Proteomic and Genomic Analysis of Lycopene-Overproducing *Escherichia coli* Strains

Brian E. Mickus[1], Jeffrey C. Silva[2], Johannes P. C. Vissers[2], Hal Alper[1], Joel F. Moxley[1], Gregory Stephanopoulos[1], and Charles L. Cooney[1]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA; [2]Waters Corporation, Milford, MA

ASMS 2007 (WPZ1 - 437)

Massachusetts Institute of Technology

Overview

- Global proteomic and genomic expression changes associated with increased recombinant lycopene production of mutant *E. coli* constructs identified using DNA microarrays and alternate scanning LC-MS^E
- Transcriptional data showed >4 times more differential expression for strains with *hnr* deletion compared to mutants with mainly metabolic gene deletions
- Conservation of expression trends across a majority of distantly-related mutants for specific genes including *hish* and *atpE*
- Over 500 unique protein identifications
- While most proteins didn't display differential expression, the differential expression of key proteins such as WrbA and MdoG clear from reproducible alternate scanning LC-MS^E method
- Low correlation between proteomic and transcriptional expression

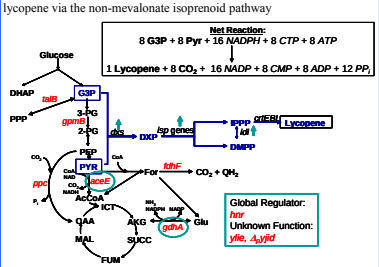
Introduction

Systems biology represents a powerful method to describe and manipulate phenotypes of interest by analyzing and incorporating biological information from various levels of cellular organization. Such an approach is illustrated from a library of both rationally-directed and combinatorial gene knockout strains of *E. coli* which have been shown to produce various levels of the small molecule lycopene when transformed with the pAC-LYC plasmid. Lycopene is an important neutraceutical of the diverse and valuable isoprenoid class, and therefore an improved description of its recombinant production has significant implications. A systems biology strategy was followed to discover global proteomic and genomic expression changes associated with increased lycopene production of mutant *E. coli* constructs.

Methods

Strains and Phenotypes

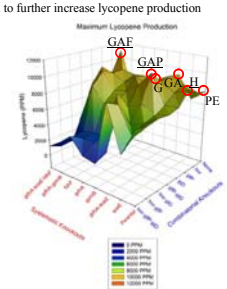
- The *E. coli* K12 pAC-LYC PT5-*dxs*, PT5-*idi*, PT5-*ispFD* pre-engineered (PE) strain of Alper *et al.* (2005) utilizes pAC-LYC *crtEBI* genes from *E. herbicola* to synthesize high amounts of lycopene via the non-mevalonate isoprenoid pathway



Modified from H. Alper *et al. (2005), Metab. Eng.* 7: 155-164.

Figure 1: Recombinant lycopene biosynthesis via the non-mevalonate isoprenoid pathway in E. coli. Chromosomal over-expressions shown by arrows; crtEBI genes on pAC-LYC plasmid; gene knockouts circled.

- Further gene knockouts identified either stoichiometrically or combinatorially and made by Alper *et al.* (2005) in PE genetic background to further increase lycopene production



Modified from H. Alper *et al. (2005), Nat. Biotechnol.* 23: 612-616.

Figure 2: Lycopene production landscape for PE and mutant strains.

Differential Gene Expression in Mutant Strains

- Full genome DNA microarray method developed and validated by ANOVA on self-self arrays
- Five knockout mutants from Alper *et al.* (2005), indicated by red below (Δ*fdhF* not yet examined), analyzed for differential expression during mid-exponential phase (OD~0.4) growth in modified M9 minimal media compared to PE strain using a Maximum Likelihood Analysis (MLA) method (T. Ideker *et al.* (2000), *J. Comput. Biol.* 7: 805-817)
- Comparison along "branches": Gene expression changes additive?
- Comparison between "branches": Significantly distinct strains accomplish common phenotype via similar gene expression patterns?

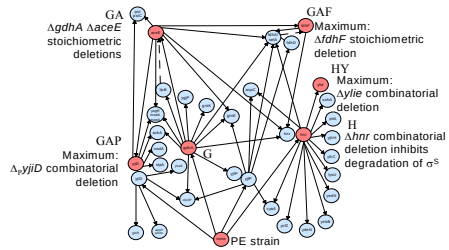


Figure 3: Relationships between deletion mutants with indicated global maximum lycopene producers. PE: Pre-engineered strain; G: ΔgdhA (glutamate dehydrogenase); A: ΔaceE (pyruvate dehydrogenase); P: ΔByjid (hypothetical protein); F: ΔfdhF (formate dehydrogenase); H: Δhnr (global regulator– degradation of σ^S); Y: Δylie (conserved inner membrane protein).

Differential Protein Expression in Mutant Strains via LC-MS^E

- Chromatography carried out on 3 µ C18 columns (300 µ ID x 15 cm). Elution was by means of an acetonitrile/0.1% formic acid gradient. Triplicate injections were made for each sample.
- Data were acquired on a hybrid quadrupole-time of flight mass spectrometer in alternating low and elevated collision energy scanning mode, utilizing a reference spray of [Glu]^1-fibrinopeptide B and erythromycin (J. C. Silva *et al.* (2006), *MCP* 5: 144-156). See also ASMS 2007 posters WPQ-275, TPZ8-571, WPQ-277, WPZ1-430, and TPZ2-475.
- Mass/retention time peak detection, charge state reduction, deisotoping, time alignment, databank searching, and absolute quantitation were carried out with Identity^E informatics software. Additional data analysis was performed with SpotFire and Microsoft Excel.

Results

Overall Gene Expression Analysis

- Deletion of *hnr* global regulator leads to the most differentially expressed genes
- Knockout of the *ylie* and $_B$*yjid* genes appears to lead to similar numbers of differentially expressed genes as the *gdhA* and *aceE* genes, suggesting metabolic roles versus global regulation
- The Δ*hnr* Δ*ylie* succession of deletions appears to increase the number of differentially expressed genes, whereas differentially expressed genes for the Δ*gdhA* Δ*aceE* Δ$_B$*yjid* succession do not appear to be additive

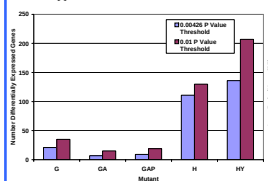


Figure 4: Total Differential Gene Expression.

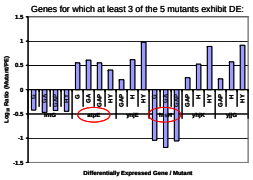


Figure 5: Conserved Differential Gene Expression.

- There are 6 genes differentially expressed in at least 3/5 mutants (for 4.26 x 10^{-3} p-value threshold) that show agreement in up- and down- regulation even for significantly distinct strains (*e.g.* GAP, H, and HY)

Δ*gdhA* Mutant Family Gene Expression Analysis

Known differentially-expressed genes conserved among G-GA-GAP mutants:

- *atpE* gene: Encodes for c subunit of F_0 domain of ATP synthase
  - c subunit is critical given the F_0 domain ratio of $a_1b_2c_{9-11}$ (D. Stock *et al.* (1999), *Science* 286: 1700-1705)
- *hisH* gene: Encodes for branch point in energy-costly histidine biosynthesis

- Higher lycopene production in mutants versus PE strain does not appear to be caused by coordinated up-regulation of glycolytic or non-mevalonate pathways
- It was hypothesized that the G-GA-GAP mutants may have more ATP and CTP available for lycopene biosynthesis than the PE strain:
  - ATP and CTP cofactor levels may be limiting in lycopene biosynthesis
  - Down-regulation of the histidine biosynthetic pathway (*hisH*) may allow for PRPP conservation and ultimately higher levels of ATP and CTP
  - Up-regulation of a critical component of ATP synthase (*atpE*) may also allow for increased production of ATP and CTP

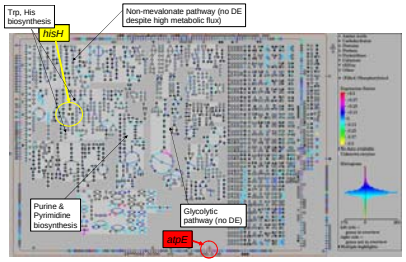


Image from EcoCyc (I. M. Keseler *et al.* (2005), *Nucleic Acids Res.* 33: D334-7)

Figure 6: GAP mutant example of log_{10}(GAP/PE) gene expression ratio metabolic map.

- Supplementing the PE, G, GA, and GAP strains with histidine and tryptophan to down-regulate the entire respective biosynthetic pathways had an insignificant effect upon lycopene production- more experimentation needed to investigate *hisH* and *atpE* gene expression

Protein Expression Analysis

- >500 identified proteins with approximately +10 to -10 fold changes
- LC-MS^E method sufficiently sensitive to detect relatively few differentially-expressed proteins
- H and HY strains exhibit greatest protein expression variation, similar to gene expression
- RpoS (σ^s), known target for increasing lycopene production (M. Becker-Hapak *et al.* (1997), *Biochem. Biophys. Res. Commun.* 239: 305-309), up-regulated in H and HY strains: It's degradation facilitated and controlled by Hnr (H) protein
- OsmY (hyperosmotically inducible periplasmic protein), positively regulated by σ^s (R. Lange *et al.* (1993), *J. Bacteriol.* 175: 7910-7917), up-regulated in H and HY and also somewhat in GAP
- MdoG (periplasmic glucan (MDO) biosynthesis protein) down-regulation consistent with OsmY increase (J. Bohringer *et al.* (1995), *J. Bacteriol.* 177: 413-422) and may provide larger periplasmic space (J. Holtje *et al.* (1988), *J. Biol. Chem.* 263: 3539-3541)
- WrbA (NAD(P)H:quinone oxidoreductase) known to decrease quinone pool in *E. coli* (E. V. Patridge *et al.* (2006), *J. Bacteriol.* 188: 3498-3506), so it's increased expression in mutants may "pull" on the *E. coli* polyisoprenoid pathway and increase lycopene production or else re-direct isoprenoid units to lycopene synthesis
- Up-regulation of GadA and GadB, glutatmate decaroxlyase acid resistance isozymes, in H and HY strains likely from RpoS up-regulation versus acid stress (Z. Ma *et al.* (2003), *Mol. Microbiol.* 49: 1309-1320)
- GatA and GatB proteins, hydrophillic components of the galacitol PTS permease, down-regulated in the H and HY strains- potentially related to lycopene membrane storage?

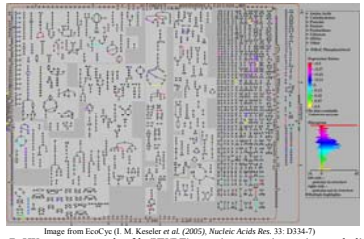


Image from EcoCyc (I. M. Keseler *et al. (2005), Nucleic Acids Res.* 33: D334-7)

Figure 7: HY mutant example of ln(HY/PE) protein expression ratio metabolic map.

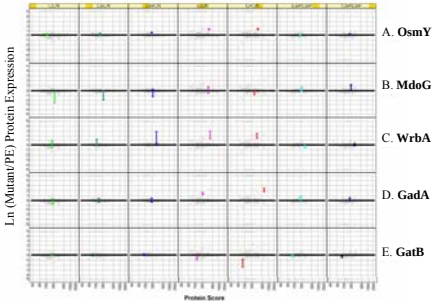


Figure 8: A. Hyperosmotically inducible periplasmic protein; B. Periplasmic glucan (MDO) biosynthesis protein B; C. NAD(P)H:quinone oxidoreductase; D. Glutamate decarboxylase; E. Galactitol PTS permease.

LC-MS^E Reproducibility

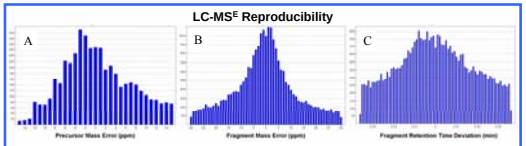


Figure 9: The tryptic peptide precursors' (A) and corresponding precursor fragments' (B) mass error distribution as well as the fragment retention time deviations (C) measured between identified fragments and the originating precursor retention time. One replicate injection shown from an LC-MS^E analysis of the GA sample. The other replicates and strains followed similar normal distributions with little bias.

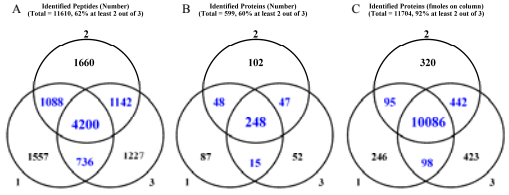


*Figure 10: GA strain (A) total **number** identified **peptides**, (B) total **number** identified **proteins**, and (C) total **fmoles** of protein identified on column for triplicate injections. The other strains displayed similar statistics.*

- The LC-MS^E method is reproducible, with about 60% of peptides and proteins and over 90% total protein mass identified in at least 2 out of 3 replicates

Correlation Between Proteomic and Genomic Results

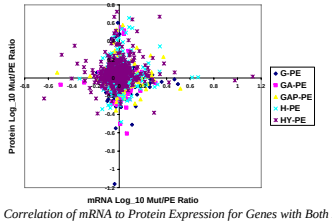


Figure 11: Correlation of mRNA to Protein Expression for Genes with Both Detected

- However, correlation of mRNA to protein very low: r~0.04: Experimental error?
- Complete Affymetrix GeneChips with PE, GAP, and HY strains to confirm previous arrays and also with the GAF and K12 strains for additional comparison

Conclusions

Conclusions:

- **The H and HY combinatorial strains have >4 times more differential gene expression than the mainly stoichiometric G, GA, and GAP strains, and some expression trends are conserved**
- **A histidine biosynthetic pathway gene is down-regulated 10-fold and a critical component of ATP synthase is up-regulated 3-fold in the Δ*gdhA* strain family**
- **>500 proteins identified by alternate scanning LC-MS^E; H and HY strains again displayed the most differential expression; gene and protein expression correlation low, though**
- **This LC-MS^E strategy is a data-independent method which is reproducible and sensitive for detecting subtle protein changes**

Future Work:

- **Verify initial microarray results with Affymetrix GeneChips on PE and global maxima strains GAP and HY, and also analyze global maxima strain GAF and K12 for comparison**
- **Fractionate PE, GAP, and HY protein extracts for greater LC-MS^E protein detection**
- **Probe transcription factor activities of commonly-expressed genes using NCA (J.C. Liao *et al.* (2003), *PNAS* 100: 15522-15527) or PCA, further integrating genomic and proteomic data**
- **Delete genes such as *hisH* and *mdoG* and/or overexpress genes such as *atpE*, *rpoS*, and *wrbA* in PE strain to determine if high lycopene-producing phenotype can be replicated**

Acknowledgements

Christopher Workman, Center for Biological Sequence Analysis, Technical University of Denmark

Alexi Goranov, Grossman Laboratory, Department of Biology, MIT

William Perry, Cooney Laboratory, Department of Chemical Engineering, MIT

This work is supported by a NIGMS/NIH Interdepartmental Biotechnology Training Grant