Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Metabolic Engineering 7 (2005) 155–164
METABOLIC ENGINEERING
www.elsevier.com/locate/ymben

# Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*

Hal Alper[1], Yong-Su Jin[1], J.F. Moxley, G. Stephanopoulos*

*Department of Chemical Engineering, Massachusetts Institute of Technology, Room 56-469, 77 Massachusetts Ave., Cambridge, MA 02139, USA*

Received 22 October 2004; accepted 10 December 2004
Available online 26 February 2005

**Abstract**

The identification of genetic targets that are effective in bringing about a desired phenotype change is still an open problem. While random gene knockouts have yielded improved strains in certain cases, it is also important to seek the guidance of cell-wide stoichiometric constraints in identifying promising gene knockout targets. To investigate these issues, we undertook a genome-wide stoichiometric flux balance analysis as an aid in discovering putative genes impacting network properties and cellular phenotype. Specifically, we calculated metabolic fluxes such as to optimize growth and then scanned the genome for single and multiple gene knockouts that yield improved product yield while maintaining acceptable overall growth rate. For the particular case of lycopene biosynthesis in *Escherichia coli*, we identified such targets that we subsequently tested experimentally by constructing the corresponding single, double and triple gene knockouts. While such strains are suggested (by the stoichiometric calculations) to increase precursor availability, this beneficial effect may be further impacted by kinetic and regulatory effects not captured by the stoichiometric model. For the case of lycopene biosynthesis, the so identified knockout targets yielded a triple knockout construct that exhibited a nearly 40% increase over an engineered, high producing parental strain.
© 2005 Elsevier Inc. All rights reserved.

*Keywords:* Metabolic engineering; Gene knockout; Lycopene; Carotenoids; *Escherichia coli*

## 1. Introduction

The central goal of metabolic engineering is the improvement of cellular phenotype, such as metabolite overproduction, by the introduction of genetic controls. To this end, metabolic engineering efforts have considered the properties of the *overall metabolic network*, in sharp contrast to the single-gene focus that characterizes typical applications of genetic engineering. This is a formidable task considering the fact that molecular and genetic interactions are complex, nonlinear and rather poorly understood. Owing to the lack of extensive knowledge about molecular interactions and their kinetics, the dissection and optimization of metabolic pathways is an outstanding issue of central importance to metabolic engineering (Stephanopoulos et al., 2004). A notable exception to the dearth of kinetic information is the constraints imposed upon metabolic function by the stoichiometry of the reaction network.

During the past decade, metabolic engineering has produced an impressive portfolio of results that were guided by rational analysis of well understood systems from a kinetic and regulatory standpoint (Koffas et al., 2003; Padilla et al., 2004; Stafford et al., 2002; Stephanopoulos et al., 1998). Significantly, fewer applications resulted from lesser-understood systems, or from the manipulation of genes not directly connected with the product-synthesizing pathway (Hemmi et al., 1998). These lesser-understood systems have traditionally been handled by sequential approaches whereby a single gene is found (usually from a screen of a combinatorial search) to affect the phenotype of interest, and

*Corresponding author.
*E-mail address:* gregstep@mit.edu (G. Stephanopoulos).
[1]These authors contributed equally to this work.

1096-7176/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2004.12.003

subsequent searches are conducted in a genetic background defined by the deletion or overexpression of said gene. Despite successes using this approach, there is no evidence that such combinatorial searches are complete or that subsequent sequential modifications would lead to the global phenotype optimum. Furthermore, it is unclear how to systematically identify gene targets when considering the entire bioreaction network.

We address these issues here computationally and experimentally in the context of lycopene synthesis in *Escherichia coli*. Our computational search makes use of a stoichiometrically balanced, genome-wide bioreaction network of *E. coli* metabolism whose fluxes are computed such as to maximize cell growth yield in the framework of flux balance analysis (FBA) (Edwards and Palsson, 2000; Segre et al., 2002). Yields and rates of product synthesis (such as lycopene) can be obtained from the calculated fluxes. Although this model is genome-wide and global for most metabolic reactions, it is important to note that it is a strictly stoichiometric model, totally devoid of any kinetic or regulatory information. Consequently, targets identified by this model improve product synthesis solely on the basis of increased availability of metabolic precursors and cofactor balancing. This beneficial effect may be negatively impacted by non-predictive, adverse kinetic and/or regulatory effects.

We employed this formalism to investigate the effect of gene deletions, the most common means of introducing genetic perturbations, on lycopene production. As such, we sought to identify genes whose elimination might lead to an increase in product yield. A major limitation of the FBA approach is that fluxes so calculated are actually those that support maximum growth for the particular genotype. As such, these calculated, in silico, fluxes are not necessarily the same as the actual, in vivo, fluxes in the organism and this is especially so for genetically perturbed systems, (gene knockouts or over-expressions), where the resulting phenotype is often suboptimal in growth and metabolite levels. To correct for this shortcoming, flux profiles were computed for recombinant suboptimal systems by optimizing a different objective function, the minimization of metabolic adjustment (MOMA) requirement between the wild type and a potential single gene knockout mutant (Segre et al., 2002). This calculation yielded flux profiles that are intermediate between the wild-type optimal and gene knockout mutant optimal. While single gene knockout phenotypes have been studied using this method (Edwards and Palsson, 2000), a systematic analysis of multiple gene knockout systems has received less attention. Furthermore, often target identification of multiple knockouts is performed through search algorithms (Burgard et al., 2003), rather than a comprehensive search of all possible multi-gene targets. Nevertheless, these predicted gene targets and identification methods have not been verified with experimental results.

We investigated the issues of gene target identification in the context of heterologous lycopene production in *E. coli* using the non-mevalonate pathway (Adam et al., 2002). Production of secondary metabolites is typically an expensive and complex cellular process; therefore it provides a good platform for testing concepts of pathway optimization (Mijts and Schmidt-Dannert, 2003). Lycopene production in *E. coli* utilizes glycolytic intermediates to form precursor monomers, which subsequently undergo polymerization to form the 40 carbon biopolymer (Fig. 1). The isoprenoid pathway and downstream reactions to create a diverse library of carotenoids have received significant attention recently (Mathews and Wurtzel, 2000; Misawa and Shimada, 1998; Sandmann, 2002; Smolke et al., 2001; Wang et al., 1999). Initial attempts for improving carotenoid production in *E. coli* targeted the expression of genes coding for enzymes that catalyze pathways upstream of the enzymes coded by the *crtEBI* operon (Farmer and Liao, 2000, 2001; Lee and Schmidt-Dannert, 2002). Even with the over-expression of *dxs* and *idi* genes (Kim and Keasling, 2001; Kajiwara et al., 1997; Mathews and Wurtzel, 2000), cellular production and accumulation of carotenoids were limited by regulatory networks and precursor supply (Farmer and Liao, 2000, 2001; Jones et al., 2000; Lee and Schmidt-Dannert, 2002; Wang et al., 1999). Therefore, these studies suggest a need to generate an enhanced production phenotype. Cellular optimization and metabolic engineering are necessary to obtain maximal production rates since lycopene is not an endogenous product and the cell lacks necessary

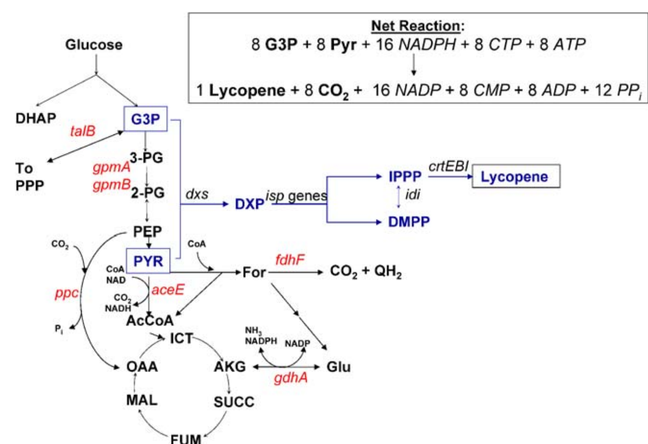


Fig. 1. Lycopene Production Pathway and Identified Gene Targets. Lycopene synthesis begins with the condensation of the key glycolytic intermediates, glyceraldehyde 3-P (G3P) and pyruvate (PYR) and continues in a nearly linear pathway. The genes encoding for *idi* and *dxs* are typical targets for lycopene over-expression along this pathway. In the engineered strain used in this study, the *idi*, *ispFD*, and *dxs* genes are overexpressed. Identified gene targets are also shown as they are connected to lycopene production.

control mechanisms and the infrastructure to produce these molecules.

The computational search identified several putative single- and multiple-gene knockouts yielding increased lycopene production. While most gene knockout mutants exhibited experimental behavior in qualitative agreement with the model predictions, a notable exception illustrates the limitations of using a purely stoichiometric model of metabolism. We also compared an iterative, *sequential* approach, whereby multiple gene targets are determined by following a path of maximal phenotype increase resulting from sequential single gene modulation, with a *simultaneous* multiple-gene modulation that examines phenotypic response over the entire space of all possible single- and multiple-gene expression modifications. This comparison led to interesting observations regarding a potential mutant optimization strategy. Application of this strategy ultimately yielded an *E. coli* mutant strain that accumulated 6600 ppm (6.6 mg/g dcw) of lycopene, an increase of approximately 40% relative to the engineered parental strain.

## 2. Methods

### 2.1. Strains and media

*E. coli* K12 PT5-*dxs*, PT5-*idi*, PT5-*ispFD*, provided by DuPont, was used as the lycopene expression strain when harboring the pAC-LYC (Cunningham et al., 1994) plasmid containing the *crtEBI* operon. Overexpressions of *dxs*, *idi*, and *ispFD* were chromosomally incorporated without an antibiotic marker through promoter delivery (Suh et al., 2003) of the PT5 promoter. Gene deletions were conducted using PCR product recombination (Datsenko and Wanner, 2000) using the pKD46 plasmid expressing the lambda red recombination system and pKD13 as the template for PCR (see supplemental information for primer designs). Gene knockouts were verified through colony PCR. The *aceE* knockout was designed to be closest to the end of the gene to reduce expression, but prevent a strict acetate requirement (Langley and Guest, 1978). Strains were grown at 37 °C with 225 RPM orbital shaking in M9-minimal media (Maniatis et al., 1982) containing 5 g/L D-glucose and 68 μg/ml chloramphenicol. All cultures were 50 ml grown in a 250 ml flask with a 1% inoculation from an overnight 5 ml culture grown to stationary phase. All experiments were performed in replicate to validate data and calculate statistical parameters. Glucose monitoring was conducted periodically using a YSI2300 glucose analyzer to verify complete usage of glucose. Cell density was monitored spectrophotometrically at 600 nm. Transposon libraries were generated using the pJA1 vector. According to the protocol (Badarinarayana et al., 2001), cells were transformed with the plasmid, then grown in 50 ml LB under 20 μM IPTG induction conditions until an OD600 of 0.4. The culture was centrifuged and reduced to a volume of 10 and 2 ml 15% glycerol stock aliquots were created. These aliquots were used for inoculation of the pJA1 transposon library cultures. All PCR products were purchased from Invitrogen and utilized Taq polymerase. M9 Minimal salts were purchased from US Biologicals and all remaining chemicals were from Sigma-Aldrich.

### 2.2. Lycopene assay

Intracellular lycopene content was extracted from 1 ml of bacterial culture at the point of total glucose exhaustion. The cell pellet was washed, and then extracted in 1 ml of acetone at 55 °C for 15 min with intermittent vortexing. The lycopene content in the supernatant was quantified through absorbance at 475 nm (Kim and Keasling, 2001) and concentrations were calculated through a standard curve. The entire extraction process was performed in reduced light conditions to prevent photo-bleaching and degradation. Cell mass was calculated by correlating dry cell with OD600 for use in ppm (mg lycopene/g dry cell weight) calculations.

### 2.3. Flux balance analysis calculations

Application of FBA to the carotenoid system required including the non-endogenous reactions (*crtEBI*) required for the production of these molecules on the background of previously published stoichiometric models (Edwards and Palsson, 2000; Segre et al., 2002) with alterations in the isoprenoid biosynthesis pathway (see supplemental information). The resulting model consisted of 965 fluxes (including the exchange fluxes) involving 546 metabolite intermediates. This model was solved subject to MOMA using the linear and quadratic programming methods through a PERL script interface (Edwards and Palsson, 2000; Segre et al., 2002) provided by Dr. Daniel Segre. An additional script was used to perform the genome-wide knockout searches. For the calculation parameters, values for the glucose uptake, oxygen uptake and nitrogen uptake were set at 5, 200, and 1000, respectively. The base steady state solution required for the quadratic programming protocol for MOMA are calculated for each simulation by first solving the linear programming problem by maximizing for growth, accounting for any gene knockouts under consideration. These values allow for glucose to be the limiting substrate in these calculations. Single knockout calculations were performed on a Pentium IV Linux platform while the exhaustive double knockout search was performed on six Power PC 1.5 GHz microchips on an AIX platform.

## 3. Results

### 3.1. Gene knockout simulations

The *E. coli i*JE660a GSM model (Reed et al., 2003) served as the basis for this stoichiometric network. Furthermore, the *crtEBI* operon was added to the model along with updated isoprenoid synthesis reaction details discovered after the formulation of this model (Adam et al., 2002; Hecht et al., 2001) (see supplemental information for details). Using this updated model, a total of 965 metabolic fluxes (included exchange fluxes) were calculated such as to: (a) balance the rates of synthesis and depletion of 546 metabolites, (b) maximize cell growth yield subject to an MOMA alteration for suboptimal systems, and (c) utilize glucose as the sole carbon source (Edwards and Palsson, 2000; Segre et al., 2002) (see Section 2 for details). Initially, the maximization of lycopene production was set as the objective function to extract characteristic phenotype behavior. Simulations revealed a direct inverse relationship between the stoichiometric maximum lycopene yield and growth yield, and a direct relationship between stoichiometric maximum lycopene yield and glucose uptake (data not shown). Therefore, these two relationships suggested the need to reduce the growth yield and maintain a relatively high glucose uptake rate to support enhanced lycopene production.

Using the stoichiometric model along with a maximum growth objective function subject to an MOMA alteration, in silico genome-wide gene knockout simulations were conducted. The phenotype of specific gene knockouts was simulated by deleting the corresponding enzyme (i.e., reaction) from the stoichiometry matrix and calculating the resulting flux profile. When multiple enzymes encode the same reaction (as is the case with isoenzymes), all instances of that reaction were removed from the stoichiometric matrix. To avoid selecting mutants with extremely low growth, a minimum growth requirement was enforced. Knockout candidates were compared on the basis of predicted production level after invoking the growth requirement.

#### 3.1.1. Single gene knockout simulations

Fig. 2 summarizes the results of a genome scan for single gene knockout mutants using the aforementioned criteria. It identifies eight single gene knockouts, which would produce a higher yield of lycopene by direct enhancement of the lycopene pathway and, indirectly, by lowering growth yield. Fig. 2 illustrates that most knockouts are not predicted to increase lycopene yield. Furthermore, all of the candidate knockouts show a reduction in the predicted growth yield. Of the eight predicted gene targets, two were eliminated from further consideration: *glyA*, due to a very low predicted improvement and *cyoA* that has been shown to exhibit

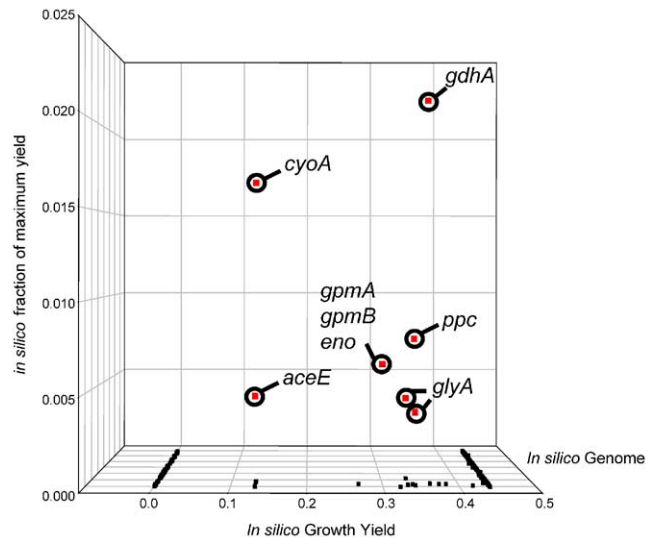


Fig. 2. *E. coli* Genome Scan for Single Gene Knockout Target Identification. The phenotype of every possible single gene knockout was simulated using FBA with MOMA as an additional constraint. The above genotype-phenotype plot illustrates the effect of single gene deletions on lycopene yield as measured by the fraction of the stoichiometric maximum yield (approximately 0.1 g Lycopene/g Glucose). Shown is also the in silico predicted growth yield of these knockouts. A single knockout scan predicted eight genes whose deletion yielded enhanced product synthesis while satisfying a minimum growth requirement. The gene *glyA* appears twice since its function can be classified as both amino acid biosynthesis and vitamin/cofactor metabolism. The enzymes encoded by these genes are as follows: *aceE* (Pyruvate dehydrogenase), *cyoA* (Cytochrome oxidase bo3), *eno* (enolase), *gdhA* (Glutamate dehydrogenase), *glyA* (Glycine hydroxymethyltransferase), *gpmAB* (Phsophoglucomutase), and *ppc* (Phosphoenolpyruvate carboxylase).

a limited range of substrate utilization (Au et al., 1985). After these exclusions, *gdhA*, *gpmA*, *gpmB*, *aceE* and *ppc* were selected as candidates for experimental validation. While *eno* also appeared as a candidate, it was not selected since the predicted phenotype was similar to the *gpm* isoenzymes and no prior strain containing the single knockout of *eno* was found in a literature search. Furthermore, while *gpmA* is the more prevalent isoenzyme form of the phosphoglycerate mutases in *E. coli* during the growth phase (Fraser et al., 1999), the actual function and interaction of all phosphoglycerate-related genes has not been fully determined. Additionally, *ppc* knockouts were found to be not viable in non-supplemented glucose-based media. Of the five selected genes, all but *gdhA* apparently directly impact the supply of lycopene precursors while the *gdhA* knockout appears to increase the availability of NADPH, an important cofactor for lycopene synthesis. Furthermore, in silico predictions indicated a reduced growth phenotype for each of these knockouts ranging between 40% and 75% of the maximum yield as is evident in Fig. 2. These results are consistent with existing strategies for increasing secondary metabolite production aiming at

the reduction of byproduct formation, balancing of precursors and, in some cases, lowering the growth rate.

### 3.1.2. Multiple gene knockout simulations

Optimization of a secondary-metabolite phenotype, such as lycopene production, obviously depends on the modulation of several genes. Hence, multiple gene knockouts need to be similarly evaluated. The difficulty here is that exhaustive investigation of all possible gene knockout combinations leads very quickly to combinatorial explosion: ${}_{965}C_2$ combinations of all possible double mutants, and so on. Hence, *sequential* and iterative optimization approaches are often invoked whereby single gene knockouts are investigated in the genetic background of deletion mutants identified for their improved phenotype from previous iterations. Such procedures emulate optimization routines of the type of steepest descent for non-linear optimization problems. However, while properties of continuity and convexity assure of a certain degree of success in the solution of mathematical problems, no such properties have been demonstrated for metabolic networks. Consequently, there can be no assurance about the results of such sequential optimization procedures.

We first investigated multiple gene knockout mutants following a sequential approach: A gene was first identified whose deletion yielded maximum lycopene improvement and double mutants were subsequently sought by scanning the effect of additional gene knockouts in the genetic background of the single gene knockout, and so on for higher mutants. Certain combinations of gene knockouts yielded extremely reduced growth phenotypes in silico, so a growth rate minimum was required of all mutants equal to 5% of the maximum, wild-type prediction. Fig. 3 summarizes the results of several multiple knockout constructs of considerable interest. As a double knockout construct, *gdhA/gpmA* or *gdhA/gpmB* is predicted to outperform the other candidate combinations. However, all triple knockout constructs based upon *gdhA/gpmA* or *gdhA/gpmB* are predicted to have an extremely low growth rate (less than the 5% threshold), which warrants their removal from further consideration on the basis of the minimum growth rate requirement despite their predicted high product yield. Additionally, *talB* and *fdhF*, although absent as single knockout candidates, become key gene targets in the construction of double or triple knockout mutants. Of further interest, the gene *talB* is predicted to improve production in a *gdhA/aceE* background, yet it decreases the yield in the *gdhA* background when the enzymatic activity of *aceE* is present. Fig. 1 depicts the reaction network along with these candidate gene targets. It underlines the rationale for specific combinations, such as *aceE* followed by *fdhF* as a knockout scheme. In this case, the knockout of *aceE* would presumably increase formate production, whose flux may then be redirected through an *fdhF* knockout. These features illustrate the need for the invoked systematic approach to identify gene knockout targets.

Next, we sought to compare the above results to those obtained by an exhaustive investigation of all possible

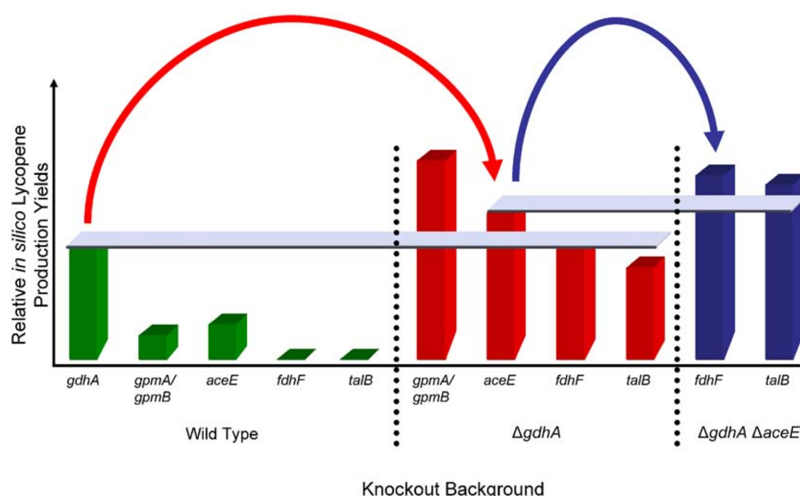


Fig. 3. In silico multiple gene knockout search strategy. The process of maximal sequential phenotype increase is illustrated. The production yields for each genetic background are simulated similarly to the method followed in Fig. 2, but in a different genetic background for the starting strain. A triple knockout construct based on the double mutant *gdhA/gpmB* was excluded as it violated the minimum growth rate requirement. On the other hand, predicted triple constructs in *gdhA/aceE* background continue to show an increase in lycopene yield. The path of maximal phenotype increase is given by the solid lines. However, since a *gdhA/gpmB* knockout has been excluded for triple knockouts consideration due to growth rate, the next highest optimal path is followed. These results indicate that novel gene targets arise as the genotype is altered as result of gene knockouts. This is especially evident in the case of *talB*. Although *talB* increases the production level in a *gdhA/aceE* knockout background, it is detrimental in a *gdhA* only knockout background.

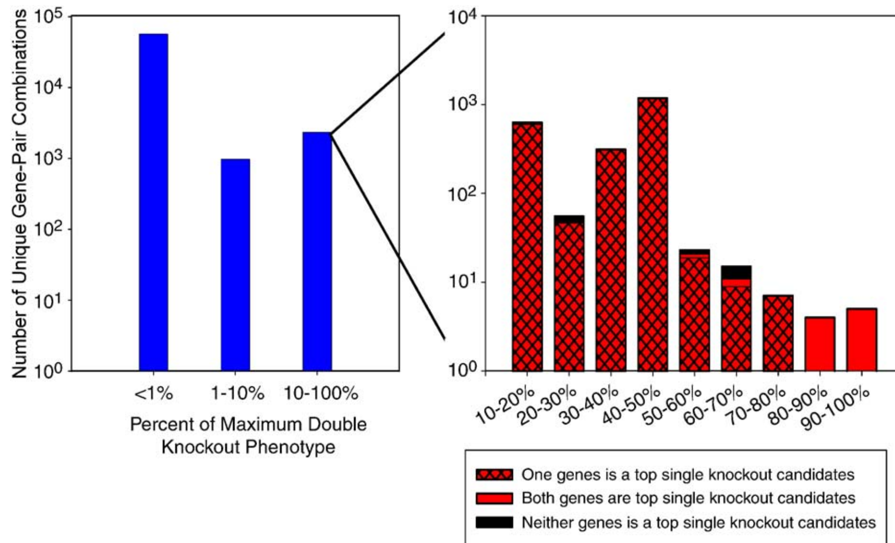


Fig. 4. Simulated product phenotype of all possible double gene knockout mutants. This figure depicts the results of the exhaustive double knockout search. In the first graph, it is evident that the majority of double knockouts have little impact on lycopene yield (A). The usual 5% minimum growth requirement has been imposed. Double knockout constructs are scored as a percent of the production level of the highest producing double knockout. The second graph (B) stratifies the top 90% of the double knockout phenotypes found by this exhaustive search. These results indicate that 98.6% of all possible combinations consist of either one or more genes from top candidates found from the single knockout illustrated in Fig. 2. More importantly, all maximum phenotypes can be identified through the sequential search.

double mutants. An exhaustive simulation is computationally expensive, however provides insight into the topology and behavior of the metabolic space. At the end of this simulation, the highest yielding viable knockout (two *cyoA* combinations were excluded, as this gene was eliminated previously) was predicted to be a *gdhA*/*gpmA* or *gdhA*/*gpmB* construct, which is similar with the result obtained using the sequential approach. Additionally, this analysis predicted that most double gene knockouts would not significantly increase the calculated lycopene yield.

Following the path of maximal sequential phenotype increases for predicting yields of double knockout constructs cannot, obviously, identify combinations of two synergistic genes, which have no phenotype impact individually as single knockouts. As shown in Fig. 4, of the top 90% phenotypes in the double knockout metabolic space, 98.6% contain at least one gene, which elicits a high increase in lycopene yield as a single knockout. Additionally, all of the desirable phenotypes (those exhibiting the highest lycopene yield) reside in this subset of genes. Only 1.4% of the top 90% of double knockout phenotype constructs would be unattainable following a sequential approach to target identification, however, the highest resulting phenotype in this subset of combinations is only 60% of the maximum yielding predicted double knockout construct. These results suggest that, *for this particular system*, the most desirable phenotypes are attainable using a sequential genome search strategy.

### 3.2. Experimental investigation of single and multiple gene knockouts

Gene knockout experiments were conducted along with shaker-flask fermentations to experimentally test the predictions of the previous simulations. Knockout constructs were created using PCR product mediated inactivation (Datsenko and Wanner, 2000) in a recombinant *E. coli* strain already engineered to produce lycopene at high yields through chromosomal overexpressions of the *dxs*, *ispFD*, and *idi* genes (see Section 2). Since this strain simply contains chromosomal overexpressions of endogenous genes (which does not change the bioreaction network), the stoichiometric model used to identify gene targets is still suitable for this host. This strain was expressing the heterologous *crtEBI* operon on a pAC-LYC plasmid encoding for the additional genes required to produce lycopene (Cunningham et al., 1994). Table 1 summarizes the results from eleven so constructed knockout mutants and presents the lycopene production at the point of glucose exhaustion when strains were grown in an M9-Minimal media with glucose as the sole carbon source.

Five major conclusions arise by comparing this experimental data to the results of the simulations. First, the trend of actual mutant growth rates compares qualitatively to the predicted values. Second, there is continuing improvement of lycopene yield with an increased number of selective gene knockouts. This trend reflects the selection criteria applied in the identification of gene targets. Third, the *gdhA*/*gpmB*

Table 1
Experimental results of single and multiple gene knockouts

| Knockout construct | Growth rate ($h^{-1}$) | Actual percentage of parental (%) | Predicted percentage of parental (%) | Percent increase in lycopene content (PPM) |
|---|---|---|---|---|
| None | 0.67 | 100 | 100 | 0% (4700 PPM) |
| *Single knockouts* | | | | |
| gdha | 0.55 | 82 | 75 | 13% (±4) |
| gpma | 0.44 | 66 | 40 | −8% (±3) |
| gpmb | 0.55 | 82 | 40 | 7% (±2) |
| acee | 0.52 | 78 | 68 | 9% (±4) |
| fdhf | 0.57 | 85 | 100 | 4% (±3) |
| *Double knockouts* | | | | |
| gdhA, aceE | 0.52 | 78 | 56 | 13% (±4) |
| gdhA, gpmA | 0.37 | 55 | 9 | 12% (±3) |
| gdhA, gpmB | 0.49 | 73 | 9 | 18% (±3) |
| gdhA, talB | 0.46 | 68 | 62 | 3% (±4) |
| *Triple knockouts* | | | | |
| gdhA, aceE, talB | 0.44 | 65 | 44 | 19% (±4) |
| gdhA, aceE, fdhF | 0.38 | 56 | 54 | 37% (±3) (6600 PPM) |

Mutant growth rates and lycopene production (shown in ppm) are compared with the corresponding levels obtained in the non-mutated parental strain with zero knockouts. Growth rate data are compared as a percentage of the parental strain and juxtaposed with the predicted values. It is important to note the differing effects of the two *gpm* isoenzymes (*gpmA* and *gpmB*), as the knockout of *gpmA* appears to give the greater impact. Total lycopene content increases with multiple knockouts obtained along the path of highest production, with the exception of *gpmA*. Overall, this data qualitatively follows the trends of the simulations presented in Fig. 3. Numbers in parenthesis indicate the standard deviations among replicate culture experiments. Different batches of media caused the absolute value of lycopene production to vary slightly. As a result, all trials were conducted along with the parental strain as an internal control.

double knockout construct produced the highest yield among double knockouts at 18% above the parental strain, as predicted by the simulations. Fourth, following the path of highest product yield in combination with the minimum growth requirement yielded a triple knockout construct of *gdhA/aceE/fdhF* which produced the highest yield of 37% above that of the parental strain. Finally, gene targets selected as being important in triple knockout constructs, matched computational predictions to be either ineffective as a single knockout (*fdhF*) or detrimental as double a knockout (*talB* in *gdhA* background). Overall, the experimental results followed the trends suggested by the simulations.

One notable exception to the qualitative adherence of experimental data to computational results is the impact of the *gpmA* knockout. As a single knockout, this construct resulted in a decrease in lycopene yield, and was consistently lower than any *gpmB* construct. Since *gpmA* is a more dominant isoenzyme than *gpmB*, the expected metabolic consequences are higher and its impact is most evident in the substantial growth rate decrease in *gpmA* knockouts. Phosphoglycerate mutase knockouts could lead to the accumulation of 3-phosphoglycerate, which is known to have regulatory functions within the cell, especially as it serves as a precursor to amino acids, and may be negatively interacting with lycopene production in this experiment. The results from this gene knockout construct illustrate that gene knockouts can increase precursor availability and lead to increased lycopene production *to the extent that regulatory effects elicited by the deletion of the gene* do not interfere with product synthesis.

Since the predictions were based on glucose as the sole carbon source, nonviable knockouts such as *ppc* were excluded from further analysis. Despite this fact, *ppc* was still found to impart an increase in lycopene yield compared to the parental strain when grown in minimal media with 0.3% Casamino acid supplementation (data not shown).

The significance of these results should be examined relative to the lycopene yield improvements afforded by the deletion of other genes not identified by the flux balance simulations. To this end, libraries of random genome transposon knockouts using the pJA1 vector (Badarinarayana et al., 2001) were constructed. Such randomized libraries do not show any significant increase in the overall lycopene yield, which is illustrated by Fig. 5. The results of Fig. 5 should not be construed to imply that there is no random gene knockout that can increase significantly lycopene yield. Only a small fraction of all transposon strains were analyzed individually and, in fact, an efficient and more exhaustive screening or selection process could identify high yielding knockouts targeting critical regulatory elements within the cell. For the purpose of this comparison, we examined 8 heterogeneous cultures and 25 individual

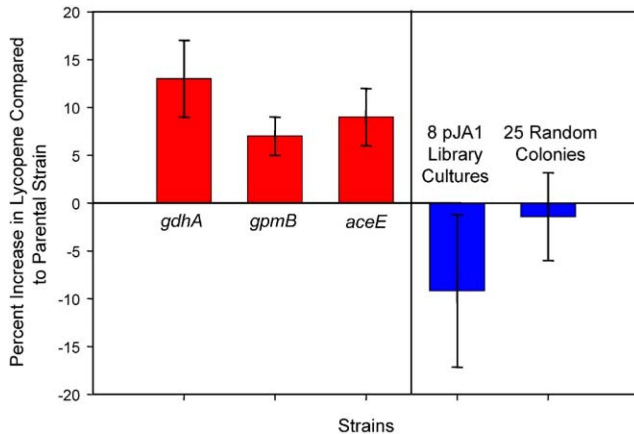


Fig. 5. Comparison of single knockout constructs to random transposon libraries relative to the parental strain. The results of the selected single knockout targets are juxtaposed to random transposons. In particular, 8 heterogeneous cultures and 25 individual colonies from the random transposon mutagenesis library were cultured separately in shaker flasks. Both the heterogeneous culture of a random transposon mutants and the average of the randomly selected colonies appeared to decrease the lycopene production. None of these cultures compared favorably with the systematically selected single gene knockout mutants of *gdhA*, *gpmB*, or *aceE*.

colonies of the random transposon mutagenesis library cultured separately in shaker flasks. None of these mutants compared favorably with the selected knockout mutants. In essence, these results indicate that the particular knockout targets identified bring about a measurable effect on lycopene accumulation that is above any lycopene change impacted by a random gene knockout. Fig. 5 juxtaposes the results of the systematically selected gene targets with the random library strains showing that gene knockouts of targets identified through stoichiometric modeling perform better than average random gene knockouts. We note that it is not yet possible to compare multiple knockout constructs due to the inadequacy of currently available genetic tools to create multigenic knockout libraries. A properly guided sequential search provides an efficient approach to multigenic modifications required to produce a desired phenotype.

## 4. Discussion

A systematic, exhaustive computational or experimental search of all feasible gene knockouts in *E. coli* to determine the genotype yielding the optimal lycopene production phenotype is a tedious and often difficult process. The effects of individual gene knockouts are not necessarily additive in determining the effect(s) of combinations of knockouts, which successfully enhance a phenotype. As the search space increases to include all possible double or higher (triple) knockouts, systematic and exhaustive searches, both computationally and experimentally, become almost infeasible. This research suggests that this non-linear process may be initially optimized, to ensure the proper supply of precursors, in a manner analogous to the method of steepest descent for nonlinear function optimization. Since this approach may often yield a local, as opposed to global production maximum, the results of other possible trajectories must be compared to identify the one with the most promising end point along the phenotype contour. In the method followed here, single gene knockout targets are first identified, of which the highest producer is selected. With this mutant as the new background, new knockouts are determined and the highest mutant is selected again. Through this process, knockout mutants are constructed with progressively increasing production phenotype.

This search technique generated novel single and multiple gene targets for increasing the production level of lycopene in *E. coli*. Furthermore, combinations of gene knockout targets for multiple knockout constructs were identified. The genes *talB* and *fdhF* exemplify the unique aspects extracted from stoichiometric modeling. As the cellular genotype changes, new stoichiometric targets arise. These new targets both computationally and experimentally illustrate the intrinsic link between cellular genotype and phenotype. Single gene modifications may not be additive in nature and thus a more systematic analysis is required to extract the optimal combinations of gene modifications. Likewise, inferences about the impact of perturbations in one strain may not be immediately transferable to another strain possessing a modified genetic background. This difficulty is enhanced by the inability of current models to capture the regulatory effects, which could negatively impact product formation.

A final issue involves determining an endpoint for this analysis. The path of maximum phenotype increase was followed to predict a quadruple knockout mutant. However, this resulted in the selection of *ppc* as the next target, which is infeasible in a glucose-based medium. Furthermore, this mutation was predicted to impart only a marginal increase in the overall lycopene yield.

An exhaustive genome scan for all possible double knockouts failed to provide any unique or interesting targets yielding desirable phenotype characteristics. This simulation also generated insights about the potential and limitations in transferring cellular information between strains of different genotypes. An undirected combination of the top single knockout gene targets could result in suboptimal combinations, yielding as low as 50% of the lycopene production in the predicted highest attainable phenotype. Additionally, the best phenotypes could be extracted by following a sequential phenotype enhancement, which saves computational and experimental efforts. Although these results cannot

be generalized for other products and strains, they nevertheless underline the ability of sequential approaches to reach very interesting phenotypes in certain cases.

Most of the gene targets identified in this study could be superimposed on a simple network diagram modeling central carbon metabolism. Discussion of the need for global as opposed to local metabolic models has indeed been brought to the forefront of research efforts after genome sequencing and a refocusing on systems biology approaches to cellular systems. As such, it is difficult to a priori determine the size of a model needed to capture the behavior of a system. In fact, the key utility of these large, global models is the extensive linking of distant reactions and metabolites through cofactor, energetics, and precursor balancing.

It should be further noted that this type of analysis is not limited to gene deletions only. It is possible to similarly explore other genetic modifications, such as gene expression amplifications, to identify putative parameters impacting cellular phenotype. As long as positive interactions exist of fundamentally stoichiometric nature, they can be uncovered by this approach generating additional promising genetic targets that can influence positively the cellular phenotype.

Neither flux balance analysis nor the *i*JE660a GSM accounts for genetic regulation and other possible cellular interactions. It is conceivable and quite possible that genetic regulations outweigh stoichiometric effects. As the former are notably absent from the stoichiometric model used, neither the latter nor alternative search methods can capture other possible gene targets arising from these complex regulatory interactions. When such advanced models of cell function become available, identifying optimal gene targets will still be a demanding undertaking that can be facilitated by the findings of this study. These more comprehensive models will allow a more detailed mapping of the phenotypic landscape along with a thorough evaluation of various search methods for promising genetic targets for metabolic engineering. Absent such models, a *sequential*, iterative optimization provides a reasonable and feasible alternative that can yield promising targets, as in the case of lycopene synthesis examined in this work.

## Acknowledgments

We acknowledge financial support of this work by the DuPont-MIT Alliance. In particular, we would like to thank Dr. Wonchul Suh for providing the parental *E. coli* strain. Dr. Daniel Segre provided the software used (with some modifications) for flux determination. Materials for constructing *E. coli* knockout strains were obtained from the *E. coli* Stock Center.

## Appendix A. Supplementary Materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/S1096-7176(04)00084-9.

## References

Adam, P., et al., 2002. Biosynthesis of terpenes: Studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase. Proc. Natl. Acad. Sci. USA 99, 12108–12113.

Au, D.C., Lorence, R.M., Gennis, R.B., 1985. Isolation and characterization of an *Escherichia coli* mutant lacking the cytochrome o terminal oxidase. J. Bacteriol. 161, 123–127.

Badarinarayana, V., et al., 2001. Selection analyses of insertional mutants using subgenic-resolution arrays. Nat. Biotechnol. 19, 1060–1065.

Burgard, A., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657.

Cunningham Jr., F.X.S.Z., Chamovitz, D., Hirschberg, J., Gantt, E., 1994. Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942. Plant Cell 6, 1107–1121.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97, 6640–6645.

Edwards, J.S., Palsson, B., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. USA 97, 5528–5533.

Farmer, W.R., Liao, J.C., 2000. Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat. Biotechnol. 18, 533–537.

Farmer, W.R., Liao, J.C., 2001. Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*. Biotechnol. Prog. 17, 57–61.

Fraser, H.I., Kvaratskhelia, M., White, M.F., 1999. The two analogous phosphoglycerate mutases of *Escherichia coli*. FEBS Lett. 455, 344–348.

Hecht, S., et al., 2001. Studies on the nonmevalonate pathway to terpenes: the role of the GcpE (IspG) protein. Proc. Natl. Acad. Sci. USA 98, 14837–14842.

Hemmi, H., Ohnuma, S., Nagaoka, K., Nishino, T., 1998. Identification of genes affecting lycopene formation in *Escherichia coli* transformed with carotenoid biosynthetic genes: candidates for early genes in isoprenoid biosynthesis. J. Biochem. 123, 1088–1096.

Jones, K.L., Kim, S.-W., Keasling, J.D., 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab. Eng. 2, 328–338.

Kajiwara, S., Fraser, P., Kondo, K., Misawa, N., 1997. Expression of an exogenous isopentyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. Biochem. J. 324, 421–426.

Koffas, M., Jung, G.Y., Stephanopoulos, G., 2003. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. Metab. Eng. 5, 32–41.

Kim, S.-W., Keasling, J.D., 2001. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. Biotechnol. Bioeng. 72, 408–415.

Langley, D., Guest, J.R., 1978. Biochemical genetics of the alpha-keto acid dehydrogenase complexes of *Escherichia coli* K12: genetic characterization and regulatory properties of deletion mutants. J. Gen. Microbiol. 106, 103–117.

Lee, P.C., Schmidt-Dannert, C., 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. Appl. Microbiol. Biotechnol. 60, 1–11.

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mathews, P.D., Wurtzel, E.T., 2000. Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. Appl. Microbiol. Biotechnol. 53, 396–400.

Mijts, B.N., Schmidt-Dannert, C., 2003. Engineering of secondary metabolite pathways. Curr. Opin. Biotechnol. 14, 597–602.

Misawa, N., Shimada, H., 1998. Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts. J. Biotechnol. 59, 169–181.

Padilla, L., Kramer, R., Stephanopoulos, G., Agosin, E., 2004. Overproduction of trehalose: Heterologous expression of *Escherichia coli* trehalose-6-phosphate synthase (OtsA) and trehalose-6-phosphate phosphatase (OtsB) in *Corynebacterium glutamicum*. Appl. Environ. Microbiol. 70, 370–376.

Reed, J., Vo, T., Schilling, C., Palsson, B., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol. 4, R54.

Sandmann, G., 2002. Combinatorial biosynthesis of carotenoids in a heterologous host: a powerful approach for the biosynthesis of novel structures. Chembiochem. 3, 629–635.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99, 15112–15117.

Smolke, C.D., Martin, V.J.J., Keasling, J.D., 2001. Controlling the metabolic flux through the carotenoid pathway using directed mRNA processing and stabilization. Metab. Eng. 3, 313–321.

Stafford, D.E., et al., 2002. Optimizing bioconversion pathways through systems analysis and metabolic engineering. Proc. Natl. Acad. Sci. USA 99, 1801–1806.

Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering: Principles and Methodologies. Academic Press, San Diego.

Stephanopoulos, G., Alper, H., Moxley, J., 2004. Exploiting biological complexity for strain improvement through systems biology. Nat. Biotech. 22, 1261–1267.

Suh, W.-C., Yuan, L., 2003. At American Society for Microbiology, General Meeting 2003.

Wang, C.-W., Oh, M.-K., Liao, J.C., 1999. Engineered isoprenoid pathway enhances astaxanthin production in *Escherichia coli*. Biotechnol. Bioeng. 62, 235–241.