Proceedings of the
Inter-Agency Conference on
Metabolic Engineering

2006

Metabolic Engineering Working Group

February 14, 2006
The World Technology Evaluation Center, Inc. (WTEC) (previously known as the Japanese Technology Evaluation Center, JTEC) 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

WTEC, Inc.
4800 Roland Avenue, Suite 201
Baltimore, Maryland 21210 USA
INTER-AGENCY CONFERENCE ON
METABOLIC ENGINEERING
2006

PROCEEDINGS

February 14, 2006

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.
Dear Colleague:

We wish to present to you Proceedings from the recent Inter-Agency Conference on Metabolic Engineering, which was held in conjunction with the Joint Genomics: GTL Contractor-Grantee Workshop IV at the Bethesda North Marriott Hotel & Conference Center, North Bethesda, Maryland on February 14, 2006.

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:

- Processes to facilitate integration and organization of genomic sequencing data into pathways and networks.
- Development of robust methods to identify common regulatory factors between pathways or networks and elucidate pathway interactions and modulations.
- Development of high throughput (HTP) computational methods, metabolic manipulation, and analysis of gene manipulation.
- Development of experimental and computational tools to evaluate metabolic flux.

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, which is available at:

http://www.metabolicengineering.gov/me2006/ReportTOC.html

has slide information from each of the presentations.

For current information on the activities of the MEWG, please refer to the MEWG web site shown below:

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
Inter-Agency Working Group on Metabolic Engineering
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.gov

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

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**
   - Peter Karp ..................................................................................................................................................................................... 4
   - Costas Maranas ............................................................................................................................................................................... 5

6. **Panel Session Summary** ............................................................................................................................................................. 7

7. **Presentations**
   - Peter Karp ..................................................................................................................................................................................... 17
   - Jay Keasling ...................................................................................................................................................................................... 21
   - Michael Betenbaugh ......................................................................................................................................................................... 25
   - Costas Maranas ............................................................................................................................................................................... 30

8. **Agency Activities in Metabolic Engineering**
   - Department of Agriculture .......................................................................................................................................................... 34
   - Department of Commerce ........................................................................................................................................................... 34
   - Department of Defense ................................................................................................................................................................. 35
   - Department of Energy ................................................................................................................................................................. 36
   - Environmental Protection Agency .................................................................................................................................................. 37
   - National Aeronautics and Space Administration ...................................................................................................................... 37
   - National Institutes of Health ......................................................................................................................................................... 38
   - National Science Foundation ......................................................................................................................................................... 38
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 – Opportunities for Leveraging Genomic Technologies for Metabolic Engineering
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 sixth Inter-Agency Conference held on February 14, 2006.

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 jointly between the DOE Genomics: GTL program and the Metabolic Engineering Working Group Inter-Agency Conference on Metabolic Engineering 2006, seeks to identify opportunities for leveraging genomic technologies for metabolic engineering. Specific topics for discussion include:

- Processes to facilitate integration and organization of genomic sequencing data into pathways and networks
- Development of robust methods to identify common regulatory factors between pathways or networks and elucidate pathway interactions and modulations.
- Development of HTP computational methods, metabolic manipulation, and analysis of gene manipulation
- Development of experimental and computational tools to evaluate metabolic flux

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

Metabolic Engineering Breakout Session

Tuesday, February 14, 2006, 2:00 to 4:30 pm

2:00 pm  * Processes to facilitate integration and organization of genomic sequencing data into pathways and networks. - Peter Karp (SRI)

As whole genome sequence information becomes available for an increasing number of organisms, there is a need to efficiently mine these genomes for specific genes, cofactors, and regulatory factors. How can we improve our ability to identify selected target genes, and organize them into functional metabolic pathways or networks?

2:30 pm  * Development of robust methods to identify common regulatory factors between pathways or networks and elucidate pathway interactions and modulations. - Jay Keasling (UCBerkeley/LBNL)

How can genomics help us move from information about components in individual pathways to discover additional constituents of related networks, or common regulators? How can we identify coordinately-controlled networks, or optimize desired metabolic outputs under specific conditions?

3:00 pm  Break

3:15 pm  * Development of HTP computational methods, metabolic manipulation, and analysis of gene manipulation. - Michael Betenbaugh (Johns Hopkins University)

Although it is possible to generate vast quantities of experimental data, there remains a need for tools to facilitate high throughput analysis of this data. How can genomics enable us to target specific genes, and how can we develop computational tools that will allow us to evaluate or predict manipulations in silico? What HTP tools are needed to experimentally evaluate or validate predicted changes in gene or metabolic manipulation?

3:45 pm  * Development of experimental and computational tools to evaluate metabolic flux. - Costas Maranas (Pennsylvania State University)

What HTP tools are needed to analyze-qualitatively and quantitatively--specific metabolites or constituents of metabolic pathways or networks? What tools are needed to evaluate or predict manipulations of metabolic flux through specific pathways? How can we predict or evaluate corresponding changes in interacting pathways or networks?

4:15 pm  General Discussion

4:30 pm  Adjournment
Abstracts

The BioCyc Collection of 200 Pathway/Genome Databases and the MetaCyc Database of Metabolic Pathways and Enzymes

Peter D. Karp1* (pkarp@ai.sri.com), Christos Ouzounis2, and Sue Rhee3

1SRI International, Menlo Park, CA; 2European Bioinformatics Institute, Hinxton, UK; and 3Carnegie Institution, Stanford, CA
* Presenting author

The BioCyc Database Collection1 is a set of 200 Pathway/Genome Databases (PGDBs) for most prokaryotic and eukaryotic organisms whose genomes have been completely sequenced to date. The BioCyc collection provides a unique resource for metabolic engineering and for global and comparative analyses of genomes and metabolic networks.

Each organism-specific PGDB within BioCyc contains the complete genome of the organism plus the following additional information inferred by the Pathway Tools2 software:

- Predicted metabolic pathways as inferred from the MetaCyc3 database
- Predicted genes to fill holes in the metabolic pathways (pathway holes are pathway steps for which no enzyme has been identified in the genome)
- Predicted operons for each bacterial PGDB
- Transport reactions inferred from the product descriptions of transport proteins by the Transport Inference Parser
- A metabolic overview diagram containing the metabolic enzymes, transport proteins, and membrane proteins of each organism is constructed automatically

The BioCyc collection can be accessed in several ways including interactive access via the BioCyc.org web site, bulk downloading in several formats including Systems Biology Markup Language (SBML) and BioPAX, and querying within SRI's BioWarehouse system for database integration. Most BioCyc PGDBs are freely and openly available to all.

We seek scientists to adopt and curate individual PGDBs within the BioCyc collection. Only by harnessing the expertise of many scientists can we hope to produce biological databases that accurately capture the depth and breadth of biomedical knowledge. To adopt a database, send email to biocycsupport@ai.sri.com.

The Pathway Tools software that powers the BioCyc Web site provides powerful query and visualization operations for each BioCyc database. For example, the Omics viewer allows scientists to visualize combinations of gene expression, proteomics, and metabolomics data on the metabolic map of an organism (see http://biocyc.org/om-expr.shtml). A genome browser permits interactive exploration of either a single genome, or of orthologous regions of multiple genomes. A newly developed set of comparative genomics tools supports many comparisons across the genomes and metabolic networks of the BioCyc collection. See http://biocyc.org/samples.shtml for an overview of BioCyc Web site functionality.

The MetaCyc database3 describes experimentally elucidated metabolic pathways and enzymes as reported in the experimental literature. MetaCyc is both an online reference source on metabolic pathways and enzymes, and a solid foundation of experimentally proven pathways for use in computational pathway prediction. MetaCyc version 9.6 describes 690 pathways from more than 600 organisms. The 5500 biochemical reactions in MetaCyc reference 4800 chemical substrates, most of which contain chemical structure information. MetaCyc describes the properties of 3000 enzymes, such as their subunit structure, cofactors, activators, inhibitors, and in some cases their kinetic parameters. The information in MetaCyc was obtained from more than 8500 research articles, and emphasizes pathways and enzymes from microbes and plants.
References

Development of Computational Tools for Analyzing and Redesigning Biological Networks

Priti Pharkya¹, Madhukar Dasika¹, Vinay Satish Kumar¹, Narayanan Veeraghavan¹, Patrick Suthers¹, Anthony Burgard², and Costas D. Maranas* (costas@psu.edu)

¹Pennsylvania State University, University Park, PA and ²Genomatica, Inc., San Diego, CA

* Presenting author

The incredible growth in recent years of biological data at all levels has provided a major impetus for developing sophisticated computational approaches for unraveling the underlying complex web of protein, DNA and metabolite interactions that govern the response of cellular systems to intracellular and environmental stimuli. Even partial knowledge of these interconnections and interactions can facilitate the targeted redesign of these systems in response to an overproduction objective. In this poster, we will highlight our progress towards the development of computational frameworks aimed at analyzing and redesigning metabolic and signaling networks.

(1) **Metabolic Network Gap Filling**: Existing stoichiometric metabolic reconstructions, even for well studied organisms such as E. coli, include “unreachable” or blocked reactions due to the inherently incomplete nature of the reconstructed metabolic maps. These blocked reactions cannot carry flux under any uptake conditions. In this project we first identify all such blocked reactions and subsequently pinpoint which reactions to add to the existing model to bridge the maximum number of such gaps. The minimal set that accomplishes this task is chosen from an encompassing list of candidate reactions constructed from databases such as KEGG and Metacyc. The developed framework is demonstrated on genome-scale metabolic models of Escherichia coli and Saccharomyces cerevisiae. Reactions with higher BLAST scores against the genome of the curated model are preferentially selected. In addition, information as to which metabolites are present (e.g., CE-MS measurements) can be integrated into the gap-filling procedure.

(2) **Assessing Objective Functions Driving Metabolic Responses to Perturbations**: Genome-scale metabolic reconstructions are increasingly being used to predict the response of metabolic networks to genetic (e.g., gene knock-outs) and/or environmental (e.g., high/low glucose) perturbations. This is accomplished by optimizing an objective function that abstracts the dominant factors driving flux reallocation. These postulated hypotheses include biomass formation maximization, minimization of metabolic adjustment (MOMA)¹, regulatory on/off minimization (ROOM)², etc. In this project, we assess the quantitative performance of these hypothesized objective functions in response to genetic and/or environmental perturbations and propose a new one based on flux ratios rather than absolute values. A comprehensive comparison using experimental data for wild-type and perturbed networks alludes to the use of composite objective functions as the best predictors.

(3) **Elucidating Fluxes in Genome-scale Models Using Isotopomer Labeling Experiments**: Isotopic label tracing is a powerful experimental technique that can be combined with the constraint-based modeling framework to quantify metabolic fluxes in underdetermined systems. The calculation of intracellular fluxes by 13C-MFA is based on the fact that when cells are fed a growth substrate with certain carbon positions labeled with 13C, the distribution of this label in the intracellular metabolites can be precisely determined based on the known biochemistry of the participating pathways. Most labeling studies focus on skeletal representations of central metabolism and ignore many flux routes that could contribute to the observed isotopic labeling patterns. In addition, often times a wide range of flux values could explain the experimentally observed labeling patterns in network areas where the experimental measurements provide low resolution. In this work, we investigate the importance of carrying out isotopic labeling studies at the genome-scale. Specifically, we explore how the activity of multiple alternative
pathways could in many cases adequately explain the experimentally measured labeling patterns and also suggest methods for improving the resolution of quantified fluxes. Finally, we investigate the effects of introducing global metabolite balances on cofactors such as ATP, NADH, and NADPH as their inclusion in labeling analysis is often neglected but may be important for obtaining biologically realistic flux distributions.

(4) Optimal Redesign: Our research group developed the OptKnock and Optstrain procedures for microbial strain redesign through targeted gene additions and deletions. Both procedures use the maximization of biomass to predict flux reallocations in the face of genetic perturbations. Here we will present how to extend these optimization frameworks to account for popular quadratic objective functions such as MOMA and contrast the obtained results. In addition, we will discuss how to computationally integrate modulations (i.e., up or down gene regulations) in addition to knockin/outs in the palette of allowed genetic manipulations for microbial strain optimization.

(5) Signaling Networks: The same pathway modeling concepts that have been extensively applied to analyze and optimize metabolite flows in metabolic networks can also be used to analyze and redirect information flow in signaling networks. Here we describe optimization-based frameworks for elucidating the input-output structure of signaling networks and for pinpointing targeted disruptions leading to the silencing of undesirable outputs while preserving desirable ones. The frameworks are demonstrated on a large-scale reconstruction of a signaling network composed of nine signaling pathways. Results reveal that there exist two distinct types of outputs in the signaling network that either can be elicited by many different input combinations or are highly specific requiring dedicated inputs. Furthermore, identified targeted disruptions are not always in terminal steps. Many times they are in upstream pathways that indirectly negate the targeted output by propagating their action through the signaling cascade.

References
Summary of Metabolic Engineering breakout session at GTL-MEWG 2006

Session Objectives: As described above the session topics were as follows:

- Processes to facilitate integration and organization of genomic sequencing data into pathways and networks.- Peter Karp (SRI)
  
  As whole genome sequence information becomes available for an increasing number of organisms, there is a need to efficiently mine these genomes for specific genes, cofactors, and regulatory factors. How can we improve our ability to identify selected target genes, and organize them into functional metabolic pathways or networks?

- Development of robust methods to identify common regulatory factors between pathways or networks and elucidate pathway interactions and modulations. - Jay Keasling (UCBerkeley/LBNL)
  
  How can genomics help us move from information about components in individual pathways to discover additional constituents of related networks, or common regulators? How can we identify coordinately-controlled networks, or optimize desired metabolic outputs under specific conditions?

- Development of high throughput (HTP) computational methods, metabolic manipulation, and analysis of gene manipulation.- Michael Betenbaugh (Johns Hopkins University)
  
  Although it is possible to generate vast quantities of experimental data, there remains a need for tools to facilitate high throughput analysis of this data. How can genomics enable us to target specific genes, and how can we develop computational tools that will allow us to evaluate or predict manipulations in silico? What HTP tools are needed to experimentally evaluate or validate predicted changes in gene or metabolic manipulation?

- Development of experimental and computational tools to evaluate metabolic flux.- Costas Maranas (Pennsylvania State University)
  
  What HTP tools are needed to analyze-qualitatively and quantitatively--specific metabolites or constituents of metabolic pathways or networks? What tools are needed to evaluate or predict manipulations of metabolic flux through specific pathways? How can we predict or evaluate corresponding changes in interacting pathways or networks?

Proceedings from last year’s meeting were provided at the session and access to them via the MEWG website was also noted.

Fred Heineken (NSF) welcomed participants to the workshop noting that each of the speakers would make a 5 minute presentation and lead discussion from the audience on that topic. The goals of the session were to provide information on where ME is headed and what issues need to be considered. What will be the next research opportunities? Ultimately MEWG would like to issue another call for proposals, based on information from this meeting, and would like to revise/update future calls for proposals.
Mark Segal (EPA) moderated the session. He explained that this is the first Metabolic Engineering Awardee workshop that has not been held at NSF. He encouraged audience participation and questions. The format used was that each speaker made a brief presentation on a particular topic or question followed by audience questions and discussion. A table discussion with all speakers followed the presentations and again, audience participation and discussion was an important component.

Summaries of the presentations:

#1 Peter Karp — SRI

*Processes to facilitate integration and organization of genomic sequencing data into pathways and networks.*

Organization of genome data into pathways and networks

- Summarize state of the art, existing approaches
- Opportunities for new research directions
- Discuss limitations in current approaches

Assigning genes to pathways traditionally involves metabolic pathways. Some kind of score for each metabolic pathway, leading to prediction of likelihood of that pathway in that organism. Genome level inferences and assignments, Biocyc, KEGG, Reactome, VIMSS—leads to inference of pathway hole fillers. Which genes within the genome might encode for missing steps in these pathways? There’s often an ordering of inference steps. Often other high throughput (HTP) data (gene expression, proteomics, metabolomics) are not included in inference of operons, pathways and pathway hole fillers. As mentioned by Christophe Schilling (Genomatica), it can take months to develop a flux balance model. We also need to infer a full kinetic model (not just steady state, as used in FBA) Will be using literature mining tools to extract kinetic parameters of enzymes, etc, to putting together data required to automatically build a kinetic model for an organism from its genome. Also would like to infer transcriptional regulatory relationships—still more work than can be done with contributions from gene expression data alone. Finally, need to integrate regulation at other modes, substrate, degradation, etc.—bioinformatics community has almost completely ignored this problem. There’s a lot of extant data on substrate-level regulation of enzymes—EcoCyc is working on this in E. coli, developing gold standard dataset that can be used to train other types of organisms' regulation.

Limitations in existing algorithms—

- Quality of genome annotations
  - False positives
  - False negatives (ORFs and missing multiple functions)
  - Lack of controlled vocabulary in many genome annotations
  - Lack of probability values in genome annotations
  - Many enzymes within pathways can never be present in a genome annotation, because none of them have been sequenced. Only know about them due to biochemical characterization.

There was a workshop last summer, talking about how to get partnership to improve genome annotations for prediction of genes of unknown function.
Inference of metabolic pathways
  o Prediction of novel pathways
  o Pathway databases don’t yet contain all experimentally elucidated pathways
  o Choosing among multiple pathway variants
  o Lack of experimental testing of predicted pathways; results would likely lead to improvements in prediction algorithms. There’s a limit to how far you can go with predictions without feedback (experimental testing).
  o Pathway curations. What is the best way to do this—one group or multiple groups?

Comments and Questions

Costas Maranas — issue of incorporating regulatory information—implying Boolean relation, quantitative/weighting relation? Possibly both—in the cell it’s quantitative, but Boolean may be the best first approximation.

Connecting gene regulation and metabolic engineering—feedback regulation. An audience member asked if there are there any methods where we can achieve metabolite measurements similar to the level of gene expression measurements. A need for metabolic engineering is high throughput (HTP) enzyme or metabolite production. Peter Karp suggests using sequence information and 3D structure analysis to predict substrates and feedback inhibition, etc. for a given enzyme. Transfer that information by analogy? If two enzymes display sequence similarity, can you assume they have substrate inhibition analogy? If we could do this computationally, it would save a lot of experimental effort.

An audience member offered: Talking about a need for a repository for genetic information. Many times it’s not quite clear what the physiological conditions are when data is being collected. Information on developed mathematical techniques for extrapolating.

Mike Betenbaugh—what are the issues of genome annotation? What are the pitfalls of the current systems? Multifunctional genes are often underannotated and the second, third, fourth, etc, functions are not noted. Consequently, we don’t see all the evidence for a pathway. Matching enzyme functions in genome annotation with function in enzyme name. He has never seen a genome annotation score and the probability of prediction for a specific enzyme.

Brian Davison—what about incorporating “fuzzy” numbers in kinetic models? There are some programs in development—would an investment accelerate this development? Would make more sense to slice up into individual functions, shoot for smaller kinetic models and connect experimental information to constrain values of parameters.

Pat Dennis—what progress in detection of metabolites? Can now detect up to 1000 metabolites/organism. Might take 5 years guesstimate.

Michaelis-Menten may not be most appropriate mechanism to design large scale models—there are alternatives Ln-Log approximation (used in Europe).

Comment about gene families. A lot of times with gene families, it’s very difficult to make specific assignments and need to determine function experimentally. For example, sometimes you end up with 20 sugar kinases—particularly if you’re not working in an organism closely related to E. coli. Can’t assign a particular function; 200 ABC transporters, difficult to knockout one and assign function according to phenotype.
Annotation challenges—we see numbers that are misassigned. Focus on something using just sequence information. How reliable are these public-accessible sequences?

How to extract pathways from microarrays, interactions.

Annotation is a problem but lack of a unifying hypothesis on what controls metabolic flux another important problem. What are we actually looking for in controlling metabolic flux? Series of enzymes that might constitute a pathway—unclear if we have 3 enzymes acting on a same substrate in a single organism—can’t predict Kms from structure or proteomics or gene expression.

#2 Jay Keasling — UC Berkeley/LBNL

*Development of robust methods to identify common regulatory factors between pathways or networks and elucidate pathway interactions and modulations.*

Extracting from talks given by Michael Laub—specificity of kinases even when sequences very similar, Lucy Shapiro—profiling to look at protein-protein interactions. Why should we care?: Because of (1) Basic science and (2) production of metabolites. For the Science aspect, and how organisms regulate subcellular processes. Production of metabolites can be either natural or unnatural to a cell; regardless, an interaction (and regulation) occurs. For engineers, regulation is something that comes up unexpectedly—even if it’s a normal metabolite, where or when it’s being made constitutively.

- Discovery
- Bioinformatics
- Sequence homology
- Network learning approach’

Bioinformatics: sequence analysis, network learning approach, metabolic model

Experimental: “-omics”, transcript profiling, protein-DNA interactions, protein-protein interactions

Methods have been developed for deducing protein-protein interactions, protein-DNA interactions. How can we design coordinately controlled network, or optimize desired metabolic outputs under specific conditions? Little information on interactions of proteins in metabolic pathways. Few methods for protein-metabolite interactions, which may be most important for regulatory metabolism.

Natural language processing, How can we move beyond genomics? Typically, biotech is performed in the absence of the type of information available in his artemisinin project—because the information is so new, don’t yet know how to use it well.

Dupont work on 3PG—Jay Keasling is making an intermediate that has a toxic effect, but wouldn’t have been able to figure this out until he used transcriptional profiling to see what are the enzymes that are affected in those pathways.

What about bag of tricks useful in engineering organisms?? Use of heterologous pathways to get around limitations in homologous host? Can help, don’t have to worry about some of the native regulations, interactions with native host regulatory repertoire. But now the flip side is the total lack of balance, lack of regulation—HMG-CoA not an issue in yeast, but bottleneck for E.
coli. Ideal approach would be metabolomics hopefully but until techniques are robust and widely used, not very useful. Need “instruction set” for pathway to function properly.

Tom Jeffries mentions taking genes from Pichia and putting them into yeast.

We need Tools for integrating gene sequence with transcriptome and proteome analysis and with metabolic function—take protein profiling, metabolite profiling, flux profiling and expression profiling—plug into (e. g.) Peter Karp’s system.

Can account for cis-regulatory elements—in principle, these algorithms can be applied here.

#3 Michael Betenbaugh — Johns Hopkins University

_high throughput computational methods, metabolic manipulation, and analysis of gene manipulation._

Although it is possible to generate vast quantities of experimental data, there remains a need for tools to facilitate high throughput analysis of this data. Need both experimental and mathematical tools to predict and to test predictions.

1. How can genomics enable us to target specific genes? What are the experimental tools of genomics? How can these specify particular genes (or groups of genes) of interest? How to identify genes to up regulate, down regulate? Are the tools appropriate? Fischer: Biotech. Annual Review, 2005—some examples of tools
   a. Shotgun sequencing to provide genome
   b. Microarrays, oligont chips transcriptome
   c. 2D gels, ICAT, MS proteome
   d. GC/MS metabolome
   e. Yeast 2-Hybrid screens, TAP interactome
   f. Gene inactivation, knockouts, RNAi phenome

2. What computational tools will allow us to evaluate and/or predict manipulations in silico? a. Databases—genomic-proteomic-microarray-metabolome
   b. Analytical tools—to demonstrate quality and importance of data; how to link these tools together?
      i. Normalization and data quality
      ii. Inference
      iii. Classification
   c. Computational models
      i. Regulatory networks
      ii. Metabolic models

3. What HTP tools are needed to experimentally evaluate or validate predicted changes in gene or metabolic manipulation? Case study, reverse transvection (96 well format) number of different genes, on top of this lay down collection of cells to analyze collection of different genes. Analyze for particular expression of marker protein or functional assay. This is just one example of a gene and its function.
   a. cDNA microarrays
   b. Protein chips and proteomics
   c. Metabolic assays
d. Cell-based functional assays
   i. siRNA
   ii. Overexpression
   iii. Functional assays
   iv. Electron microscopy and localization assays

Questions:

Primary tool is DNA microarray data—what can you do with microarray data? What are the limits? They can provide function, elucidate regulatory mechanisms. They cannot translate to protein, so we need to get to an integrated genomics system where we don’t just rely on a single technology platform. Need to include proteomics, interactions. Suggests that RNA can be a good first measure, particularly for metabolites (?)

Tom Jeffries: Where we have looked at it, generally metabolites present in low concentrations, if you increase enzyme activities, you will increase levels of metabolites.

Everything is focused on we have a hammer, how can we wedge it into E. coli. What kinds of tools are available for the native organism? Metabolomics can be used on any growing system, as an example. What if you don’t have an E. coli, yeast, or Shewanella?

Are there specific other tools that we should be looking at besides what’s listed here? Enzyme activity in vivo and in vitro—emphasize that functional activity is a glaring absence, specifically with respect to Km and Vmax, enzyme kinetics, in HTP format.

Mark Segal asked about the prospects for looking at complex systems via reverse engineering as touched on by Tim Gardner. How do you determine novel pathways that may not be complete in single organisms but may be functional through cooperation within a close knit community? That is, reverse engineering to deduce sequence from function—what are the prospects for that in this context? What is the future of that approach? Betenbaugh answered that we are having a hard enough time working with single organisms to consider looking at community functions, but that the use of reverse engineering using microarrays as a tool was something worth exploring to help us with pathway elucidation. Crossing platforms between organisms—what makes it work, then try to find homologs elsewhere. That’s the analogy to what Peter Karp mentioned earlier—filling in holes in metabolic map. That’s the reverse genetics using this to fill in functional holes—then back out what genes are possible. Gets to the question of how do you get to a novel pathway? How do you discover this? Discovering novel pathways is a great challenge for bioinformatics. Looking at covarying metabolites? That’s the way a lot of pathways were discovered historically—looking for a specific metabolite and then figuring out where it came from. Look at it as a form of Artificial Intelligence—reverse engineering from a metabolomics vantage.

Pat Dennis—if you can identify all the end products of pathways, then you know there has to be a pathway. Then you can work backwards. If you know how much an end product is there, and you can grow cells and determine flux, combine information about how much of each product you should be making due to estimates of cell growth and relative abundance. Measure how much glucose is used per cell division. You can build a comprehensive metabolic flow chart just based on this. Now say we want to look at minor components of cell. Isn’t that FBA? If you double the cell mass, you assume you need to double all those metabolites. But biomass
production is an optimizing principle, says that configuration of best adapted bacterium maximizes output. But you can use labels, feed labeled metabolites (C13-labeled glucose in the C-1 position)—yes, still need basic biochemistry.

If you propose to go to organism X and look at all the metabolites. Then we’ll back-engineer and look at the pathways. This doesn’t sound fundable. How can you phrase this in a scientific way?

Are the computational and modeling tools sufficient?

Paul Schlosser and J Bailey—1991 paper. FBA assumes that none of the outputs in the pathway can affect inputs back in the pathway. This means people build metabolic pathway models, but unless they’re in the context of whole cell model, may not give you a realistic prediction. Need a coarse grain metabolic cell model. Also, kinetic descriptors—an approximation (easier to handle than specific parameters for each enzyme)—for whole cell might be sufficient. Sensitivity analysis of which parameters are important is the key to analysis.

What about other models other than metabolic models? Consistent approach to coarse graining—focusing on specific functions of the cell, trying to split up into pieces that can be modeled separately. Is that being done in a fashion such that everyone agrees on which are the pieces? There is a long tradition of building small metabolic models—whole range of glycolytic models, photosynthetic models, amino acid biosynthetic models—viable strategy. Many of these are very good. Piecemeal or modular modeling is successful—coarse grain model with embedded fine-grain models. Our old E. coli model is an example of coarse grain model—modules embedded in it. Whole cell model provides framework for smaller, fine-grain models.

What HTP tools are needed?

Metabolite screening—not specifically genomic tools. Bottom line is that complexity of the system is so great that it’s difficult to predict from first principles where the outputs will be. Sounds as if not that much progress has been made compared to last year when this question was first brought up. It’s much easier to write a proposal that’s directed in a specific fashion—you need to develop selections, strain adaptation evolution. Need for development of general genetic selections for metabolites; from an engineering approach, seems most useful method. Still, even if you use rational selection to identify something that works, fits more in the center-based framework than in the individual laboratory.

Christophe—this might be driven by specific products and the assays you can develop for them. Quantitative physiology, fermentation studies—to be able to assay results of quantitative predictions and see how well they work. Most of the time we can only perform qualitative comparisons—perhaps there’s not as much training in this area (less popular now).

What about microbio reactors—quantitative physiology requires a platform to achieve a large amount of data in a short amount of time. Not exactly HTP, but higher throughput than previously possible. Somewhat scalable—separation processes, disposable chips. Microreactor, commercial product, is here at the poster session. If you had an SMA20, running on a 96well chip and pulling off metabolites, could probably do a fair amount of screening.
Development of experimental and computational tools to evaluate metabolic flux

Beyond genomics and transcriptomics—what the cell has actually managed to accomplish, vs. what you would like it to do. Dominant method is C13 labeling—feed labeled molecule (either all carbons or fraction) to cell, undergoes metabolism, then deconvolute where the labeled carbons end up in the slew of compounds you subsequently isolate. GC/MS or NMR are traditional analytical methods—most models are biased and you have to interrogate them, perform single deletions, feed different substrates—ask whether flux predictions of your FBA or kinetic model are correct and measurable. This is very important in ME projects—heart of metabolism. This is not a new area, publications in 1994 for initial modeling approaches by Stephanopoulos, Christensen & Neilsen 1999 first to combine labeling and GC/MS, Schmidt et all, 1999 incorporates isotopometer analysis using NMR spectra. A number of people contributed optimization algorithms—this is possible to generate model that combines limited scope metabolic models (30-50 rxns). For genome-scale models (1000 rxns), measurables do not always elucidate unique flux distributions. Many relevant pathways are absent—how much ambiguity still exists? All of a sudden adding in a new set of players/metabolites, affects utilization of cofactors. A couple of orders of magnitude makes things much more difficult.

Large-scale non-linear programming problem, Burgard & Van Dien) have already created Isotopomer mapping matrices for genome-scale models of E. coli. How is error propagated, especially as you scale. IMM you have a multiple of different reactants and products—IMMs map all the different variants, tell you that this specific variant of a reactant will generate a specific variant of a product. How is genome-scale IMM constructed? Automatically or by hand? Mixture of both. Starting point is purely computational, but models inadequate so manual curation contributed.

Question
1. What are the bottlenecks in broadening the use of HT data for the quantitative estimation of metabolic fluxes? Hundreds of papers—flux measurements are far more useful than microarray measurements, so why are they not used at the same frequency? What are the barriers to their use?
2. Given HT data, how can we use computations to reliably elucidate fluxes in metabolic networks?
3. How can we anticipate the effect of environmental and/or genetic manipulations on metabolic fluxes? Too expensive to perform all iterations, so how can you empower your model to make better predictions than you’ve seen before?

Can IMMs be seen as a stoichiometry matrix in which every species appears in every labeled fashion? View it as a roadmap for the flavor of conversion of A to B—if first carbon is labeled, what kind of product mixture will you get? Instead of a model with 1000 reactions, it’s a model with 1 million reactions. Problem: the production rates of each reaction are not independent—depends on initial pool, different labeled substrates will have to react the same way—if not, non-linearity is introduced. So do these labeling experiments resemble steady state? Unsteady state is not trivial.

Wim Vermaas—You have GC/MS, NMR—there are other ways of monitoring metabolites (CE-MS, just to see what’s there). GC/MS depends on what’s in the database, situation with Arabidopsis was that they saw 5000 metabolites that they had no clue what they are. How to solve this? You can’t just look at metabolites that you know about—CE-MS gives you the entire
mix of things that are there. This is especially relevant for things that are species specific and present in low concentration. What methods are available on the horizon for this? One thing that’s possible is what a group in Japan is doing—they identified all metabolites (qualitative) in B. subtilis, then used this information in flux elucidation. Nontrivial, not a ready answer. That might be approachable using very large (open?) chemical databases becoming available? PubChemDB, for example. We see an interesting compound with a wonderful GC-MS spectrum, doesn’t resemble anything in the databases—there’s a group in Northwestern that’s looking at identifying novel molecules with a Lego block kind of free energy minimization to predict stable possibilities—upper bound on the number of species. But 20 years ago at Stanford they built a system to predict how a small molecule would fragment in a MS—could envision doing a similar thing to predict spectrum of the given molecule in other kinds of devices. Going back to IMM—computational challenge is to create automatic IMM at genome-scale. Issues with carbon scrambling/flipping ends up with a cocktail of variants—doesn’t happen all the time. Can be pushed, but the curation step is still necessary. Perhaps you could devise a program to predict this scrambling, if it’s occurring at key spots? At the end of the day, you want to resolve different alternatives between different flux distributions—labeling must be done so that your end result spectra will unambiguously resolve between these two hypotheses. This is more expensive than labeling all carbons.

What is the upper bound on the number of metabolites in E. coli? Database says roughly 800—clearly must be more. At the moment, can experimentally identify 1200; B. subtilis is over 1000 so reasonable to expect that E. coli is in that range as well. We went through inflationary period trying to generate ever larger models—many fewer reactions are actually traversed in practice—need to eliminate unnecessary reactions, collapse it down to relevant reactions.

Regarding expanding the complexity of the model—is the limitation in the model or in the experimental side. Are you ever going to be able to verify this expanding model? Is it making bigger model for subsystems? The moment you use a large model, computationally it becomes intractable to do FBA. Even if your algorithm works perfectly, there are many different flux equations that explain the data equally well—next step is what will be the experiment to eliminate half of the remaining candidates. How to design next set of experiments? Label glucose at a specific position? Acetate? We have a number from looking at the number of genes in pathways—is there a number to predict about unstable intermediates that are not dependent on enzyme activity? How to correct this? Nonenzymatically driven reactions—difficult to include these in your model. Palsson has the only one who does a bit on these spontaneous reactions; Christophe Schilling confirms this.

Can try to feed in your model all the experimental observation, tell your model these metabolites must be present. What’s the minimal number of reactions that allow all these metabolites to be present—constants of ratios? Maximization of biomass as driver for FBA—is there anything better? Preservation of certain fluxes or genetic manipulations is a better descriptor when compared to biomass drivers. You minimize the ratio of violations.

**Future Directions**

To summarize where we need to go—what should be in the next MEWG solicitation?

Suggestions:
1. algorithm for predicting novel metabolic pathways(s) from genomics and other –omics
2. algorithms for automatic generation of FBA and kinetic models
3. how to elucidate genome-scale fluxes
4. methods for metabolomics—large scale
5. anything that allows us to measure fluxes cheaper and faster
6. ways of using HTP to understand metabolism and steer metabolism
7. metabolomics tools
8. making downstream omics more workable with analogy to microarrays systems so we have same level of detail at metabolite, enzyme, functional scale as we do for genomics/microarrays
9. reverse engineering (reverse genomics)
10. does in vitro kinetics apply to in vivo kinetics? Would make life easier for experimentalists. There have been publications that show they are not the same, at least in yeast. If you have to tweak your models too much, you’re not being consistent with the original parameters of the in vitro model.
11. how do you engineer a regulatory matrix on enzymes you want to introduce into a pathway, to effect the output you want (how do you introduce promoters, etc to get regulation to result in desired product)? Large scale synthetic biology—5 different promoters on 6 different genes—how can you do this in a cost-effective manner? Need better tools. Chip synthetics? How to generate background hosts into which you’d put these constructs (deletion modifications on a large scale)?
12. How to create novel pathways?
13. How to perform large scale adaptive evolutions? Why are some organisms better to use?
14. predictive process modeling? Integrating biological models with process models
15. infrastructure incentives—examples are EcoCyc, BioCyc databases. Becoming indispensable tools—how to develop more of these kinds of tools.
16. facilitating assembling modules into a whole cell model
Organization of Genome Data into Pathways and Networks

Peter D. Karp, Ph.D.
Bioinformatics Research Group
SRI International
pkarp@ai.sri.com

- Summarize state of the art, existing approaches
- Opportunities for new research directions
- Limitations in current approaches
Assigning Genes to Pathways

- Genome Sequence
- Gene Expression
- Proteomics
- Metabolomics

- BioCyc
- KEGG
- Reactome
- VIMSS

- Infer metabolic pathways
- Infer pathway hole fillers
- Infer FBA model
- Infer kinetic model

- Infer operons
- Infer transcriptional regulatory relationships
- Infer other modes of regulation

Limitations in Pathway Assignment

- Inference of metabolic pathways
  - Quality of genome annotations
    - False positives
    - False negatives (ORFs and missing multiple functions)
    - Lack of controlled vocabulary in many genome annotations
    - Lack of probability values in genome annotations
  - Many enzymes within pathways can never be present in a genome annotation – never sequenced
**Experimental/Computational Partnership**

*To Improve Genome Annotations*

- Focused effort proposed to
  - Experimentally verify computational predictions of functions for genes of unknown function
  - Seek which genes encode functions with no associated sequence
  - Capture computational and experimental results in common database

- Roberts, R.J., Karp, P.D., Kasif, S., Linn, S., and Buckley, M.R.  

  http://biology.plosjournals.org/plosonline/?request=get-document&doi=10.1371/journal.pbio.0020042

- Karp, P.D., "Call for an enzyme genomics initiative" Genome Biology 5:401.1-3  

---

**Limitations in Pathway Assignment**

- Inference of metabolic pathways
  - Prediction of novel pathways
  - Pathway databases don't yet contain all experimentally elucidated pathways
  - Choosing among multiple pathway variants
  - Lack of experimental testing of predicted pathways; results would likely lead to improvements in prediction algorithms
Curation of Organism-Specific Pathway Models

- Centralized in a single group?
- Distributed across many groups?
- Automated mining of pathways from the literature
Development of robust methods to identify common regulatory factors between pathways or networks and elucidate pathway interactions and modulations

Jay Keasling
UCBerkeley/LBNL

Questions

• How can genomics help us move from information about components in individual pathways to discover additional constituents of related networks, or common regulators?

• How can we identify coordinately-controlled networks, or optimize desired metabolic outputs under specific conditions?
You saw the following talks:

- Michael Laub – protein phosphorylation
- Lucy Shapiro – cell cycle and regulation
- Andrew Emili – functional proteomics
- David Hill - interactome
- Tim Gardner – transcription and metabolic networks

Why should we care?

- Basic science
  - Organism-organism interactions and interactions of organisms with their environment
  - Stress response
  - Cell cycle
- Production of metabolites
  - Makes pathway regulation difficult
  - Reduces product formation
    - Understanding interactions could improve product titers/yields
Discovery

• Bioinformatics
  – Sequence homology
  – Network learning approach
  – Metabolic modeling?

• Experimental methods
  – Omics
    • Transcript profiling
      – E.g., Lucy Shapiro’s array work on cell cycle
      – E.g., Tim Gardner’s work on *E. coli*
    • Protein profiling
    • Metabolite profiling

• Experimental methods (continued)
  – Protein-DNA interactions
    • DNA footprinting
    • Chromosome Immuno-Precipitation (ChIP)
  – Protein-protein interactions
    • Protein tagging with fluorescent proteins
    • Protein complex pull downs (TAP tags)
    • Knock-outs and overexpression
    • Protein phosphorylation
    • Two-hybrid
Missing data/methods important for metabolic engineering

• Methods have been developed for deducing protein-protein interactions, protein-DNA interactions, etc.
  – Little information on interactions of proteins in metabolic pathways

• Few methods for protein-metabolite interactions, which may be the most important for regulating metabolism.
  – How would you deduce these interactions in high throughput … or even in low throughput?

Questions

• How can genomics help us move from information about components in individual pathways to discover additional constituents of related networks, or common regulators?

• How can we identify coordinately-controlled networks, or optimize desired metabolic outputs under specific conditions?
High Throughput Computational methods, Metabolic manipulation, and Analysis of gene manipulation.

Although it is possible to generate vast quantities of experimental data, there remains a need for tools to facilitate high throughput analysis of this data.

MEWG Workshop Topic 3
Feb. 14th 2006

Moderator: Michael J. Betenbaugh
Johns Hopkins University

Question 1: How can genomics enable us to target specific genes?

What are the experimental tools of genomics?

How can these specify particular genes of interest?
Question 1: How can genomics enable us to target specific genes?

Fischer: Biotech. Annual Review, 2005

Question 2:
What computational tools will allow us to evaluate and/or predict manipulations in silico?

- Databases
  - Genomic-Proteomic-Microarray-Metabolome

- Analytical Tools
  - Normalization and Data quality
  - Inference
  - Classification

- Computational Models
  - Regulatory Networks
  - Metabolic models
Question 2:
What computational tools will allow us to evaluate and/or predict manipulations in silico?

Question 3:
What HTP tools are needed to experimentally evaluate or validate predicted changes in gene or metabolic manipulation?

- cDNA microarrays
- Protein chips and proteomics
- Metabolic assays
- Cell based functional assays
  - siRNA
  - Overexpression
  - Functional assays
  - Electron microscopy and localization assays
- In vivo models
Question 3:
What HTP tools are needed to experimentally evaluate or validate predicted changes in gene or metabolic manipulation?


**Bibliography**

Cell based assays using reverse transfection of RNAi
4. Development of experimental and computational tools to evaluate metabolic flux

Labeled Isotopes
- e.g. C\textsubscript{13} glucose

\begin{itemize}
\item \textbf{E.Coli cell}
\item \textbf{C\textsubscript{13} destinations}
\item \textbf{Observations}
\end{itemize}

\begin{itemize}
\item \textbf{Flux Elucidation}
\item \textbf{Publications}
\begin{itemize}
\item Isotopomer analysis using GC/MS (Christensen & Nielsen, 1999; Fischer & Sauer, 2003)
\item Isotopomer analysis using NMR spectra (Schmidt et al., 1999)
\item Computational models for flux elucidation (Zupke et al., 1994; Wiechert & Graff, 1996; Wiechert et al., 1996; Mollney et al., 1999)
\item Optimization algorithms (Phukompikul et al., 2001; Ghosh et al., 2004; Rascos et al., 2005, Zamboni et al., 2005)
\end{itemize}
\end{itemize
**Limitations**

- Employed metabolic models are of limited scope (i.e. 30-50 rxns)
- For genome-scale models (~1,000 rxns), measurables do not always elucidate unique flux distributions

Many relevant pathways are absent...

---

**Elucidating Fluxes in Genome-Scale Models using Isotopomer Labeling Experiments** *(Poster no: 92)*

- Enable flux elucidation for genome-scale models

Isotopomer mapping matrices (IMM) have been constructed genome-scale models of *E coli* (Burgard & Van Dien)

Large-scale non-linear programming problem

Minimize

\[
\sum_{i} \sum_{k} (I_{ik} - I_{ik}^{exp})^2
\]

s.t.

\[
\sum_{j \geq 0} \sum_{i \geq 0} \sum_{k \geq 0} \text{IMM}_{ijk} I_{ik} = 0 \quad \forall i, k
\]

Isotopomer balance

\[
\sum_{j} S_{ij} v_j = 0 \quad \forall i
\]

Mass balance

\[
\sum_{k} I_{ik} = 1 \quad \forall i
\]

Isotopomer fraction balance
Question 1: What are the bottlenecks in broadening the use of HT data for the quantitative estimation of metabolic fluxes?

Microarrays ➔ GC-MS ➔ Glucose ➔ High Throughput Data ➔ Computational Approaches ➔ Elucidation of Metabolic Fluxes ➔ Environmental and/or Genetic Perturbations

Question 2: Given HT data, how can we use computations to reliably elucidate fluxes in metabolic networks?

Microarrays ➔ GC-MS ➔ Glucose ➔ High Throughput Data ➔ Computational Approaches ➔ Elucidation of Metabolic Fluxes ➔ Environmental and/or Genetic Perturbations
Question 3:
How can we anticipate the effect of environmental and/or genetic manipulations on metabolic fluxes?
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:

1. An agricultural production system that is highly competitive in the global economy.
2. A safe, secure food and fiber system.
3. Healthy, well-nourished population.
4. Greater harmony between agriculture and the environment.
5. 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.
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 Cl20, 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 fits 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.