

# Importance of systems biology in engineering microbes for biofuel production

Aindrila Mukhopadhyay, Alyssa M Redding, Becky J Rutherford and Jay D Keasling

## ABSTRACT

Microorganisms have been rich sources for natural products, some of which have found use as fuels, commodity chemicals, specialty chemicals, polymers, and drugs, to name a few. The recent interest in production of transportation fuels from renewable resources has catalyzed numerous research endeavors that focus on developing microbial systems for production of such natural products. Eliminating bottlenecks in microbial metabolic pathways and alleviating the stresses due to production of these chemicals are crucial in the generation of robust and efficient production hosts. The use of systems-level studies makes it possible to comprehensively understand the impact of pathway engineering within the context of the entire host metabolism, to diagnose stresses due to product synthesis, and provides the rationale to cost-effectively engineer optimal industrial microorganisms.

## INTRODUCTION

Global warming, energy security, and increased energy needs in the face of dwindling petroleum reserves are driving forces behind the renewed interest in production of fuels and other petroleum-derived chemicals from lignocellulosic biomass [ 1–3]. Microorganisms from diverse environments naturally produce a wide variety of chemicals that have found use as fuels, commodity and specialty chemicals, polymers, and drugs [ 4,5]. While increases in the production of these natural products were limited to mutagenesis and selection before the advent of recombinant DNA technology, metabolic engineering has been used extensively over the past two to three decades to increase production of these useful chemicals [ 5,6]. More recently, there has been much interest in the microbial production of chemicals found naturally in plants and other organisms and the production of unnatural chemicals not found in any organism. Production of advanced biofuels (e.g. alkanes, alkenes, and aromatics) by microorganisms will require a significant retooling of their metabolism [ 4,7,8].

The ability to manipulate microbial metabolism requires a systems-level understanding of metabolism and the ability to monitor and manipulate many variables simultaneously. The ever-increasing databank of sequenced genomes provides the metabolic toolkit needed to build large numbers of metabolic pathways, producing thousands of chemicals (in the first quarter of 2008 alone, 47 new microbial genome sequences were released, bringing the tally of sequenced microbial genomes to 674 since *Haemophilus influenzae* was completed in 1995 [ 9]). The ability to manipulate and evolve metabolic enzymes also serves to increase the number and types of chemicals that can be produced microbially.

Increasing the yield of a natural product or introducing genes that allow the production of unnatural products can have severe consequences on host growth and thereby impact product titers. Having the ability to monitor only one or a few, often gross, parameters, such as growth rate, meant that problem diagnosis and repair was time-consuming, if not impossible. With the advent of functional genomics techniques, we now have the ability to monitor thousands of parameters simultaneously [10,11<sup>33</sup>]. However, diagnosing and fixing such problems requires system-wide models of genetic regulation and metabolism and computational tools to analyze the vast amounts of data, an area commonly referred to as systems biology [ 12]. In this review we discuss the importance of this powerful emerging science for engineering microbial hosts in industrial applications and the unique challenges encompassed therein.

## Methods in systems biology

The study of gene expression levels using high-density DNA microarrays has become a cornerstone of many functional genomics pipelines, as such information provides context for interpreting genomic sequence data. A staggering amount of array data now exists for hundreds of different microbial and non-microbial systems, providing cell-wide data across a dizzying collection of stressors, toxins, mutations, adaptations, and growth conditions. More recently, microarray data have been used in combination with mutation studies to provide more perspective for changes observed [ 13]. High-density DNA array technology has also been used to enable rapid profiling of microbial diversity [ 14] and gene function [ 15] from complex, environmental samples. Attempts to process and use the tremendous amount of data

generated from transcriptional array studies have given rise to entirely new disciplines within the fields of bioinformatics and computational biology [ 12], and necessitated the development of rigorous reporting protocols [ 16].

Proteomics also has seen significant advances in the past decade and is the primary method to capture regulation of cellular response beyond the mRNA level. While traditionally completed with two-dimensional electro-phoresis [ 17], high-throughput identification of proteins using mass spectrometry has become the foundation of much proteomics research for functional genomics studies. Multiple strategies now exist for the identification of proteins [ 17,18], and numerous workflows to quantify protein abundance have also been developed [ 19]. Notably, the iTRAQ technique, which uses isobaric tags to differentially label samples from up to eight conditions simultaneously [ 20], has been used to survey both protein changes [ 21] and to determine members of protein complexes [ 22]. Improvements in computational analysis are important to reduce artifacts in information and false positives [ 23,24]. However, standardized methods that incorporate statistically sound analyses for reporting and handling proteomics data are still evolving [ 25,26].

Transcript and protein data have been demonstrated to correlate relatively well for specific pathways [ 12] but often correlate poorly in cell-wide analyses [ 27,28]. These results highlight the importance of regulatory processes beyond transcriptional control. An elegant set of experiments demonstrated that in *Saccharomyces cerevisiae*, both mRNA synthesis and degradation significantly impact protein levels [29]. Studies have also demonstrated that stressors likewise affect mRNA stability [ 30]. Advances in proteomics technologies have revealed the importance of post-translational modifications [ 31], protein localization [32,33], and protein-protein interactions as some of the most important cellular responses to stress [ 34]. Functional analysis of *S. cerevisiae* genes showed that candidates essential for growth under a particular stress condition did not show the most significant change [ 13], further emphasizing the importance of responses beyond differential expression.

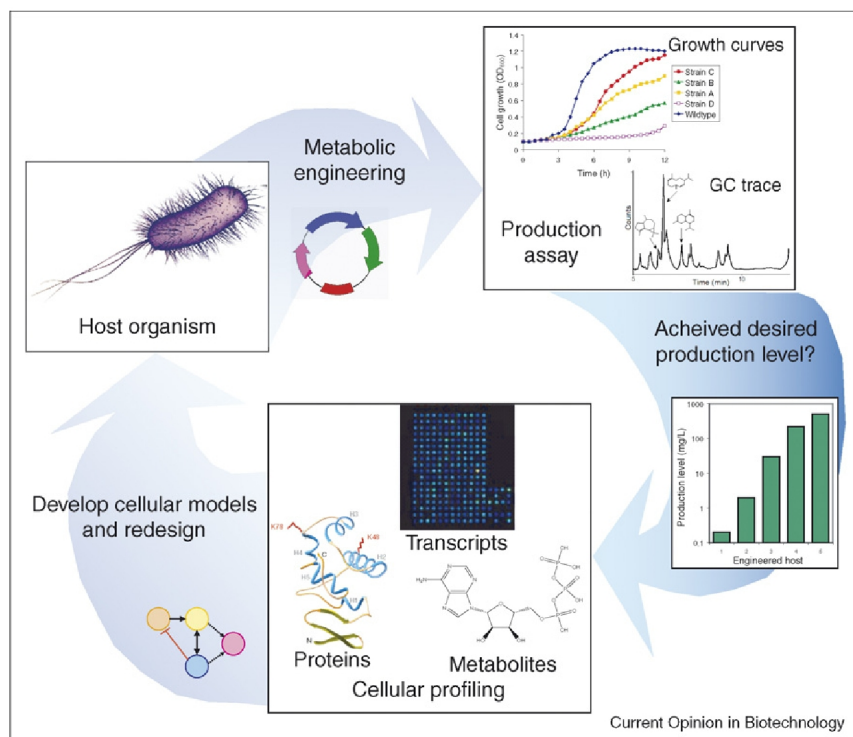
In light of these complexities, it is becoming widely understood that monitoring mRNA and protein levels alone may not be sufficient to capture information about enzyme activity and substrate turnover. Additional cell-wide studies, such as metabolomics [ 35] and flux analysis [ 36], are fast growing into well-established workflows that provide a truly orthologous measurement of the cellular response to stress and manipulation. Advances in the field of metabolomics have made it possible to identify thousands of metabolites [ 37] and quantify tens to hundreds of metabolites in targeted sets [ 38]. Similarly, metabolic engineering based on flux measurements has been used to improve the yield of metabolic products [ 39]. For metabolic pathway engineering specifically, the measurement of carbon flux through central biosynthetic pathways was identified as a key requirement. New techniques are being developed to model pathway fluxes without relying on steady-state assumptions [ 40]. Flux analysis has also found surprising importance in the discovery of new pathways in organisms thought to be fully annotated [ 41].

The complexity of cellular regulation poses a great challenge for the integration of these system-wide data into models that will accurately represent and eventually predict physiology. However, integrated models have been used to identify essential genes [ 42], transcriptional elements [11<sup>33</sup>], regulatory circuits [ 43], and stress response [ 10] in many microorganisms. Such existing models provide valuable information about cellular activities and may be used to pinpoint responses that prevent optimal production. Some very promising recent computational methods have attempted to address issues with data integration [ 44]. Open source tools that allow researchers to analyze, compare, and mine genomics data [ 45,46] are likewise valuable. Model development forces a more complete understanding of cellular regulation, beyond what is able to be deduced solely on the basis of genome annotation.

### **Applicability of functional genomics to metabolic engineering**

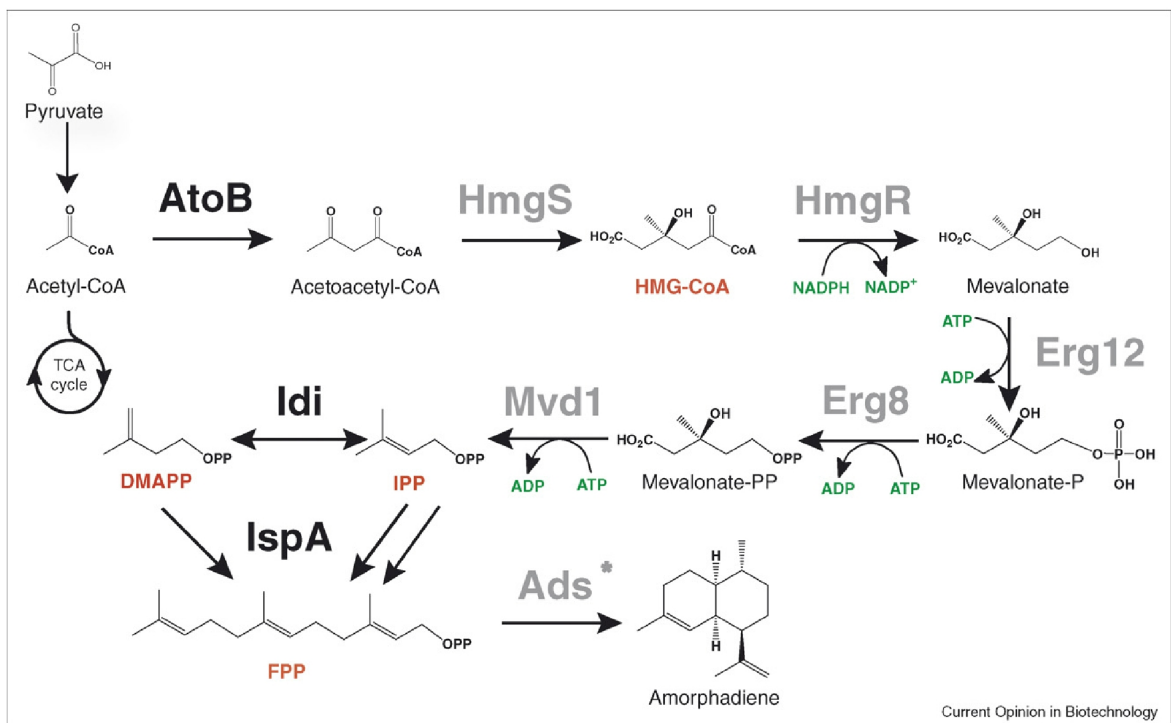
The desire to increase product titers and minimize the cost of precursors provides the primary driving force behind most engineering efforts, particularly for low-margin products like commodity chemicals. Commercially viable titers for bio-products can range from several mg/L in the case of pharmaceuticals [ 47] to hundreds of g/ L for commodity chemicals and biofuels [ 48,49]. Reaching these production levels requires a significant amount of pathway optimization after constructing the pathway and eliminating or minimizing undesirable effects of the pathway or production on the cell. Strain development becomes an iterative process whereby pathway manipulation is followed by a system-level study to identify potential bottlenecks and reveal detrimental side effects ( Figure 1). Once it has been successfully demonstrated that a product of interest can be produced in vivo, achieving economically viable production

levels requires mini-mizing the generation of less desirable side products and maximizing carbon flux toward the target product. For example, improvements in bio-ethanol production in *E. coli* have utilized many such steps, and this progress has been very well reviewed [ 6].



**Figure 1.** Flowchart of systems biology applied to achieving a production target. A host organism must be selected and metabolically engineered to produce the molecule(s) of interest. The results of metabolic engineering must be evaluated, for example monitoring phenotypes by performing growth curves and product titer using gas chromatography (GC). Cellular profiling, using measurements at all levels of the system includes mRNA, proteins, metabolites, and flux analysis to identify bottlenecks in the pathway. Once bottlenecks are identified, another round of engineering must take place to overcome the limitations. The cycle is completed once the desired production level has been achieved.

Most metabolic engineering efforts use a combination of native and heterologous genes, such as the production of the anti-malarial drug precursor artemisinic acid [50], 1,3-propanediol [49], and, very recently, isobutanol production in *E. coli* [51]. Understanding how the incorporation of an engineered exogenous pathway perturbs the host system is important for overcoming pathway bottlenecks. For example, overexpression of heterologous enzymes can create stress by depleting the pool of available amino acids, especially when the genes originate from an organism having a different codon usage compared with the production strain. Consequently, many heterologous proteins are now codon-optimized to draw predominantly on the largest available tRNA pools and to alleviate potential problems with mRNA structure [52]. Additionally, the burden of expressing both native and non-native pathways can cause imbalances in the cellular redox state by altering the  $\text{NAD}^+/\text{NADH}$  and  $\text{NADP}^+/\text{NADPH}$  ratios or levels of ATP, which can lead to overflow metabolism [53]. Imbalances in enzymatic activity can also result in the accumulation of toxic or inhibitory pathway intermediates, which may drastically reduce cellular growth as well as production levels. Functional genomics can be important in troubleshooting such issues [54]. As an example, the accumulation of 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) was found to be a bottleneck in the production of iso-prenoids in *E. coli* and was corrected by overexpression of tHMG1 [55] or by downregulating the synthesis of HMG-CoA [56]. The various considerations of metabolic engineering as they apply to the production of amorpha-4-diene in *E. coli* have been summarized (Figure 2).



**Figure 2.** An example of an engineered pathway for the production of the anti-malarial drug precursor, amorphadiene. This pathway demonstrates several key points that are pertinent to any metabolic engineering project. Namely, there are a large number of genes that must be co-expressed; all enzymes in gray were derived from *S. cerevisiae*, with the exception of *ads*, which was obtained from *Artemisia annua*, while enzymes in black were overexpressed but were native to *E. coli*. Toxic intermediates are shown in red. The toxicity of these intermediates was overcome by balancing the expression levels of the corresponding enzymes. The requirements for energy equivalents are shown in green.

An excellent example of the successful application of systems biology is the microbial production of 1, 3-propanediol [49] by DuPont-Genencore. As early as 2000, Chotani et al. recognized the importance of identifying bottlenecks using functional genomics studies [57]. At the time, techniques other than microarray were still in their infancy; however, their research suggests a clear systematic approach toward engineered systems using functional genomics studies. To achieve a viable titer, many different hurdles had to be overcome. The original pathway, which utilized glycerol, was replaced by a pathway capable of using the less expensive substrate, glucose. Pathway and host optimization eventually resulted in production levels of 135 g/L [49]. This engineered system continues to be the focus of cell-wide studies to characterize its phenotype [58].

In metabolic engineering, the product itself may prove to be a bottleneck if it is not exported out of the cell efficiently. The presence of high concentrations of the product in the medium can also have a toxic effect on cellular metabolism, which is especially true for biofuels, such as ethanol, longer chain alcohols, and long chain hydrocarbons [59,60,61]. The knowledge of mechanisms that confer fitness in response to detrimental production factors, such as changes to pH, the presence of salts, and accumulation of toxic byproducts, can also be useful in developing a robust production host. The important aspect of optimizing strains for industrial application was emphasized in recent studies by Alper et al. [62<sup>33</sup>], which illustrates the necessity of simultaneously altering multiple traits in a host organism. Engineering a single trait, such as ethanol resistance, frequently involves manipulating multiple genes. Obtaining strains with multiple traits involves an even larger set of genes, rendering traditional sequential screening strategies ineffective. By mutagenizing global transcriptional machinery, Alper et al. successfully identified strains that were simultaneously resistant to high concentrations of both ethanol and SDS. These studies also emphasize the role of regulatory proteins in evolving strains that contain multiple desirable traits as well as for phenotypes that remain stable and predictable during scale up and long-term growth. Transcriptional regulators present a difficult set of genes to study both via microarrays and proteomics as they often show little change and

are typically present in low abundance. However, many emerging technologies can improve our ability to conduct focused, high-throughput studies for specific gene classes. One example is that of the nCounter [ 63] system, which allows highly sensitive measurement of mRNA changes for 524 transcripts at a time using a newly developed, direct counting method. While many engineering efforts are being attempted to design optimal production hosts, the inherent complexity of microbial systems presents a formidable challenge to cellular engineering. Systems biology provides the opportunity to study the cell from a global perspective to gain a snapshot of the systems that are being affected during production and give clues about where pathway bottlenecks lie. It is important to note that existing studies with model systems such as *E. coli* and *S. cerevisiae* provide a wealth of information that is highly pertinent to many of the conditions important in industrial processes [ 64,65]. While not reviewed here, an important knowledgebase for microbial engineering also comes from non-model organisms that display traits desirable in an industrial host.

## CONCLUSIONS

Several important fuel production systems have been demonstrated in microbes recently and include biodiesel production [ 66] and isobutanol production [51<sup>3</sup>]. While these studies are promising, extensive engineering will be necessary to increase the titers to enable production of commercially viable fuels. It will not be surprising if systems-level analyses reveal many strategies that can be combined to optimize production. Many recent cell-wide studies have, in fact, focused on understanding the cellular physiology of production microbes exposed to industrial processing conditions [ 67], stress and starvation [ 65], and heterologous metabolic pathways with their potentially toxic intermediates [ 54] and are beginning to find use in engineering better hosts. While the majority of functional genomics studies to date have been targeted toward understanding cellular physiology, the expectation is that this information can now be used to deduce pathway bottlenecks and optimize cellular circuit design. As the technology moves beyond the discovery phase and into application, implementation of knowledge from systems biology approaches in metabolic engineering holds the promise of creating robust production host organisms for industrial applications.

## ACKNOWLEDGEMENTS

The authors would like to thank Hector Garcia Martin (LBNL) and Rajat Sapra (Sandia National Laboratory) for critical reading of this manuscript. This work was funded by the Joint BioEnergy Institute (JBEI) through a grant from the US Department of Energy and by the Synthetic Biology Engineering Research Center (SynBERC) through a grant from the National Science Foundation.

## REFERENCES

1. Energy Information Administration (2007) International Energy Outlook 2007, US Department of Energy, [http://www.eia.doe.gov/oiarf/ieo/pdf/0484\(2007\).pdf](http://www.eia.doe.gov/oiarf/ieo/pdf/0484(2007).pdf).
2. Herrera S: Bonkers about biofuels. *Nat Biotechnol* 2006, 24:755-760.
3. Kintisch E: Half-billion-dollar bonanza for plant scientists. *ScienceNOW* 2007.
4. Fortman JL, Chhabra S, Mukhopadhyay A, Chou H, Lee TS, Steen E, Keasling JD: Biofuel Alternatives to ethanol: pumping the microbial well. *Trends Biotechnol* 2008.
5. Chang MCY, Keasling JD: Production of isoprenoid pharmaceuticals by engineered microbes. *Nat Chem Biol* 2006, 2:674-681.
6. Dien BS, Cotta MA, Jeffries TW: Bacteria engineered for fuel ethanol production: current status. *Appl Microbiol Biotechnol* 2003, 63:258-266.
7. Wackett LP: Biomass to fuels via microbial transformations. *Current Opinion in Chemical Biology* 2008, 12:187-193.
8. Blanch HW, Adams PD, Andrews-Cramer KM, Frommer WB, Simmons BA, Keasling JD: Addressing the need for alternative transportation fuels: the joint bioenergy institute. *ACS Chem Biol* 2008, 3:17-20.
9. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al.: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995, 269:496-512.

10. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A et al.: Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 2007, 316:593-597.
- Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC et al.: A predictive model for transcriptional control of physiology in a free living cell. *Cell* 2007, 131:1354-1365.
12. Ideker T, Galitski T, Hood L: A new approach to decoding life: systems biology. *Annu Rev Genom Hum Genet* 2001, 2:343-372.
13. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B et al.: Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 2002, 418:387-391.
14. DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL: High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* 2007, 53:371-383.
15. He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, Huang Z, Wu W, Gu B, Jardine P et al.: GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 2007, 1:67-77.
16. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC et al.: Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 2001, 29:365-371.
17. Domon B, Aebersold R: Review—mass spectrometry and protein analysis. *Science* 2006, 312:212-217.
18. Scherperel G, Reid GE: Emerging methods in proteomics: top-down protein characterization by multistage tandem mass spectrometry. *Analyst* 2007, 132:500-506.
19. Mueller LN, Brusniak MY, Mani DR, Aebersold R: An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* 2008, 7:51-61.
20. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, Guertin S, Pribil P, Lee KH: 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* 2007, 7:3651-3660.
21. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S et al.: Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 2004, 3:1154-1169.
22. Dong M, Yang LL, Williams K, Fisher SJ, Hall SC, Biggin MD, Jin J, Witkowska HE: A "tagless" Strategy for Identification of Stable Protein Complexes Genome-wide by Multidimensional Orthogonal Chromatographic Separation and iTRAQ Reagent Tracking. *J Proteome Res* 2008, 7:1836-1849.
23. Huttlin EL, Hegeman AD, Harms AC, Sussman MR: Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. *J Proteome Res* 2007, 6:392-398.
24. Elias JE, Gygi SP: Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007, 4:207-214.
25. Bradshaw RA, Burlingame AL, Carr S, Aebersold R: Reporting protein identification data—the next generation of guidelines. *Mol Cell Proteomics* 2006, 5:787-788.
26. Wilkins MR, Appel RD, Van Eyk JE, Chung MCM, Gorg A, Hecker M, Huber LA, Langen H, Link AJ, Paik YK et al.: Guidelines for the next 10 years of proteomics. *Proteomics* 2006, 6:4-8.
27. Trauger SA, Kalisak E, Kalisiak J, Morita H, Weinberg MV, Menon AL, Poole FL, Adams MWW, Siuzdak G: Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J Proteome Res* 2008, 7:1027-1035.
28. Corbin RW, Paliy O, Yang F, Shabanowitz J, Platt M, Lyons CE, Root K, McAuliffe J, Jordan MI, Kustu S et al.: Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. *Proc Natl Acad Sci U S A* 2003, 100:9232-9237.
29. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK: Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* 2006, 103:13004-13009.
30. Anderson KL, Roberts C, Disz T, Vonstein V, Hwang K, Overbeek R, Olson PD, Projan SJ, Dunman PM: Characterization of the *Staphylococcus aureus* heat shock, cold shock, stringent, and SOS

- responses and their effects on log-phase mRNA turnover. *J Bacteriol* 2006, 188:6739-6756.
31. Eichler J, Adams MWW: Posttranslational protein modification in Archaea. *Microbiol Mol Biol Rev* 2005, 69:393-425.
  32. Sibbald MJJB, Ziebandt AK, Engelmann S, Hecker M, de Jong A, Harmsen HJM, Raangs GC, Stokroos I, Arends JP, Dubois JYF et al.: Mapping the Pathways to Staphylococcal Pathogenesis by Comparative Secretomics. *Microbiol. Mol. Biol. Rev.* 2006, 70:755-788.
  33. Thanbichler M, Shapiro L: Getting organized—how bacterial cells move proteins and DNA. *Nat Rev Microbiol* 2008, 6:28-40.
  34. Verma S, Xiong YJ, Mayer MU, Squier TC: Remodeling of the bacterial RNA polymerase supramolecular complex in response to environmental conditions. *Biochemistry* 2007, 46:3023-3035.
  35. Oldiges M, Lutz S, Pflug S, Schroer K, Stein N, Wiendahl C: Metabolomics: current state and evolving methodologies and tools. *Appl Microbiol Biotechnol* 2007, 76:495-511.
  36. Sauer U: Metabolic networks in motion: C-13-based flux analysis. *Mol Syst Biol* 2006:2.
  37. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T: Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* 2003,2:488-494.
  38. Bajad SU, Lu WY, Kimball EH, Yuan J, Peterson C, Rabinowitz JD: Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography– tandem mass spectrometry. *J Chromatogr A* 2006, 1125:76-88.
  39. Alper H, Jin Y-S, Moxley JF, Stephanopoulos G: Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metabol Eng* 2005, 7:155-164.
  40. Noth K, Groenke K, Luo B, Takors R, Oldiges M, Wiechert W: Metabolic flux analysis at ultra short time scale: isotopically non-stationary <sup>13</sup>C labeling experiments. *J Biotechnol* 2007, 129:249-267.
  41. Tang Y, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, Keasling JD: Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using gas chromatography-mass spectrometry and Fourier transform–ion cyclotron resonance mass spectrometry. *J Bacteriol* 2007, 189:940-949.
  42. Gerdes S, Edwards R, Kubal M, Fonstein M, Stevens R, Osterman A: Essential genes on metabolic maps. *Curr Opin Biotechnol* 2006, 17:448-456.
  43. Wecke T, Veith B, Ehrenreich A, Mascher T: Cell envelope stress response in *Bacillus licheniformis*: integrating comparative genomics, transcriptional profiling, and regulon mining to decipher a complex regulatory network. *J Bacteriol* 2006, 188:7500-7511.
  44. Shannon P, Reiss D, Bonneau R, Baliga N: The Gaggle: an open-source software system for integrating bioinformatics software and data sources. *BMC Bioinform* 2006, 7:176.
  45. Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP: The Microbes Online Web site for comparative genomics. *Genome Res* 2005, 15:1015-1022.
  46. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A: The EcoCyc and MetaCyc databases. *Nucleic Acids Res* 2000, 28:56-59.
  47. Das A, Yoon SH, Lee SH, Kim JY, Oh DK, Kim SW: An update on microbial carotenoid production: application of recent metabolic engineering tools. *Appl Microbiol Biotechnol* 2007, 77:505-512.
  48. Lin Y, Tanaka S: Ethanol fermentation from biomass resources: current state and prospects. *Appl Microbiol Biotechnol* 2006, 69:627-642.
  49. Nakamura C, Whited GM: Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* 2003, 14:454-459.
  50. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J et al.: Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 2006, 440:940-943.
  51. Atsumi S, Hanai T, Liao JC: Non-fermentative pathways for synthesis of branched-chain higher Alcohols as biofuels. *Nature* 2008, 451:86-89.
  52. Gustafsson C, Govindarajan S, Minshull J: Codon bias and heterologous protein expression. *Trends Biotechnol* 2004, 22:346-353.
  53. Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA: Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol* 2006, 72:3653-3661.
  54. Kizer L, Pitera DJ, Pfleger BF, Keasling JD: Application of Functional Genomics to Pathway Optimization for Increased Isoprenoid Production. *Appl Environ Microbiol* 2008, 74:3229-3241.

55. Pitera DJ, Paddon CJ, Newman JD, Keasling JD: Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metabol Eng* 2007, 9:193-207.
56. Pflieger B, Pitera DJ, Smolke CD, Keasling JD: Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* 2006, 8:1027-1032.
57. Chotani G, Dodge T, Hsu A, Kumar M, LaDuca R, Trimbur D, Weyler W, Sanford K: The commercial production of chemicals using pathway engineering. *Biochim Biophys Acta (BBA)—Protein Struct Mol Enzymol* 2000, 1543:434-455.
58. Antoniewicz MR, Kraynie DF, Laffend LA, Gonzalez-Lergier J, Kelleher JK, Stephanopoulos G: Metabolic flux analysis in a nonstationary system: Fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metabol Eng* 2007, 9:277-292.
59. Sikkema J, de Bont JA, Poolman B: Mechanisms of membrane toxicity of hydrocarbons. *Microbiol Rev* 1995, 59:201-222.
60. Jarboe LR, Grabar TB, Yomano LP, Shanmugan KT, Ingram LO: Development of ethanologenic bacteria. *Adv Biochem Eng Biotechnol* 2007, 108:237-261.
61. van Voorst F, Houghton-Larsen J, Jønson L, Kielland-Brandt MC, Brandt A: Genome-wide identification of genes required for growth of *Saccharomyces cerevisiae* under ethanol stress. *Yeast* 2006, 23:351-359.
62. Alper H, Stephanopoulos Gregory: Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metabol Eng* 2007, 9:258-267.
63. Geiss G, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T et al.: Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 2008, 26:317-325.
64. Han M-J, Lee SY: The *Escherichia coli* proteome: past, present, and future prospects. *Microbiol Mol Biol Rev* 2006, 70:362-439.
65. Gasch A, Werner-Washburne M: The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genom* 2002, 2:181-192.
66. Kalscheuer R, Stolting T, Steinbuchel A: Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology* 2006, 152:2529-2536.
67. Pham TK, Chong PK, Gan CS, Wright PC: Proteomic analysis of *Saccharomyces cerevisiae* under high gravity fermentation conditions. *J Proteome Res* 2006, 5:3411-3419.

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technology, State, and Community Programs, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.