Available online at www.sciencedirect.com
SCIENCE DIRECT®
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.

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

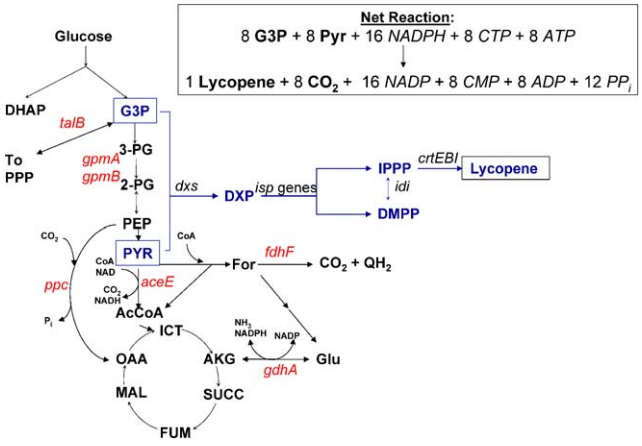

Fig. 1. Lycopene Production Pathway and Identified Gene Targets. Lycopene synthesis begins with the condensation of the key glycolytic intermediates, glyceraldehyde 3-P (G3P) and pyruvate (PYR) and continues in a nearly linear pathway. The genes encoding for *idi* and *dxs* are typical targets for lycopene 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.