic flux analysis using carbon labeling experiments: bondomer balancing and Boolean function mapping.' Metab. Eng., 2004

- In the *tal* reaction, let us assume we have S_{61} (0111101) and G_3 (011) as the reactants. From inspection, we should expect to get **E_7 (1101)** and **F_{27} (011011)**.



- When the Boolean/arithmetic functions are executed, we get:

Fragmentation: $R1^L = 61 / 2^4 = 3$ (011), $R1^R = 61 \% 2^4 = 13$ (1101);

$R2^L$ = NULL, $R2^R = 3$ (011)

Transposition: $R2^R = 3$ (011) becomes $P2^R = 24$ (00001); 3 (<<) 3 = 24

Note: Now $P1^L = R1^R = 13$, $P1^R = 0$; $P2^L = R1^L = 3$, $P2^R = 24$.

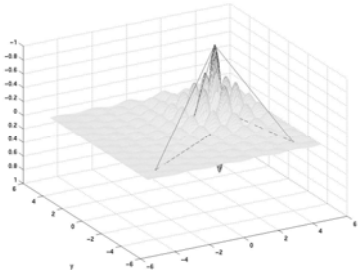
Condensation: $P1 = P1^L + P1^R = 13 + 0 = 13$ **(E_{13} = 1101)**;

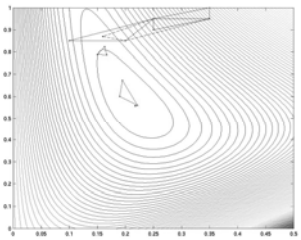
$P2 = P2^L + P2^R = 24 + 3 = 27$ **(F_{27} = 011011)**.

return

Optimization module

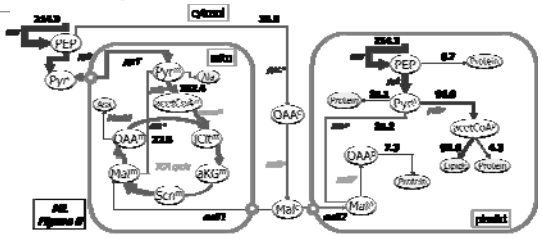
- Minimization of error between experimental and simulated intensities.
- Simulated annealing routine was developed (global optimization).
- Downhill simplex algorithm was used near the optimum (local optimization).



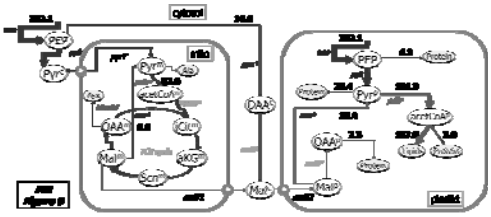


return

Plastidic Pyruvate Node



Pyr crossroad for ML temp. condition



Pyr crossroad for MH temp. condition

Agency Activities in Metabolic Engineering

Department of Agriculture (USDA)

The Cooperative State Research, Education and Extension Service (CSREES) is the USDA agency that participates in the Interagency Metabolic Engineering Working Group. In the CSREES Strategic Plan, five goals are listed:

- An agricultural production system that is highly competitive in the global economy.
- A safe, secure food and fiber system.
- Healthy, well-nourished population.
- Greater harmony between agriculture and the environment.
- Enhanced economic opportunity and quality of life for Americans.

These goals reflect the goals of the overall USDA strategic plan (enhancing economic opportunities for agricultural producers, supporting increased economic opportunities and improved quality of life in rural America, enhancing protection and safety of the nation's agriculture and food supply, improving the nation's nutrition and health, and protecting and enhancing the nation's natural resource base and environment).

Metabolic Engineering (ME) can enhance competitiveness of the US agricultural system through the production of commercially useful products such as chemicals, biofuels, and biomolecules from agricultural commodities. Through modification of plants, animals, and microorganisms, ME can also result in new uses for existing crops and animals, added value to traditional agricultural products, and improved quality of agriculturally derived foods and materials. It is also possible through ME to produce plants with enhanced nutritional value or to modify plants and microorganisms for remediation of polluted environments.

The participation in MEWG has allowed CSREES to leverage funding for support of several research projects that address one or more of CSREES' and USDA's goals. Funding is supporting research on metabolic engineering of biofuels that may lead to maximized ethanol production as well as reduced costs. Another funded project involves production of flavor compounds in microbes that may eventually lead to improvements of metabolic function for processing of agricultural biomass and manufacture of bio-based industrial products. Funded metabolic engineering research projects in plants have the potential to produce fruits and vegetables with increased nutritional value and extended shelf-lives, to increase natural product-based disease and pest resistance, to enhance oil production in oilseeds, and to modify plants for production of pharmaceuticals and other economically important compounds. Thus, metabolic engineering, through both basic and applied research, is of vital importance for achieving the strategic goals of CSREES and USDA.

Department of Commerce (DOC)

The MEWG supports the DOC mission by advancing research and development of new commercial and industrial processes. As an emerging technology whose scientific basis is developing rapidly, ME is important to DOC'S NIST and especially its Biotechnology Division. NIST is especially interested in ME projects that support the development of biological and metabolic models, measurement methods and standards.

Department of Defense (DOD)

The Department of Defense (DoD) currently supports a broad range of research in the area of metabolic engineering through the Army Research Office (ARO) and other Army research activities, the Air Force Office of Scientific Research (AFOSR), the Office of Naval Research (ONR), and the Defense Advanced Research Projects Agency (DARPA). The specific focus of the ARO, AFOSR, ONR, and DARPA efforts will be summarized and future directions in metabolic engineering research and technology development will be addressed.

The broad needs for the DoD that can be served through research efforts in metabolic engineering are summarized below. These science and technology targets will provide enhanced and expanded capabilities for the missions of the services and provide greatly expanded capabilities for the civilian sector.

- Materials
- Processes
- Devices
- Fabrication Schemes
- Information Processing

Current interests in metabolic engineering at ARO are focused on the characterization of biochemical pathways, inter- and intra-cellular signaling, and enzymatic mechanisms, and the genetic basis for manipulation of protein expression, structure and function, and cell fate, in systems with potential relevance to the Army. The goal is to develop a detailed understanding of how macromolecules and cells execute their designated functions and how they interact with other cells and macromolecules. With this information, it will be possible to design and engineer particular sub-cellular elements and metabolic pathways and cell systems to exhibit a set of specific functions and properties, according to Army needs, and to identify and non-invasively correct molecular deficiencies to optimize and maintain cognitive and physical performance under normal and extreme conditions. ARO currently supports research in several areas, including: how molecular transport, subcellular compartmentalization, and reaction sequences are involved in enzymatic regulation and superstructure formation; understanding and manipulating aminoacylation of tRNAs and genetic code expansion to produce new polymeric peptides containing non-natural amino acids; biologically based means for fabrication of functional nanostructures; systems engineering of cell differentiation processes; the role and regulation of classes of proteins differentially expressed in response to environmental or external stimuli; molecular genetics and genomics of human cognition, performance and function; and the design and implementation of unique biomolecular and cell based strategies for economically and environmentally favorable manufacturing, as well as the biodegradation of environmental pollutants.

AFOSR's metabolic engineering efforts focus on elucidating the fundamental science to advance miniature biofuel cells for sensor and micro UAV applications. To this end, they are exploring mechanisms for metabolism of complex biofuels (mixtures of various sugars, cellulose, etc.) either in vivo or in vitro for energy production. Characterization of electron and proton transfer in enzymatic redox reactions, and optimization of these reactions at an electrode surface, is also of interest.

One of the metabolic engineering foci at ONR, currently, is the microbial synthesis of energetic materials (EM) and EM precursors for the purposes of cost and environmental impact.

Practically all such materials are non-natural products and their biosynthesis therefore requires the re-engineering of existing pathways and/or the assembly of new or hybrid pathways in one or more host organisms. An example of a simple EM precursor now under study is 1,2,4-butanetriol, which as its energetic trinitrate is used as a plasticizer in propellant and explosives formulations. More advanced EM targets, such as RDX, HMX and CI20, involve high density fused ring cores with multiple nitramino (C-N(NO2)) substituents. While these are very difficult targets, they suggest worthwhile research goals such as the biosynthesis of highly electron withdrawing substituents on carbon (as in C-nitramino) or the assembly of strained heterocyclic rings. Clearly, a theoretical/experimental approach to the prediction of the true scope of enzyme reaction specificity, with energetic boundaries, would be particularly valuable in the design of pathways for EM biosynthesis. Other non-polymeric targets, besides EM, would include novel photonic/electronic/optical materials.

DARPA's metabolic engineering programs are driven by an interest in protecting human assets against biological threats and using biology to maintain human performance. The general concept of this thrust is to understand how nature controls the metabolic rate of cells and organisms (e.g., extremophiles, hibernation) and apply this understanding to problems of interest to DoD. Examples of current investments in metabolic engineering include efforts to develop technologies for engineering cells, tissues and organisms to survive in the battlefield environment so they can be used as sensors. Related basic research on biochemical circuit engineering in laboratory model organisms is also supported. In addition, DARPA is developing technologies that permit the long-term storage of cells including human blood. More complete descriptions of current DARPA programs and solicitations in these areas can be viewed at http://www.darpa.mil/dso.

Department of Energy (DOE)

The Department of Energy is supporting research in metabolic engineering research, largely through the Offices of Science (SC), Energy Efficiency and Renewable Energy (EE), and Environmental Management (EM). The research falls in two main categories: 1) basic research, which involves the advancement of metabolic engineering fundamental knowledge and capabilities, and 2) applied research, which employs metabolic engineering techniques in development of target products. The basic research efforts of the Department reside within SC, whereas most of the applied research in this area is conducted within EE. In general, these research efforts are conducted by universities, national laboratories, and industry.

The Department's goals related to metabolic engineering research are to:

- To expand the level of knowledge and understanding of metabolic pathways and metabolic regulatory mechanisms related to the development of novel bio-based systems for the production, conservation, and conversion of energy.
- Apply metabolic engineering techniques to enhance and develop plants and microorganisms for use in the production of chemicals and fuels or for environmental remediation of waste sites.

Environmental Protection Agency (EPA)

The mission of the Environmental Protection Agency is to protect human health and the environment from adverse effects of anthropogenic activity. Included in this mission are various elements for which metabolic engineering can play a useful role.

One prominent concern is the introduction of chemicals to the environment, which may have detrimental effects on humans and other biota. As mandated by statute and implemented by rule, the Agency routinely conducts evaluation of chemicals intended for use, currently in use, or determined to exist at significant levels in the environment. From these evaluations, the Agency may decide to implement management strategies designed to limit the potential for adverse effects.

The application of novel technologies such as the use of biotechnology as a substitute to conventional manufacturing and processing of raw materials into final products is consistent with the mission of the Agency. EPA implements this by supporting development of technologies which 1) use chemical substitutes that are less toxic; 2) produce more efficient activity resulting in decreased requirement for the chemical or; 3) develop engineering procedures which produce little or no toxic end products. Finally, consistent with the pollution prevention ethic is the reevaluation of chemical stewardship from one of "cradle to grave" to a more multigenerational philosophy in which a chemical may be utilized successively in different forms prior to final disposal. Metabolic engineering has a role to play by enabling the development of biological mechanisms for production or use that meet one or more of these criteria.

While it is generally accepted that chemical-based technologies have evolved to provide a higher standard of living for the general population, it is also recognized that the use of some chemicals, either through the chemical characteristics or the handling, synthesis or disposal, have produced negative effects on human health and/or the environment. Advances in technology allow scientists to better predict the potential for adverse effects from exposure to chemicals as well as mechanisms to diminish the negative effects of chemical production such as production of toxic byproducts and disposal of the chemical. The approach, which strives to identify synthetic pathways that are less polluting than existing pathways and that encourages the development of nontoxic chemical products, is referred to as "Green Chemistry". The use of metabolic engineering to evaluate the potential for increased risk from chemicals, by allowing the study of responsible metabolic pathways and by permitting modification of such pathways to reduce risk, is another way in which metabolic engineering firs within the EPA mission.

Finally, basic research, which utilizes methods of metabolic engineering, can provide longer-range approaches to assist EPA in its overall mission of protecting human health and the environment. The EPA supports extramural metabolic engineering research through the Technology for a Sustainable Environment (TSE) program, which awards grants in the area of pollution prevention. Since 1995, the TSE program has funded metabolic engineering research related to methanol conversion, solvent tolerance, biopolymer production and pesticide production-all focused on the elimination of pollution at the source.

National Aeronautics and Space Administration (NASA)

One of NASA's strategic goals is to extend the duration & boundaries of human space flight to create new opportunities for exploration & discovery. To prepare for and hasten the journey, the NASA Office of Biological and Physical Research must address the following questions through its research:

- How can we assure the survival of humans traveling far from Earth?
- What technology must we create to enable the next explorers to go beyond where we have been?

NASA's efforts in the area of metabolic engineering are on approaches and applications that will have a significant impact on the reduction of required mass, power, volume, crew time, and on increased safety and reliability, beyond the current baseline technologies. The targeted and purposeful alteration of metabolic pathways found in an organism may play a key role in the development of biological approaches and technologies that enable efficient use of spacecraft resources for long-duration space missions.

National Institutes of Health (NIGMS/NIH)

The National Institute of General Medical Sciences (NIGMS) supports metabolic engineering research, usually in the form of grants to investigators in universities (R01s) or in small businesses (SBIRs). These grants support basic research in two general areas: (1) the development of microbial or plant-based metabolic routes to useful quantities of small molecules such as polyketides; (2) the development of a much better understanding of the control architecture that integrates the genetic and catalytic processes in normal and aberrant cells.

National Science Foundation (NSF)

The mission of NSF is to:

- Promote the Progress of Science
- Advance the National Health, Prosperity, and Welfare
- Secure the National Defense
- Provide for Other Purposes

Support of ME research allows NSF to address specific goals within its mission. These include, but are not limited to; development of technologies integrating theoretical, computational, and experimental approaches to the study of metabolic processes; the targeted and purposeful alteration of metabolic pathways in living organisms in order to better understand and utilize these pathways for chemical transformation, energy transduction, and supramolecular assembly; providing a framework for studying the dynamics of interactions and interconversions of biological molecules in order to understand how organisms regulate specific physiological processes at the cellular and sub-cellular levels and the "cross-talk" between pathways; measurement and control of *in vivo* metabolic fluxes; metabolic control analysis of pathway groups or networks; development of *in vivo* techniques to accomplish these goals.

Metabolic Engineering has been supported in all interagency competitions by three Directorates within NSF. There is a recognition at NSF that this Activity has been beneficial to NSF and that NSF would like to continue with this Activity.