Multi-scale engineering and modeling of heterologous natural product biosynthesis in *Escherichia coli*

A dissertation

submitted by

Brett Adam Boghigian

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Chemical Engineering

TUFTS UNIVERSITY

February 2011

Adviser: Blaine Alan Pfeifer

Table of Contents

Abstract

The engineering of biological systems for industrial applications is a complex process. Optimization of a cellular phenotype is complicated by the sheer number of genetic and environmental variables. With regards to heterologous natural product biosynthesis, this is further complicated by the foreign nature of the metabolic pathways and structurally complex products involved. To address this problem, heuristic and systematic approaches were employed for the engineering of two heterologous natural products (a polyketide and an isoprenoid) in *Escherichia coli*. The methods developed and applied herein are critical to advancing the field of heterologous natural product biosynthesis to the scale of competitive industrial bioprocesses.

Stoichiometric modeling was applied to survey heterologous hosts for supporting polyketide biosynthesis. Simulations under different host and environmental conditions revealed multiple gene knockouts that were capable of improving product titer. Work has shown that multiple pathways exist in nature for producing the two precursors necessary for polyketide production; however, *E. coli* does not possess these. These heterologous pathways were expressed, and with concurrent substrate feeding experiments, their effects were analyzed on polyketide production. Native gene over-expressions and deletions also improved polyketide titer.

Due to an inability to thoroughly search genomic space with the aforementioned computational method, a new algorithm was developed to identify knockout targets based on network topology and applied to isoprenoid production. By using a genetic algorithm, this method identified a four knockout strain capable of improved titer, while reducing computation time by several orders of magnitude. When constructed in the laboratory using an accelerated genome evolution method, isoprenoid yield improved nearly 3-fold in some cases. The aforementioned algorithm was reformulated in an attempt to identify over-expression targets for improving isoprenoid titer. This method identified four targets, three of which improved titer when implemented genetically, though failed to meet the predicted levels of improvement. Upon over-expression of the isoprenoid biosynthetic pathway genes, one gene improved titer to a higher extent than the predicted targets (almost 4-fold), showing that the rate-limiting step lies within the pathway itself. Applying heuristics for isoprenoid production, heterologous gene promoter strength, strain background, and process-related parameters were varied and allowed for a 240-fold improvement in titer.

Acknowledgments

I am a full believer in not forgetting where you come from and who helped you get where you are today. The following dissertation represents a significant investment of my time and emotion of not only my Ph.D., but my entire chemical engineering career thus far. Moreover, I never would have excepted to learn as much about myself as I did throughout this Ph.D. process. At the end of the day, it would not have been possible with a great supporting cast, so thank you to:

- My advisor, Blaine Pfeifer, for allowing a (then) bright-eyed nineteen year old into his lab, inspiring me, and later, being willing to learn as much from me as I learned from him.
- Kyongbum Lee, for supplying the groundwork for my theoretical background. When I was beginning my B.S. in the fall of 2003, he told me that I was "joining the Marines of engineering". You had me hooked with that line and I never looked back.
- Greg Stephanopoulos, for investing his valuable time in serving on my dissertation committee. His pioneering research over the last twenty-plus years has shaped how I think as a chemical engineer today and it is truly an honor to have him serve on my committee.
- Linc Sonenshein, for also investing his valuable time in serving on my dissertation committee. His kind demeanor and

intellectual curiosity in my work was a refreshing reminder of why I chose to pursue an advanced degree through graduate research.

- The entire ChBE department faculty and staff at Tufts for their various forms of support since 2003. I will miss the interaction and the relationships I have formed with all of the department's faculty members over the years.
- The NIH, for funding during my last three semesters of my research.
- Amy Manocchi, for her unconditional support and for the endless number of memories (both good and bad…ha!) over the past few years. I would not have survived a Ph.D. without her.
- Yong Wang, for being an immense influence in teaching me how to think like a biologist, sharing an infective enthusiasm for research, and for always being willing to help.
- Haoran Zhang, for being a great desk-mate, travel companion, and researcher for the past five years. His advice and funny jokes have been invaluable for completing my Ph.D.
- All of other the members of the Pfeifer and Lee labs for their various forms support and advice over the years.

- My four “mentees”, Daniel Salas, Melissa Myint, Sterling Wall, and John Armando. I was blessed to have such compassionate and intelligent undergraduates work with me.
- To the above students and all of the other students I TA’d, for teaching was by far the most gratifying part of my time as a graduate student.
- David Wilbur, for putting up with all of the equipment issues in the chemistry department over the years.
- My mentors throughout the years that helped shaped me as the researcher and engineer I am today:
    - Mary Lopez and Wayne Patton then at PerkinElmer
    - Jon Coffman and Misha Shamashkin then at Wyeth BioPharma
    - Brian Kelley and Bob Kiss at Genentech
- Collaborators Ajikumar Parayil at MIT and Harris Wang at Harvard Medical School for their valuable conversations and advice.
- Gautham Sridharan and Amanda Baryshyan for also doing their best to keep me sane during my graduate years at Tufts.
- My undergraduate Tufts ChBE crew: Anthony Dennis, Brandon Bosacker, Nick Curato, and Matt Dallas, for making the long,

long hours of problem sets enjoyable…I actually miss them. And of course the rest of my Tufts friends.

- The Bedford crew: Adam Chan, Alex Iwanchuk, Brian Shaw, Chris Costa, Chris Matheson, Corey Baggett, Dave Arzumanyan, Derek Castellana, Greg Gannon, Greg Hughes, Ian Sullivan, Jeff Tate, Pat DePriest, Peter Stewart, Richie Volpicelli, Steve McLane, Tim Brady, Tyler Gray, and all of the girls, for all of the good times and welcomed (although sometimes unwelcomed) distractions.
- My high school teachers who showed me that math and science were interesting and believing in me: Mary Morris, Colleen Irving, Fran Messmer, and Karen Haswell.
- The late John Baronian, who first sparked my interest in Tufts University.
- My Aunt Karen, Uncle Bill, Jeff, and Scott for their continued support and encouragement for as long as I can remember.
- My brother, Todd, for out of everyone on this last, doing his best to try to understand what I do.
- Mom and Dad, for doing their best to supply me with a set of tools to push me onto a path to success, and for their continued support long after they could understand what I do. They reminisce that when I was a toddler, I would walk around and

ask “What’s this?” so much, that they would get angry. Well, thank you for feeding my intellectual curiosity because I am convinced that trait led me on the path of being a Ph.D. engineer.

List of Tables

List of Figures

Introduction

Background & Motivation

Metabolic engineering has been defined as "the directed improvement of product formation or cellular properties through the modification of specific biochemical reactions or introduction of new ones with the use of recombinant DNA technology" (Stephanopoulos 1999). The prospect for the production of chemicals and (bio)pharmaceuticals through biological systems has gained immense interest in recent years due to the vast range of products that can be derived from them, as well as the ability to produce complex compounds with enantiomeric purity (Bailey 1991). Moreover, the robustness of biological systems enables the utilization of inexpensive, complex substrates (such as biomass hydrolysate) for catalyst and product synthesis. Although enzyme-catalyzed biological reactions offer the potential for diverse product range and excellent regioselectivity, these methods typically have tradeoffs in productivity due to the need for large-scale culture of microbial or mammalian cells. The requirement for a biological process is also dependent upon the type of final product, with large biopharmaceuticals such as proteins and vaccines requiring biological host systems, while many small-molecule pharmaceuticals must compete with methods that employ

chemical-synthetic (or semi-synthetic) production. A key challenge in improving pharmaceutical production is to apply the fundamentals of metabolic engineering particularly towards the optimization of small-molecule pharmaceutical production.

Natural products are a class of small-molecule chemical compounds that originate from the secondary (non growth-associated) metabolic processes of living organisms (Clardy et al. 2006; Cragg et al. 1997; Demain 2006; Gershenzon and Dudareva 2007; Gulder and Moore 2009; Handelsman et al. 1998; Harvey 2008; Jones et al. 2009; Lam 2007; Leeds et al. 2006; Li and Vederas 2009; Newman 2008; Newman and Cragg 2007; Newman et al. 2003; Paterson and Anderson 2005; Salas and Mendez 2007; von Nussbaum et al. 2006; Walsh and Fischbach 2010; Zhou et al. 2008). They possess a wide range of useful pharmacological activities which include antibacterial, anticancer, immunosuppressant, immunostimulant, and hypocholesterolemic (Newman and Cragg 2007). Of all small-molecule new chemical entities (NCE's) between 1981 and 2006, 34% were natural products or semi-synthetic derivatives of such molecules (Newman and Cragg 2007). In fact, of the 109 antibacterial NCEs and the 83 anticancer NCEs, 74 and 45, respectively, were derived from natural products (Newman and Cragg 2007). Therefore, even with the "blockbuster model" adopted by current pharmaceutical companies, natural

products continue to provide successful drugs and drug leads (Li and Vederas 2009).

Most natural products have evolved independently of their useful therapeutic applications, and consequently, their native hosts tend to produce them in relatively small quantities (Khosla and Keasling 2003). As a result, production of natural products from their original native hosts is oftentimes an expensive and slow process. While increasing production titers in native hosts is usually accomplished by random mutational strategies, this is an inherently slow process. To address this setback, researchers have explored the use of heterologous host organisms to produce such compounds, using similar recombinant strategies that spurred biopharmaceutical protein production. Through metabolic and genetic engineering of heterologous hosts (which is usually difficult to achieve in a native host), the complex natural product biosynthetic machinery can be redesigned with the goal of optimizing final product formation. Furthermore, the bioprocess engineering options with heterologous hosts may also lead to higher production titers or productivities that may be unachievable with native producers. Compared with strain improvement in native hosts, the upfront work involved with reconstituting heterologous biosynthesis will likely be overcome by the speed at which these hosts can be rationally engineered for improved titer. Heterologous natural

product biosynthesis can be broken down with the following workflow of natural product discovery to heterologous reconstitution to development (Figure 1 shows this workflow and supporting technologies for each step). Although the heterologous natural product biosynthetic field has seen ample research dedicated to discovery (Harvey 2008; Leeds et al. 2006; Li and Vederas 2009; Zerikly and Challis 2009) and reconstitution themes (Binz et al. 2008; Mutka et al. 2006b; Pfeifer and Khosla 2001; Pfeifer et al. 2003; Wenzel et al. 2005), the relatively recent advent of the field has seen less emphasis placed on development research initiatives that are crucial for sufficient and economical commercialization (Wattanachaisaereekul et al. 2007). To emphasize this point, there are very few examples of heterologous production schemes resulting in product titers capable of supporting subsequent commercialization attempts (Boghigian and Pfeifer 2008). Moreover, there have been a limited number of studies using model-driven approaches for the metabolic engineering of heterologous natural product biosynthesis (Alper et al. 2005b; Chemler et al. 2010; Fowler et al. 2009). As a result, there is a strong existing need for research within the development theme so as to complement the growing efforts in natural product discovery and heterologous reconstitution. When viewed comprehensively, it will be the

combination of these three themes that will facilitate continued access to a known source of therapeutic potential.

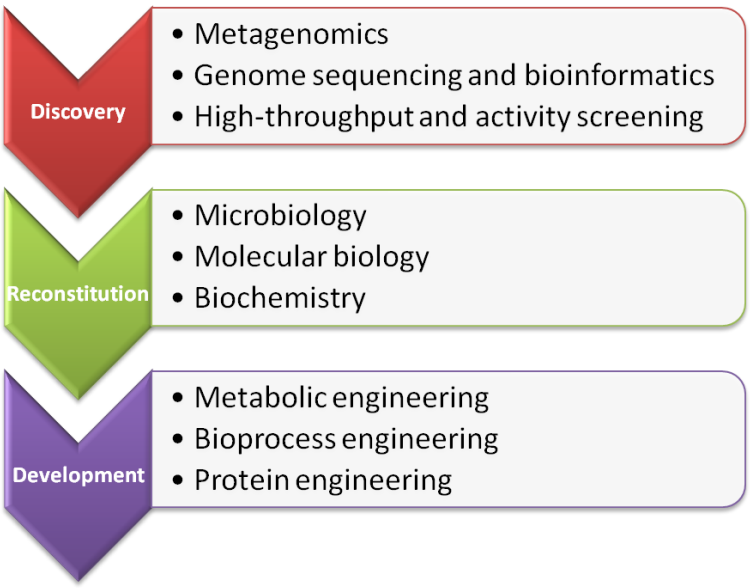

Figure 1 The workflow for natural product biosynthesis through a heterologous host.
First, discovery of a gene cluster or product can be undertaken using metagenomic approaches, or can be identified from genomic sequence data using bioinformatic approaches. If compound material is available, a pharmacological activity can be screened in a high-throughput manner to warrant further investigation of the compound in question. Secondly, with gene sequence information available, biotechnological techniques can be utilized to accomplish functional expression of the gene(s) in question and establish metabolite production through a heterologous host. Lastly, if there is significant biomedical interest in a drug in question, metabolic, bioprocess, and protein engineering can be used for further optimization of the strain/cell-line in question to develop a high titer, economical production process.

Objectives

To accomplish the goals set out for this dissertation, two major objectives were proposed:

- Develop and evaluate the applicability of systematic approaches (model-based) for improving heterologous polyketide and isoprenoid titers in *Escherichia coli*
    - Evaluation of stoichiometric modeling as a means of surveying heterologous host capabilities
    - Evaluation of stoichiometric modeling as a means of identifying knockout targets
    - Evaluation of stoichiometric modeling as a means of identifying over-expression targets
- Evaluate the utilization of heuristic approaches (knowledge-based) for improving heterologous polyketide and isoprenoid titers in *Escherichia coli*
    - Application towards the metabolic scale
    - Application towards the process scale

These objectives were formulated to address both basic (understanding cellular physiology and its relationship to heterologous natural product biosynthesis) and applied (improving titer or specific titer) facets of this research field.

Approach

To address the problem of improving titer in heterologous natural product bioprocesses, this particular dissertation will be focused on the application of the abovementioned engineering strategies to generate precursors of two relevant, potent, and commercially successful molecules: erythromycin and Taxol (Taxol® is a registered trademark of Bristol-Meyers Squibb, having a generic name of paclitaxel, however, the first naming of this compound was of "Taxol", so it will be henceforth referred to by this name (Wani et al. 1971)). These molecules encompass numerous characteristics that make them attractive as targets for heterologous biosynthesis and titer improvement. Erythromycin is a polyketide-based antibacterial compound isolated from a slow growing and genetically less tractable soil-dwelling bacterium, while Taxol is an isoprenoid-based anticancer compound isolated from the Pacific yew tree. The methods developed, evaluated, and/or applied here are designed to aid production once reconstitution has been achieved in a heterologous host. However, more broadly, many of the methodologies described and used in this dissertation can be applied to other products or other host systems.

The approaches described in this dissertation can be divided into two categories: systematic (driven by modeling) and heuristic (driven by previous knowledge), applied to metabolic and process scales.

Systematic methods were applied to the metabolic-scale in an effort to tailor the *E. coli* metabolic network for improved product formation. Heuristic methods were applied towards analysis of specific pathways supporting product formation, as well as process parameters thought to influence final product formation.

Dissertation Architecture

Chapter 2 presents background on polyketides and isoprenoids and their similar logic in biosynthesis. Information is then presented for the erythromycin and Taxol biosynthetic pathways, past and current production schemes, and challenges associated with each molecule. Next, a background of heterologous natural product biosynthesis and using *Escherichia coli* as a host is presented. The background section concludes with an extensive review of metabolic model development and metabolic flux analysis with respect to pharmaceutical production. Chapter 3 presents the use of genome-scale metabolic models for surveying potential hosts for heterologous polyketide biosynthesis, as well as the identification of knockout targets to improve titer under varying environmental conditions. Chapter 4 presents the engineering of multiple metabolic pathways for the improvement of polyketide titer based on heuristics and static metabolic maps. Chapters 5 and 6 address the development of novel computational algorithms for indentifying knockout and over-

expression targets (respectively) for improving isoprenoid production titers. Chapter 7 presents a similar multi-scale engineering strategy, in which genetic and process parameters were varied to improve isoprenoid titer. Chapter 8 presents the development of a platform *E. coli* strain and bioprocess that allows for simultaneous, high-level production of polyketide and isoprenoid compounds with strong partitioning in a two-phase bioprocess. Finally, Chapter 9 presents a summary of the dissertation, its main conclusions, and prospects for future directions of research based on this body of work.

Chapter 2 – Natural Products, Heterologous Biosynthesis, and Metabolic Modeling

The Canonical Logic of Polyketide and Isoprenoid Natural Product Biosynthesis

Polyketides and isoprenoids represent two of the largest classes of natural products, each with a wide range of pharmacological activities (Fischbach et al. 2008; Walsh 2004). Despite being derived from divergent portions of cellular metabolism, they are produced in a canonical logic: "upstream", "elongation and cyclization", and "tailoring" pathways to generate the final products (Chen and Baran 2009; Morrone et al. 2010; Walsh 2004) (Figure 2).

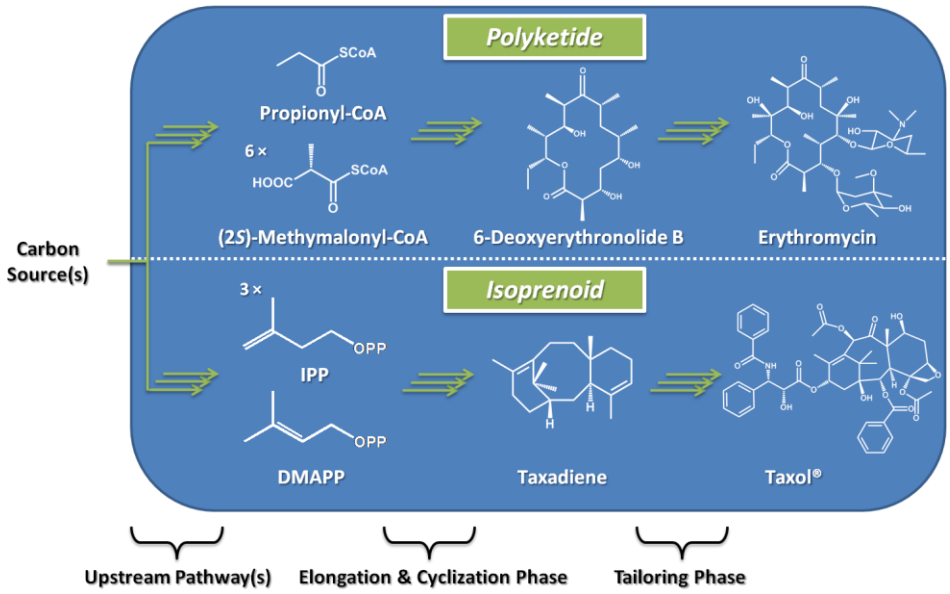

Figure 2 A map showing the canonical logic of polyketide and isoprenoid biosynthesis.
The examples shown are erythromycin and Taxol. A variety of carbon sources can be utilized to generate the metabolic precursors for elongation and cyclization of the carbon chains to the first dedicated intermediate in final product formation. From there, a tailoring phase decorates the macrocycle to generate the final product. Abbreviations: IPP = isopentenyl diphosphate; DMAPP = dimethylallyl diphosphate.

The most widely studied sub-class of polyketides are produced by bacterial type I polyketide synthases (PKS's). While there also exists type II and type III PKS's, this dissertation only focuses on the heterologous biosynthesis of type I PKS's. These polyketides are constructed by successive rounds of decarboxylative Claisen condensation reactions between an acyl thioester and thioesterified malonate derivatives (and are generated by "upstream" pathways from primary metabolism). These molecular assembly lines undertake the "elongation and cyclization" portions of polyketide formation. They create an impressive diversity of products, which are dictated by the identity and order of each domain, in turn specifying: 1) the sequence of monomer units incorporated into the growing chain, 2) the chemistry at each carbon in the chain, and 3) the total chain length of the polyketide before cyclization (Cane et al. 1998). As a result of the existence of a large number of catalytic steps within a PKS, many PKS enzymes are often greater than 300 kDa in size, and these enzymes can associate into higher-order super structures which often exceed 1 MDa in size. Release of the covalent thioester linkage of the chain most often occurs by the action of a thioesterase domain at the end of the PKS, allowing for cyclization by an intramolecular nucleophilic attack, and ultimately regenerations of the catalyst system (Hu et al. 2003; Sharma and Boddy 2007). The "tailoring" step in this logic creates: 1)

even greater product diversity, and 2) a product with specific and potential biological activity. Tailoring of the nascent polyketide can be O-, C-, or N-linked glycosylation (Blanchard and Thorson 2006; Salas and Mendez 2007), acylation, cytochrome P450-like oxygenation, O-methylation (Gruschow et al. 2007), or even halogenation (Neumann et al. 2008).

Isoprenoids are the largest class of natural products, of which more than 55,000 have been isolated (Ajikumar et al. 2008). The "upstream" pathway for isoprenoid biosynthesis generates dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) as products. Two general pathways exist that produce these two intermediates, known as the mevalonate (Martin et al. 2003; Pitera et al. 2007) or the 1-deoxy-D-xylulose 5-phosphate (DXP) pathways (Eisenreich et al. 2004) (the DXP pathway can be seen in Figure 3). The mevalonate pathway exists in eukaryotes (mammals, plants, and fungi) while the DXP pathway exists in bacteria and plant plastids (Ajikumar et al. 2008). The mevalonate pathway begins with acetyl-CoA, while the DXP pathway begins with the condensation of pyruvate and glyceraldehyde 3-phosphate. The "elongation and cyclization" step then occurs by head to-tail condensation of the DMAPP and IPP monomers resulting in C_{10}, C_{15}, C_{20}, or higher carbon chain lengths. Cyclization occurs through terpene synthases (or terpene cyclases) and

depending on the carbon chain length, isoprenoids can be acyclic, monocyclic, or polycyclic molecules, and are are characterized by their chain length (Table 1) (Wink 2010). As with polyketides, post-cyclization terpene C-H functionalization leads to final products with pharmacological activity in these pathways.

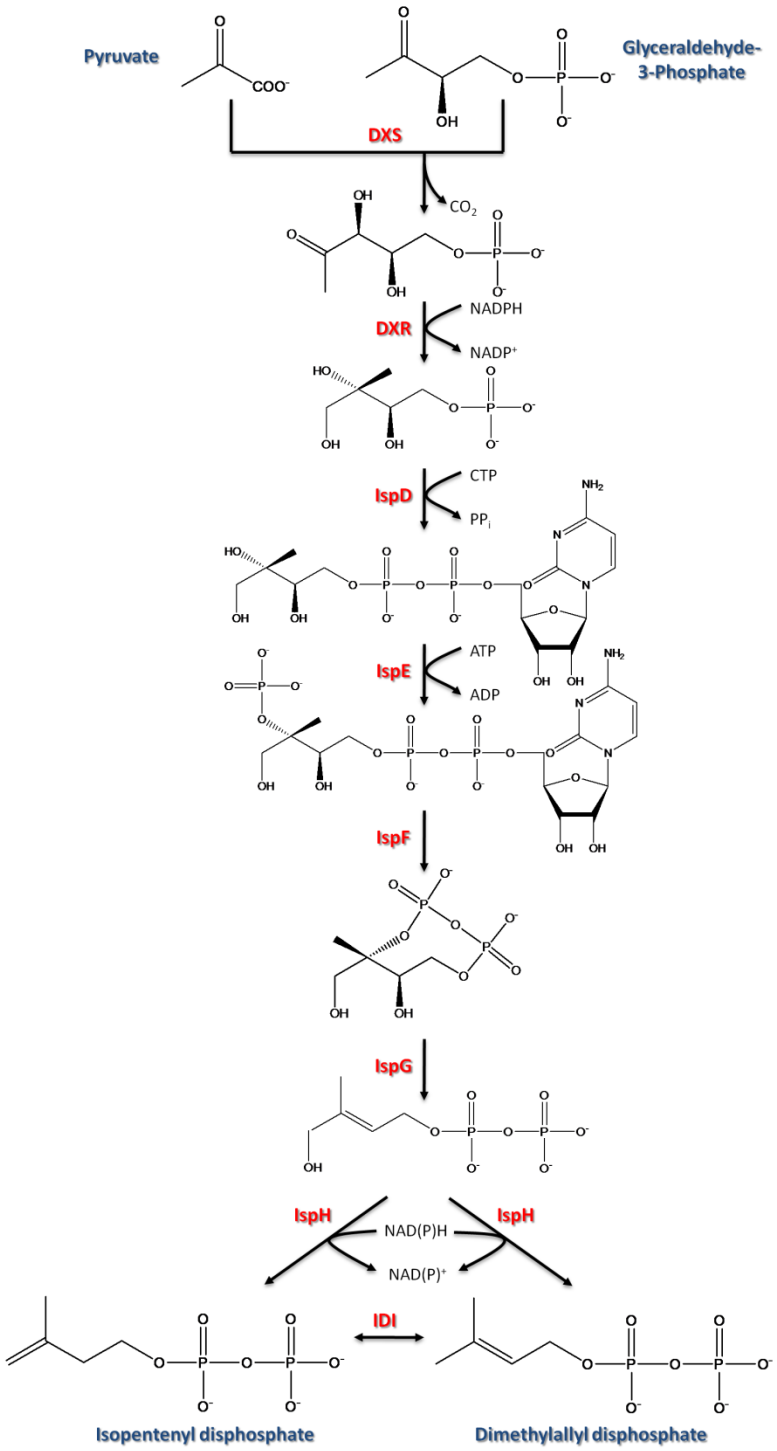

Figure 3 The DXP-based isoprenoid biosynthetic pathway, native to *E. coli*. Abbreviations: DXS = 1-deoxy-D-xylulose-5-phosphate synthase, DXR = 1-deoxy-D-xylulose-5-phosphate reductoisomerase, IspD = 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, IspE = 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, IspF = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, IspG = (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase, IspH = 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, IDI = isopentenyl-diphosphate isomerase.

Table 1 Characterization of isoprenoid molecules by the number of isoprene units they contain.
Examples of each class are given.

Isoprene Units	Carbon Atoms	Name	Example
1	5	Hemiterpenes	Isoprene
2	10	Monoterpenes	Menthol
3	15	Sesquiterpenes	Artemisinin
4	20	Diterpenes	Taxol
5	25	Sesterterpenes	Neomangicol
6	30	Triterpenes	ß-Amyrin
8	40	Tetraterpenes	ß-Carotene
9-30,000	> 40	Polyterpenes	Rubber

Erythromycin

Erythromycin is a broad range macrolide antibiotic which has had much clinical and commercial success, spawning several next generation antibiotics such as azithromycin . In 1952, erythromycin was isolated from a soil bacterium eventually designated *Saccharopolyspora erythraea* (Heilman et al. 1952; McGuire et al. 1952). Production processes were built around this organism (Smith et al. 1962) and were steadily improved over the years to allow titers of roughly 7 g l^{-1} in a rationally engineered industrial strain (Minas et al. 1998). Nevertheless, the titer of the erythromycin process is roughly an order of magnitude lower than similar classically optimized antibiotic fermentation processes such as penicillin (70 g l^{-1}), salinomycin (60 g l^{-1}), or cephalosporin C (30 g l^{-1}) (Adrio and Demain 2006) (Figure 4). Though its chemical synthesis was completed in 1981 (Woodward et al. 1981a; Woodward et al. 1981b; Woodward et al. 1981c), the many steps and low yield necessitates fermentation as a primary means of production.

Recent efforts have been dedicated towards improving the genetic technologies (Wang et al. 2007b), understanding the regulation of erythromycin production (Chng et al. 2008), and improving the erythromycin production titer (Reeves et al. 2007) and purity (Chen et al. 2008) of *S. erythraea*-based erythromycin fermentation. However, to

truly harness the combinatorial potential of the erythromycin modular biosynthetic process for the purpose of generating next-generation derivatives (Menzella and Reeves 2007; Menzella et al. 2005; Nguyen et al. 2006; Weissman and Leadlay 2005), a heterologous production host will likely be necessary.

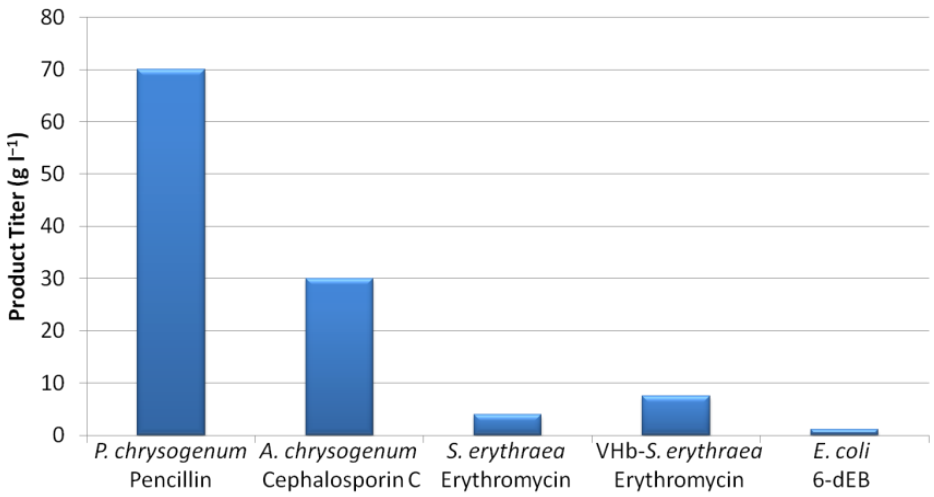

Figure 4 Examples of product titers in classical antibiotic fermentations. Production titers comparing: 1) classical β-lactam antibiotic fermentations, 2) the industrial strain of the native erythromycin producer, as well as a an engineered form of this strain expressing a *Vitreoscilla* sp. Hemoglobin gene, and 3) the highest report titer of 6-dEB in lab-scale fed-batch fermentation. Data taken from: (Adrio and Demain 2006; Lau et al. 2004; Minas et al. 1998).

The 14-membered aglycone macrocyclic core (the polyketide precursor) of erythromycin is 6-deoxyerythronlide B (6-dEB) and is constructed a type I PKS called the deoxyerythronolide B synthase (DEBS), from one propionyl-CoA starter molecules and six (2*S*)-methylmalonyl-CoA extender molecules. After identification of the *eryA* genes in the early 1990's (Cortes et al. 1990; Donadio et al. 1991), the genes and DEBS complex has been the subject of many seminal studies focused on modular polyketide biosynthesis (Gokhale et al. 1999; Kao et al. 1994; Kao et al. 1996; Luo et al. 1996; Menzella et al. 2005; Pfeifer et al. 2001; Pieper et al. 1996; Pieper et al. 1995; Tang et al. 2006). Three genes *eryA1*, *eryA2*, and *eryA3* produce three enzymes DEBS1 (370 kDa), DEBS2 (380 kDa), and DEBS3 (332 kDa), which become covalently linked. This >1 MDa complex contains 28 catalytic domains organized into a loading didomain, six extension modules (each composed of several domains), and a terminal thioesterase (TE) domain that cyclizes and releases the final product. This DEBS complex exists as a homodimer, meaning that the entire quaternary structure of this complex is greater than 2 MDa. The resulting product, 6-dEB, contains a total of 10 stereocenters, corresponding to three D-methyl, three L-methyl, one D-hydroxy, and three L-hydroxy substituents (Baerga-Ortiz et al. 2006; Valenzano et al. 2009). After two glycosylation reactions, two hydroxylation reactions, and one

methylation reaction, the 6-dEB macrocycle is converted to the full erythromycin A molecule. This process can be seen in Figure 5.

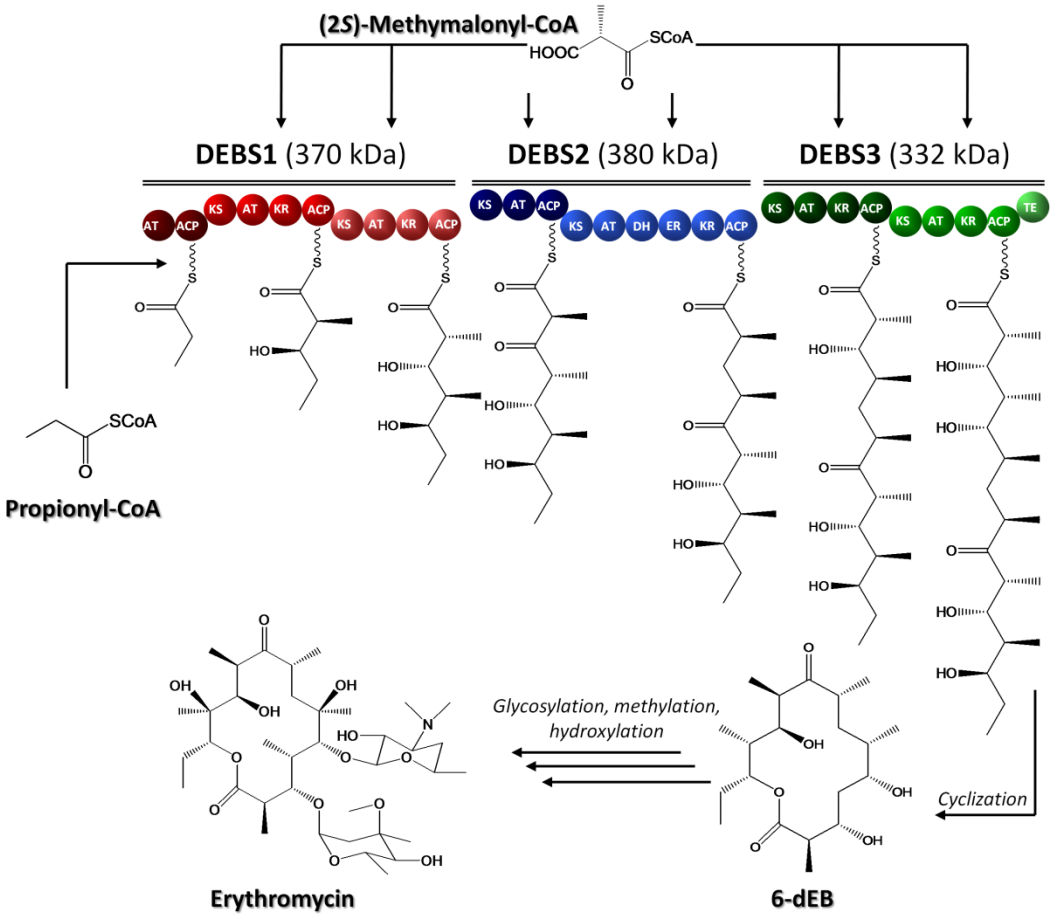

Figure 5 A figure of the three DEBS enzymes and their respective catalytic domains. One molecule of propionyl-CoA is primed on DEBS1, while six molecules of (2*S*)-methylmalonyl-CoA are used as extender units. The terminal thioesterase domain cyclizes and releases the polyketide, creating the macrolactone 6-dEB. Post-PKS tailoring creates the full erythromycin molecule. Abbreviations: AT = acyltransferase, ACP = acyl carrier protein, KS = ketosynthase, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase, TE = thioesterase.

Taxol

The diterpenoid natural product Taxol possesses impressive anticancer properties and has shown efficacy against carcinomas of the ovary, breast, lung, head and neck, bladder and cervix, melanomas, and AIDS-related Karposi's sarcoma (Skeel 1999). Taxol was discovered in the 1960s as part of an interagency program between the National Cancer Institute and the United States Department of Agriculture developed to obtain and screen the anti-proliferative properties of plant extracts. Extracted from the bark of *Taxus brevifolia* (the Pacific Yew tree), Taxol's structure was elucidated in 1971 (Wani et al. 1971) and the compound showed promising activity against leukemia and melanoma models, prompting the continued isolation from *T. brevifolia* bark (Cragg et al. 1993).

Early stage production of Taxol resulted in extremely low yields (Cragg et al. 1993; Cragg and Snader 1991). The sacrifice of a 100-year old tree generated approximately 3 kg of bark, yielding approximately 300 mg of purified Taxol or the equivalent of approximately a single dose (Horwitz 1994). In 1988, a semi-synthetic method for producing Taxol from a common precursor (10-deacetylbaccatin III) was developed (Denis et al. 1988; Witherup et al. 1990). The semi-synthetic route offered two advantages. First, relatively higher quantities of 10-deacetylbaccatin III could be isolated from the needles of other more

prevalent *Taxus* species (specifically *Taxus baccata* l, the common Yew) (Denis et al. 1988). Second, isolation from the renewable Yew needle presented an environmental friendly alternative to the destructive production processes dependent upon Yew bark harvest. A fully synthetic route to Taxol was independently accomplished in 1994 by two groups (Holton et al. 1994a; Holton et al. 1994b; Nicolaou et al. 1995a; Nicolaou et al. 1995b; Nicolaou et al. 1995c; Nicolaou et al. 1995d; Nicolaou et al. 1994), and spurred much interest in the synthetic chemistry community (Danishefsky et al. 1996; Morihira et al. 1998; Mukaiyama et al. 1999; Wender et al. 1997a; Wender et al. 1997b). In the end, the compound's complex molecular architecture containing eleven choral centers and a notably rare oxetane ring complicated efforts to establish an efficient and economic synthetic production process (Nicolaou et al. 1994).

To date, the most efficient synthesis of Taxol resulted in a 0.4% overall yield after 37 steps (Kingston 2001). Taxol has been extracted from *T. brevifolia* bark at a roughly 1000 µg g biomass^{-1} (Horwitz 1994) and produced in *Taxus canadensis* suspension cell culture at roughly 120 µg g biomass^{-1} (Ketchum et al. 1999). Meanwhile, heterologous production of the first committed step to Taxol biosynthesis, taxa-4(5),11(12)-diene (henceforth referred to as taxadiene) has only reached 600 ng g biomass^{-1} for *Arabidopsis*

thaliana (Besumbes et al. 2004), 471 µg g biomass^{-1} for transgenic tomato fruit (Kovacs et al. 2007), and 402 µg g biomass^{-1} for *S. cerevisiae* (Engels et al. 2008) (Figure 6).

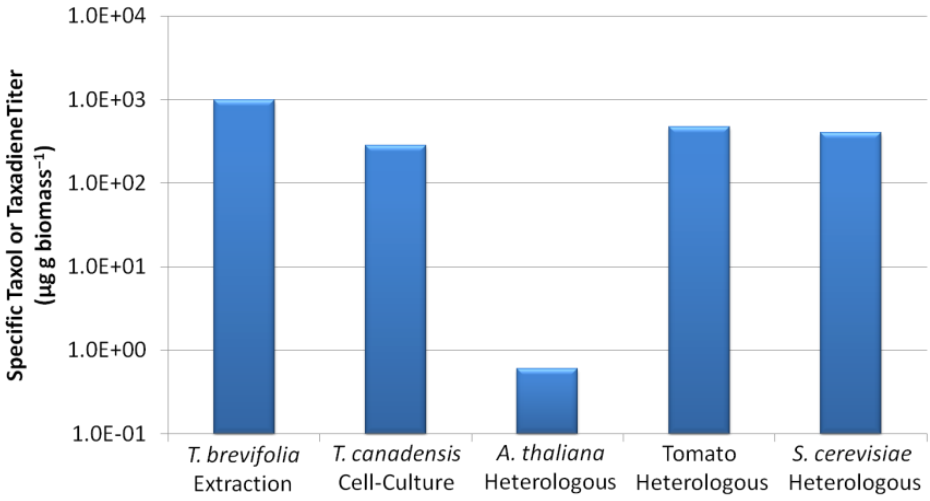

Figure 6 Data comparing specific production titers of Taxol and taxadiene. Shown is Taxol production from native hosts (extraction from *T. brevifolia* tree bark and *T. canadensis* suspension cell-culture) and taxadiene production from heterologous hosts (*A. thaliana*, tomato fruit, and *S. cerevisiae*). Data taken from: (Besumbes et al. 2004; Engels et al. 2008; Horwitz 1994; Ketchum et al. 1999; Kovacs et al. 2007).

A central difference between Taxol and erythromycin is that the entire Taxol biosynthetic pathway has not been elucidated (Frense 2007; Renneberg 2007). While the DXP pathway produces the two universal precursors for isoprenoid biosynthesis (Figure 3), Figure 7 shows a simplified metabolic map of "elongation and cyclization" and "tailoring" pathways with relation to Taxol biosynthesis. In this pathway, there are multiple steps that are not elucidated to generate the intermediate 10-deacetylbaccatin III, as well as multiple tailoring steps after this metabolite. However, like the case for erythromycin, heterologous production of Taxol is viewed with great interest due to the potential to rationally control or alter the biosynthetic process towards improved titer and/or bioactivity.

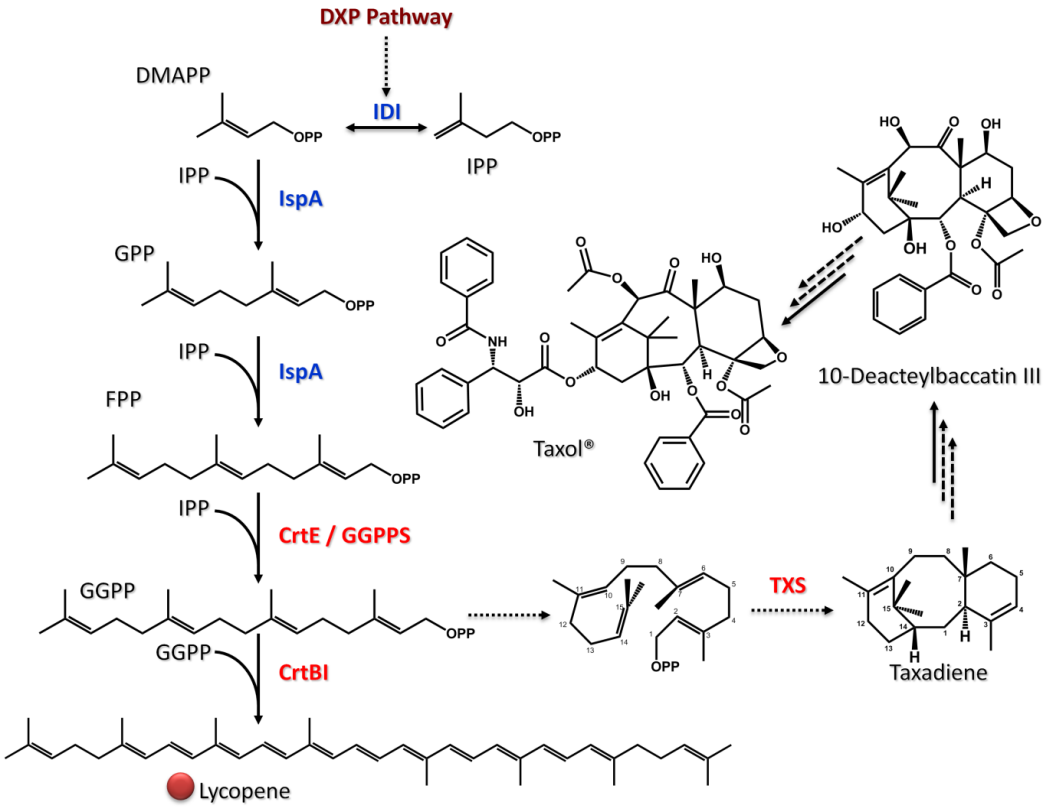

Figure 7 The "Extension & Cyclization" phase of isoprenoid and Taxol biosynthesis. Dotted lines indicate partial reaction steps, while dashed lines indicate multiple unknown steps in Taxol biosynthesis. Enzyme abbreviations in red are heterologous to *E. coli*. The red dot next to lycopene indicates that it has a red chromophore. Abbreviations: IDI = isopentenyl diphosphate isomerase, IspA = prenyltransferase, CrtE / GGPPS = geranylgeranyl diphosphate synthase, CrtB = phytoene synthase, CrtI = phytoene dehydrogenase.

Heterologous Biosynthesis using *Escherichia coli* & Considerations

As stated previously, many natural products are produced in relatively low titers in their native hosts, or the native hosts are uncultivable in laboratory and/or manufacturing conditions. *E. coli* has served as a model organism for basic microbiology studies as well as molecular biological tool development. It was also the organism that spurred heterologous production of biopharmaceuticals (protein) products, such as recombinant insulin (by Eli Lilly, approved in 1982) and recombinant human growth hormone (by Genentech, approved in 1985) (Papini et al. 2010). More recently, a joint venture between DuPont and Tate & Lyle commercialized a process for using *E. coli* as a heterologous host (with genes from *Saccharomyces cerevisiae* and *Klebsiella pneumoniae*) for producing 1,3-propanediol (propylene glycol) at a yield of 51% (g 1,3-propanediol g glucose^{-1}), titer of 135 g l^{-1}, and productivity of 3.5 g l hr^{-1} (Nakamura and Whited 2003), with an annual production of 120,000 tons. Moreover, there is an incredible body of work recently dedicated towards the production of *E. coli*-derived natural and non-natural alcohol fuels (Atsumi et al. 2008a; Atsumi et al. 2008b; Atsumi and Liao 2008a; Cann and Liao 2008; Connor et al. 2010; Connor and Liao 2008; Hanai et al. 2007; Shen and Liao 2008; Zhang et al. 2008b), reviewed recently here (Alper and

Stephanopoulos 2009; Atsumi et al. 2008b; Atsumi and Liao 2008b; Connor and Liao 2009; Jarboe et al. 2010; Keasling and Chou 2008; Lynd et al. 2008; Mukhopadhyay et al. 2008; Savage et al. 2008; Stephanopoulos 2007; Wackett 2008).

Classical heterologous hosts for secondary metabolite production are different than classical hosts for protein and biofuel production. They are actinomycetes such as *Streptomyces coelicolor*, *Streptomyces lividans*, and *Streptomyces albus*, particularly due to their existing ability to produce a wide variety of therapeutic natural products, and their developing genetic engineering technologies (Baltz 2010). More recently developed hosts for heterologous natural product biosynthesis include *E. coli* (Mutka et al. 2006b; Pfeifer et al. 2001), *Bacillus subtilis* (Eppelmann et al. 2001), *Pseudomonas putida* (Fu et al. 2008), *Saccharomyces cerevisiae* (Kealey et al. 1998), *Aspergillus nidulans* (Kennedy et al. 1999), and even a genome-minimized version *Streptomyces avermitilis* (Komatsu et al. 2010). In general, the ideal heterologous host should have high specific growth rates, minimal nutrient requirements, established genetic engineering protocols, a fully sequenced and well annotated genome, and easily scalable bioprocesses.

Heterologous production of pharmaceuticals offers advantages over production in native hosts in several ways: 1) the genetic and

metabolic knowledge of a heterologous host is typically much greater than the native host, 2) as a result, the computational and experimental application of metabolic engineering is much easier than the same steps in a potentially uncharacterized native host, and 3) the knowledge of existing bioprocesses and scale-up protocols for heterologous hosts generally have been determined, leading to platform processes and reduced development times (Zhang et al. 2008a). While the production of protein pharmaceuticals often relies on maximal forced over-expression of one to a few genes, the heterologous production of small-molecules often requires reconstitution of a variety of enzymes to catalyze the multi-step transformation of either an exogenously fed metabolite or a native metabolite to a more complex product. In this case, the enzymes should only be produced in sufficient quantities such that they are not rate-limiting, but also not at too high a level such that their production acts as major sinks of nucleotides or amino acids for mRNA or protein production, respectively (Keasling 2008). At the same time, the introduction of foreign metabolites into a cellular environment can cause stress-responses potentially detrimental to both cell growth and product production. Concurrently with heterologous expression for catalyzing these transformation steps, one must attempt to tailor the metabolic network to increase carbon-flux from native metabolism to heterologous metabolism such that

more product can be made (Stephanopoulos and Vallino 1991). This is far from trivial in that cellular systems, as opposed to engineered systems (such as an electrical circuit), can evolve themselves to overcome cellular perturbations (for example, a gene-knockout) by reverting its metabolism in such a way that may have been unexpected to the researcher. As a recent example of successful heterologous small-molecule biosynthesis, *E. coli* was utilized to produce the sesquiterpene amorphadiene (complex precursor to the antimalarial compound artemisinin) at impressive titers of nearly 30 g l^{-1} with an overall volumetric productivity of approximately 185 mg l^{-1} hr^{-1} (Tsuruta et al. 2009). It should be noted that both the 1,3-propanediol and amorphadiene examples were engineered through heuristic methods, and as far as the author is aware, no systematic methods were used to aid in strain development.

Metabolic Modeling for Driving Metabolic Engineering

As a result of the large-scale and often unpredictable nature in engineering biological networks (Alm and Arkin 2003; Almaas 2007; Jeong et al. 2000), quantification of metabolic fluxes is of utmost importance to understand cellular regulation, identify bottlenecks in product formation, and gain insight to the fundamental processes of biological systems. With the development of modern "-omics" techniques such as genomics, transcriptomics, proteomics (Aebersold

and Mann 2003; Aggarwal and Lee 2003; Aldor et al. 2005; Brotz-Oesterhelt et al. 2005; Han et al. 2001; Hancock et al. 2002; Lee and Reardon 2003; Lee and Lee 2003; Wang et al. 2005), and metabolomics (Arita 2009; Garcia et al. 2008; Lee et al. 2006; van der Werf et al. 2007; Wishart 2007; Zamboni and Sauer 2009), along with the methodologies of systems biology (Aggarwal and Lee 2003; Barrett et al. 2006; Church 2005; Daniels et al. 2008; Dhurjati and Mahadevan 2008; Kell 2004; Mukhopadhyay et al. 2008; Nielsen and Jewett 2008; Novere et al. 2009; Otero and Nielsen 2010; Panagiotou et al. 2009; Papini et al. 2010; Park et al. 2008; Rochfort 2005; Rokem et al. 2007; Souchelnytskyi 2005; Stephanopoulos et al. 2004; Strange 2005; Wang et al. 2006; Westergaard et al. 2007) gaining popularity among both engineers and biologists, the wealth of information on an increasing number of organisms is advantageous for metabolic engineering (Durot et al. 2009; Feist et al. 2009; Medini et al. 2008). The sequencing of many genomes of industrially-relevant organisms (Bentley et al. 2002; Blattner et al. 1997; Durfee et al. 2008; Goffeau et al. 1996; Ikeda et al. 2003; Jeffries et al. 2007; Kunst et al. 1997; Nolling et al. 2001; Oliynyk et al. 2007) and improved multidimensional genome annotation (Maillet et al. 2007; Reed et al. 2006a; Reed et al. 2006b) has led to a better understanding of the structure, connectivity, and capabilities of many metabolic networks. Contrasting with

metabolomics, which aims to quantitatively measure the concentrations of every metabolite within a cell population at a certain time (Kell 2004; Nielsen and Oliver 2005; van der Werf et al. 2007), metabolic flux analysis (MFA) aims to quantify the flow of primarily carbon and nitrogen throughout a metabolic network. As a result, metabolic fluxes are the functional output of the transcriptome, proteome, and metabolome.

Within MFA, there exist data-driven studies (often using isotopically labeled substrates) and optimization-driven studies for the quantification of fluxes. In the first method, the stoichiometric matrix (to be introduced later) is reduced to an over-determined form (more independent equations than unknown variables) and then either least squares linear regression (in the case where isotope labeling data is not available) or least squares non-linear regression (in the case where labeling data is available) is used to determine the flux distribution (Tsai and Lee 1988; Vallino and Stephanopoulos 1993). In the second method, the stoichiometric matrix is under-determined (more variables than unique equations) and then optimization is used to determine the flux distribution (Edwards et al. 2002). The first method relies more heavily on experimental measurements and, therefore, often requires in the utilization of a smaller stoichiometric model. In contrast, the solution space in the second method can be constrained with relatively

few measurements (usually, carbon-source uptake rate and biomass composition), and even large models by today's standards (over 1,000 reactions) can be solved quite rapidly with modern optimization packages on single-processor systems. The aim of this portion of Chapter 2 is to familiarize the reader with a variety of techniques that are currently used to quantify metabolic fluxes, how pharmaceutical and biopharmaceutical strain or cell-line development can be enhanced with these studies, and the recent success stories achieved thus far in applying MFA for pharmaceutical and biopharmaceutical production (Figure 8).

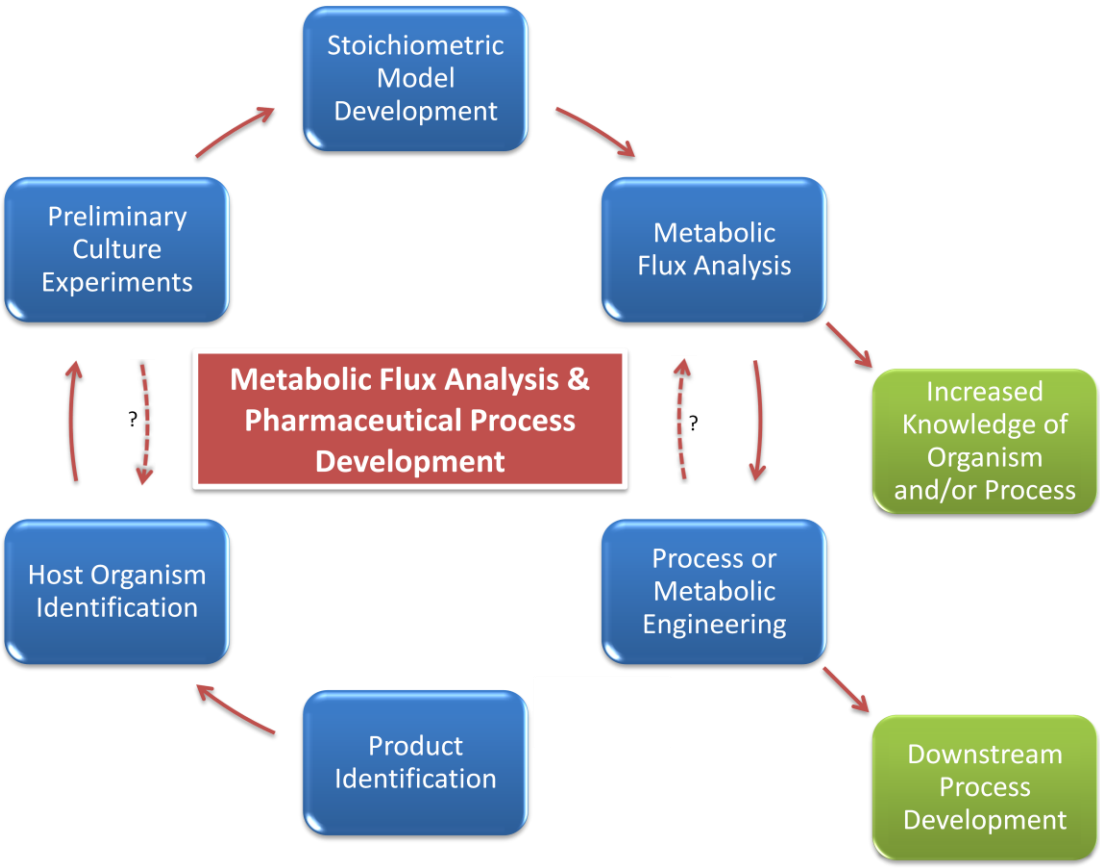

Figure 8 A schematic overview of how metabolic flux analysis fits into the development of pharmaceutical production processes.
After identification of a therapeutic molecule in drug discovery, a development group must choose a production host organism. After which, vectors, strains, or cell-lines are created which are able to generate the product of interest and preliminary culture experiments are conducted. If the product titer or quality is inferior, another host organism may be selected. Next, a stoichiometric model containing pertinent portions of metabolism can be developed, as has been described in the text. Metabolic flux analysis can be applied to this model to: 1) gain more information regarding the metabolism of the organism in question and how it pertains to product titer or quality; and 2) drive further metabolic or process engineering to improve product titer or quality. As highlighted in this review, metabolic flux analysis can be utilized to facilitate upstream strain and process development.

Model Development

Although much work has been conducted on the metabolic regulation of microbial systems, particularly in the case for secondary metabolite production (Martin and Demain 1980; Smith et al. 1962), kinetic information is not available to sufficiently model these systems beyond individual compartments of the *Escherichia coli* cell (Castellanos et al. 2007; Castellanos et al. 2004) and the human erythrocyte (Joshi and Palsson 1989a; Joshi and Palsson 1989b; Joshi and Palsson 1990a; Joshi and Palsson 1990b). Conversely, knowledge regarding the stoichiometry of the biochemical reactions that comprise these microbial systems (and higher organisms as well) is well-established and largely unambiguous.

If it is assumed that the cells are well-mixed reaction systems, the dynamic mass-balance around the system leads to a non-homogeneous system of coupled ordinary differential equations:

Equation 1

$$\frac{dX_{met}}{dt} = r_{met} - \mu X_{met}$$

In Equation 1, X_{met} is a vector of metabolite concentrations, r_{met} is a vector of the net rates of formation of those corresponding metabolites, μ is the specific growth rate, and t is time. It has been cited that there is very high turnover of pools of intracellular metabolites, even after significant environmental perturbations,

resulting in a pseudo-steady state approximation for these metabolites, as described in Equation 2:

Equation 2

$$0 = r_{met} - \mu X_{met}$$

Next, if it is assumed assume that there is no dilution of the metabolite pool due to cell growth, this term is also dropped from, yielding Equation 3:

Equation 3

$$0 = r_{met}$$

Being that r_{met} is a vector expressing the net rates of production of all intracellular metabolites, it can be expressed as a system of coupled linear algebraic equations:

Equation 4

$$0 = Sv$$

In Equation 4, S is an m by n dimensional matrix is comprised of m metabolites and n reactions, whereas the contents of each entry in the matrix represent the stoichiometric coefficient of the m^{th} metabolite participating in the n^{th} reaction. Here also, v, is a vector of metabolic fluxes (reaction rates) for all n reactions in the system. As such, the stoichiometric matrix contains chemical information (reaction stoichiometry) and network information (metabolite connectivity). Stoichiometric matrices are sparse, meaning that a large fraction of the matrix's entries are zero. Most metabolites only take place in a small-number of reactions, while a small number of

metabolites take place in many reactions, which are referred to as currency metabolites (such as ATP, NAD(P)(H), and H_2O). Stoichiometric matrices vary in size and, as a result, their application. As will be discussed later, isotopic-labeling studies require the utilization of smaller-scale models (on the order of 10^1 to 10^2 reactions), while optimization-based analyses such as flux balance analysis can be quite large (over 10^3 reactions) and solved rapidly using either free or commercially-available optimization software.

Developed genome-scale stoichiometric models of organisms of biotechnological or industrial importance include: *E. coli* (Feist et al. 2007), *Saccharomyces cerevisiae* (Herrgard et al. 2008), *Bacillus subtilis* (Henry et al. 2009), *Streptomyces coelicolor* (Borodina et al. 2005), *Aspergillus niger* (Andersen et al. 2008), *Aspergillus nidulans* (David et al. 2008), *Aspergillus oryzae* (Vongsangnak et al. 2008), *Lactococcus lactis* (Oliveira et al. 2005), *Mannheimia succiniciproducens* (Kim et al. 2007b), *Methanosarcina barkeri* (Feist et al. 2006), *Geobacter sulfurreducens* (Mahadevan et al. 2006), *Geobacter metallireducens* (Sun et al. 2009), *Rhodoferax ferrireducens* (Risso et al. 2009), *Corynebacterium glutamicum* (Kjeldsen and Nielsen 2009), *Clostridium acetobutylicum* (Lee et al. 2008; Senger and Papoutsakis 2008; Shinfuku et al. 2009), *Pseudomonas putida* (Puchalka et al. 2008), *Arabidopsis thaliana* (de Oliveira Dal'Molin et

al. 2010; Poolman et al. 2009), *Zymomonas mobilis* (Widiastuti et al. 2010), *Mus musculus* (Sigurdsson et al. 2010), *Dehalococcoides* sp. (Ahsanul Islam et al. 2010), *Clostridium cellulolyticum* (Salimi et al. 2010), *Pichia pastoris* (Chung et al. 2010; Sohn et al. 2010), and *Clostridium thermocellum* (Roberts et al. 2010) with many currently being updated, expanded, and developed. In addition to these organisms, there have been a variety of other organisms reconstructed to better understand physiological, medical, and evolutionary problems (AbuOun et al. 2009; Baart et al. 2007; Becker and Palsson 2005; Beste et al. 2007; Boyle and Morgan 2009; Duarte et al. 2007; Gomes de Oliveira Dal'molin et al. 2010; Gonzalez et al. 2010; Heinemann et al. 2005; Jamshidi and Palsson 2007; Kim et al. 2010; Oberhardt et al. 2008; Pastink et al. 2009; Plata et al. 2010; Resendis-Antonio et al. 2007; Schilling and Palsson 2000; Sheikh et al. 2005; Suthers et al. 2009; Teusink et al. 2006; Thiele et al. 2005; Tsoka et al. 2004; Vanee et al. 2010). Figure 9 is a pictorial description of these models as a function of the year they were developed with their corresponding size in terms of total reaction number. As has been the case with much experimental work, *E. coli* has served as the model organism for stoichiometric model construction and MFA in general (Feist and Palsson 2008). While the use of genome-scale microbial models for metabolic engineering has been reviewed, genome-scale modeling of

mammalian cell systems is only more recently being developed (Selvarasu et al. 2009b).

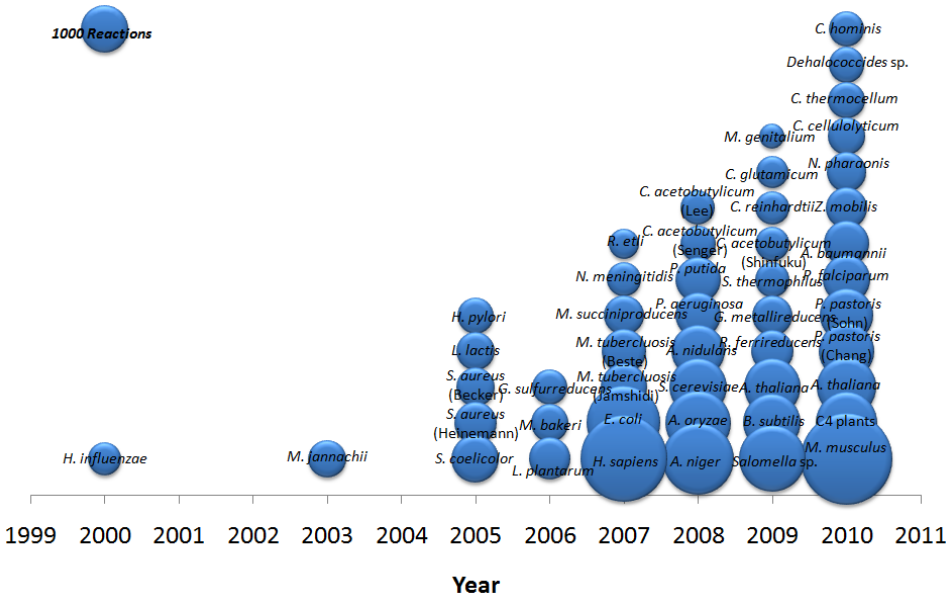

Figure 9 Genome-scale stoichiometric model development.
A figure representing all of the most updated genome-scale stoichiometric models of prokaryotic and eukaryotic cellular metabolism as a function of the year they were developed. Models that were developed concurrently have been parenthetically annotated with the first author's name. The bubble size represents the relative size of the metabolic model in terms of reactions. A reference bubble is shown which corresponds to 1000 reactions.

Model refinement, especially at the genome-scale, is important to correct for inconsistencies between model predictions and experimental observations. This trait has most recently been examined at the level of cell lethality with respect to single gene-knockouts or alternative carbon sources. Two types of inconsistencies can arise: 1) the model predicts growth when no growth is observed experimentally, or 2) the model predicts no growth when growth is observed experimentally. An automated algorithm, GrowMatch (Kumar and Maranas 2009), has been developed to reconcile these inconsistencies either by removing functionalities in the model (in the first case), or by adding functionalities in the model (in the second case). This algorithm was applied to the latest genome-scale *E. coli* model, *i*AF1260 (Feist et al. 2007), and suggested consistency-restoring hypotheses for 69 of the 110 single gene-knockout inconsistencies identified. This algorithm could be particularly useful in the development and rapid refinement of new genome-scale models.

Most recently, a high-throughput method for automated generation, optimization, and analysis of genome-scale metabolic networks was developed (Henry et al. 2010). This method, called Model SEED, can generate a draft metabolic model in approximately 48 hr from an assembled genome sequence. This method was applied to generate 130 genome-scale metabolic models to a taxonomically

diverse set of bacteria. Twenty-two of the models were validated against available experimental gene essentiality and Biolog (growth phenotype arrays) data, improving the average model accuracy from 66% to 87% after the optimization model refining method using GapFill and GapGen, developed previously (Satish Kumar et al. 2007).

Constraints-Cased Methods

Flux Balance Analysis & Constraints-Based Modeling

Rather than measuring extracellular metabolites and their isotopic distributions to determine metabolic fluxes, other strategies have employed optimization frameworks with relatively few measurements (usually, the rate of carbon-source uptake) to determine metabolic fluxes. The most widely-used framework in this context is the flux balance analysis (FBA) approach (Edwards et al. 2002). FBA is a linear optimization problem which assumes a metabolic objective of biomass formation which, at steady-state, is the growth rate. This metabolic objective is based on the theory that a cellular system will maximize its resources (primarily a carbon-source) to make biomass. While FBA has been shown to be fairly accurate when it comes to predicting growth rates and exchange fluxes in numerous experimental settings (at steady-states), it has shown to have less predictive power in other settings.

Before FBA was introduced, linear programming was used to determine maximum yields for acid and solvent production in clostridia (Papoutsakis 1984), as well as to study aerobic overflow of acetate in *E. coli* (Majewski and Domach 1990). Their network was comprised mainly of the TCA cycle and used ATP and GTP production as the metabolic objective subject to additional constraints, as their stoichiometric equations did not include a reaction for biomass formation (Majewski and Domach 1990). Early studies utilizing FBA on small-scale models of *E. coli* metabolism were used to predict phenotypes such as growth-rate, byproduct formation, and amino acid production (Varma et al. 1993a; Varma et al. 1993b; Varma and Palsson 1994a) under different oxygenation rates. Drawing upon knowledge that RNA content increases and DNA and protein contents decrease with increasing growth-rate, the incorporation of growth-rate dependent equations of biomass formation and energy requirements were utilized in an *E. coli* model (of 300 reactions and 289 metabolites) (Pramanik and Keasling 1997; Pramanik and Keasling 1998). This importance was highlighted in the observation that, with the correct biomass composition, the predicted fluxes differed by 16% from experimental measurements, while the predicted and experimental fluxes differed by 80% when using the incorrect biomass composition (Pramanik and Keasling 1997). It has been recognized that these

linear programming problems will find only a single solution; whereas, multiple flux distributions may result in the same objective function value. This led researchers to develop mixed integer linear programming (MILP) frameworks that enumerate multiple flux distributions that satisfy the given constraints and result in the same objective function value (Lee et al. 2000). More recently, this framework was applied to a genome-scale model of *E. coli*, revealing that only a small subset of metabolic reactions have variable fluxes across optima(Reed and Palsson 2004).

This FBA work was then extended to simulating batch and fed-batch culture by discretizing time into finite elements (Varma and Palsson 1994b; Varma and Palsson 1995). The modeling results were then directly compared to experimental results with overall good accuracy (Varma and Palsson 1994b). However, it was found that these models were sensitive to parameters (perhaps not surprisingly) such as glucose and oxygen uptake rate, while relatively insensitive to other constraints such as growth and non-growth associate maintenance energy requirements (Varma and Palsson 1995). Later, this problem was reformulated as dynamic flux balance analysis (dFBA) to better simulate and understand *E. coli* diauxic growth on glucose and acetate (Mahadevan et al. 2002). In this work, a nonlinear, dynamic objective function was developed to prevent the calculation of

physiologically unrealistic metabolite concentrations or metabolic fluxes. As a result, this algorithm proved to be a better predictor of the reutilization of acetate by *E. coli* in batch culture.

While FBA and the remainder of the "constraints-based" modeling strategies are solely governed by stoichiometry, the inclusion of thermodynamic principles can further constrain solution-space of the optimization problem without the need for additional measurements. Two such methodologies have been developed to address the incorporation of thermodynamic calculations into optimization-based MFA calculations: energy balance analysis (EBA) (Beard et al. 2002) and thermodynamic metabolic flux analysis (TMFA) (Henry et al. 2007). It is recognized that FBA only generates a single solution given a set of constraints (for example, an FBA solution may be stoichiometrically feasible but not thermodynamically feasible). Both EBA and TMFA aim to distinguish between stoichiometrically and thermodynamically feasible states, allowing for only solutions that satisfy both to be valid. EBA is based on chemical potential differences associated with reactions (Beard et al. 2002), while TMFA utilizes group contribution theory (Mavrovouniotis 1991) to estimate the standard Gibbs free energy change associated with reactions (Henry et al. 2007).

Another method for reducing the solution-space of the optimization is to incorporate regulatory information, most often in the form of Boolean operators “AND”, “OR”, and “NOT” (Covert et al. 2001; Cox et al. 2005). While the kinetic parameterization of many of these processes has not been explored thoroughly, a vast amount of knowledge has accumulated regarding gene and regulatory networks in a variety of organisms. Unlike the chemical and thermodynamic constraints described previously, regulatory constraints are generally organism-specific and therefore self-imposed and perhaps explained by evolution (Covert et al. 2001). While gene-expression arrays can be used to quantify the amount of transcripts with a large-coverage of the genome for some organisms, those reaction fluxes associated with non-transcribed genes can be set to zero. Regulatory information can be particularly applicable in simulating dynamic experiments. Interfacing regulatory constraints through Boolean operators with FBA has been used to simulate two or more carbon-sources and aerobic/anaerobic diauxie (Covert et al. 2001), the Arc two-component system, a cellular anoxic redox controller (Cox et al. 2005), and more recently, the quorum sensing *lux* system (Anesiadis et al. 2008), all in *E. coli*.

Knockout Simulation & Identifying Metabolic Engineering Targets

Although traditional FBA has been quite successful at simulating optimal phenotypes in many different cellular systems, this algorithm only provides cursory information with regards to implementing metabolic engineering strategies to improve the production of a product. One of the most significant advances in this respect was the development of the minimization of metabolic adjustment (MoMA) optimization (Segre et al. 2002). It was hypothesized that mutant strains with single- or multiple-gene knockouts would not perform in an "optimal" state as the wild-type strain would, therefore nullifying the FBA biomass objective function. The basis for this hypothesis assumes evolution has accounted for "designing" a microbial network that makes biomass, while a mutant strain would not possess the same qualities. As a result, the MoMA algorithm uses the same stoichiometric constraints as FBA, but relaxes the maximization of the biomass metabolic objective (Segre et al. 2002). The linear objective function was replaced with a nonlinear objective function (which is reformulated as a quadratic programming problem) of minimizing the Euclidean distance between the simulated wild-type flux vector (as calculated by an FBA of the wild-type model) and the knockout-strain flux vector. This algorithm was then capable

of improving the predictive capability of actual fluxes (over FBA) in carbon-limited, carbon-rich, and nitrogen-limited media in a small-scale *E. coli* model. A slight alteration of the MoMA algorithm, which has been called linear MoMA (Becker et al. 2007), was developed with two distinct differences: 1) the objective to minimize the 1-norm distance between the wild-type flux vector and the knockout flux vector (rather than the Euclidean, or 2-norm distance as proposed in MoMA); and 2) solving the wild-type and the knockout fluxes simultaneously, as described below (Burgard et al. 2003).

A variety of bi-level optimization algorithms have built upon the FBA and MoMA approaches (Burgard et al. 2003; Pharkya et al. 2003; Pharkya et al. 2004; Pharkya and Maranas 2006). It is well understood that there is a tradeoff between growth (biomass formation) and product overproduction in cellular systems. By utilizing a multilayered optimization structure, the OptKnock framework (Burgard et al. 2003) optimizes for one competing objective within another: a cellular objective (biomass production) within an industrial objective (chemical production) subject to the same stoichiometric constraints and carbon-source uptakes rates as described earlier. The OptKnock framework was applied to succinate, lactate, and 1,3-propanediol production on a genome-scale metabolic model of *E. coli* (Edwards and Palsson 2000), and identified multiple combinations of double-, triple-, and quadruple-

knockout strains capable of producing the three products (separately) while still sustaining growth.

Expanding upon OptKnock, the OptStrain algorithm was developed (Pharkya et al. 2004). This framework involves: 1) automated curation of databases to compile a pathway or multiple pathways to produce a product from a substrate, 2) identification of the maximum-yield pathway while minimizing the use of non-native reactions, and 3) incorporation of the identified pathway into the parent-strain model followed by the OptKnock procedure to eliminate reactions competing with the target product formation. The authors examined the production of both native and non-native products, including hydrogen production capabilities of *E. coli* (Reed et al. 2003), *Clostridium acetobutylicum* (a known hydrogen producer) (Desai et al. 1999; Papoutsakis 1984), *Methylobacterium extorquens* (a known methanol consumer) (Van Dien and Lidstrom 2002), and the production of vanillin from glucose in *E. coli*.

Another extension of OptKnock is the OptReg framework (Pharkya and Maranas 2006). Rather than examining the sole inclusion or exclusion ("on" or "off") of a reaction from a metabolic network and its impact on product formation, this framework aims at the modulation ("up-regulation" or "down-regulation") of fluxes relative to a base case. In some cases, a proposed knockout to improve

production of a biochemical may be lethal to the cellular system, while down-regulation of the gene to a basal level of expression may allow the cell to function at near-wild-type ability, while still reverting flux towards production of the biochemical. Gene-additions and knockouts do not constitute the entire array of genetic engineering tools available. In addition, engineering promoters (Alper et al. 2005a; Cox et al. 2007) as well as a variety of other cloning and expression strategies (Chou 2007; Sorensen and Mortensen 2005) can accomplish the fine-tuning of both native and heterologous gene expression.

Regulatory on/off minimization (ROOM) was developed as an alternative to MoMA (Shlomi et al. 2005). Rather than minimizing the Euclidean distance between the simulated wild-type and the knockout-strain flux vectors, ROOM minimizes the number of flux changes. An argument for this methodology is that ROOM implicitly accounts for regulatory changes by identifying significant flux changes of a few reactions, rather than identifying small flux changes in many reactions, of which MoMA favors. While, in many cases, fluxes calculated by both MoMA and ROOM correlated well (much better than FBA) with experimental fluxes, MoMA predicts less true-positive lethal knockouts than both FBA and ROOM (Shlomi et al. 2005). A possible explanation for this is that, for central metabolism, for which labeling experiments can determine flux distributions, both MoMA and

ROOM are accurate predictors. Whereas, for reactions not involved in primary metabolism, MoMA may incorrectly predict lethality.

Lastly, although not the subject of this review, evolutionary algorithms have also been applied to model cellular metabolism and identify knockout targets for improved product yield. Of particular note here is the OptGene framework (Patil et al. 2005), which interfaces a genetic algorithm (GA) for exploring genotypic-space with an optimization algorithm (FBA, MoMA, or ROOM) as a scoring method. As an application, the authors explored glycerol, succinate, and vanillin production in a genome-scale model of *S. cerevisiae*. They found that the optimal yield for succinate was obtained with a four gene-knockout. This was identified by searching only 0.03% of the total solution space, making the OptGene framework a computationally tractable option (Patil et al. 2005). This framework was used to drive metabolic engineering of *S. cerevisiae* for improving the production of heterologous cubebol (a sesquiterpene natural product) by 85% (Asadollahi et al. 2009). A recent study compared the performance of a genetic algorithm to the performance of a simulated annealing algorithm (SAA) (Rocha et al. 2008). A comparison of these two stochastic algorithms showed that their performance in identifying optimal genotypes for producing lactate or succinate in *S. cerevisiae* or *E. coli* was quite similar. In yet another method developed to identify

gene knockout targets for metabolite over-production, a local search method was utilized (Lun et al. 2009). The optimal three-knockout strategy for acetate over-production was found in 2,566 seconds using a local search while a global search took 36,924 seconds. The MILP local search strategy, called Genetic Design through Local Search (GDLS) is based on the idea that knockouts that aid in metabolite over-production are close to each other with respect to metabolite or reaction connectivity.

Figure 10 shows a pictorial timeline of important methodological advances and applications for both data-driven and optimization-driven approaches to MFA.

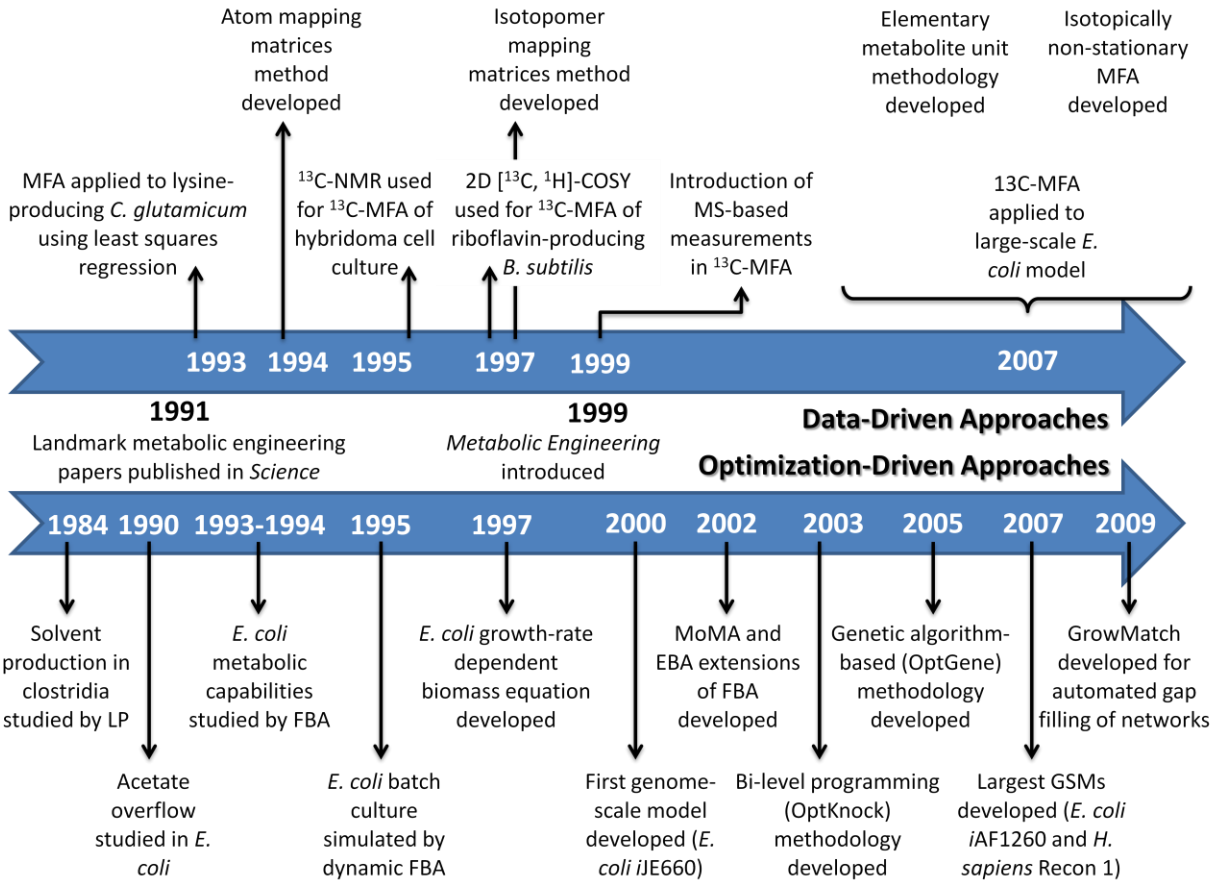

Figure 10 A timeline showing important methodological advances and applications for both data-driven and optimization-driven approaches to MFA.

Applications in Pharmaceutical Strain Development

The importance of small-molecules as pharmaceuticals and therapeutic agents is well-recognized (Barrett 2005; Clardy et al. 2006; Clardy and Walsh 2004; Cragg et al. 1997; Newman 2008; Newman and Cragg 2007). The term "small-molecule" is typically taken to mean molecules of less than 1,000 daltons and of type non-protein and non-peptide. Of the nearly 380 billion (US$) in pharmaceutical revenue in 2007, 81% was derived from small molecule sales. Another 17.5% was derived from biologics, or biopharmaceuticals (therapeutic proteins and monoclonal antibodies) sales, and 1.1% from vaccine sales. The most significant growth rate in pharmaceutical sales from 2007-2012 will be that of the monoclonal antibody segment, at more than 11% (Datamonitor 2008). Natural products are small-molecules made by a variety of organisms that have widespread clinical relevance. Examples are the ß-lactam penicillin-G from *P. chrysogenum*, the macrolide erythromycin A from *Saccharopolyspora erythraea*, and the diterpenoid Taxol® from *Taxus brevifolia* (the Pacific yew tree). Between 1981 and 2002, of 1,031 FDA approvals for clinical usage, 5% were natural products while another 23% were natural product-derived small-molecules (Newman et al. 2003). Although some therapeutic proteins were historically derived from animal or human sources (for example, porcine insulin, human transferrin, human

serum albumin, or cadaver-derived human growth hormone), contamination of such products with adventitious agents including viruses and prions led to the establishment of recombinant protein production processes based on the recombinant engineering of well-characterized and safe microbial or mammalian host cells.

While the production of penicillin from *P. chrysogenum* has served as the classical example for traditional strain improvement through medium/bioprocess optimization and random mutagenesis (Adrio and Demain 2006; Demain 2006), the rational engineering of native and heterologous strains for the production of natural products is becoming increasingly popular (Boghigian and Pfeifer 2008; Chemler and Koffas 2008; Leonard et al. 2009; Zhang et al. 2008a). The titers of these high-value products can often be quite low, resulting in poor process productivities and, therefore, high costs. In an effort to improve production, the application of modern metabolic engineering techniques is essential. As stated previously, elucidation of metabolic fluxes through labeling methods and simulation of fluxes and identification of metabolic engineering targets through optimization frameworks is essential for improving the production. While most of the efforts in this area have revolved around the production of non-therapeutic small-molecules such as amino acids (Lee et al. 2007; Park et al. 2007) and organic acids (Burgard et al. 2003; Sanchez et al.

2006b), this review will highlight some of the successes in applying the previously described MFA techniques for the production of both small-molecule and protein pharmaceuticals.

Small-Molecule Pharmaceutics

One of the earliest examples in applying MFA for secondary metabolite or complex small-molecule production was in riboflavin-producing *B. subtilis* (Sauer et al. 1997; Sauer et al. 1996). Utilizing ^{13}C-labeling in a glucose-limited chemostat and two-dimensional proton detected [^{13}C, ^{1}H]-correlation spectroscopy (2D [^{13}C, ^{1}H]-COSY), the authors quantified fluxes throughout glycolysis, the pentose phosphate pathway, and the TCA cycle. Ultimately, it was concluded that *B. subtilis* has an unusually high-capacity for NADPH-regeneration and that riboflavin biosynthesis was limited by neither the supply of precursors (ribose-5-phosphate and 3-phosphoglycerate) nor the supply of cofactors (ATP and NADPH), but more likely the kinetics of the riboflavin biosynthetic pathway. One of the most difficult problems to overcome was developing a model large enough to encompass the relevant secondary metabolite pathways without producing inaccurate flux estimates with large confidence intervals. Moreover, for optimization-based studies, secondary metabolite production presents a novel challenge in defining the objective function since secondary metabolites are non-growth associated.

Applications of the optimization frameworks of FBA and MoMA were used to predict knockouts that would improve heterologous production of the natural product lycopene through *E. coli* (Alper et al. 2005b; Alper et al. 2005c). Lycopene, a C_{40} bright red carotenoid with antioxidant properties, shares precursor molecules of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) with a variety of other isoprenoids, such as two plant natural products: the anticancer-compound Taxol, and the antimalarial-compound artemisinin (from the plant *Artemisia annua*, otherwise known as sweet wormwood) (Keasling 2008; Klein-Marcuschamer et al. 2007; Withers and Keasling 2007). Lycopene's red color eases quantification through spectrophotometric methods and therefore facilitates high-throughput screening. This property then allowed or helped verify both rational and combinatorial improvements to biosynthesis (Alper et al. 2005b; Alper et al. 2005c; Alper et al. 2006a; Alper and Stephanopoulos 2008; Farmer and Liao 2001; Jin and Stephanopoulos 2007; Kang et al. 2005; Klein-Marcuschamer et al. 2007; Yoon et al. 2008; Yoon et al. 2007; Yoon et al. 2006). Given that lycopene shares early pathway intermediates with numerous other isoprenoid natural products, it was assumed that the genetic modifications responsible for improved lycopene production would translate to these other natural product

systems (once the heterologous components had been transferred to *E. coli*).

As the interplay between primary and secondary metabolism has not been elucidated in many cases for native natural products, even less is known of this relationship between primary and secondary metabolism in a heterologous host. To test this interaction, multiple knockouts were predicted and implemented, resulting in increased lycopene production while still sustaining growth. Starting with a genome-scale model of *E. coli* metabolism (Reed et al. 2003), the reactions required for lycopene biosynthesis (coded by *crtEBI*) were added before a multi-dimensional (single-, double-, and triple-knockouts) search of genotypic-space using the MoMA framework was conducted. This sequential search showed knockouts that experimentally improved lycopene production including: Δ*gdhA* resulted in 13% increased production (above 4,400 ppm), Δ*gdhA*/Δ*gpmB* resulted in 18% increased production, and Δ*gdhA*/Δ*aceE*/Δ*fdhF* resulted in 37% increased production (Alper et al. 2005b). As such, this method identified a triple-knockout strain that would increase lycopene production with little to no *a priori* knowledge of the system, a very important characteristic when attempting to improve the production of heterologous products in which there is little to no information on biosynthetic regulation. Global transcriptional

machinery engineering (gTME) (Alper et al. 2006b) further improved production (Alper and Stephanopoulos 2007), while combining predicted knockouts with over-expression targets resulted in a strain capable of a 400% increase in production (Jin and Stephanopoulos 2007). An important point to take away from this study is that an exhaustive search technique is computationally infeasible, but a greedy search method (a search method that aims to find the global optimum by iteratively searching local optima) can be successful at finding improved predicted phenotypes. In the time since this study, better search strategies (such as the GDLS and OptGene algorithms previously discussed) have been applied to the *E. coli* genome-scale metabolic network that allow for identification of greater than three or four knockout strains with improved predicted phenotypes.

Due to the highly under-determined nature of the stoichiometric matrices of genome-scale metabolic models, ^{13}C-MFA was not applied until recently. Production of a precursor to artemisinin, amorphadiene, has been accomplished in both *E. coli* (Martin et al. 2003) and *S. cerevisiae* (Ro et al. 2006). To determine fluxes in the amorphadiene-producing *E. coli* strain, a model was reduced from a genome-scale model (Reed et al. 2003) to contain 238 reactions (350 fluxes) and 184 metabolites (also including the reactions needed to make amorphadiene), a model at least four times larger than existing models

for ^{13}C-MFA. An isotopomer model was developed and GC-MS was utilized to measure the amino acid mass isotopomer distributions (MID's) arising from ^{13}C-glucose in a chemostat culture, along with three external flux measurements (glucose, cell-density, and amorphadiene). Unfortunately, due to the large scale of the model and the limited number of measured metabolites (thirteen amino acids), the authors had to reduce the model's size to achieve an acceptable χ^2 distribution. Even with this reduction, multiple local minima were statistically indistinguishable, suggesting that many flux distributions could have led to identical labeling patterns. As described earlier, utilizing the OptMeas framework on this system led to tightened confidence intervals of flux values. However, much ambiguity in the system remained due to the choice of uniformly-labeled glucose ([U-^{13}C]glucose) (Chang et al. 2008). The use of a larger model needed to encompass the relevant cellular metabolism for complex small-molecule (such as a natural product like amorphadiene) production was addressed here. However, although this effort was successful in utilizing a large-scale model for an isotopomer analysis, the study also highlights several remaining issues: 1) the model needed to be reduced in size, 2) the choice of isotope and its effect on flux estimation, 3) the significant time needed to generate a large-scale isotopomer model, and 4) the significant computation time needed to solve large-scale.

Some of the methods discussed in this paper developed more recently to address these issues.

More recently, an integrated, systems-level analysis was performed on a phosphofructokinase (PFK)-knockout mutants of *S. coelicolor* A3(2) (this organism has three highly similar PFK-encoding genes) for improved production of two native pigmented antibiotics, actinorhodin (a polyketide) and undecylprodigiosin (Borodina et al. 2008). Starting with a genome-scale model of *S. coelicolor* A3(2) (Borodina et al. 2005), the authors constrained fluxes through the $NADP^+$-dependent glucose-6-phosphate 1-dehydrogenase, phosphofructokinase, and isocitrate dehydrogenase using data from ^{13}C-MFA (from labeled glucose) and a glucose-uptake rate from extracellular measurements in the wild-type strain. Next, using MoMA, the authors predicted new flux distributions resulting in decreased (or zero) flux through the PFK-mediated reaction. In addition to the ^{13}C-MFA and MoMA calculations, the authors measured uptake and production rates of glucose, glycerol, succinate, ethanol, acetate, and pyruvate by cation-exchange HPLC coupled to refractive index (RI) and ultraviolet detectors (UV). To monitor gene-expression, the authors used microarrays for genome-wide transcriptional analysis as well as quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR) for analysis of the genes

responsible for actinorhodin and undecylprodigiosin biosynthesis. Lastly, the authors performed enzyme activity assays on PFK and glucose-6-dehydrogenase. With a wealth of data generated on the transcript and metabolic-levels, the authors concluded a variety of details which helped determine the relationship between primary and secondary metabolism in antibiotic-overproducing *S. coelicolor* A3(2). One of the three mutants, Δ*pykA2*, showed approximately a five-fold increase in both actinorhodin and undecylprodigiosin titers in both complex and minimal medium. The mutant also had a decreased maximal specific growth rate and a decreased biomass yield on glucose, while at the same time having an increased specific glucose uptake rate. At a metabolic level, deletion of *pyk* led to increased flux to the pentose phosphate pathway (7% of the imported glucose in the wild-type as compared to 50% in the Δ*pykA2* mutant). This result shows that even in this mutant, glycolysis was not blocked completely. However, metabolite quantification showed an accumulation of glucose-6-phosphate and fructose-6-phosphate, suggesting that knocking-out genes leading to glycolysis is a better way of diverting flux to the pentose phosphate pathway (PPP) than over-expression of genes leading to the PPP. While the stability of these engineered strains was not examined in particular, it is likely that the decreased growth rates is due to imbalance of intracellular metabolites (such as

the glucose-6-phosphate and fructose-6-phosphate discussed) and not toxicity of the product metabolites, due to the relatively low levels of antibiotics produced in general (approximately 50 μM at the highest levels).

Chapter 3 – Metabolic surveying of heterologous hosts for polyketide biosynthesis

Introduction

As cited in Chapter 2, heterologous hosts for polyketide biosynthesis range from classical hosts such as *Streptomyces* sp. to newer hosts such as *E. coli*, *B. subtilis*, and *S. cerevisiae*. Although these newer hosts have many advantages over the classical ones (faster growth rates, lack of complex morphologies), the new heterologous hosts may also present limitations that include an intracellular environment not evolutionarily optimized for polyketide production (Fischbach et al. 2008). Polyketide biosynthesis requires short-chain acyl-coenzyme A (CoA) monomers such as acetyl-CoA, propionyl-CoA, malonyl-CoA, methylmalonyl-CoA, and benzoyl-CoA (Chan et al. 2009; Walsh 2008). These monomers take part in sequential NADPH-dependent condensation reactions catalyzed by biosynthetic enzymes termed polyketide synthases (PKSs) (Khosla et al. 2007). Certain polyketide precursors are available from primary metabolism (for example, acetyl-CoA); however, when this is not the case, dedicated precursor metabolic pathways must be re-engineered or introduced to tailor to the specific polyketide product of interest. For example, the heterologous biosynthesis of 6-deoxyerythronolide B (6-dEB), the 14-membered aglycone macrocyclic core of the wide-

spectrum antibiotic erythromycin, was accomplished in *E. coli* by combining native and heterologous metabolism to supply the required acyl-CoA substrates needed by the similarly heterologous deoxyerythronolide B synthase (DEBS) polyketide synthase (Khosla et al. 2007; Pfeifer et al. 2001). The metabolic design allowed exogenously fed propionate to be converted to propionyl-CoA and (2*S*)-methylmalonyl-CoA. These substrates support DEBS-catalyzed 6-dEB biosynthesis and are commonly incorporated into other polyketide compounds; yet, they are normally present at low concentrations within *E. coli*. Though heterologous biosynthesis was achieved, only 5-10% of propionate used to initiate intracellular precursor supply was eventually converted to 6-dEB; hence, just as a heterologous host may provide certain precursors/cofactors to aid the heterologous biosynthetic effort, separate metabolism can also act as a drain away from the desired polyketide compound (Pfeifer et al. 2002). The situation would be aided by a better understanding of the relationship between native and heterologous metabolism such that subsequent metabolic engineering can more effectively improve the heterologous production of the desired polyketide product.

In this chapter, FBA was used to computationally characterize three potential heterologous hosts in the context of polyketide biosynthesis. More specifically, the 6-dEB pathway was chosen because

of its prior and current use as a model for heterologous polyketide biosynthesis and because biosynthesis depends on non-abundant acyl-CoA precursors that require more effort and insight to provide during the heterologous metabolic engineering needed to establish biosynthesis. In doing so, the goal was to use well-established metabolic modeling methods to determine: 1) which commonly used heterologous host would be best suited for 6-dEB biosynthesis and 2) which genotypic alterations would result in improved 6-dEB production under certain environmental conditions. More generally, the study is an initial effort to better characterize polyketide biosynthesis within several heterologous host systems, understand the interplay between native and heterologous metabolism, and provide testable hypotheses for experimental metabolic engineering to improve heterologous biosynthesis.

Materials & Methods

Model Construction

Genome-scale stoichiometric models were downloaded from Professor Bernhard Ø. Palsson's website (http://gcrg.ucsd.edu/); *i*AF1260 (Feist et al. 2007), *i*YO844 (Oh et al. 2007), and *i*MM904 (Herrgard et al. 2008) were utilized as base models for *E. coli*, *B. subtilis*, and *S. cerevisiae*, respectively (Table 2). To account for the reactions needed to make 6-dEB, a DEBS-catalyzed biosynthetic

reaction and a 6-dEB transport reaction (assumed to be accounted for by diffusion) were added to the *E. coli*, *B. subtilis*, and *S. cerevisiae* models. The 6-dEB biosynthetic reaction was assumed to be irreversible and held the following stoichiometry: 1 propionyl-CoA + 6 (2*S*)-methylmalonyl-CoA + 6 NADPH → 1 6-dEB + 6 CO_2 + 7 CoA + 1 H_2O + 6 $NADP^+$. A 6-dEB transport reaction (from the cytosolic to the extracellular compartments) was added as a sink to balance the 6-dEB metabolite and is consistent with what is observed experimentally for *E. coli* (Pfeifer et al. 2001).

Table 2 Information on the genome-scale models of *E. coli*, *B. subtilis*, and *S. cerevisiae*.

Organism	Name	Genes	Metabolites	Reactions	Cellular Compartments	Year	Citation
E. coli	*i*AF1260	1,261	1,668	2,382	3	2007	(Feist et al. 2007)
B. subtilis	*i*YO844	844	992	1,250	2	2007	(Oh et al. 2007)
S. cerevisiae	*i*MM904	904	713	1,402	8	2008	(Herrgard et al. 2008)

In addition to the biosynthetic reaction, pathways were added to the *E. coli* and *S. cerevisiae* networks to facilitate acyl-CoA precursor supply. While the *B. subtilis* model has the ability to synthesize (2*S*)-methylmalonyl-CoA from propionyl-CoA (through the action of a propionyl-CoA carboxylase), *E. coli* does not have such ability. This particular pathway was chosen because it has previously been experimentally implemented into *E. coli* to support 6-dEB production (Pfeifer et al. 2001). A propionyl-CoA carboxylase reaction was therefore added to the *E. coli* model (and was assumed to be reversible(Reszko et al. 2003)) and held the following stoichiometry: ATP + HCO_3^- + propionyl-CoA ↔ ADP + H^+ + (2*S*)-methylmalonyl-CoA + P_i. Both propionate and non-propionate pathways were introduced to the *S. cerevisiae* model, which were all assumed to occur within the cytoplasm. The propionate-dependent pathway required the addition of a propionate transport reaction in addition to the propionyl-CoA synthetase and carboxylase reactions. For the non-propionate pathway, a methylmalonyl-CoA mutase was added to convert succinyl-CoA to (2*R*)-methylmalonyl-CoA. Next, a methylmalonyl-CoA epimerase was added to convert (2*R*)-methylmalonyl-CoA to (2*S*)-methylmalonyl-CoA and, lastly, a methylmalonyl-CoA decarboxylase was added to derive propionyl-CoA from (2*S*)-methylmalonyl-CoA (Haller et al. 2000). Because succinyl-CoA only existed within the

mitochondrion of the base model, a transport reaction was added to the network to allow for the exchange of succinyl-CoA from the mitochondrion to the cytosol. Lastly, the lower bound of the oxygen uptake rate (OUR) was set to –7.4 mmol $gDCW^{-1}$ hr^{-1}, as was observed experimentally (Van Hoek et al. 1998). This bound was set to 25 mmol $gDCW^{-1}$ $hr^{-1}$for *B. subtilis* (Sauer et al. 1996).

Calculations were made in MATLAB® 7.4 (Mathworks Inc.; Natick, MA) utilizing the SMBL Toolbox (version 2.0.2, http://sbml.org/software/sbmltoolbox/) (Keating et al. 2006; Schmidt and Jirstrand 2006) and the COBRA Toolbox (version 1.3.3, http://gcrg.ucsd.edu/) (Becker et al. 2007). Optimization was undertaken using the CPLEX (version 11.0) algorithm of the TOMLAB™ Optimization Environment (TOMLAB™/CPLEX) interfaced with the COBRA Toolbox and MATLAB® 7.4.

Flux Balance Analysis

Stoichiometric, steady-state balances on all metabolites are imposed as linear constraints on the basic equation (as described previously):

Equation 5

$$S \cdot v = 0$$

In Equation 5, S is an $m \times n$ matrix where m is the number of metabolites and n is the number of reactions in the model; while v is a column vector of length equal to the number of reactions ($n \times 1$). The

steady-state assumption yields that the dot product of S and v is equal to a zero row- vector of $1 \times m$. As the systems dealt with here are all underdetermined and their relationships are linear, the problem can be formed as a linear optimization problem subject to a metabolic objective:

Equation 6

$$\text{Maximize: } z = c^T v$$

$$\text{Subject to: } S \cdot v = 0$$

and

$$\alpha_i \leq v_i \leq \beta_i$$

In this optimization framework, c is a row vector containing weighting factors for individual fluxes on the objective function, z. a_i and b_i are the lower and upper bounds, respectively, of each flux as determined by either thermodynamics or experimental measurements. All flux units are in mmol $gDCW^{-1}$ (grams of dry cell weight) hr^{-1}, except for the biomass formation flux, which has units of hr^{-1}. For reactions without experimental data to provide flux information, reversible reactions have lower bounds of –1000 while irreversible reactions have lower bounds of 0; the upper bounds for both reversible and irreversible reactions are 1000. Heterologous production of 6-dEB has not been accomplished in *B. subtilis* or *S. cerevisiae*, and there is no information available on chemostat cultures of *E. coli* engineered to produce 6-dEB, In an effort to choose carbon-source (either glucose or

glycerol) uptake rates that would allow for a fair comparison between hosts, the growth-rates of all three organisms were plotted as a function of either glucose or glycerol uptake rate. A propionate uptake rate (PUR) of 0.02 was chosen based upon *E. coli* 6-dEB fed-batch bioreactor data previously published (Gonzalez-Lergier et al. 2006; Pfeifer et al. 2002). As a result, this value was used across all three organisms. ATP maintenance fluxes (v_{ATPM}) were unchanged from the base models for all three organisms.

In glucose-limited aerobic *E. coli* cultures in both batch and chemostat modes, the maximization of the biomass objective function provided high predictive fidelity of fluxes in central metabolism (as determined by comparison with ^{13}C-labeling studies) without the use of additional constraints (Schuetz et al. 2007). Although this was conducted on a small-scale model of *E. coli* metabolism, the authors state that the best predictive objective functions on small-scale models were also the best for two older (Edwards and Palsson 2000; Reed et al. 2003) genome-scale models of *E. coli* metabolism (Schuetz et al. 2007). Bi-level optimization frameworks have been developed to identify distributed metabolic objectives (to predict the weights of the objective function vector *c*) (Burgard and Maranas 2003; Gianchandani et al. 2008; Nolan et al. 2006); however, due to the lack of flux measurements, this method cannot be applied here. These two reasons

prompted use of the biomass formation equation as the objective function for optimization.

Calculations were made on an 32-bit Microsoft® Windows Vista Ultimate system with an Intel® Core™ 2 Duo T7300 processor running at 2.00 GHz with 4 GB RAM. Single FBA calculations using the TOMLAB™/CPLEX algorithm took approximately 70ms, 30ms, and 40ms for *E. coli*, *B. subtilis*, and *S. cerevisiae*, respectively. All linear programming problems that did not converge were removed from further analysis.

Altered Medium Formulations

Most FBA simulations to date have relied on a "computational minimal medium" in an attempt to mimic an experimental minimal medium typically containing a variety of salts and a single carbon source (for example, glucose). Previous experimental studies with *E. coli* showed that using a complex medium markedly improved both 6-dEB titers and cell-densities in shake-flask and bench-scale bioreactor cultures (Lau et al. 2004). Glycerol was chosen as an alternative carbon-source since, experimentally, many heterologous gene expression systems (including the one engineered for 6-dEB biosynthesis) rely on *lac* operators susceptible to catabolite repression by glucose. As a result, computational studies were performed in four types of media: minimal medium with glucose (the conventional

choice), minimal medium with glycerol, complex medium with glucose, and complex medium with glycerol. Medium composition can be approximated by setting uptake rates of specific chemical components known to exist in the medium of interest. The "computational complex medium" contained all twenty naturally-occurring amino acids (L-isomers) (Oh et al. 2007). The lower bounds of the amino acid uptake reactions were set to –0.1 mmol $gDCW^{-1}$ hr^{-1} and were chosen based upon previous literature values and because they satisfied the relative biomass differences experimentally observed between media (Oh et al. 2007; Selvarasu et al. 2009a).

Minimization of Metabolic Adjustment

It has been shown that the minimization of metabolic adjustment (MoMA) framework has better predictive power than FBA for calculating the flux distribution of gene-knockout mutants (Segre et al. 2002). The basic hypothesis underlying the MoMA framework is that in perturbed metabolic networks such as a network with a deleted reaction, the strain will perform in a suboptimal state because it has not evolved its perturbed network, therefore, nullifying the biomass maximization assumption. Though this method is similar to FBA, it replaces the objective function with one of minimizing the Euclidean distance between the wild-type strain flux vector and the flux vector of

the single-gene knockout mutant, and can be formulated as a standard quadratic programming problem:

Equation 7

Minimize: $z = (x - w)^T (x - w)$

Subject to: $S \cdot v = \mathbf{0}$

and

$\alpha_i \leq v_i \leq \beta_i$

Here, w is the wild-type flux vector, and x is the knockout strain's flux vector. The remaining variables are as described previously. For both single- and double-knockout MoMA calculations, first-dimension knockouts that led to no growth-rate were removed from the next dimension of knockouts to avoid trivial solutions (where the growth-rate would again be zero) and reduce computation time. All quadratic programming problems that did not converge were removed from further analysis. MoMA was chosen over other knockout identification frameworks such as regulatory on/off minimization (ROOM, which aims to minimize the *number* of flux changes) due to its ability to more accurately predict phenotypes shortly after mutation (Shlomi et al. 2005).

Multiple knockout flux distributions were also calculated using the MoMA framework. With the newest *E. coli* reconstruction, an exhaustive library of *only* double-knockout mutants would result in over one million mutants, which would be difficult to conduct in a

timely manner on a single-processor computer. Even as such, previous studies have demonstrated the value of computationally evaluating the impact of double-knockout mutants on cellular phenotypes (Alper and Stephanopoulos 2007; Alper and Stephanopoulos 2008). Due to the size of the models used in this study, a greedy search method was employed as opposed to an exhaustive search. In this respect, the global optimum hopefully found by iteratively searching local optima. At each iteration, a full knockout search was conducted, and the top performing strains were identified (as determined by the phenotype fraction, to be described later) and those genetic backgrounds were used to conduct an exhaustive search to identify possible synergistic effects of multiple gene-knockouts. This greedy search proved to be effective at finding the optimum at each stage in a timely manner (as supported by conducting an exhaustive double-knockout search and comparing results in one case).

Results

Model Construction

In addition to the heterologous metabolism needed for 6-dEB production, the models were analyzed for native pathways that could provide the needed substrates for biosynthesis. Our model showed that *E. coli* succinate metabolism could synthesize propionyl-CoA from succinate through the action of an adenosylcobalamin-dependent

pathway coded by the *ygf* operon (Haller et al. 2000). Alternatively, the *E. coli* network (Figure 11a) could produce propionyl-CoA through exogenously fed propionate, and (2*S*)-methylmalonyl-CoA through the heterologous propionyl-CoA carboxylase. The *B. subtilis* network (Figure 11b) is capable of natively producing both propionyl-CoA and (2*S*)-methylmalonyl-CoA from exogenously fed propionate.

Native *S. cerevisiae* is not known to have extensive propionate metabolism. When glucose limited, the cell cannot integrate monocarboxylic acids such as propionate into primary metabolism through conversion to propionyl-CoA (Mollapour et al. 2008). At neutral intracellular pH, these acids become charged and impermeable to internal membranes, invoking a "weak organic acid stress" response because they are left unmetabolized (Mollapour et al. 2008). However, computationally, it is possible to introduce precursor pathways to *S. cerevisiae* to enable 6-dEB production, and both propionate and non-propionate pathways were added to facilitate precursor supply.

In both *E. coli* and *B. subtilis*, an irreversible methylcitrate pathway can consume propionate and produce succinate for biomass production through the TCA cycle. Suspecting that this would be a considerable sink of propionate from 6-dEB production, the operon coding for the methylcitrate pathway (*prpRBCD*) was experimentally deleted from the *E. coli* chromosome in an effort to improve 6-dEB

production (Pfeifer et al. 2001). In calculations with the methylcitrate synthase gene present, all exogenously fed propionate was directed through this pathway and utilized for cell-growth; the resulting DEBS flux was zero. In the absence of the methylcitrate pathway, propionate is channeled towards 6-dEB production. To better mimic the *E. coli* experimental system currently utilized and to avoid computational results with zero DEBS flux, the methylcitrate synthase reaction was removed from both the *E. coli* and *B. subtilis* models. While it is not entirely certain that *B. subtilis* has a dedicated propionyl-CoA synthase, two gene products have been shown to catalyze this reaction (*ytcI* and *acsA*).

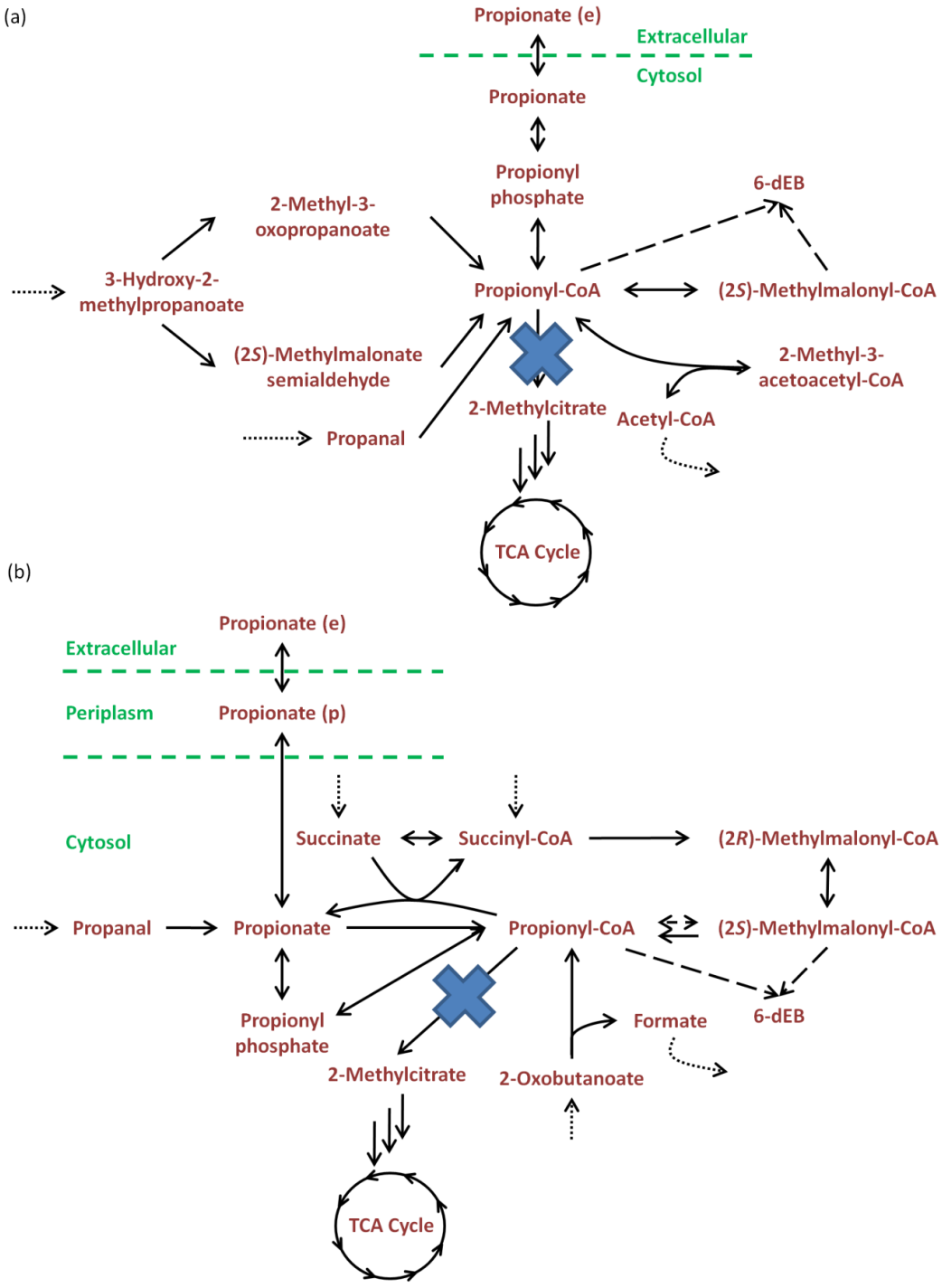

Figure 11 Metabolic networks of propionate metabolism and 6-dEB biosynthesis. Networks shown are for (a) *B. subtilis* and (b) *E. coli*. The relevant portion of the *S. cerevisiae* metabolic network is entirely heterologous and is described in the text. Metabolites are labeled in dark red and cellular compartments are labeled in green. Heterologous reactions are shown with long-dashed lines, while the deletion of the methylcitrate synthase is shown with a blue "X".

Flux Balance Analysis

All FBA simulations were conducted with the biomass equation as the objective function with the logic that the cell would still attempt to maximize biomass formation during 6-dEB biosynthesis. To identify carbon source uptake rates that would allow for direct comparison of phenotypes between the three organisms, the growth-rate was plotted as a function of substrate uptake rate for either glucose or glycerol. The simulations utilized to identify substrate uptake rates suitable for comparisons between hosts revealed that the biomass yield on moles of glucose is higher than the biomass yield on moles of glycerol for all three organisms (Figure 12). This is due to the higher molecular weight of glucose relative to glycerol, however, the mass yield of cell-mass on glucose (for *E. coli*) is 0.472 gDCW g glucose^{-1} while it is slightly higher for glycerol (at 0.480 gDCW g glycerol^{-1}). It is also apparent that *B. subtilis* has the ability to actively utilize both glucose and glycerol at higher uptake rates than both *E. coli* and *S. cerevisiae* due to its relatively higher maximum specific OUR. The simulations revealed that an uptake rate of 3 mmol gDCW^{-1} hr^{-1} for both glucose and glycerol would allow for comparisons to be drawn across host species due to the fact that all three organisms are carbon-source limited under these conditions (as can be seen by the constant slope of the growth-rate for these regions). As a result, for simulations on both

minimal and complex medium, carbon source uptake rates were set at either 3 mmol glucose $gDCW^{-1}$ hr^{-1}or 3 mmol glycerol $gDCW^{-1}$ hr^{-1}. All of these simulations were undertaken using the "wild-type" models (which did not include any of the engineered reactions or propionate as an additional carbon source).

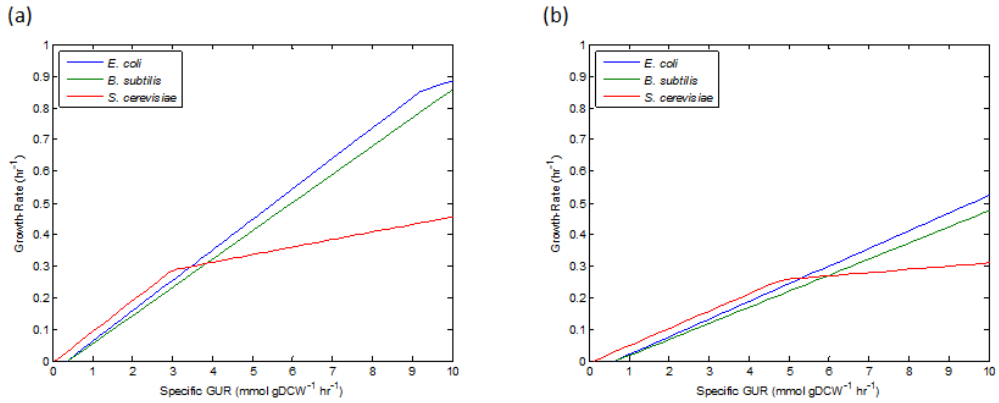

Figure 12 Growth phenotypes under glucose and glycerol simulations.
Plots of the growth rate as a function of carbon-source uptake rate in parent strains of *E. coli*, *B. subtilis*, and *S. cerevisiae*. All simulations were performed with either (a) glucose or (b) glycerol in minimal medium.

After utilizing the given conditions identified through the simulations in Figure 12, all of the additional reactions were included in the model (as described previously). Single FBA simulations were performed on all three hosts under the four different medium compositions (Table 3). As can be seen, the DEBS flux (and therefore the 6-dEB production rate) is the same for all three organisms under all medium compositions, a consequence of the identically imposed propionate uptake rates. Also, 6-dEB production rate observed represents the theoretical yield of 6-dEB on propionate. This shows that there are no catabolic pathways for propionate besides the methylcitrate series of reactions (which were removed as previously described). Although the production rates are higher than what is observed experimentally, going forward, knockouts that increase this flux would likely increase 6-dEB titers experimentally.

Table 3 Growth and 6-dEB production phenotypes under varying conditions. Host comparison for parent models growing on either a minimal or complex medium with either glucose or glycerol as the primary carbon source (at an uptake rate of 3.0 mmol $gDCW^{-1}$ hr^{-1}). The PUR was set at 0.02 mmol $gDCW^{-1}$ hr^{-1}. All flux units are in mmol $gDCW^{-1}$ hr^{-1}, while growth-rate has units of hr^{-1}.

Host	Carbon-Source	Medium	Growth-Rate	v_{DEBS}
E. coli	Glucose	Minimal	0.250	0.0029
B. subtilis	Glucose	Minimal	0.233	0.0029
S. cerevisiae	Glucose	Minimal	0.287	0.0029
E. coli	Glycerol	Minimal	0.132	0.0029
B. subtilis	Glycerol	Minimal	0.119	0.0029
S. cerevisiae	Glycerol	Minimal	0.158	0.0029
E. coli	Glucose	Complex	0.503	0.0029
B. subtilis	Glucose	Complex	0.438	0.0029
S. cerevisiae	Glucose	Complex	0.323	0.0029
E. coli	Glycerol	Complex	0.358	0.0029
B. subtilis	Glycerol	Complex	0.310	0.0029
S. cerevisiae	Glycerol	Complex	0.267	0.0029

Single-Gene Knockouts

The MoMA framework was applied to the *E. coli* and *B. subtilis* models to identify single-gene knockouts that increased 6-dEB production. The models used here have gene-protein-reaction (GPR) associations, such that the removal of a gene coding for a portion of an enzyme complex will render the entire reaction inactive (for example, the succinate dehydrogenase complex to be discussed later). At the same time, removal of a gene coding for an enzyme that has an isozyme will result in no difference in flux distribution, as a result, the isozyme will acquire the flux held by the removed reaction. Calculations were again conducted for minimal and complex medium formulations with either glucose or glycerol as the primary carbon-source. To aid in the analysis of the multi-dimensional data generated, the phenotype fraction, f_{ph} is defined:

Equation 8

$$f_{ph} \equiv (f_{biomass})(f_{DEBS}) = \left(\frac{v_{biomass,knockout}}{v_{biomass,parent}}\right)\left(\frac{v_{DEBS,knockout}}{v_{DEBS,parent}}\right)$$

The phenotype fraction equally weights both the biomass flux and the DEBS reaction flux. Knockouts that have the highest f_{ph} are the best performing strains in terms of both biomass and 6-dEB production and are the top candidates of knockouts to perform experimentally. This parameter filters out knockout strains that have low growth-rates, even if the strain's production of 6-dEB is higher. As

can be seen in Figure 13, many knockouts (of the entire genome) had no effect on the growth-rate or the DEBS flux (when f_{ph} = 1) or led to no growth-rate or DEBS flux (when f_{ph} = 0).

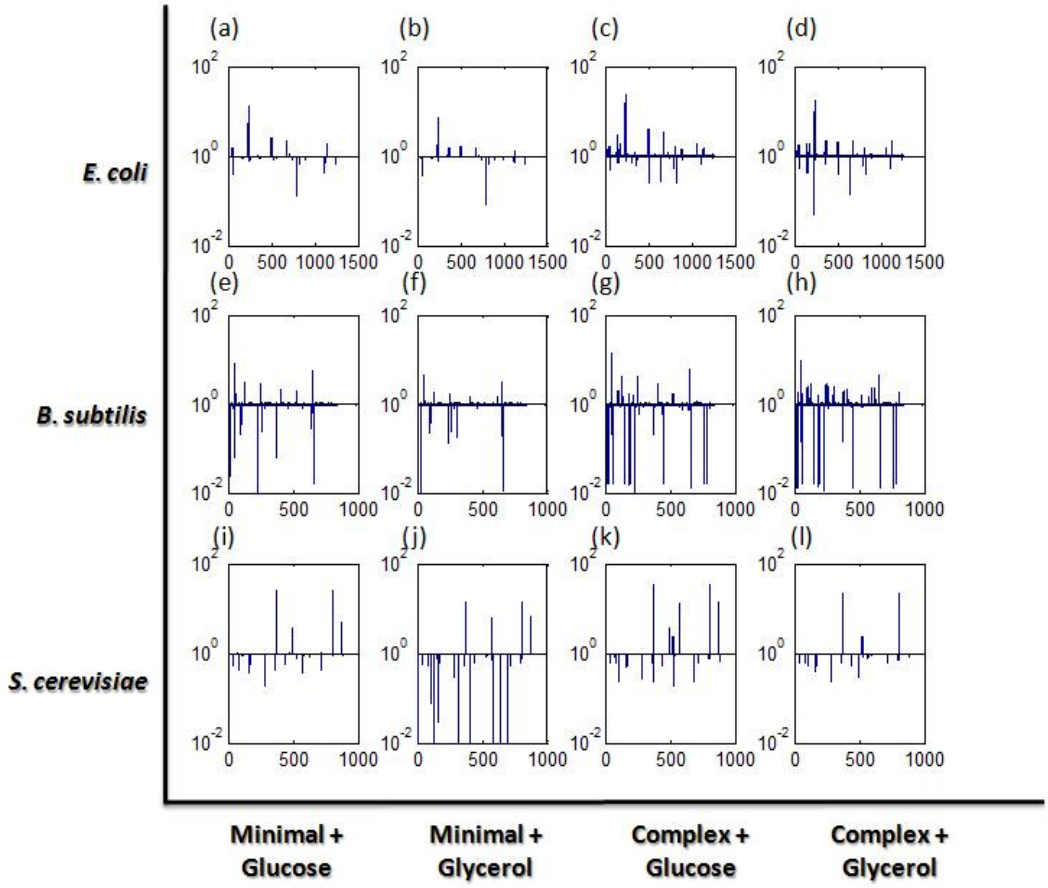

Figure 13 Single gene-knockout simulations.
Summary of single-gene knockouts using minimal or complex medium with glucose or glycerol as carbon sources for (a-d) *E. coli*, (e-h) *B. subtilis*, and (i-l) *S. cerevisiae*. Plotted on the *y*-axes is the phenotype fraction f_{ph} as a function of the *x*-axes gene number.

E. coli

The most marked improvements across all media compositions revolved around the succinate metabolite node. In all four cases, deletion of any of the genes coding for a portion of the succinate dehydrogenase complex (*sdhABCD*: catalyzing UQ + succinate → fumarate + UQH_2) and succinyl-CoA synthetase (*sucCD*: catalyzing ATP + CoA + succinate ↔ ADP + P_i + succinyl-CoA) led to the highest f_{ph}. The glutamate dehydrogenase gene (*gdhA*: catalyzing L-glutamate + H_2O + $NADP^+$ ↔ α-ketoglutarate + ammonia + NADPH) was also identified as a top candidate knockout across all four media. Glutamate dehydrogenase was previously identified as the best single-gene knockout for improving lycopene production in *E. coli*, both computationally and experimentally (Alper et al. 2005b; Alper et al. 2005c). The suggested explanation was that this mutant significantly increased the availability of NADPH for lycopene biosynthesis, which requires sixteen molecules of NADPH per molecule of lycopene (Alper 2006). A similar argument could be made for the improvement observed here since 6-dEB biosynthesis requires the reducing power of six molecules of NADPH.

Another notable identified target was a subunit of *E. coli*'s phosphotransferase system (PTS, *ptsH*), which phosphorylates imported sugars by producing pyruvate from phosphoenolpyruvate.

Under minimal medium with glucose, the cell overcomes the lack of a PTS system by uptaking glucose through two mechanisms: 1) an ABC transporter system, and 2) a proton symport system, both of which acquire no flux in the *ptsH*$^+$ strain. Glucose then becomes phosphorylated by hexokinase (utilizing ATP), rather than the PTS system. In addition to the periplasmic sugar transport reactions being inactivated by the deletion of *ptsH*, the dihydroxyacetone phosphotransferase (dihydroxyacetone + phosphoenolpyruvate → dihydroxyacetone phosphate + pyruvate) is also inactivated as a result of this deletion. Numerous reactions in glycolysis and the TCA cycle are down-regulated (in the case of fructose-6-phosphate aldolase, succinate dehydrogenase, and succinyl-CoA synthetase among others) or silenced (in the case of fumarate reductase). Pyruvate is generated through activation of oxaloacetate decarboxylase, driven by up-regulation of phosphoenolpyruvate carboxylase. It is therefore apparent that down-regulating succinate dehydrogenase or succinyl-CoA synthetase has a similar effect on improving 6-dEB production as the succinate dehydrogenase or succinyl-CoA synthetase mutants, albeit less pronounced.

Interestingly, this knockout improves the f_{ph} even when glycerol was the main carbon-source. Even though no sugars are being imported by the PTS system with glycerol as the carbon-source, the

inactivation of the dihydroxyacetone phosphotransferase has a more pronounced effect. The dihydroxyacetone phosphotransferase acquires a flux of 1.938 mmol $gDCW^{-1}$ hr^{-1} when glycerol is used as the carbon source, whereas it has a flux of 0.677 mmol $gDCW^{-1}$ hr^{-1} when glucose is the carbon source. Though, the overall effect is quite similar in this case further downstream in metabolism, where down-regulation of succinyl-CoA synthase (although not succinate dehydrogenase) allows for increased 6-dEB production. In the case for glycerol as the carbon source, glycerol kinase is up-regulated to phosphor late the imported carbon-source, up-regulating phosphofructokinase I and fructose bisphosphatase. Down-regulation of pyruvate dehydrogenase allows for up-regulation of both pyruvate kinase and phosphoenolpyruvate carboxylase. Taken together, these results suggest that there are multiple routes to improving 6-dEB production: reverting precursor supply, improving cofactor availability, and engineering a cellular regulatory system (such as the PTS system).

B. subtilis

As was the case for the succinate dehydrogenase and succinyl-CoA synthetase genes in *E. coli*, the top five-performing single gene-knockouts are consistent across all four medium compositions for *B. subtilis*. The α-ketoglutarate dehydrogenase (*citK*: catalyzing α-ketoglutarate + CoA + NAD^+ → CO_2 + NADH + succinyl-CoA) leads to

f_{ph} values between 4.57 and 13.61 (depending on the medium formulation). The second best single gene-knockout was the succinyl-CoA synthetase (*sucCD*: catalyzing ATP + CoA + succinate ↔ ADP + P_i + succinyl-CoA) exhibiting f_{ph} values between 3.22 and 5.97. Next, pyruvate dehydrogenase (*pdhABC*: catalyzing CoA + NAD^+ + pyruvate → acetyl-CoA + CO_2 + NADH) yields f_{ph} values between 2.40 and 5.00. The fourth best single gene-knockout across the medium formulations was aconitase (*citB*: catalyzing: citrate ↔ isocitrate) leading to f_{ph} values between 1.86 and 4.38. Lastly, the isocitrate dehydrogenase (*citC*: catalyzing isocitrate + $NADP^+$ ↔ α-ketoglutarate + CO_2 + NADPH) leads to f_{ph} values between 1.76 and 4.29.

As is evident from these results, engineering primary metabolism (the TCA cycle, specifically) has a significant influence on 6-dEB production. Reverting even a portion of carbon flow from a high flux pathway (such as the TCA cycle), to a low flux secondary metabolite pathway (such as the 6-dEB biosynthetic pathway) could have a significant positive effect on the secondary metabolite pathway. Interestingly, four of the five top performing single gene-knockouts occur in series in the TCA cycle (aconitase → isocitrate dehydrogenase → α-ketoglutarate dehydrogenase → succinyl-CoA synthetase), while the other, the pyruvate dehydrogenase complex, is responsible for

priming the TCA cycle with an acetyl-CoA molecule to react with an equivalent of oxaloacetate through the action of the citrate synthase.

S. cerevisiae

As compared to *E. coli* and *B. subtilis*, the exhaustive single gene-knockout analysis revealed far fewer knockouts that would improve 6-dEB production in *S. cerevisiae*. Not counting genes coding for individual subunits of an enzyme complex, only twelve knockouts showed f_{ph} values greater than 1.1 across all media conditions. This is likely due to the highly "engineered" nature of the substrate pathways needed to provide for the precursors for 6-dEB. However, the single gene-knockout of the succinyl-CoA ligase (LCS1, LCS2: catalyzing ATP + CoA + succinate ↔ ADP + P_i + succinyl-CoA) did produce the highest f_{ph} values (between 13.87 and 32.82) observed after the exhaustive single gene-knockout analysis for all three organisms. Fumarate hydratase (FUM1: catalyzing fumarase + H_2O ↔ malate) was the next best knockout for three of the four medium compositions; however, the same knockout was lethal for complex medium with glycerol as the main carbon source. Interestingly, it has been shown that a *S. cerevisiae* strain in which FUM1 replaced with a thienylalanine-resistance gene was able to aerobically grow on a glucose complex medium (2% peptone, 1% yeast extract, 2% glucose), while it was not

able to aerobically grow on glycerol complex medium (2% peptone, 1% yeast extract, 4% glycerol) (Arikawa et al. 1999).

Multiple-Gene Knockouts

Utilizing the same MoMA framework used in the genome-wide single-gene deletions, it was extended to conduct a search of the genotypic space to find double gene-knockout mutants which increased 6-dEB production. The QP MoMA algorithm and a greedy search, could identify multiple gene knockouts that increased f_{ph} further than the single dimension of knockouts. MoMA for a double-knockout first relies on solving the FBA of the parent strain, then the first-dimension of MoMA on a small set of knockouts (as determined from the top performing strains in the previous section), and finally a second-dimension of MoMA on the entire genome. Here, f_{ph} is defined relative to the parent strain and not the previous knockout. The results are shown in Figure 14. The first round of knockouts made a much larger impact on the f_{ph} than the second-round of knockouts; however, there were several second-round knockouts that improved f_{ph}.

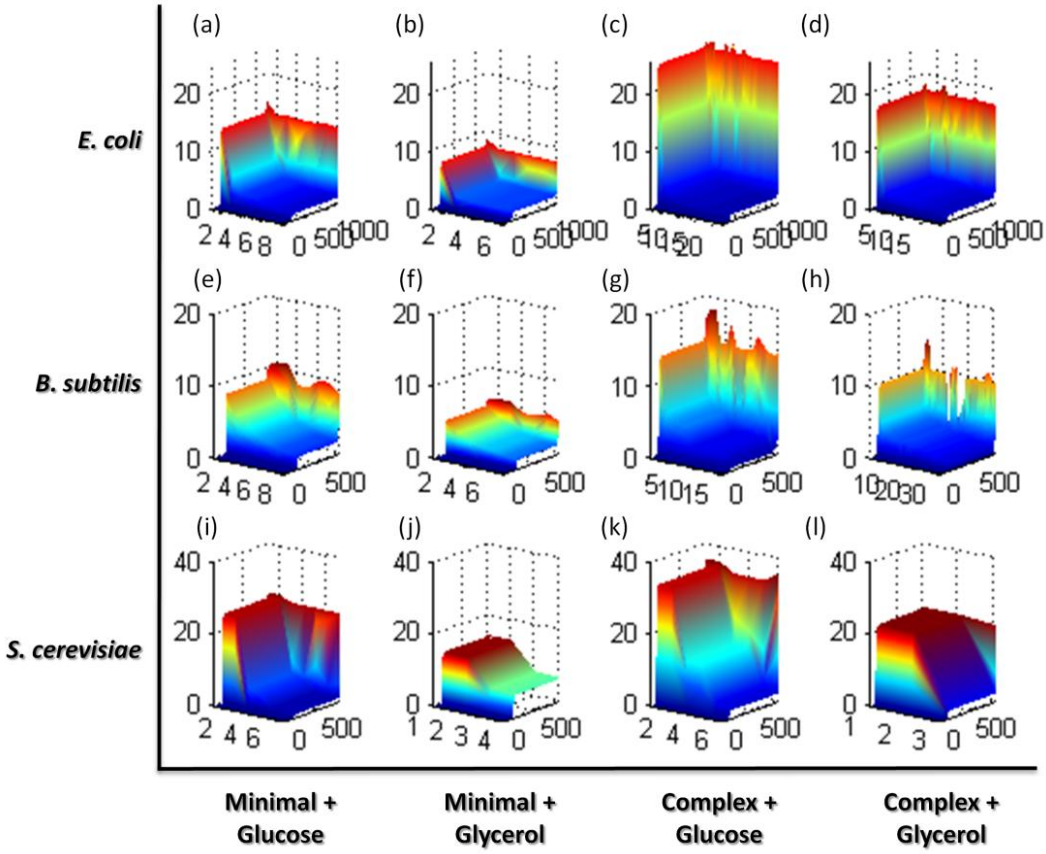

Figure 14 Double gene-knockout simulations.
f_{ph} is plotted as a function of two dimensions of knockouts. The simulated conditions were growing on minimal or complex medium with glucose or glycerol as carbon sources for (a-d) *E. coli*, (e-h) *B. subtilis*, and (i-l) *S. cerevisiae*. The first series of knockouts which had the highest f_{ph} are plotted on the left-hand axis, while the second series of knockouts are sorted from lowest to highest and plotted on the right-hand axis.

The addition, modulation, or deletion of more than two or three genes is often required to reach the theoretical yield of a particular intermediary or secondary metabolite (Alper and Stephanopoulos 2008). Unfortunately, an exhaustive or even greedy multi-dimensional knockout search beyond three genes becomes a very time consuming operation on a single-processor system.

Discussion

In this study, flux analysis was utilized in an attempt to better understand the interaction between cellular metabolism and heterologous polyketide biosynthesis in three common heterologous hosts, with the goal of identifying genotypic and bioprocess alterations which can be implemented experimentally to improve heterologous polyketide production. Although variable space for bioprocess conditions (for example, medium composition) can be explored fairly efficiently with the use of automated liquid-handling systems and microplate cultures or miniature bioreactors, it is more difficult to thoroughly explore genotypic space. Genetic techniques such as λ-Red mediated homologous recombination have improved the speed and efficiency of removing genes from the *E. coli* chromosome (Datsenko and Wanner 2000); however, genome-wide applications of this approach have not been studied beyond single gene-knockouts (Baba et al. 2006) and have not been applied for the improvement of complex

natural product production. In general, a comprehensive, experimental evaluation of single gene-knockouts generated to improve cellular phenotypes is not undertaken due to the cost, time, and labor associated with such an effort. This is certainly true in the case of improving 6-dEB production, and therefore, computational analyses offer a cost- and time-effective way to probe the metabolic network prior to undertaking experimental changes predicted to improve a phenotypic outcome.

To this end, three common heterologous hosts were examined, the Gram-negative bacterium *E. coli*, the Gram-positive bacterium *B. subtilis*, and the simple eukaryote *S. cerevisiae*, and their abilities to produce 6-dEB, a complex polyketide precursor to the antibiotic erythromycin. In addition to annotated genome sequences which allowed for model construction, these host systems were chosen because of their relatively simple and extensive molecular biology protocols, their established metabolic and bioprocess engineering strategies, and consequently, their future potential as hosts to support complex natural product biosynthesis. At the same time, 6-dEB production was chosen because of the intracellular requirements for propionyl- and (2*S*)-methylmalonyl-CoA and because it has been studied as a model PKS system for heterologous biosynthesis. The required substrates are common precursors to a number of polyketide

products but are typically not found in abundant quantities in *E. coli*, while they may be so in *B. subtilis*. Therefore, 6-dEB biosynthesis is a better test of a heterologous host's capabilities to more broadly support polyketide biosynthesis.

When applying FBA to compare the *E. coli*, *B. subtilis*, and *S. cerevisiae* hosts, results suggest that *S. cerevisiae* may be inferior for heterologous 6-dEB polyketide biosynthesis, given that it does not natively make the short-chain acyl-CoA monomers needed as substrates. Nonetheless, efforts have been directed towards the heterologous production of polyketides in *S. cerevisiae*. In an early effort, the 6-methylsalicylic acid synthase (6-MSAS) gene was cloned from *Penicillium patulum* and inserted into *S. cerevisiae* which enabled the production of the polyketide, 6-methylsalicylic acid (6-MSA) (Kealey et al. 1998).The titers observed with *S. cerevisiae* (approximately 1.7 g l^{-1}) were approximately an order of magnitude higher than those produced from a similar heterologous production effort using *E. coli*. The precursors required for 6-MSA are one molecule of acetyl-CoA and three of malonyl-CoA. Such impressive results for *S. cerevisiae* suggest the relative, if not unexpected, cellular abundance of these substrates for biosynthesis, pointing to potential success for similar compounds with the same precursor requirements. More recently, two separate pathways for (2*S*)-methylmalonyl-CoA

production (a propionate-dependent and independent pathways similar to those described in this modeling work) were introduced to *S. cerevisiae* for production of the polyketide triketide lactone (TKL) (Mutka et al. 2006a). Though impressive in demonstrating proof of principle production, uptake rates of propionate and methylmalonate were not measured and low TKL levels forced LC-MS/MS detection for semi-quantification of titers. It is acknowledged, however, that *S. cerevisiae* may indeed offer advantages for heterologous biosynthesis of polyketide compounds from other eukaryotic hosts. However, when focusing the study to a polyketide product dependent on propionyl- and (2*S*)-methylmalonyl-CoA substrates, our results support the use of *E. coli* and *B. subtilis* as better host choices. While these two organisms are similar from a metabolic standpoint, they are genetically quite different. For example, *B. subtilis* does not support the use of multi-copy extra-chromosomal plasmids, which are often used for heterologous gene-expression in *E. coli* and *S. cerevisiae*. However, being that *B. subtilis* does natively produce polyketides and non-ribosomal peptides, it has an active phosphopantetheinyl transferase (required for catalytically functional polyketide and non-ribosomal peptide synthase clusters).

Given that the initial FBA results supported *E. coli* and *B. subtilis* as better heterologous hosts for 6-dEB biosynthesis, these two

hosts were studied using: 1) simulations of growth on complex medium formulations and 2) an exhaustive single-gene knockout search and a greedy double-gene knockout search to identify gene deletions that would improve 6-dEB production. The use of complex medium (with either glucose or glycerol as a carbon source) increases the growth-rate and DEBS flux in all three hosts, a result that has been similarly observed experimentally in *E. coli*. Not surprisingly, this result occurs because the cell does not have to generate its own amino acids and, instead, gathers these nutrients from the medium, allowing more metabolic resources to be dedicated to cell growth and 6-dEB production.

The MoMA results identified single- and double-gene knockout strains with improved capacity for 6-dEB biosynthesis. Many knockouts had no effect on the growth-rate or the DEBS flux; however, several knockouts had a positive effect on DEBS flux while still sustaining cellular viability. Knockouts associated with the succinate and succinyl-CoA metabolite nodes serve as highly-connected metabolites located between the TCA cycle and the heterologous 6-dEB biosynthetic pathway (Table 4). Two separate studies have shown that *E. coli* exhibits robust behavior despite perturbations to genes in the TCA cycle (Ishii et al. 2007; Kim et al. 2007a). Therefore, engineering even a small portion of native, high-flux pathways such as the TCA

cycle to a low-flux heterologous pathway has the potential, as demonstrated here computationally, to increase heterologous pathway flux while only impacting the growth-rate moderately. In such a situation, a reduced growth-rate can often be overcome through the use of adaptive evolution strategies (Fong et al. 2003; Fong and Palsson 2004).

Table 4 Summary of the succinate dehydrogenase (*sdhABCD*) and succinyl-CoA synthetase (*sucCD*) reaction removal in *E. coli* and the effect on growth-rate and 6-dEB biosynthesis.

Carbon-Source	Medium	Gene(s) Removed	f_{ph}
Glucose	Minimal	*sdhABCD*	5.23
		sucCD	13.01
Glycerol	Minimal	*sdhABCD*	1.81
		sucCD	7.25
Glucose	Complex	*sdhABCD*	14.90
		sucCD	24.14
Glycerol	Complex	*sdhABCD*	9.77
		sucCD	16.94

Many of the top-performing double-knockout strains were combinations of the top-performing single-knockouts. For *E. coli* grown on minimal medium with glucose, the f_{ph} for the single-knockouts of *sucCD* and *sdhABCD* were 13.01 and 5.23, respectively, while the double-knockout of both of these operons led to an f_{ph} of 13.45. While these values appear to be quite high, if experimental results follow the simulations, culture titers will increase to over 1 g 6-dEB l^{-1}, levels only previously accomplished with optimized bench-scale bioreactors (Lau et al. 2004).

To our knowledge, the first to use these computational approaches to study polyketide biosynthesis was Hatzimanikatis's group, in both combinatorial biosynthetic (Gonzalez-Lergier et al. 2005) and metabolic (Gonzalez-Lergier et al. 2006) aspects. The prior research examined the theoretical yield of 6-dEB on both glucose and propionate in *E. coli*, and while pioneering in applying flux analysis to heterologous polyketide biosynthesis, the work was different in several regards to the study conducted in this report. First, the prior study assumed a metabolic objective of 6-dEB biosynthesis rather than growth-rate, justified since the goal was to determine the maximum theoretical yield of 6-dEB under different imposed constraints. In contrast, our approach maintains growth-rate as the objective function since, experimentally, cell growth is still observed during 6-dEB

biosynthesis implying that the cell does not revert all of its resources to producing 6-dEB and instead biomass formation may be a more accurate depiction of the experimental setting. The previous study also did not account for the methylcitrate pathway being deleted from the 6-dEB producing strain. As calculated here, when the methylcitrate pathway was present, all exogenously fed propionate was shuttled through this pathway to the TCA cycle and used for biomass production. Finally, the previous study separately examined glucose and propionate as precursors for 6-dEB biosynthesis; whereas, fed-batch experiments indicate that both substrates are taken up simultaneously and, therefore, our approach attempted to account for this dual substrate uptake.

Although the application of metabolic flux analysis for improving heterologous polyketide biosynthesis has only just begun, the approach has been successful for heterologous isoprenoid biosynthesis (Alper et al. 2005b; Alper et al. 2005c) and a variety of other non-therapeutic small-molecules (Lee et al. 2007; Park et al. 2007). The same FBA/MoMA optimization approaches utilized in this study were combined with experimental validation to improve lycopene production in *E. coli* (Alper et al. 2005b). Though the lycopene report utilized a smaller genome-scale model (*i*JE660a) (Edwards and Palsson 2000), it was found that most of the knockout targets predicted to

improve lycopene flux could have been identified using a much smaller network based generally around primary metabolism (Alper et al. 2005b). While it was found that this was the case for growth on minimal media with glucose, the same situation did not hold for growth on complex medium where a fair fraction of the knockouts predicted to improve production do not appear to significantly alter primary metabolism but instead affect the metabolism of the exogenously fed metabolites or the redox environment. For example, inactivation of *E. coli*'s branched-chain amino acid transferase (*ilvE*) leads to f_{ph} values of 1.74 and 1.50 for complex medium with glucose and glycerol, respective, whereas inactivation of these genes is lethal in minimal medium.

Although in both the complex and minimal media cases, reactions surrounding the highly-connected succinyl-CoA metabolite node proved to be important. These parallels to previous efforts and the prior success in other metabolic engineering scenarios support the continued use of computational analyses to improve the understanding of intracellular environments for heterologous polyketide biosynthesis towards the eventual improved production of therapeutic polyketide compounds.

Though recognized as a powerful tool for exploring the cellular metabolism of selected host organisms, there are also limitations to

using the previously described computational techniques for predicting intracellular flux distributions. Primarily, neither the stoichiometric model nor the FBA or MoMA frameworks account for genetic regulatory mechanisms such as feedback inhibition or quorum sensing, which can account for regulation at a variety of levels (translational, transcriptional, and metabolic). Secondly, due to the steady-state nature of the model and simulations, cellular adjustments to overcome gene-deletions (for example, intracellular accumulation of a particular metabolite or decreasing a specific carbon-source uptake rate) cannot be reflected in this type of modeling.

In summary, FBA and MoMA frameworks were used to simulate the interaction between native and heterologous metabolism as it relates to heterologous polyketide biosynthesis. This task was undertaken in three well-characterized hosts previously recognized as experimental options for heterologous polyketide biosynthesis with the goal of identifying genotypic alterations which would improve production titers. Results obtained now present new hypotheses to be tested experimentally. Such potential reveals yet another facet of heterologous polyketide biosynthesis: the range of metabolic engineering tools available with well-characterized hosts such as *E. coli*, *B. subtilis*, and *S. cerevisiae*. Now, the same experimental tools available for introducing and reconstituting heterologous polyketide

pathways can be used to implement genetic changes computationally predicted to improve polyketide production. This next level of metabolic engineering is an attempt to move beyond the heterologous introduction of a polyketide pathway and to start engineering the optimal activity of the new pathway within the new host. By combining approaches that include metabolic modeling, genetic engineering, and systems-level analyses (transcriptomics, proteomics, and metabolomics) (Askenazi et al. 2003; Bailey 2001; Barrett et al. 2006; Jewett et al. 2006; Kell 2004; Kim et al. 2008; Koffas and Stephanopoulos 2005; Nielsen and Jewett 2008; Nielsen and Oliver 2005; Rokem et al. 2007; Stafford et al. 2002; Stephanopoulos et al. 2004; Wang et al. 2006), the goal will be to match the cellular production of the new host to that of the native host. Yet, with the added availability of bioprocess and emerging metabolic engineering tools such as global transcriptional machinery engineering (gTME) (Alper et al. 2006b; Alper and Stephanopoulos 2007) and signal pathway engineering (Wang et al. 2007a), the eventual goal of heterologous natural product biosynthesis will be to surpass the cellular production titers of native host systems (Lee et al. 2005; Tyo et al. 2007).

Chapter 4 – Multi-scale engineering of 6-dEB production

Introduction

The PKS responsible for synthesizing the erythromycin macrocycle, the deoxyerythronolide B synthase (DEBS), has been the study of numerous seminal efforts in producing polyketides and understanding their biosynthesis including: cell-free synthesis (Pieper et al. 1995), heterologous biosynthesis (Kao et al. 1994; Pfeifer et al. 2001), the analysis of intermodular communication (Gokhale et al. 1999), combinatorial biosynthesis (Menzella et al. 2005), and structural analyses (Tang et al. 2006). The product of DEBS, 6-deoxyerythronolide B (6-dEB), has been produced in three different heterologous hosts, *Streptomyces coelicolor* (Kao et al. 1994), *Streptomyces lividans* (Xue et al. 1999), and *Escherichia coli* (Pfeifer et al. 2001), as well as a functional portion of the DEBS PKS in *Saccharomyces cerevisiae* (Mutka et al. 2006a). *E. coli* strain BAP1 has been developed previously for the heterologous production of polyketide and nonribosomal peptide natural products (Pfeifer et al. 2001) and is used as the base production system in this study. To generate BAP1, the *Bacillus subtilis* surfactin phosphopantetheinyl transferase gene (*sfp*) (Quadri et al. 1998) was inserted into the *prpRBCD* location of the BL21(DE3) chromosome (removing *E. coli*'s primary propionate

catabolic pathway (Haller et al. 2000; Textor et al. 1997)), under the control of an inducible T7 promoter (Pfeifer et al. 2001; Studier and Moffatt 1986). During this same genetic insertion, a T7 promoter was inserted before the native *prpE* gene (coding for a propionyl-CoA synthetase) to increase flux towards the production of propionyl-CoA, a direct precursor of 6-dEB.

There have been a number of studies focused on improving the stability of the large plasmids harboring the PKS genes (Murli et al. 2003), utilizing alternative substrate pathways for production (Dayem et al. 2002), and high-cell density bioprocess optimization (Lau et al. 2004) towards improving 6-dEB BAP1 production. Previously in our laboratory, metabolic modeling strategies were utilized for surveying heterologous hosts and medium compositions with respect to improving 6-dEB biosynthesis (Boghigian et al. 2010). Further, the *ygf* operon was analyzed by systematically deleting and over-expressing individual operon genes to understand their effect on 6-dEB biosynthesis (Zhang et al. 2010a). While most of the individual deletions and over-expressions led to either the same or decreased 6-dEB production titers under the conditions tested, deletion of *ygfH* (propionyl-CoA:succinate CoA transferase), led to an approximately two-fold increase in production titer. In an effort to further understand the effect of these pathways on polyketide formation, and examine the

interactions between these pathways, a multi-scale engineering strategy was applied to incorporate metabolic pathway engineering along with different bioprocess-related conditions (substrate feeding strategies, temperature). The results have implications for improving titers of both 6-dEB and other polyketides which utilize one or both of the acyl-CoA precursors examined here.

Materials & Methods

Background Strains & Plasmids

E. coli BAP1 was used as previously described (Pfeifer et al. 2001). TB3 is a derivative of BAP1 (Zhang et al. 2010a) constructed by P1 transduction with a Δ*ygfH*::*kan* (propionyl-CoA:succinate CoA transferase) mutant of BW25113 as a donor (Baba et al. 2006).

The genes required for the production of 6-dEB from propionate were cloned into plasmids pBP130 and pBP144, constructed previously (Pfeifer et al. 2001). Briefly, pBP130 (approximately 26kb) contains the *eryA2* and *eryA3* genes (coding for the DEBS2 and DEBS3 enzymes) under a single T7 promoter, on a pET21c background. Plasmid pBP144 (approximately 19kb) contains *eryA1* under a T7 promoter and genes coding for the two subunits of the *Streptomyces coelicolor* propionyl-CoA carboxylase enzyme (*accA1* and *pccB*) (Rodriguez and Gramajo 1999) under the control of second T7 promoter, on a pET28 background. All three *eryA* genes were cloned from the native

erythromycin producer, *Saccharopolyspora erythraea* (Cortes et al. 1990; Donadio et al. 1991). pYW7317 is a derivative of pBP144 without the *accA1* and *pccB* genes (Zhang et al. 2009).

Plasmid pACYCDuet-*matBC* was kindly provided by Prof. Mattheos A.G. Koffas and contains *matB* (coding for a malonyl-CoA synthetase) and *matC* (coding for a dicarboxylate carrier protein) from the nitrogen fixing soil bacterium *Rhizobium trifolii*, under the control of two separate T7 promoters (An and Kim 1998; An et al. 1999; Leonard et al. 2008).

Plasmid Construction

Standard molecular biology protocols were conducted according to Sambrook (Sambrook and Russell 2001). GeneHogs (Invitrogen) or XL-1 Blue (Stratagene) strains were used depending on the resistance marker of the plasmid to be constructed. The endonucleases used in this study were all purchased from New England Biolabs (Ipswich, MA, USA). All genes native to *E. coli* were PCR amplified from the BL21(DE3) (Novagen) genome. Oligonucueltoides were purchased from Eurofins MWG Operon (Table 5).

Table 5 Oligonucleotide primers utilized in this chapter.
All sequences are 5'→3' and restriction sites are denoted with an underline.

Name	Sequence (5'→3')
BamHI_prpE_for	GGGGGATCCATGTCTTTTAGCGAATTTTATCAGCGTTC
HindIII_prpE_rev	GGGAAGCTTACCTACGGTTCAGGTCC
NdeI_atoC_for	GGGCATATGACTGCTATTAATCGCATCC
XhoI_atoC_rev	GGGCTCGAGTTATACATCCGCCGGATCG
NdeI_sbm_for	GGGCATATGTCTAACGTGCAGGAGTG
XhoI_sbm_rev	GGGGTCGAGTTAATCATGATGCTGGCTTATCAG
pKD13_operon_for	AATACCCTCATTTTGATTGCGTTTTACGGA GCAAATAATGATTCCGGGGATCCGTCGACC
pKD13_operon_rev	ATTGCTGAAGATCGTGACGGGACGAGTCAT TAACCCAGCATGTAGGCTGGAGCTGCTTCG
k2	CGGTGCCCTGAATGAACTGC
ver_operon_rev	CGCCCAGCCAGTTGAGTTCA

The propionyl-CoA synthetase (*prpE*) was amplified and cloned into MCS1 of pACYCDuet-1 utilizing *Bam*HI and *Hind*III restriction sites, generating pACYCDuet-*prpE*. The transcriptional activator of the ATO system (*atoC*) was amplified and cloned into MCS2 of pACYCDuet-1 and pACYCDuet-*prpE* utilizing *Nde*I and *Xho*I restriction sites, generating pACYCDuet-*atoC* and pACYCDuet-*prpE-atoC*, respectively. An *E. coli* codon-optimized version of the *Streptomyces coelicolor* A3(2) methylmalonyl-CoA epimerase gene (*mce*) was synthesized by Operon (Huntsville, AL, USA) designed with flanking *Eco*RI and *Hind*III sites. This construct was blunt-cloned with *Sma*I into a modified pBluescript-II vector and inserted between the *Eco*RI and *Hind*III sites in MCS1 of pCDFDuet-1, generating pCDFDuet-*mce*. The methylmalonyl-CoA mutase gene (*sbm*, encoding a "sleeping beauty mutase") was PCR amplified from *E. coli* BL21(DE3) and cloned into MCS2 of pCDFDuet-*mce* utilizing *Nde*I and *Xho*I restriction sites, generating pCDFDuet-*mce-sbm*. All plasmids were verified by Sanger sequencing at the Tufts University Core Facility. All the plasmids used in this study are listed in Table 6.

Table 6 Plasmids constructed or used in this chapter.

Name	Description	Source
pACYCDuet-1	*cat;* P15A *ori lacI T7lac*	Novagen
pCDFDuet-1	*aadA;* CloDF13 *ori lacI T7lac*	Novagen
pBP130	*bla*; T7prom-*eryA2-eryA3*-T7term	(Pfeifer et al. 2001)
pBP144	*kan*; T7prom-*pccB-accA1*-T7prom-*eryA1*-T7term	(Pfeifer et al. 2001)
pYW7317	*kan*; T7prom-*eryA1*-T7term	(Zhang et al. 2009)
pACYCDuet-*matBC*	*cat*; T7prom-*matB*-T7term-T7prom-*matC*-T7term	(Leonard et al. 2008)
pACYCDuet-*prpE*	*cat*; T7prom-*prpE*-T7term	This chapter
pACYCDuet-*atoC*	*cat*; T7prom-*atoC*-T7term	This chapter
pACYCDuet-*prpE-atoC*	*cat*; T7prom-*prpE*-T7prom-*atoC*-T7term	This chapter
pCDFDuet-*mce*	*aadA*; T7prom-*mce*-T7term	This chapter
pCDFDuet-*mce-sbm*	*aadA*; T7prom-*mce*-T7prom-*sbm*-T7term	This chapter
pKD13	*bla*, *cat*; template for chloramphenicol cassette amplification	(Datsenko and Wanner 2000)
pKD46	*bla*; encodes γ, ß, and *exo* under the control of a pBAD promoter	(Datsenko and Wanner 2000)

Strain Construction

The entire *ygf* operon (*sbm-ygfDGH*) was deleted from the BAP1(*araBAD*:*tet*) chromosome by λ-Red recombination. First, BAP1(*araBAD*:*tet*) was transformed with plasmid pKD46 and expression of the γ, ß, and *exo* genes was induced with 10 mM L-arabinose at 30°C. A kanamycin resistance gene (*kan*) flanked by two flipase recognition target (FRT) sites was PCR amplified from pKD13 (using the pKD13_operon_for and pKD13_operon_rev primer pair) containing 50bp of homology arms upstream of *sbm* and downstream of *ygfH*. This PCR reaction was digested with *Dpn*I, gel purified, and approximately 100 ng of DNA was transformed into induced cells. Cells were then plated on LB-agar supplemented with 25 mg l^{-1} kanamycin. Successful recombinants were verified by PCR (using the k2 and ver_operon_rev primer pair). This strain was stored as a glycerol stock, prepared electrocompetent, and transformed with pCP20 (Cherepanov and Wackernagel 1995) to excise the *kan* gene between the FRT sites, generating a kanamycin sensitive strain, BAB2 (all strains are listed in Table 7).

Table 7 Strains constructed or used in this chapter.

Name	Description	Source
BL21(DE3)	F$^-$ *ompT hsdSB* (r_B^-, m_B^-) *gal dcm* (DE3)	Novagen
GeneHogs	F$^-$ *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *ara*D139 Δ(*ara-leu*)7697 *galU galK rpsL* (Str^R) *endA*1 *nupG fhuA*::IS2	Invitrogen
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*q*ZΔM15* Tn*10* (Tet^R)]	Stratagene
BAP1	BL21(DE3); Δ*prpRBCD*::T7prom-*sfp*-T7prom-*prpE*	(Pfeifer et al. 2001)
TB3	BAP1; Δ*ygfH*::FRT	(Zhang et al. 2010a)
BAB2	BAP1; Δ*sbm-ygfDGH*::FRT	This chapter

Initial Screening Cultures

All production cultures contained 5 g l^{-1} yeast extract, 10 g l^{-1} tryptone, 10 g l^{-1} sodium chloride, 15 g l^{-1} glycerol, 3 ml l^{-1} 50% (v v^{-1}) Antifoam B, 100 mM HEPES, and were adjusted to pH 7.60 with 5 M sodium hydroxide. For the initial screening study and the temperature modulation study, 3 ml cultures were conducted in 16 × 100 mm culture tubes.

For the initial screening study, the culture medium previously described was prepared supplemented with 60 mM sodium propionate, 60 mM disodium malonate, or 60 mM disodium methylmalonate. These were mixed with the production medium lacking the additional carbon source to create the various substrate concentrations desired. For precultures, a stab of glycerol stock was inoculated into 2 ml LB medium with appropriate antibiotics and grown overnight at 37°C and 250 rpm. For production cultures, 3 ml production medium was inoculated into 16 × 100 mm culture tubes with the precultures to an OD_{600nm} = 0.1. These production cultures were grown for 72 hr at 22°C and 250 rpm. At the end of the culture period, cell-density was measured spectrophotometrically at 600 nm and a single, 1 ml aliquot was stored at –20°C for subsequent analyses. When needed, antibiotics were supplemented at concentrations of 100 mg l^{-1} for carbenicillin, 50 mg l^{-1} for kanamycin, and 34 mg l^{-1} for chloramphenicol. Induction of

heterologous gene-expression was accomplished by supplementing 100 µM isopropyl ß-D-1-thiogalactopyranoside (IPTG) to the culture medium.

Shake-Flask Production Cultures

Shake-flask cultures (15 ml in 125 ml Erlenmeyer flasks) containing production medium were used for subsequent production tests. Single colonies were picked from freshly streaked plates and inoculated into 1 ml production medium containing necessary antibiotics. These cultures were grown at 37°C and 250 rpm until $OD_{600nm} \approx 0.6$ and were used to inoculate 15 ml production medium with necessary antibiotics and 100 µM IPTG at a volumetric ratio of 5%. Cultures were then incubated at 22°C and 250 rpm for 120 hr. At the end of the culture period, cell-density was measured spectrophotometrically at 600 nm and a single 1 ml aliquot was stored at –20°C for subsequent analyses. As before, antibiotics were supplemented at concentrations of 100 mg l^{-1} for carbenicillin, 50 mg l^{-1} for kanamycin, 34 mg l^{-1} for chloramphenicol, and 5 mg l^{-1} for tetracycline.

6-dEB Quantification by RP-HPLC-ELSD

The HPLC method for 6-dEB separation and quantification has been described previously (Wang et al. 2007a). Briefly, quantification of 6-dEB was carried out on an Agilent 1100 series HPLC coupled with

an Alltech 800 series evaporative light-scattering detector (ELSD). The guard column used was an Inertsil ODS3 C_{18} 5 µm, 4.6 mm × 10 mm while the analytical column used was an Inertsil ODS3 C_{18} 5 µm, 4.6 mm × 150 mm (GL Sciences). Ultra-high purity grade nitrogen gas (AirGas East) was used as the mobile phase for the ELSD at a pressure of 3.00 ± 0.05 bar, while the ELSD drift tube temperature was maintained at 55°C and the gain setting was set at 16. Culture samples were first centrifuged for 10 min at 10,000 × g to remove insolubles. A 20 µl supernatant injection volume was then applied to the column. The mobile flow rate was 1 ml min^{-1} and 6-dEB was eluted at 7.92 ± 0.05 min. Quantification was carried out against a five-point calibration curve of purified 6-dEB (kindly provided by Kosan Biosciences).

6-dEB Quantification by Mass Spectrometry

When 6-dEB was not detectable by RP-HPLC-ELSD (with a limit of detection of approximately 5 mg l^{-1}), the production titer was quantified by mass spectrometry. Erythromycin was used as an internal standard during the MS analysis to account for internal measurement drift of the instrument. Clarified culture medium (750 µl) was extracted with an equal volume ethyl acetate and dried. The extract was dissolved in 50 µl of HPLC-grade methanol containing 5 mg l^{-1} erythromycin for analysis. To prepare a suitable calibration

curve, the culturing procedure was repeated using *E. coli* BAP1 without any plasmids and 6-dEB standards added at different concentrations at the end of the culture period prior to ethyl acetate extraction. The standard samples were then used to prepare the calibration curve to correlate 6-dEB MS peak intensity with 6-dEB concentration. The calibration curve was generated for every sample set and experiments were repeated three separate times using a Thermo Electron Corporation LTQ XL Linear Ion Trap Mass Spectrometer (Waltham, MA, USA).

Metabolite Quantification

Medium and byproduct organic acids were quantified by the previously mentioned HPLC system coupled to a Refractive Index Detector (RID). Clarified culture supernatant (20 µl) was applied to a Bio-Rad Aminex® HPX-87H Ion Exchange (300 mm × 7.8 mm, 9 µm) column, preceded by a 30 mm guard column of the same resin. The isocratic analysis used a 9.5 mM H_2SO_4 solvent held at a flow rate of 0.3 ml min^{-1}. These conditions were identified by using an iterative stochastic search HPLC optimization program based on the compounds anticipated to be present in the culture medium (Dharmadi and Gonzalez 2005). A five-point standard calibration curve was created and used for quantification of propionate, malonate, methylmalonate, glycerol, pyruvate, acetate, ethanol, succinate, formate, and lactate.

The elution order was as follows: pyruvate (16.7 min), malonate (18.9 min), methylmalonate (20.9 min), succinate (22.7 min), lactate (24.2 min), glycerol (25.1 min), formate (26.8 min), acetate (29.1 min), propionate (34.5 min), and ethanol (41.3 min). All specific production or consumption rates presented are averaged over the course of the culture period.

Results

Initial Screening Study

Upon inspection of generalized metabolic maps, and given the DEBS (2*S*)-methylmalonyl-CoA requirement, constructed of a metabolic pathway capable of converting exogenous methylmalonate to (2*S*)-methylmalonyl-CoA (Figure 15) was undertaken. This pathway would include *E. coli*'s native YgfG or the heterologous reversible PCC to provide propionyl-CoA. The MatB-MatC pathway responsible for methylmalonate-methylmalonyl-CoA conversion from *R. trifolii* was then reconstructed in *E. coli*. While these MatB and MatC synthases have a preference for malonate as a substrate, it has been shown that they can also activate methylmalonate at 20.4% the *in vitro* rate of malonate (An and Kim 1998).

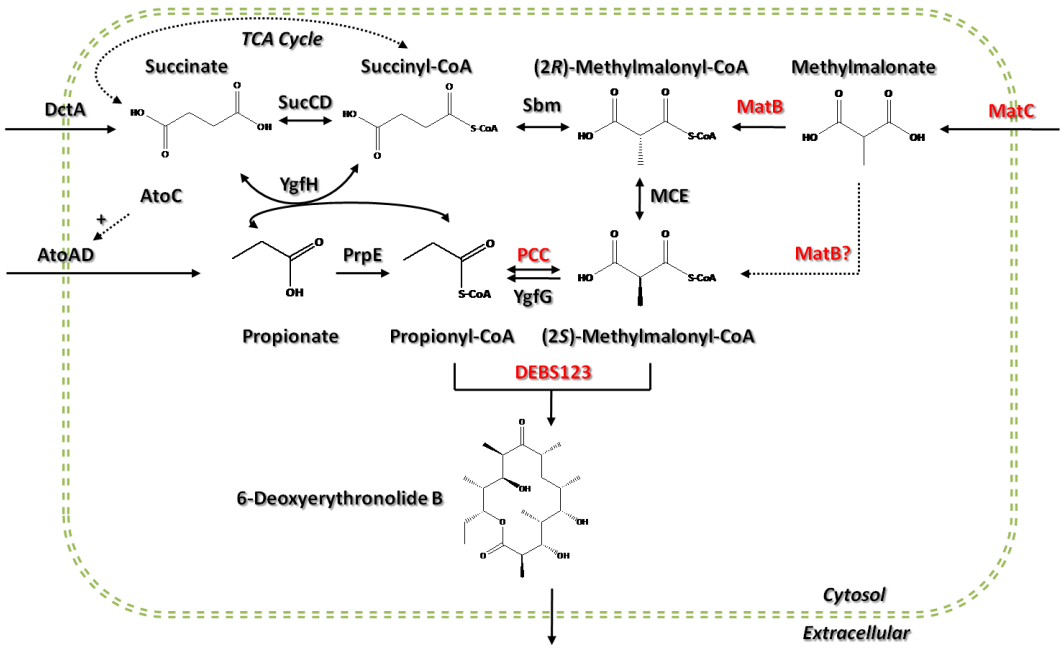

Figure 15 Propionate, methylmalonate, and succinate metabolism and their relation to 6-dEB production in *E. coli*.
Heterologous enzymes are shown in red text. Abbreviations: SucCD = succinyl-CoA synthetase; Sbm = sleeping beauty mutase = methylmalonyl-CoA mutase; MatB = malonyl-CoA synthetase; MatC = *R. trifolii* dicarboxylate carrier protein; MCE = methylmalonyl-CoA epimerase; PCC = propionyl-CoA carboxylase; YgfG = methylmalonyl-CoA decarboxylase; PrpE = propionyl-CoA synthetase; YgfH = propionyl-CoA:succinate CoA transferase; DEBS123 = deoxyerythronolide B synthase; AtoC = transcriptional activator of the ATO system; AtoAD = acetyl-CoA:acetoacetyl-CoA transferase; DctA *E. coli* dicarboxylate carrier protein.

A three-variable (propionate, malonate, and methylmalonate), two-level (0 mM or 20 mM) full-factorial supplementation experiment across nine strains was designed over a variety of plasmid combinations (containing the propionate pathway, the malonate pathway, and the 6-dEB biosynthetic pathway). No 6-dEB (<5 mg l^{-1}) was made in the absence of propionate supplementation (Figure 16). The addition of malonate did not improve 6-dEB titers under any conditions tested. When propionate and methylmalonate were both supplied, 6-dEB production increased approximately two-fold compared to propionate supplementation alone. However, the overall 6-dEB titers were lower in the presence of the MatBC pathway, presumably due to the increased metabolic burden of maintaining the third plasmid and any resulting undesired effects caused by gene-expression and enzymatic activity.

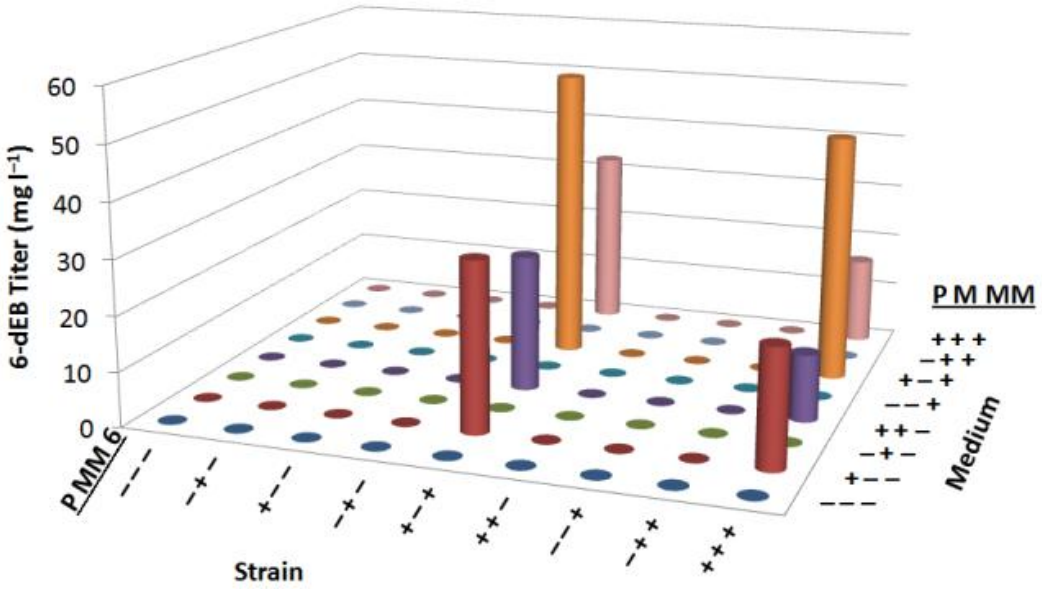

Figure 16 Initial screening study, 6-dEB production.
Data from the three variable (propionate, malonate, and methylmalonate) two-level (0 mM or 20 mM) full-factorial supplementation experiment across nine strains with a variety of plasmid combinations. For all figures, on the left-hand axis –'s and +'s indicate whether the propionate pathway, the methylmalonate pathway, or the DEBS complex is included, respectively. On the right hand axis, –'s and +'s indicate whether propionate, malonate, or methylmalonate is supplemented in the medium at a concentration of 0 or 20mM, respectively. The *y*-axis shows 6-dEB titer as a function of these two parameters.

In terms of precursor consumption, propionate was favored in the strains that did not contain the MatBC pathway; whereas when MatBC was present, malonate consumption was preferred and increased dramatically even in the presence of multiple substrates (Figure 17a, Figure 17b, and Figure 17c). This is most likely due to the MatBC possessing a natural preference for malonate. Propionate uptake did not change dramatically in any of the conditions tested, even in the cases where 6-dEB production was observed (Figure 17a). As can be seen in Figure 17c, in general, methylmalonate consumption was minimal (often <2 mM consumed) in all cases, although the MatBC pathway did stimulate consumption slightly (to approximately 4 mM in some cases). Acetate overflow was significant, reaching approximately 80 mM in some cases; however, this overflow metabolism was retarded with the addition of propionate, malonate, and methylmalonate (Figure 17e).

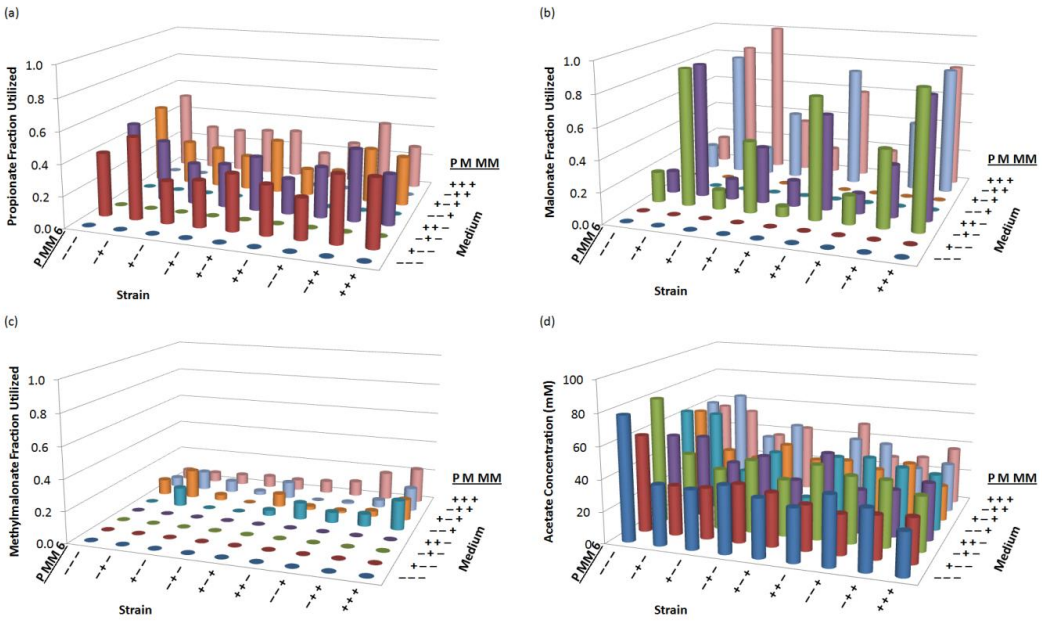

Figure 17 Initial screening study, metabolite profiling.
Data from the three variable (propionate, malonate, and methylmalonate) two-level (0 mM or 20 mM) full-factorial supplementation experiment across nine strains with a variety of plasmid combinations. For all figures, on the left-hand axis –'s and +'s indicate whether the propionate pathway, the methylmalonate pathway, or the DEBS complex is included, respectively. On the right hand axis, –'s and +'s indicate whether propionate, malonate, or methylmalonate is supplemented in the medium at a concentration of 0 or 20mM, respectively. The *y*-axis shows (a) fraction of propionate utilized, (b) fraction of malonate utilized, (c) fraction of methylmalonate utilized, and (d) amount of acetate produced as a function of these two parameters.

Temperature Influence on 6-dEB Production

Temperature has a significant influence on *E. coli* 6-dEB biosynthesis, particularly with respect to the active production of the DEBS enzymatic complex (Pfeifer et al. 2001; Wang and Pfeifer 2008). As a result, 6-dEB production was analyzed at 5°C intervals between 12°C and 37°C. Figure 18a presents cell-density after 72 hr of culture as a function of temperature. The lowest cell-density achieved was at 12°C, and due to the very low specific growth rate at that temperature, stationary phase was never reached (data not shown). Cell-density is also low after 72 hr at 37°C. The cell-densities are not significantly different between 22°C, 27°C, or 32°C (ANOVA $p = 0.611$), while it is slightly lower at 17°C ($p < 0.001$ when compared to 22°C).

Under the temperatures tested, 6-dEB was only produced at 22°C, 27°C, and 32°C. While the titers were not different between 22°C and 27°C ($p = 0.333$), they were lower at 32°C ($p < 0.001$ when compared to 22°C). The 6-dEB production titers had no correlation to the specific propionate uptake rates (Figure 18b) which show a nearly linear decrease with respect to increasing temperature.

Byproduct production rates are shown as a function of temperature in Figure 18c. Acetate was produced at all temperatures; however, the production rate was highest at 37°C. Lactate has not been previously observed at detectable levels (>1 mM) in the culture

medium for 6-dEB production through *E. coli* (Wu et al. 2010); however, at the higher temperatures tested here (27°C, 32°C, and 37°C) this byproduct was generated. The specific production rate of lactate was also higher at 37°C. Formate, succinate, pyruvate, and ethanol were not observed at significant quantities (<1 mM) in the medium at the end of the culture period (data not shown).

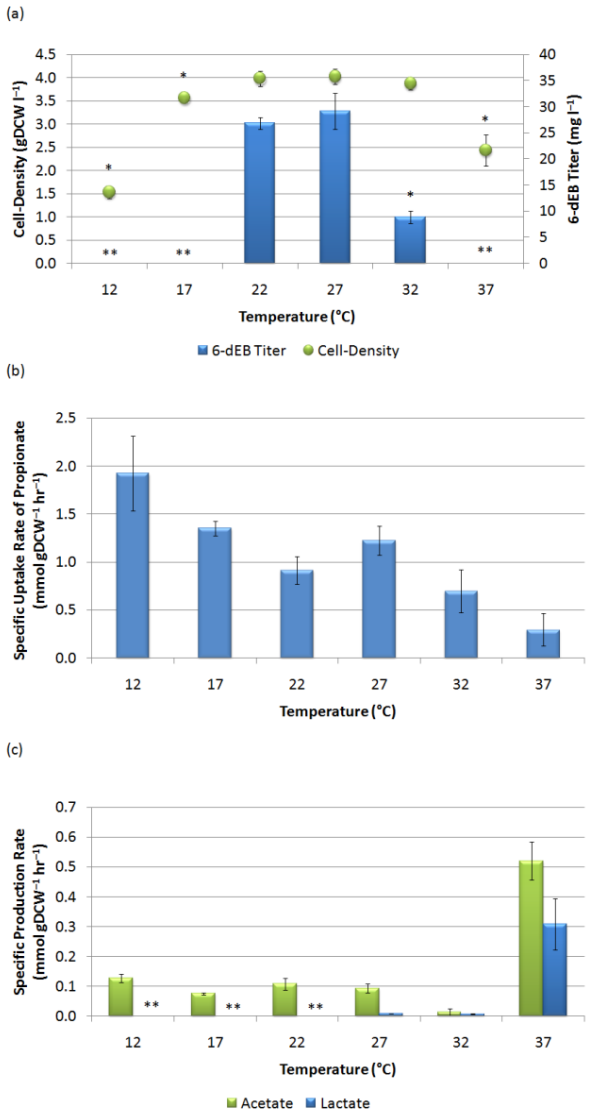

Figure 18 Temperature modulation study.
The *y*-axes show (a) cell-density (left *y*-axis) and 6-dEB titer (right *y*-axis), (b) specific uptake rate of propionate, and (c) specific production rates of acetate and lactate as a function of temperature between 17°C and 37°C at intervals of 5°C. Error bars represent ± one standard deviation of four replicates. For panel (a), * indicates statistically significant results ($p < 0.05$, as determined by paired Student's *t*-test) when compared to the 22°C data set. ** indicates that 6-dEB was not detectable (<5 mg l^{-1}). For panel (c), ** indicates that the metabolite was not detectable (<1 mM).

Propionate Pathway Engineering

The propionate pathway has previously been designed to provide the precursors for 6-dEB production in *E. coli* (Wang et al. 2007a; Zhang et al. 2010a). In the current production system, *atoC*, encoding an activator of *E. coli* short-chain fatty acid metabolism, is natively expressed from the chromosome. Whereas, *prpE,* encoding a propionyl-CoA synthetase, is inducibly expressed from the chromosome and the *pcc* genes, encoding a propionyl-CoA carboxylase, are expressed from a multi-copy plasmid (15-20 copies per cell). As a result of the variation in expression design, this pathway is likely 'unbalanced' for 6-dEB heterologous production as *prpE* and *atoC* are expected to be expressed to a smaller extent than the *pcc* genes. Balancing of expression levels may then improve production. Three different vectors were generated to over-express *E. coli*'s native *prpE* and *atoC* separately and in combination. These constructs were then co-transformed with the 6-dEB production plasmids (pBP130 and pBP144) in two strains (BAP1 and TB3).

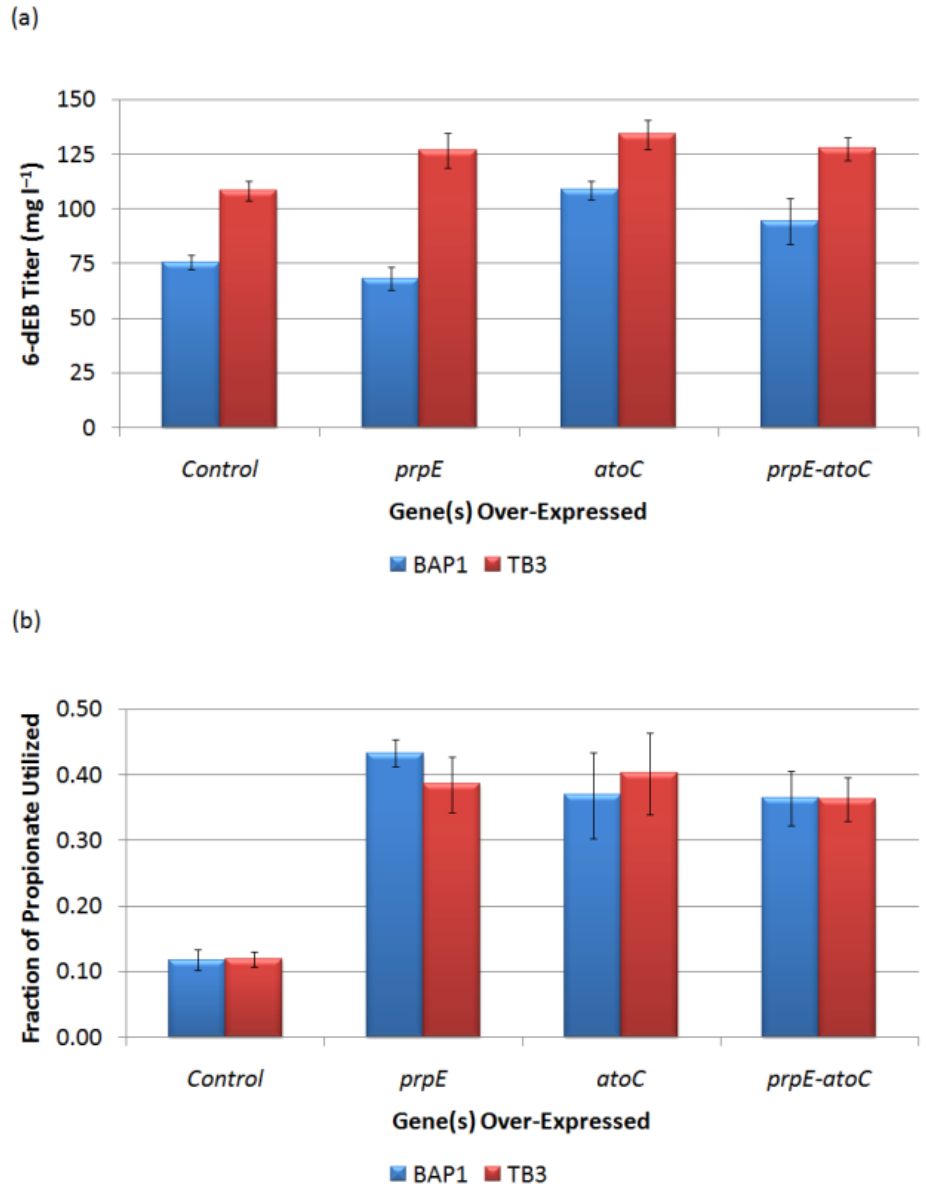

Figure 19 Propionate pathway engineering.
Data from the propionate pathway engineering study, as a function of two strains (BAP1 and TB3) and four plasmid systems (a control with only pBP130 and pBP144, an additional pACYCDuet-prpE, an additional pACYCDuet-atoC, and an additional pACYCDuet-prpE-atoC). Panel (a) shows the 6-dEB titer while panel (b) shows the fraction of propionate uptake as a function of these cellular parameters. Error bars represent ± one standard deviation of three replicates.

In BAP1, over-expression of *prpE* by itself had no effect on 6-dEB production (p = 0.214). Over-expression of *atoC* improved 6-dEB production 1.44-fold (Figure 19a). Similar titers were observed when both *prpE* and *atoC* were over-expressed. Compared to BAP1, TB3 improved titer approximately 1.5-fold (a slightly lower improvement than observed previously (Zhang et al. 2010a), which is likely due to a different inoculation method). In TB3, all three plasmid constructs improved 6-dEB production similarly (ANOVA p = 0.416). In this case, the over-expression of *prpE* alone did not help improve the 6-dEB production in strain TB3 when compared to the co-expression of *prpE* and *atoC*. This is probably because the deletion of *ygfH* in TB3 resulted in sufficient propionyl-CoA and further accumulation through PrpE activity did not aid 6-dEB biosynthesis. The 6-dEB titers presented here increased from 75.6 ± 3.2 mg l^{-1} to 134.1 ± 6.6 mg l^{-1} after varying these plasmid and strain systems.

Figure 19b shows the fraction of propionate consumed by the strains after 120 hr of shake-flask culture across all of the conditions tested. The *atoC* gene encodes for the regulatory controller of the ATO operon (responsible for short-chain fatty acid degradation (Jenkins and Nunn 1987)), and *prpE* expression is needed for the production of propionyl-CoA, a direct precursor for 6-dEB biosynthesis. It was shown

that over-expression of *prpE* and *atoC* both stimulate propionate uptake to the same effect. For both BAP1 and TB3, regardless of whether *prpE* and *atoC* were over-expressed separately or together, the fraction of propionate utilized increased between 3-4 fold (ANOVA $p = 0.653$). In all plasmid systems, propionate uptake between BAP1 and TB3 was the same ($p > 0.05$), meaning that the yield of 6-dEB on propionate was higher in the TB3 strain as TB3 produced higher 6-dEB titers.

Methylmalonate Pathway Engineering in the Presence of Propionate

The initial screening study revealed that methylmalonate supplementation could improve 6-dEB production in the presence of propionate, prompting us to further investigate this interaction. It was first hypothesized that the lack of improvement upon incorporation of the MatBC pathway (even with the improved methylmalonate uptake) was because the malonyl-CoA synthase generated the (2*R*) stereoisomer of methylmalonyl-CoA; whereas, the DEBS complex only accepts the (2*S*) isomer of methylmalonyl-CoA. Because a methylmalonyl-CoA epimerase has not been previously identified in *E. coli*, a heterologous methylmalonyl-CoA epimerase gene was used to allow for the interconversion between isomers and to determine its effect on 6-dEB production.

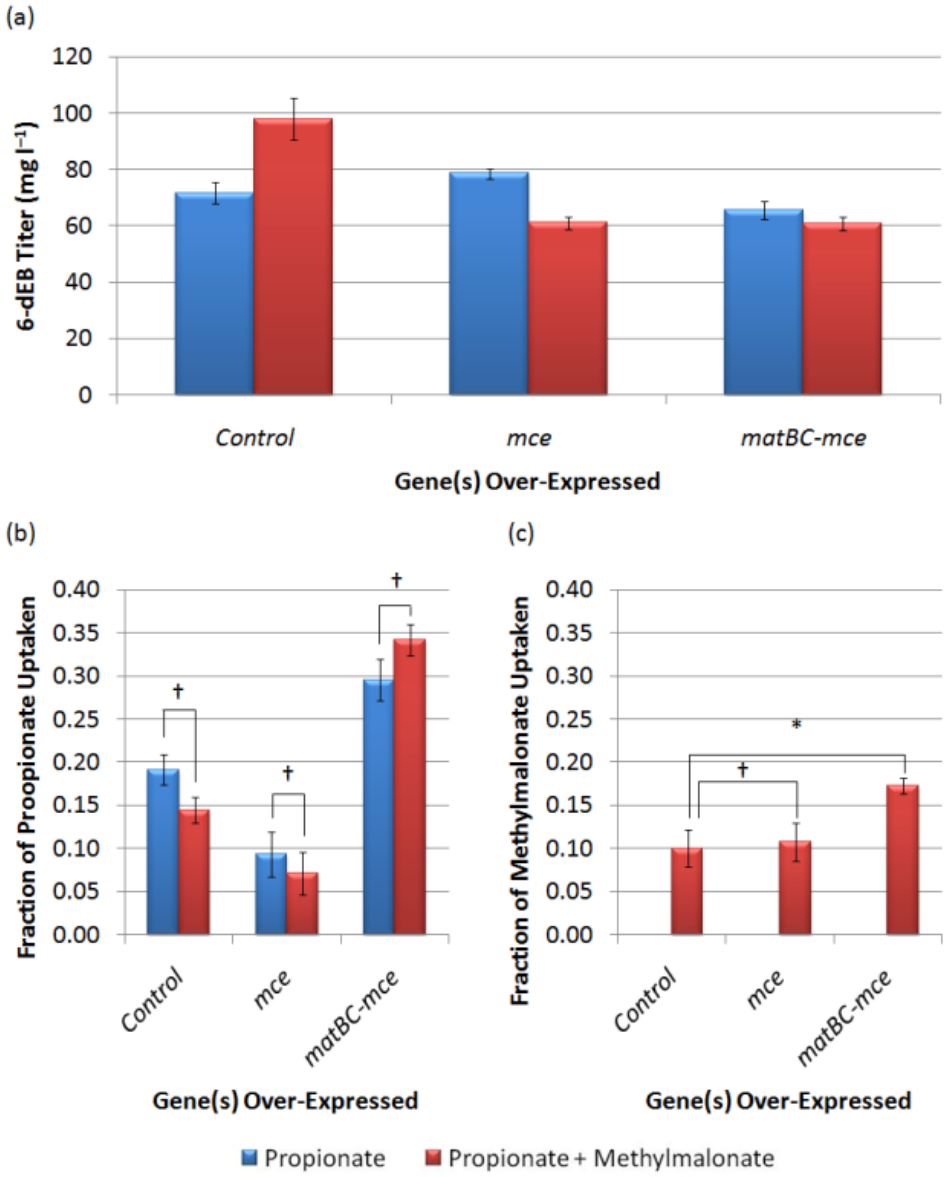

Figure 20 Methylmalonate pathway engineering in the presence of propionate for BAP1.
Cultures were fed either 20 mM propionate or 20 mM propionate and 20 mM methylmalonate. Three different plasmid systems were analyzed (a control with only pBP130 and pBP144, an additional pCDFDuet-*mce*, and pACYCDuet-*matBC*/pCDFDuet-*mce*). Panel (a) shows the 6-dEB titer, (b) shows the fraction of propionate uptake, and (c) shows the fraction of methylmalonate uptake as a function of these parameters. Error bars represent ± one standard deviation of three replicates. * indicates statistically significant results ($p < 0.05$, as determined by paired Student's *t*-test), while † indicates statistically insignificant results ($p > 0.05$) between the comparisons shown.

First, the effects of propionate and methylmalonate supplementation were tested in BAP1 in the presence of: 1) no extra plasmids, 2) a plasmid over-expressing the *S. coelicolor* A3(2) methylmalonyl-CoA epimerase gene (*mce*), and 3) a plasmid over-expressing *mce* and the *matBC* genes. Consistent with the small-scale cultures in the initial screening study, methylmalonate supplementation improves 6-dEB production in the presence of propionate in the shake-flasks from 71.7 ± 3.74 mg l^{-1} to 97.9 ± 7.32 mg l^{-1} (Figure 20a). However, 6-dEB production decreases to 61.1 ± 2.18 mg l^{-1} and 60.9 ± 2.43 mg l^{-1} when both carbon sources are used in the presence of *mce* and *matBC-mce*, respectively. As shown in Figure 20b, propionate uptake is decreased when *mce* is used. When *matBC* is included, propionate uptake is stimulated, indicating that this MatC carrier is not specific and may facilitate propionate transport as well. Compared to the control, methylmalonate uptake was not different when *mce* was included. However, methylmalonate uptake did increase after the inclusion of *matBC-mce* (Figure 20c).

The same expression and production studies were then conducted in TB3, yielding different results. As established previously, TB3 showed higher titers of 6-dEB from propionate (101.0 ± 7.5 mg l^{-1}), however, when methylmalonate was supplemented, the titer was unchanged (p = 0.730; Figure 21a). When over-expressing *mce* or both

matBC-mce, the titers decrease significantly to 58.6 ± 1.19 mg l^{-1} and 33.1 ± 7.4 mg l^{-1} with only propionate, and to 75.2 ± 3.9 mg l^{-1} and 52.1 ± 2.4 mg l^{-1} with both substrates present. This departs from the trends observed in BAP1, where when either *mce* or *matBC-mce* was expressed, 6-dEB production decreased in the presence of methylmalonate. As can be seen in Figure 21b and Figure 21c, propionate and methylmalonate uptake was not drastically different than what was observed in BAP1. Again, expression of *matBC-mce* stimulated methylmalonate uptake, while expression of *mce* alone did not.

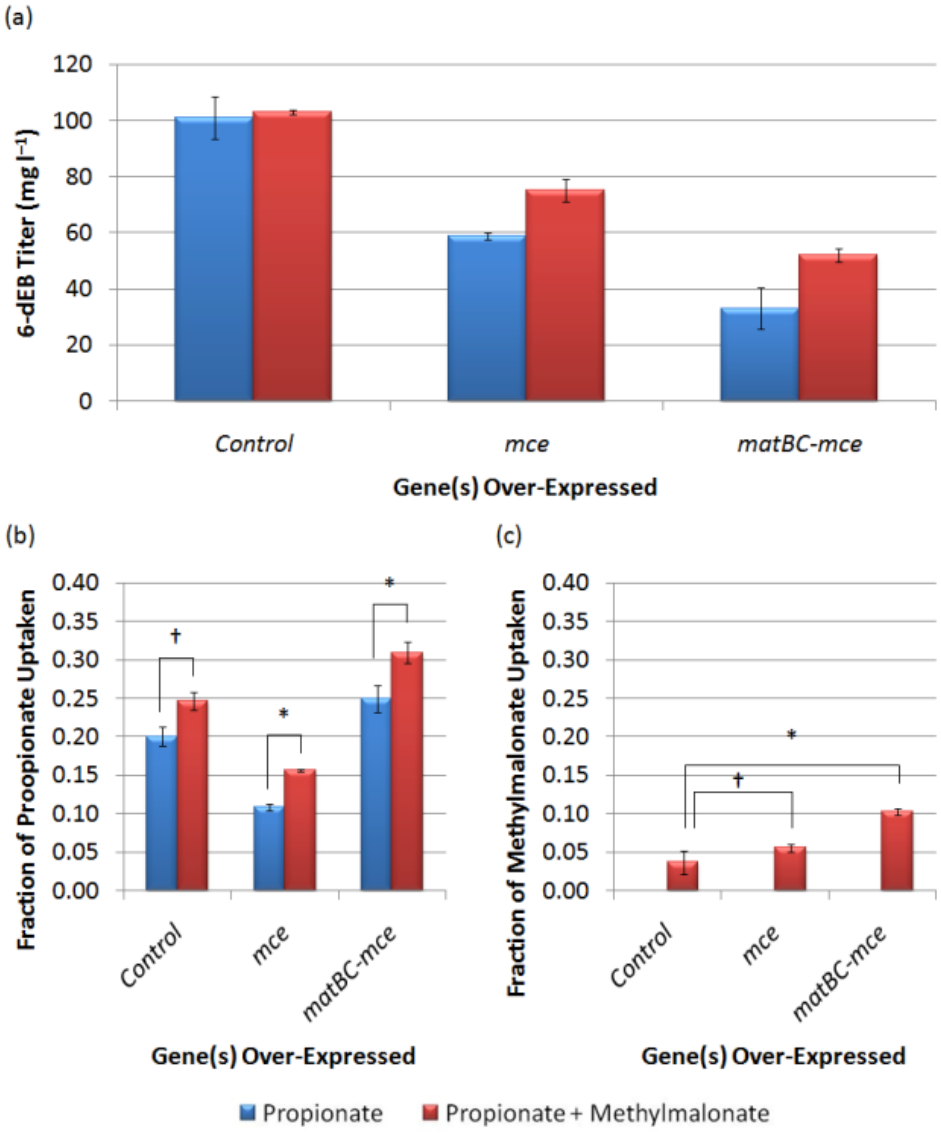

Figure 21 Methylmalonate pathway engineering in the presence of propionate for TB3.
Cultures were fed either 20 mM propionate or 20 mM propionate and 20 mM methylmalonate. Three different plasmid systems were analyzed (a control with only pBP130 and pBP144, an additional pCDFDuet-*mce*, and pACYCDuet-*matBC*/pCDFDuet-*mce*). Panel (a) shows the 6-dEB titer, (b) shows the fraction of propionate uptake, and (c) shows the fraction of methylmalonate uptake as a function of these parameters. Error bars represent ± one standard deviation of three replicates. * indicates statistically significant results ($p < 0.05$, as determined by paired Student's *t*-test), while † indicates statistically insignificant results ($p > 0.05$) between the comparisons shown.

Methylmalonate Pathway Engineering in the Absence of Propionate

It has been found that BAP1 could produce 6-dEB in the absence of propionate and methylmalonate (Zhang et al. 2010a). While it still remains unclear as to why methylmalonate supplementation in the presence of propionate improves 6-dEB production, the effects of these strains and plasmids in the absence of propionate were analyzed. With the utilization of mass spectrometry, the limit of detection of 6-dEB in the culture is approximately 0.1 mg l^{-1}, allowing 6-dEB production to be quantified in the absence of propionate supplementation. BAP1 produced 6-dEB at a titer of 0.32 ± 0.11 mg l^{-1} from 20 mM methylmalonate (Figure 22a). However, the uptake of methylmalonate was still minimal at 3.0 ± 0.2 mM (Figure 22b). The inclusion of the *mce* pathway had no effect on 6-dEB titer (p = 0.636) or methylmalonate uptake (p = 0.142). However, when the MatBC pathway was also expressed, 6-dEB titer improved to 1.27 ± 0.29 mg l^{-1} (p = 0.037) but methylmalonate uptake was not significantly different (p = 0.261).

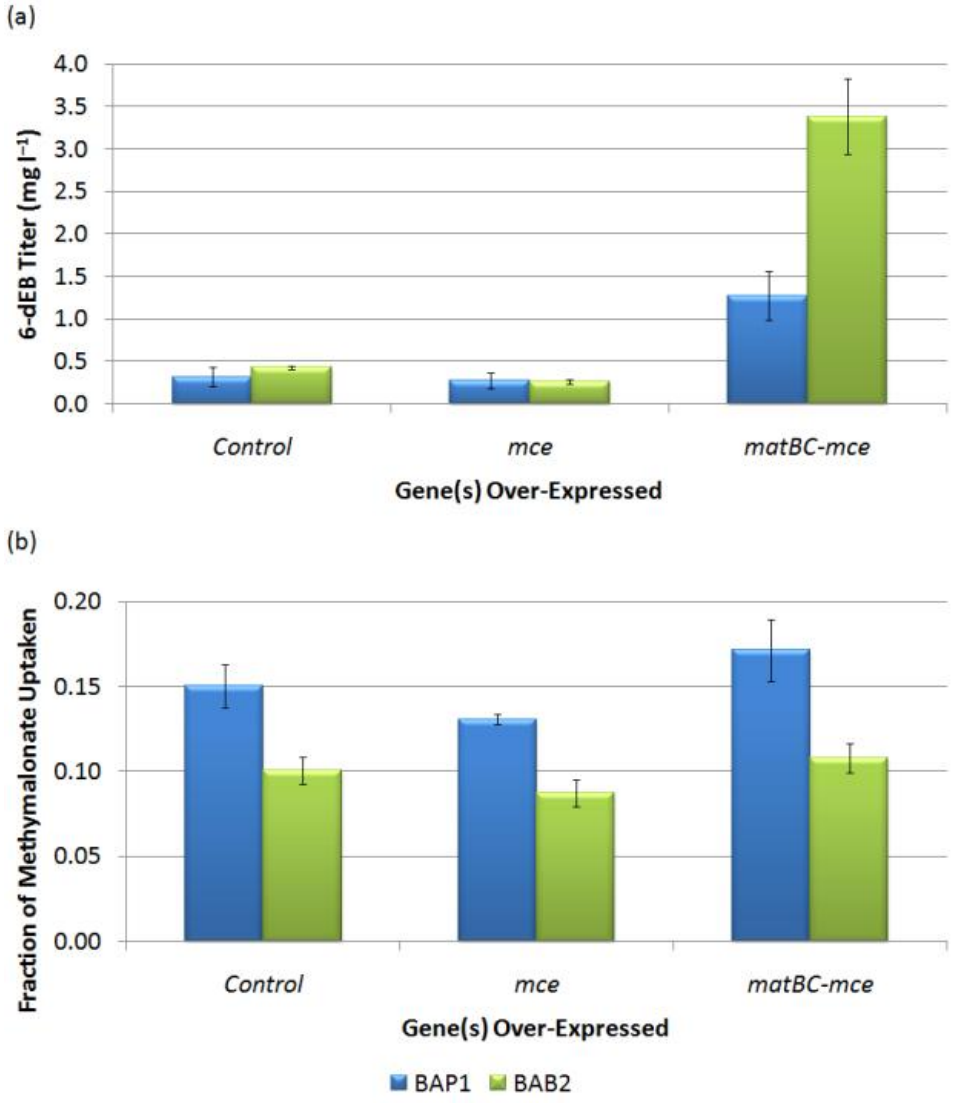

Figure 22 Methylmalonate pathway engineering in the absence of propionate. Data from engineering the methylmalonate pathway in the absence of propionate, as a function of three strains (BAP1, TB3, and BAB2) and three plasmid systems (a control with only pBP130 and pBP144, an additional pCDFDuet-*mce*, and pACYCDuet-*matBC*/pCDFDuet-*mce*). Panel (a) shows the 6-dEB titer, while panel (b) shows the fraction of methylmalonate utilized as a function of these cellular parameters. Error bars represent ± one standard deviation of three replicates.

In an effort to understand the effect of *mce* and *matBC-mce* on 6-dEB production from methylmalonate, strain was constructed that lacked the ability to convert polyketide precursors to metabolites to be used for growth (encoded by the *ygf* operon). In the *ygf* operon mutant of BAP1 (BAB2), 6-dEB production from methylmalonate was not significantly different when compared to BAP1 (p = 0.350) or BAP1 with *mce* (p = 0.770). Methylmalonate uptake was the lowest in BAB2 in all cases. Interestingly, with the inclusion of the MatB-MatC pathway, 6-dEB titer improved almost 8-fold to 3.39 ± 0.74 mg l^{-1}, even with decreased methylmalonate uptake.

Methylmalonyl-CoA Mutase-Epimerase Pathway Engineering

Last, this portion of the study aimed to further understand the role of Sbm and *E. coli*'s native succinate-to-propionate conversion cycle in 6-dEB production. Previously, *sbm* was over-expressed and deleted in BAP1 in separate experiments, to find that neither genetic modification had an influence on 6-dEB production (Zhang et al. 2010a). It was hypothesized the lack of effect of these *sbm* expression studies on 6-dEB production was due to a lack of methylmalonyl-CoA epimerase activity, since Sbm has been shown to generate the (2*R*) isomer of methylmalonyl-CoA; whereas, DEBS accepts only the (2*S*) isomer (see Figure 15). In this set of experiments, *sbm* and the *mce* gene previously described were over-expressed, thereby fully

connecting the succinate pathway with the 6-dEB production pathway. This combination was then tested in four different medium compositions which included 20 mM propionate, 20 mM succinate, 20 mM both substrates, and no substrates.

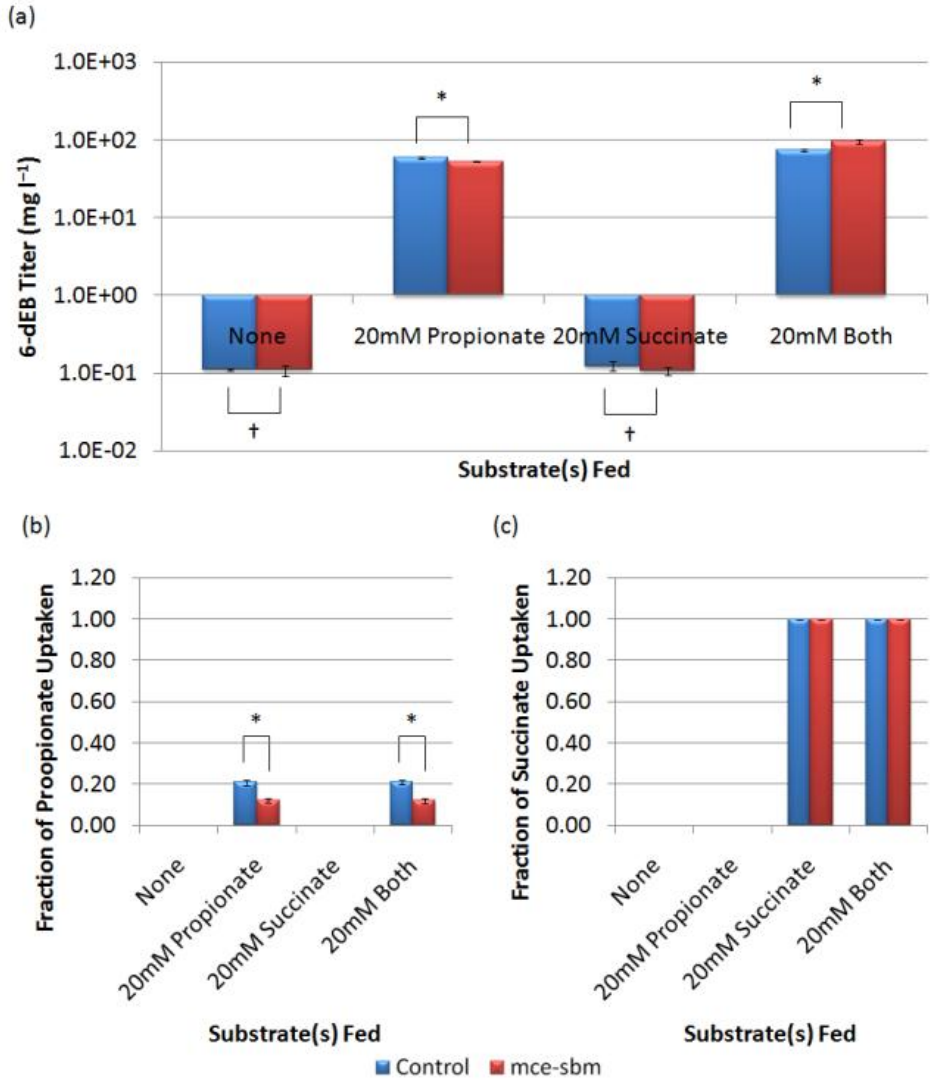

Figure 23 Methylmalonyl-CoA mutase-epimerase pathway engineering.
Data from engineering the methylmalonyl-CoA mutase-epimerase pathway as a function of two plasmid systems (a control with only pBP130 and pBP144, and pCDFDuet-*mce-sbm*) and four medium formulations (no substrates, 20 mM propionate, 20 mM succinate, or 20 mM both substrates). Panel (a) shows the 6-dEB titer, panel (b) shows the fraction of propionate utilized, and panel (c) shows the fraction of succinate utilized as a function of these cellular parameters. Error bars represent ± one standard deviation of three replicates. * indicates statistically significant results ($p < 0.05$, as determined by paired Student's *t*-test), while † indicates statistically insignificant results ($p > 0.05$) between the comparisons shown.

In the absence of either substrate, the titers with and without the Sbm-MCE were roughly 0.1 mg l^{-1}, consistent with what was observed previously (Zhang et al. 2010a) (Figure 23a). When propionate was utilized as the sole substrate, the inclusion of the Sbm-MCE pathway slightly decreased 6-dEB titer (p = 0.046). When succinate was utilized as the sole substrate, 6-dEB production decreased to levels observed in absence of a substrate; the Sbm-MCE pathway had no effect (p = 0.280). When both propionate and succinate were utilized, the titer without the Sbm-MCE pathway increased to 73.4 ± 2.0 mg l^{-1}, while the inclusion of the Sbm-MCE pathway further increased titer to 96.4 ± 5.6 mg l^{-1}. As can be seen in Figure 23b, the provision of succinate in the medium decreased propionate uptake in both the control and with expression of *sbm-mce*. In all cases, succinate was absent in the medium at the end of the culture period (Figure 23c).

Discussion

Heterologous polyketide biosynthesis presents a significant challenge in recombinant protein production and metabolic pathway engineering. Polyketides, being significant sources of therapeutic compounds, and PKS's, being complex enzymes, are attractive systems for chemists, biologists, and engineers. This study focused on the metabolic engineering of multiple pathways for substrate provision for heterologous polyketide biosynthesis in *E. coli*. The ultimate goals of

which are to 1) understand the interactions of these pathways, 2) identify the rate-limiting steps in polyketide biosynthesis, and 3) rationally engineer a system with improved titer. A multi-scale engineering strategy was applied through heuristic gene over-expression and deletion experiments and feeding experiments to better understand the interplay of native propionyl-CoA and methylmalonyl-CoA metabolism as well as heterologous methylmalonate and native succinate metabolism. While 6-dEB was used as a means of examining the effect of these pathways and substrates, similar strategies could be applied to improving the titer of other polyketides which use the same starter or extended acyl-CoA units. For example, *Streptomyces hygroscopicus* uses seven (2*S*)-methylmalonyl-CoA extender units for making the immunosuppressant rapamycin, while *Mycobacterium tuberculosis* uses (2*S*)-methylmalonyl-CoA for the biosynthesis of mycolic acids (Chan et al. 2009).

Our initial screen was comprised of a three variable (propionate, malonate, and methylmalonate), two-level (0 mM or 20 mM) full-factorial supplementation experiment across nine different plasmid systems, all in the base strain of BAP1. Because propionate and methylmalonate are the deactivated forms of the direct precursors for 6-dEB biosynthesis, it is expected that increased intracellular levels would improve 6-dEB biosynthesis. Malonate was used to serve as a

frame of reference for analyzing the MatBC pathway. As expected, the MatBC pathway had a stronger preference for malonate than for methylmalonate, as indicated by dramatic improvements in uptake when *matBC* was over-expressed. Malonate uptake was preferred over methylmalonate in general, presumably due to malonate being used a precursor for malonyl-CoA, the first committed step to essential fatty acid biosynthesis. However, no 6-dEB production was observed when malonate was used alone. The combination of malonate and propionate/methylmalonate did not improve production, leading us to believe that malonyl-CoA levels have no effect on propionyl-CoA or methylmalonyl-CoA levels.

The analysis of temperature influence on 6-dEB titer raises some interesting questions regarding the mechanisms of improvement. The lack of detectable 6-dEB production at 12°C and 17°C is likely a result of low cell-density, due to extremely low specific growth rates at these temperatures. (It has been previously shown that *E. coli* K-12's exponential-phase specific growth rate drops to roughly 0.01 hr^{-1} at 10°C (Ferrer et al. 2003).) Whereas, lack of production at 37°C could be due to temperature-induced protein misfolding or associated problems such as plasmid stability, which could also explain the observation of decreased 6-dEB titer at 32°C compared with 22 and 27°C. Over-expression of *E. coli*'s native chaperone systems (GroEL-GroES and

DnaK-DnaJ-GrpE (Nishihara et al. 1998; Nishihara et al. 2000)) did not improve 6-dEB production at 22°C or higher temperatures, but did slightly improve DEBS levels in some cases (data not shown). To this end, a trade-off between gene dosage, DEBS levels, and temperature appears to be critical to 6-dEB production. It has previously been shown that codon optimization of the *eryA* genes resulted in significantly increased levels of the DEBS enzymes; however, this same step decreased 6-dEB production from 19.8 ± 1.8 mg l^{-1} to undetectable (<0.1 mg l^{-1}) levels (Menzella et al. 2006). Alternatively, when the DEBS genes were integrated into the chromosome of BAP1 under the control of T7 promoters, 6-dEB production also decreased from 12.06 ± 4.42 mg l^{-1} to 0.47 ± 0.26 mg l^{-1} (Wang and Pfeifer 2008). The significant drain of free amino acids for heterologous protein production can cause deleterious effects during pathway engineering, a problem highlighted previously (Jones et al. 2000). Over-expression of cold-adapted chaperonins (such as the Cpn60-Cpn10 system from *Oleispira antartica* (Ferrer et al. 2003; Strocchi et al. 2006)) may improve 6-dEB production at these lower temperatures where the native *E. coli* chaperone system could not.

Increasing the expression of *prpE* and/or *atoC* was identified as a metabolic engineering target considering that deletion of *ygfH* was implemented to improve 6-dEB production (Zhang et al. 2010a) and

both of these enzymes are connected to this metabolite node. It has been previously demonstrated that a constitutive mutation in *E. coli*'s *atoC* allowed for transcription of the ATO genes (Jenkins and Nunn 1987) and improved the propionate uptake roughly 10-fold in M9 minimal medium supplemented with 1% (wt vol^{-1}) glucose and 10 mM propionate (Rhie and Dennis 1995). In our study, when *prpE* and/or *atoC* were over-expressed, propionate uptake increased roughly 4-fold, yet 6-dEB production only increased roughly 30%. This indicates that there is another significant sink of propionyl-CoA that is responsible for drawing from this metabolite pool, even in the absence of YgfH and PrpBCD. Possible sources could be enzymes that have preference for acetyl-CoA but are also promiscuous for propionyl-CoA (Man et al. 1995).

This study presents the first production of 6-dEB solely from methylmalonate. Inclusion of the MatBC pathway in BAP1 increased the 6-dEB titer from 0.32 ± 0.11 mg l^{-1} to 1.27 ± 0.29 mg l^{-1}, while deletion of the entire *ygf* operon further improved production to 3.39 ± 0.74 mg l^{-1}. Previously, 6-dEB had been produced in only trace quantities (0.85 ± 0.2 mg l^{-1}) from methylmalonate using *matB* and the *Streptomyces coelicolor* methylmalonyl-CoA mutase (*mutAB*), however 10 mM propionate was also added to the medium (Murli et al. 2003). The extremely low uptake rates of methylmalonate appear to be

a barrier in improved 6-dEB production from this substrate. Even in the absence of propionate and in the presence of MatBC, methylmalonate uptake was never above 20% of the fed substrate (a total of 4 mM).Whereas other substrates, more commonly used carbon sources (propionate, succinate, and malonate), were imported at significantly higher rates. Interestingly, the rate of methylmalonate uptake was 21.8% that for malonate (as determined in the initial screening study), which is strikingly similar to a previous *in vitro* analysis of MatB, which demonstrated 20.4% activity utilizing methylmalonate when compared to malonate (An and Kim 1998). There is no known transporter dedicated to methylmalonate uptake, nor is there a known methylmalonyl-CoA synthetase or CoA-ligase specific for utilizing methylmalonate as a substrate. This specificity issue may be overcome by protein engineering of the *R. trifolii matBC* system to make it specific or strongly preferential for methylmalonate/methylmalonyl-CoA. In a slightly different manner, methylmalonate uptake could be improved by a laboratory evolution experiment utilizing methylmalonate as a sole carbon substrate.

Finally, another commonly utilized cellular metabolite, succinate, could be engineered into the 6-dEB backbone through the functional expression of a methylmalonyl-CoA mutase-epimerase pathway. In all the studied cases, propionate addition was needed for

high-level production, indication that propionate was the most favorable substrate for 6dEB biosynthesis in *E. coli*. When 20 mM succinate was fed (in addition to 20 mM propionate) without the introduced Sbm-MCE pathway, 6-dEB titer decreased. However, when the methylmalonyl-CoA mutase-epimerase pathway was expressed, 6-dEB titer increased, indicating that basal levels of Sbm (and lack of an epimerase) could not incorporate succinate into (2*S*)-methylmalonyl-CoA. This appears to contrast previous work where 6-dEB was only produced at roughly 1 mg l^{-1} from 5 mM propionate, 50 mM succinate, and 50 mM glutamate using a *Propionibacterium shermanii* methylmalonyl-CoA mutase and the same *Streptomyces coelicolor* methylmalonyl-CoA epimerase (Dayem et al. 2002). Combining this new information with previous information from computational modeling (Boghigian et al. 2010), leads us to believe that metabolic engineering of the succinate and succinyl-CoA metabolite nodes could be critical to further improve 6-dEB production (and production of other propionate-dependent polyketide products) by providing access to primary metabolism.

Chapter 5 – Identification of knockout targets through elementary mode analysis and a genetic algorithm, and experimental implementation for improving lycopene production

Introduction

As stated in Chapter 2, there currently exists a variety of algorithms exist for identifying knockout targets such as OptKnock (Burgard et al. 2003), OptStrain (Pharkya et al. 2004), OptGene (Patil et al. 2005), and OptReg (Pharkya and Maranas 2006). However, these methods rely on optimization frameworks for determining metabolic fluxes, such as FBA (Edwards et al. 2002), MoMA (Segre et al. 2002), ROOM (Shlomi et al. 2005), or some variation of these techniques. Although these optimization approaches can accurately predict optimal growth and production fluxes in some cases (Edwards et al. 2001), other experimental settings produce inaccurate predictions (Schuetz et al. 2007). In this chapter, the primary objective was to predict an over-producing phenotype *strictly* from the wild-type organism's metabolic network topology (similar to the notion of a "minimal cell" (Trinh et al. 2008)), therefore alleviating the need to quantify fluxes using an optimization methodology.

In chemical process design, process operability has been defined as the extent to which there is an adequate amount of equipment over-design so that the process constraints can be satisfied (Vinson DR, personal communication). Industrial processes are often designed to incorporate a level of robustness, in which fluctuations or alterations in input variable(s) (such as a stream flow-rate, temperature, or pressure) do not affect the output variable(s) (such as a conversion, reaction rate, or purity). Drawing an analogy to the metabolic pathways of cellular systems, they are incredibly robust (Behre et al. 2008; Daniels et al. 2008; Kim et al. 2007a). In varying environmental and genetic situations (the input variables), they can still grow and produce a product of interest (Kitano 2004; Kitano 2007). For a metabolic engineer's purposes, this robustness can often be a hindrance. Unlike chemical systems, cellular systems are evolved systems, and will choose to operate at some point within their available genetic means, as a result of the cell's desire to survive in a given environment. While many methods have been utilized for simulating flux distributions (such as FBA and MoMA used in this dissertation), they require an optimality assumption (maximize growth rate), which may or may not be valid in some cases. Moreover, during a batch (or fed-batch) bioreactor process, the cell's operating point will move depending on the state of the overall process (for example,

exponential growth phase versus stationary phase). Rather than identify genetic targets based on single flux distributions, the goal of the presented algorithm was to simply reduce the robustness of the cellular system in question, and inevitably couple product synthesis with growth. In essence, the aim is to control the operability of the metabolic network.

In cellular systems, the operating space is governed by a polytopic flux cone (Figure 24), which is governed by the reactions (genes) in the system. By systematically tuning the bounds of the flux cone (which are the strain's elementary modes, or EM's), the cell's operating space can be constrained. An actual flux distribution can be represented as a linear combination of the EM's. Next, the question of how to define the desired operating space becomes important. Generally speaking, the desired operating space should be characterized by a high-yield of product on substrate, while not significantly sacrificing growth-rate of the organism. As such, the algorithm presented here aims to constrain the desired operating space to allow the organism to (1) grow and (2) produce a high-yield of a product from a given carbon-source. In this chapter, this was accomplished through the development of an algorithm called ConstrainStrain that couples elementary mode analysis (EMA) and a genetic algorithm for optimizing metabolic network structure. As a

surrogate for taxadiene production, lycopene, was used as a target molecule. As cited in Chapter 2, lycopene is a C_{40} bright red carotenoid with antioxidant properties, shares precursor molecules of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) with a variety of other isoprenoids, and facilitates screening procedures in this case.

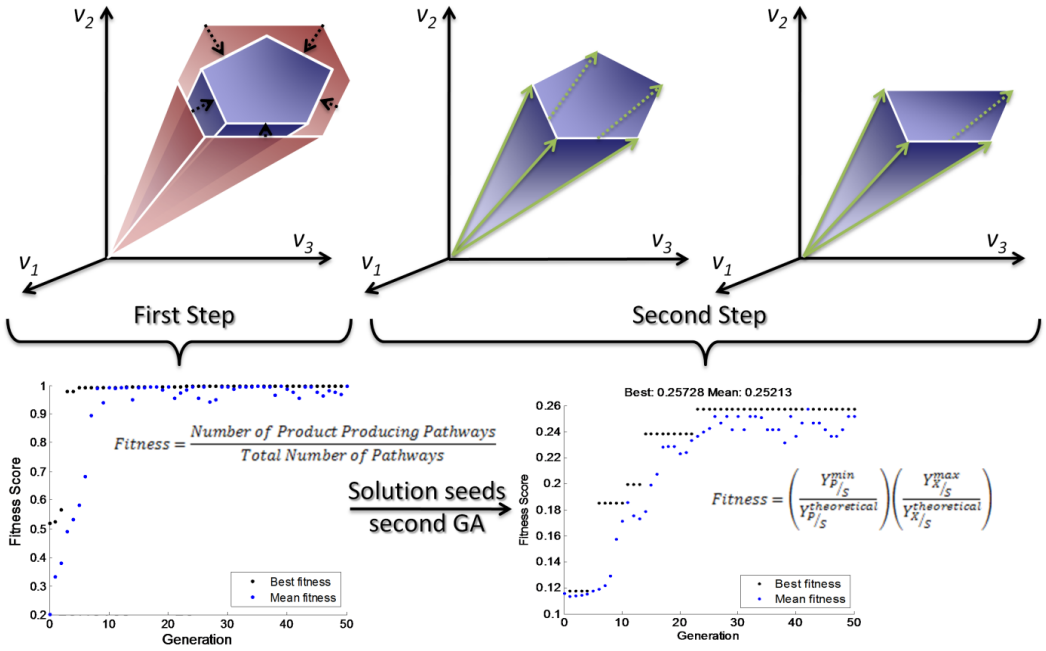

Figure 24 An overview of the ConstrainStrain algorithm.
Here, the polytopic flux cone is represented in a simplified three-dimensional space (three reactions). Elementary modes are represented by green arrows. Sample genetic algorithm fitness curves are shown as a function of generation number on the bottom.

Materials & Methods

Model Construction

The two small-scale *E. coli* stoichiometric models utilized in this study were based on one previously developed (Trinh et al. 2008). Briefly, because the previous model was developed for the utilization of multiple five- and six-carbon sugars, all of the carbon-source utilization reactions besides the glucose utilization reaction were removed; glucose was assumed to be actively imported by the phosphoenolpyruvate sugar transferase system. In the original model, an additional reaction was included due to a heterologous pyruvate decarboxylase from *Zymomonas mobilis*; this reaction is not native to *E. coli* and was therefore also removed.

Taxadiene biosynthesis was introduced into the model and coupled to the non-mevalonate pathway (native to *E. coli*) previously used to support heterologous carotenoid production (Das et al. 2007; Yuan et al. 2006). Whenever possible, linear pathways were combined into a single reaction to reduce the size of the model (see Figure 3 for the pathway). The first reaction (encoded by *dxs*, *dxr, and ispDEFGH*) held the stoichiometry: glyceraldehyde-3-phosphate + pyruvate + 2 NADPH + ATP → dimethylallyl diphosphate (DMAPP) + CO2 + 2 $NADP^+$ + ADP. The second reaction was for the reversible

isomerization of DMAPP and isopentenyl diphosphate (IPP), encoded by *idi*. The third reaction held the stoichiometry 4 IPP → geranylgeranyl diphosphate (GGPP) and is encoded by *ispA* and *crtE*. The last reaction was for taxadiene biosynthesis and held the stoichiometry GGPP → taxadiene and is encoded by *txs* (Koepp et al. 1995). To avoid the inclusion of a specific transport reaction, taxadiene was not balanced in this reaction. For the glycerol model, the glucose uptake reaction was added and replaced with two reactions to enable glycerol metabolism. The first, encoded by *glpK*, holds the stoichiometry of: glycerol + ATP → glycerol-3-phosphate + ADP. The second, encoded by *glpABCD*, holds the stoichiometry of: glycerol-3-phosphate ↔ dihydroxyacetone 3-phosphate.

Elementary Mode Analysis

Elementary mode analysis (EMA) was undertaken utilizing the bit pattern tree method (Terzer and Stelling 2008). Developed recently, this algorithm is capable of enumerating 2,450,787 EMs over ten-times faster (on a four-thread system) than the latest release of METATOOL (Pfeiffer et al. 1999; von Kamp and Schuster 2006), and is therefore currently the fastest method for EM enumeration. The mathematical rigor associated with the bit pattern tree method and other EMA algorithms has been described previously (Haus et al. 2008; Jevremovic et al. 2010; Terzer and Stelling 2008; Urbanczik and

Wagner 2005). The code was acquired from Professor Jörg Stelling's website (http://www.csb.ethz.ch/tools/efmtool) and interfaced with The MathWorks™ MATLAB software (version 7.6.0.324).

Genetic Algorithm

Chromosomal representation of the metabolic genotype for passing to the genetic algorithm is binary in nature where a "1" indicates the reaction is included in the individual and "0" indicates that the reaction is not present. For simplicity's sake, a one-to-one association between reactions in the network and genes in the GA's population was assumed. This one-to-one association decreases computation time by utilizing fewer variables for optimization and does not present a significant problem experimentally, for the gene-associations with the enzymes catalyzing the reactions are well-known for *E. coli* due to the organism's biochemical knowledge and sequenced genome (Blattner et al. 1997; Durfee et al. 2008; Hayashi et al. 2006). A binary vector of length n therefore represents a single individual in the GA population.

Initialization of a population is a critical step for determining the success of the algorithm to find the global optimum. An initial population of fifty individuals containing between two and six knockouts was seeded to the algorithm (using MATLAB's "randerr" function). This was arrived at empirically as randomly seeding

individuals with approximately 50% 0's resulted in mostly non-viable strains and did not allow for the GA to reach the optimal solution. Next, each individual in the population is evaluated and given a fitness score. A previous study on using GAs to optimize genotypic space for succinate, glycerol, and vanillin production used product flux determined by optimization (FBA and MoMA) as a scoring function (Patil et al. 2005). As stated before, this approach relies on assumptions that may or may not be valid. Here, EMA was used as the method for scoring the individuals with fitness functions as described in Equation 9 and Equation 10.

Genetic algorithms use crossover of the chromosomes (mixing of two individuals in a population to create a new individual) and mutation (change a "0" to "1" and vice-versa with a specified frequency) to evolve the solution population. The implementation here was interfaced with The MathWorks™ MATLAB software and its Genetic Algorithm & Direct Search Toolbox. For crossover, mutation, and selection of individuals, two-point, uniform, and tournament-based methods were used, respectively. These parameters were not optimized in this study. As stated, the population size was chosen as fifty individuals, with five of the top performing individuals automatically passed to the next generation of the GA. The selection function used in the GA was either roulette- or tournament-based. The GA always

terminated as a result of being below the tolerance (of the MATLAB default, 10^{-6}) which was always between 50 and 100 generations. Figure 25 shows the overall schematic of the genetic algorithm and fitness calculation.

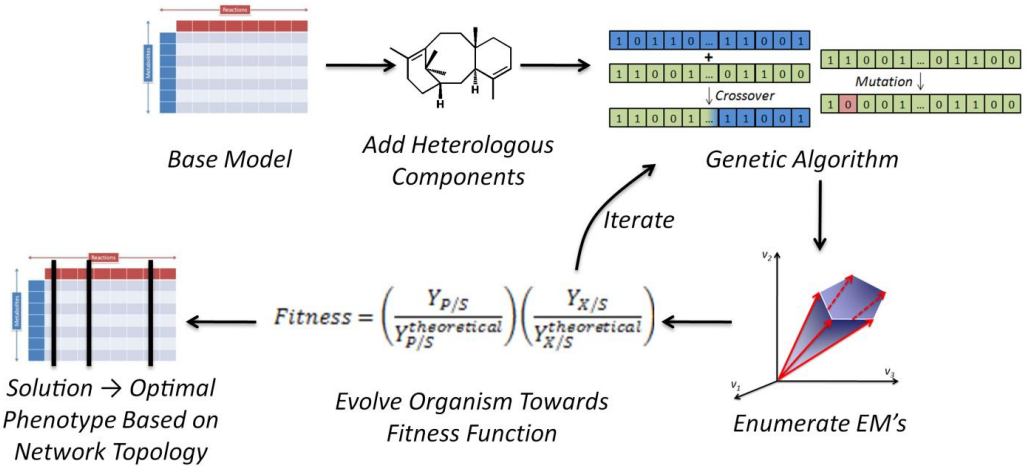

Figure 25 Schematic overview of the framework.

As a method to reduce the computation time of the GA optimization, the GA was forced to always include (through fixed inclusion of a "1" in the individual genotype) reactions that were determined to either 1) reduce maximal product yield to zero, or 2) reduce maximal biomass yield to zero (indicating a lethal knockout). The first step of the algorithm was designed to tailor the metabolic network to contain only elementary modes that produced the product of interest (either taxadiene or lycopene). As a result, the fitness score is as described in Equation 9:

Equation 9

$$Fitness\ Score = \frac{Number\ of\ Product\ Producing\ Modes}{Total\ Number\ of\ Modes}$$

The solution of this first round of the genetic algorithm was used to seed the second round of the genetic algorithm, in which the population was evolved to the following fitness score in Equation 10:

Equation 10

$$Fitness\ Score = \left(\frac{Y_{P/S}^{min}}{Y_{P/S}^{theoretical}}\right)\left(\frac{Y_{X/S}^{max}}{Y_{X/S}^{theoretical}}\right)$$

Both steps of the algorithm are represented in Figure 24.

Strain Construction by MAGE

Multiplex automated genome engineering (MAGE) is a technique for allelic replacement, designed originally for use with *E. coli* and the λ-Red recombination system (Wang et al. 2009). Briefly, this system allows climate-regulated growth of cells with real-time cell-

density monitors, fluidic systems for transfer of cells from growth chambers with different environmental conditions, fluidic systems for buffer exchange and electrocompetent cell preparation, and a system for electrotransformation of a population of cells with either single stranded (ssDNA) or double stranded DNA (dsDNA). This process can be repeated for multiple cycles to evolve a population of cells to contain multiple genetic modifications. Genetic mismatches, insertions, or deletions can be performed with this system. Oligonucleotides designed for deletion of identified were designed to be 90 bp long, contain four terminal 5' phosphorothiorated bases, and to target the lagging strand of chromosomal DNA. With this oligonucleotide design strategy at these parameters, the replacement efficiency can exceed 30% (Wang et al. 2009).

The parent strain for MAGE cycling was *E. coli* EcHW2e. Briefly, the λ prophage was obtained from DY330 (Novere et al. 2009; Sharan et al. 2009; Swingle et al. 2010; Yu et al. 2000). A carbenicillin resistance gene (*bla*) was inserted before the λ-Red genes in DY330. This region was introduced into *E. coli* MG155 mediated by P1 transduction (Masters 1977). To enable mismatch base pairings to stably exist within the *E. coli* chromosome, the MutHLS complex (the methyl-directed mismatch repair system) must be disrupted. A kanamycin resistance gene (*kan*) was inserted into the *mutS* locus

using λ-Red recombination. The constructed strain was designated EcHW1. Transformation of pAC-LYC (containing *crtEBI* genes from *Erwinia herbicola* (Cunningham et al. 1994), required to produce lycopene in *E. coli*) into EcHW1 generated EcHW2, a carbenicillin, kanamycin, and chloramphenicol resistant strain of *E. coli* capable of producing lycopene. After MAGE cycling, a strain with optimized ribosome binding site sequences for *dxs* and *idi* (to to be more similar to the canonical Shine-Dalgarno sequence, giving rise to enhanced translation efficiency) was identified as a top producer of lycopene. This strain was identified after randomly implementing any number of 24 genetic modifications (twenty over-expression targets and four knockout targets), and MAGE cycling for between 5-35 cycles. This strain produced 9,000 ppm (μg $gDCW^{-1}$) of lycopene after 24 hr of culture in LB-minimal salt medium (also known as LB-Lennox medium, containing 5 g NaCl l^{-1} instead of 10 g l^{-1}) at 30°C. Lycopene was used as a surrogate for taxadiene being that lycopene is produced from two molecules of GGPP, while taxadiene is produced from one molecule of GGPP and is colored, easing analysis.

MAGE cycling experiments were cycled for a total of twelve rounds. Selection was undertaken on LB-min plates supplemented with 34 mg l^{-1} chloramphenicol after 24 hr at 30°C and then for 48 hr at room temperature. At this point, colonies showed a red phenotype

corresponding to lycopene production. 48 clones were randomly picked from each of the three pools of knockouts and diagnostic PCR was conducted for each of the knockout targets to determine the genotype. Verification primers were designed upstream of the target gene's start codon and downstream of its stop codon, such that the PCR product would be substantially smaller if the gene was successfully removed (data not shown).

Shake-Flask Production Cultures

Shake-flask cultures (15 ml in 125 ml Erlenmeyer flasks) containing production medium were used for lycopene production tests. The medium used was comprised of 5 g l^{-1} yeast extract, 10 g l^{-1} tryptone, 10 g l^{-1} sodium chloride, 3 ml l^{-1} 50% (v v^{-1}) Antifoam B, 100 mM HEPES, and pH 7.60. Carbon sources were added at a final concentration of 15 g l^{-1}. Glucose was added post-sterilization from a filter sterilized stock solution, while glycerol was added before autoclaving. Single colonies were picked from freshly streaked plates and inoculated into 1 ml production medium containing necessary antibiotics. These cultures were grown at 30°C and 250 rpm until $OD_{600nm} \approx 0.6$ and were used to inoculate 15 ml production medium with necessary antibiotics and 100 µM IPTG at a volumetric ratio of 5%. Cultures were then incubated at 22°C and 250 rpm for 72 hr. After 24 hr and 72 hr of culture, 1 ml aliquots were taken, cell-density was

measured spectrophotometrically at 600 nm, and the lycopene concentration was assayed as to be described. At all steps, antibiotics were supplemented at concentrations of 100 mg l^{-1} for carbenicillin, 50 mg l^{-1} for kanamycin, and 34 mg l^{-1} for chloramphenicol.

Lycopene Quantification

Lycopene quantification was undertaken as described previously (Yoon et al. 2006). A frozen aliquot of cells were thawed at room temperature and then centrifuged at 13,000×g for 3 min. The supernatant was decanted and the pellet was re-suspended in an equal volume of sterile water to wash once. The cells were centrifuged again at 13,000×g for 3 min and the supernatant was decanted. The pellet was then re-suspended in an equal volume of HPLC-grade acetone, vortexed for 5 s, and incubated at 50°C for 15 min in the dark. The samples were then centrifuged 13,000×g for 10 min, diluted 10-fold in HPLC-grade acetone, and the absorbance was measured spectrophotometrically at 474 nm. The concentration of lycopene was determined against a five-point calibration curve generated using authentic lycopene derived from tomato (purchased from Sigma-Aldrich, >90% purity).

Metabolite Quantification

Medium and byproduct organic acids were quantified by the previously mentioned HPLC system coupled to a Refractive Index

Detector (RID). Clarified culture supernatant (20 µl) was applied to a Bio-Rad Aminex® HPX-87H Ion Exchange (300 mm × 7.8 mm, 9 µm) column, preceded by a 30 mm guard column of the same resin. The column temperatures were maintained at 42°C throughout the analysis. The isocratic analysis used a 9.5 mM H_2SO_4 solvent held at a flow rate of 0.3 ml min^{-1}. These conditions were identified by using an iterative stochastic search HPLC optimization program based on the compounds anticipated to be present in the culture medium (Dharmadi and Gonzalez 2005). A five-point standard calibration curve was created and used for quantification of glucose, glycerol, and acetate. The elution order was as follows: glucose (16.7 min), glycerol (25.1 min), and acetate (29.1 min). All specific production or consumption rates presented are averaged over the course of the culture period.

Microplate Growth Assay

To more accurately determine the intrinsic growth properties of the strains constructed in this study, cultures were undertaken in a 96 well-plate format. Sterile, flat-bottom 96 well-plates were inoculated with 200 µl medium containing appropriate antibiotics. A defined, minimal medium (M9; containing 12.8 g l^{-1} $Na_2HPO_4 \cdot 7H_2O$, 3 g l^{-1} KH_2PO_4, 0.5 g l^{-1} NaCl, 1 g l^{-1} NH_4Cl, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$) with 15 g l^{-1} glucose or 15 g l^{-1} glycerol and a complex medium (production medium, as described previously) with 15 g l^{-1} glucose or

15 g l^{-1} glycerol were used in this study. A stab of glycerol stock was inoculated into LB medium supplemented with appropriate antibiotics and grown overnight at 30C and 250 rpm. Cell-density was measured spectrophotometrically at 600nm, and the wells were inoculated to OD = 0.05. Well plates were inserted into a Molecular Devices VERSAmax microplate spectrophotometer preheated to 30°C and incubated for 16 hr with mixing. Every 10 min, mixing stopped and the absorbance at 600 nm was measured for all wells.

Growth Parameter Determination

Data was exported to Microsoft Excel 2007, and then to MATLAB® version 7.10.0.499 R1010a (The Mathworks). Background absorbance was subtracted from the raw data, and then fit to a logistic population model (Bailey and Ollis 1986).

Equation 11

$$X(t) = \frac{X_0 e^{\mu t}}{1 - \frac{X_0}{X_{sat}}(1 - e^{\mu t})}$$

Equation 11 describes the cell-density, X, as a function of time, t, with the following constant parameters (X_0, the initial cell-density; X_{sat}, the stationary phase cell-density; and μ, the specific growth rate). A non-linear least-squares regression of the data with the logistic population equation was undertaken in MATLAB (using the "nlinfit" function) to determine the X_0, X_{sat}, and μ parameters. 95% confidence

intervals for the cell-density estimates, and the estimated parameters were examined for accuracy.

Results

Model Construction & Elementary Mode Analysis

Information on the models and their characteristics can be found below (Table 8).

Table 8 Model information.
Model information (metabolites and reactions) for the production of taxadiene from glucose or glycerol in *E. coli*. The total number of elementary modes are given, as well as the number and percentage biomass-producing EM's, taxadiene-producing EM's, and biomass- and taxadiene-producing EM's. Lastly, the computation time for the enumeration of the EM's for each network is given in seconds (± one standard deviation, from ten simulations).

	Glucose	Glycerol
Metabolites	50	51
Reactions	64	65
Total Number of Elementary Modes	40,425	15,010
Biomass-Producing EM's (%)	31,354 (77.56%)	11,144 (74.24%)
Taxadiene-Producing EM's (%)	7,205 (17.82%)	2835 (18.89%)
Biomass- and Taxadiene-Producing EM's (%)	3,018 (7.47%)	954 (6.36%)
Computation Time (n = 10)	7.4144 ± 0.2191 s	2.7117 ± 0.1011 s

Figure 26 shows a scatter plot of all of the glycerol model's elementary modes and their corresponding yields of biomass on glycerol, and taxadiene on glycerol (normalized to the fraction of the theoretical yield).

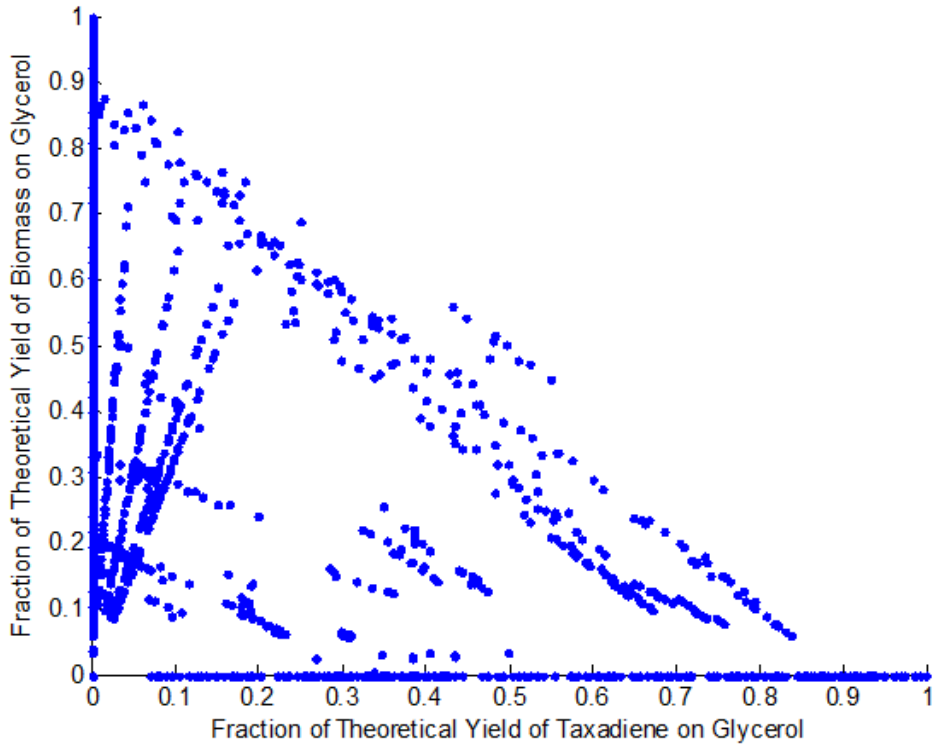

Figure 26 EM's of the parent strain.
A scatter plot of all of the parent glycerol model's 15,010 EM's and their corresponding yields of biomass on glycerol, and taxadiene on glycerol (normalized to the fraction of the theoretical yield). In some cases, one point will correspond to multiple elementary modes with the same yields.

ConstrainStrain Algorithm

ConstrainStrain was then applied to both models and repeated thirty times in an attempt to sample a significant fraction of the design space (Table 9).

Table 9 Information on the solutions of thirty simulations of first step the ConstrainStrain algorithm as applied towards taxadiene production from glycerol. The range of solution fitness scores, mean and median of these fitness scores, as well as the average computation time (± one standard deviation, from thirty simulations). The range of the solution strain's genotypes (in number of knockouts and corresponding elementary modes) are shown.

Range of Fitness Scores	0.577-0.992
Mean Fitness Score	0.947
Median Fitness Score	0.992
Average Computation Time (*n* = 30)	3,509.5 ± 788.8 s
Range of Number of Total Modes	39-352
Range of Number of Knockouts	2-7

The majority of the solution genotypes were almost optimal, many of which contained the same exact genotype. Figure 27a shows a summary of these strains with their corresponding total number of modes, biomass-producing modes, taxadiene-producing modes, and biomass- and taxadiene-producing modes. Figure 27b shows the minimum taxadiene yield of all of the strain's modes, and the maximum biomass yield of all of the strain's modes. As can be seen in this figure, most of these strains evolved to contain all modes that could produce taxadiene, however, this came at a cost of producing biomass, as many strains can now only produce, at maximum, 70% of the original yield of biomass on substrate. The same algorithm was applied to the glucose model, however, the results are not shown.

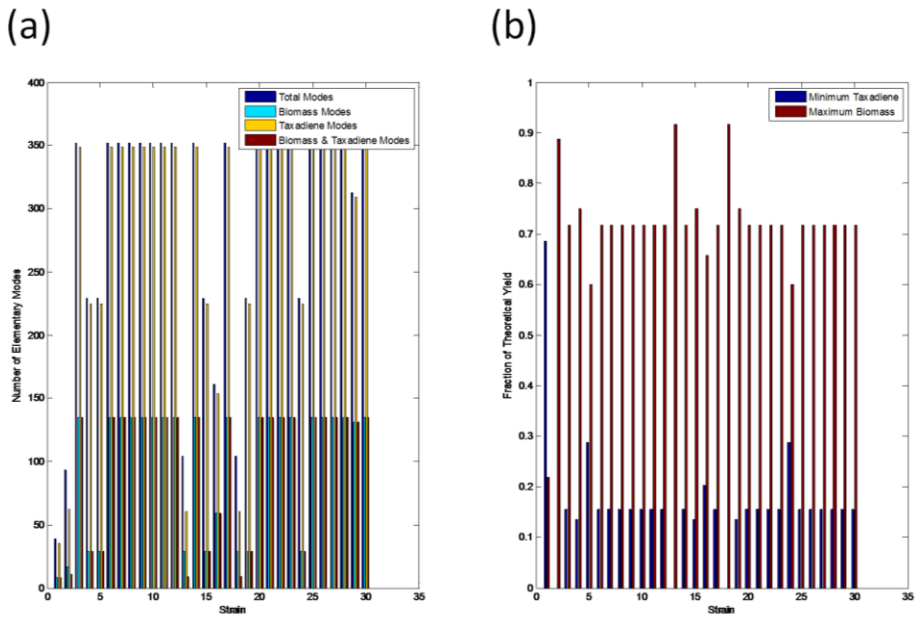

Figure 27 Results of the thirty simulations of the first step of ConstrainStrain. (a) shows the total number of EM's (dark blue), biomass-producing EM's (light blue), taxadiene-producing EM's (yellow), and biomass- and taxadiene-producing EM's (maroon). (b) shows the minimum taxadiene yields and maximum biomass yields of all of the strains EM's (normalized to the fraction of the theoretical yield)

The strain chosen to proceed onto the next step of the algorithm was one that had relatively few knockouts, and the maximal fitness score. This strain had two reaction removals, corresponding to three genes (*pntAB*, *gapA*), coding for pyridine nucleotide transhydrogenase (NAD + NADPH ↔ NADP + NADH) and glyceraldehyde 3-phosphate dehydrogenase (glyceraldehyde 3-phosphate + NAD+ ↔ 1,3-diphosphateglycerate + NADH). This strain contained 305 elementary modes. This was fed into the second round of the algorithm and run 30 times. Figure 28 again shows the distribution of elementary modes and their corresponding yields before the second step of the algorithm. It should be noted that no elementary modes correspond to zero yield of taxadiene on biomass, however, the maximum yield of biomass on substrate is roughly 72% of the original maximal yield.

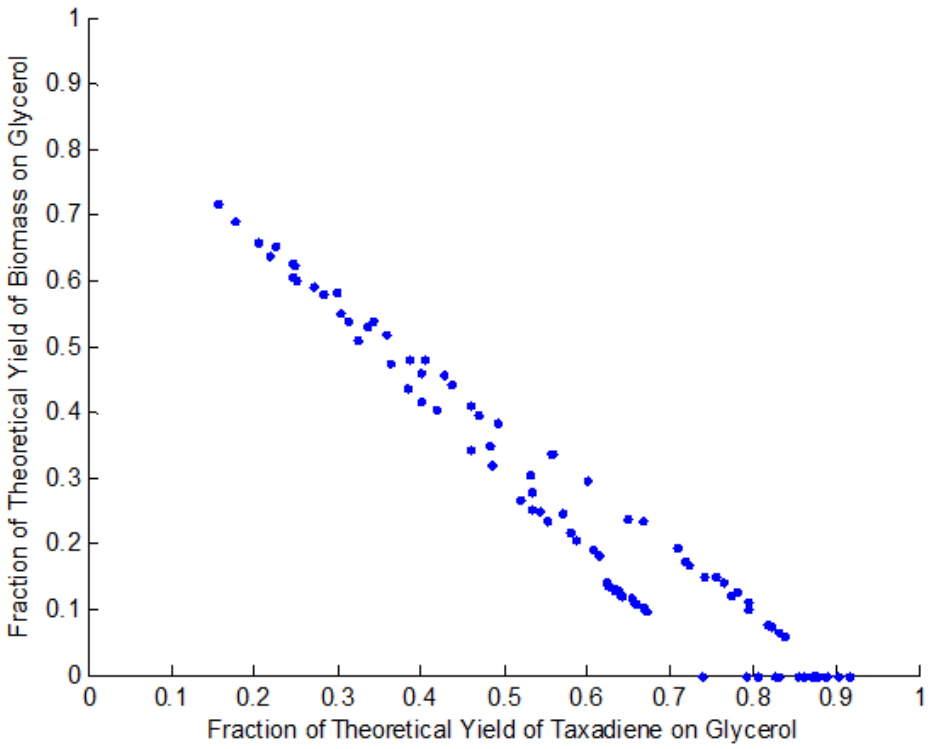

Figure 28 EM's of the *ΔpntAB ΔgapA* strain.
A scatter plot of the 305 EM's in the *ΔpntAB ΔgapA* genotype and their corresponding yields of biomass on glycerol, and taxadiene on glycerol (normalized to the fraction of the theoretical yield). In some cases, one point will correspond to multiple elementary modes with the same yields.

Table 10 shows statistics on thirty simulations on the second step of the ConstrainStrain algorithm.

Table 10 Information on the solutions of thirty simulations of second step the ConstrainStrain algorithm as applied towards taxadiene production from glycerol. The range of solution fitness scores, mean and median of these fitness scores, as well as the average computation time (± one standard deviation, from thirty simulations). The range of the solution strain's genotypes (in number of knockouts and corresponding elementary modes) as well as phenotype (range of minimum taxadiene and maximum biomass yields) are shown.

Range of Fitness Scores	0.140-0.196
Mean Fitness Score	0.176
Median Fitness Score	0.185
Average Computation Time (*n* = 30)	349.3 ± 40.7 s
Range of Number of Total Modes	21-212
Range of Number of Knockouts	4-10
Range of Minimum Taxadiene Yield	0.281-0.520
Range of Maximum Biomass Yield	0.268-0.580

The computation time of this second step in the algorithm was significantly shorter due to the smaller model size of the seeded population. This strain contained another two reaction removals, corresponding again to three genes (*adhE* and *tktAB*), coding for acetaldehyde dehydrogenase (acetyl-CoA + NADH → acetaldehyde + NAD+ + CoASH) and transketolase (erythrose-4-phospate + xylose-5-phosphate ↔ glyceraldehyde-3-phosphate + fructose-6-phosphate). This strain contained 208 elementary modes. Figure 29 again shows this strain's EM characteristics.

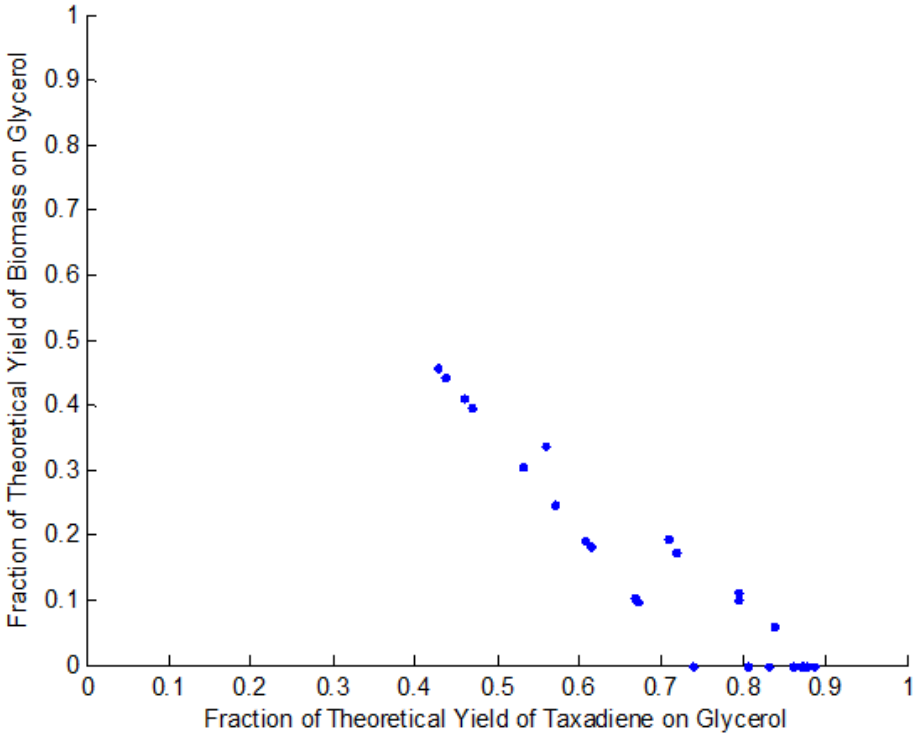

Figure 29 EM's of the Δ*pntAB*, Δ*adhE*, Δ*gapA*, Δ*tktAB* strain.
A scatter plot of the 305 EM's in the Δ*pntAB*, Δ*adhE*, Δ*gapA*, Δ*tktAB* genotype and their corresponding yields of biomass on glycerol, and taxadiene on glycerol (normalized to the fraction of the theoretical yield). In some cases, one point will correspond to multiple elementary modes with the same yields.

A total of three strains were computationally identified as strains to be constructed by MAGE cycling. Two strains designed for taxadiene production from glucose and one designed for production from glycerol (Table 11).

Table 11 Information on the three solution sets to be constructed in the laboratory. The maximum biomass yield and minimum taxadiene yield (as a percentage of the theoretical yield) are shown.

	Glucose 1 ("Set 1")	Glucose 2 ("Set 2")	Glycerol ("Set 3")
Genotype	Δ*pntAB*,Δ*pykAF*,Δ*aceA*, Δ*adhP*, Δ*adhE*, Δ*pgk*	Δ*pntAB*,Δ*lpdA*, Δ*aceEF*, Δ*pgk*	Δ*pntAB*,Δ*adhE*,Δ*gapA*, Δ*tktAB*
Maximum Biomass Yield	15.0%	33.3%	45.8%
Minimum Taxadiene Yield	75.4%	46.7%	42.7%

Strain Construction by MAGE

This portion of the study was undertaken at Harvard Medical School. Oligonucleotides were designed to knockout all of the genes described previously and MAGE cycled in separate pools of EcHW2e. 48 clones were picked from each of the three pools to query the knockouts by diagnostic PCR. The number of knockouts were abnormally low as determined by experience (Wang HH, personal communication). The experiment was designed such that at least one of the 48 clones for both "Set 2" and "Set 3" would contain all knockouts. For "Set 1", one strain was identified as having knockouts at Δ*pykF* (pyruvate kinase: pyruvate + ATP ↔ ADP + phosphoenolpyruvate + 2 H^+), Δ*aceA* (isocitrate lyase: isocitrate ↔ glyoxylate + succinate), and Δ*adhP* (ethanol dehydrogenase: acetaldehyde + NADH + H^+ ↔ ethanol + NAD^+) loci. For "Set 2", none of the 48 strains selected had any of the knockouts designed. For "Set 3", multiple colonies had a knockout at the Δ*tktA* locus (transketolase I: D-erythrose-4-phosphate + D-xylulose-5-phosphate ↔ D-fructose-6-phosphate + D-glyceraldehyde-3-phosphate; and D-sedoheptulose-7-phosphate + D-glyceraldehyde-3-phosphate ↔ D-ribose-5-phosphate + D-xylulose-5-phosphate, however, this was the only strain identified. As a result, two strains were generated, called EcHW2e(Δ*pykF*, Δ*aceA*, Δ*adhP*) and EcHW2e(Δ*tktA*), and stored as glycerol stocks with and

without pAC-LYC at −80°C. As the strains constructed do not have the exact genotype as the computationally predicted strains, it is assumed that the product yields predicted by the model do not correlate to the experimentally observed yields.

Shake-Flask Production Cultures

Lycopene production experiments were undertaken in four strains (EcHW2, EcHW2e, EcHW2e(Δ*pykF*, Δ*aceA*, Δ*adhP*), and EcHW2e(Δ*tktA*)) and two medium formulations (production medium with either glucose or glycerol as the principle carbon source). Cultures lasted 72 hr at 22°C and 250 rpm, where sampling occurred at 24 hr and 72 hr. Cell-density was measured spectrophotometrically at 600nm, lycopene was extracted and measured spectrophotometrically at 474nm, and substrates and byproducts were measured by CEX-HPLC-RID.

The cell-densities after both 24 hr and 72 hr were higher when growing on glycerol than that of growing on glucose for all strains except for Δ*tktA* (Figure 30a). Moreover, the cell-densities continued to increase significantly after 24 hr when growing on glycerol, except for the parent strain, EcHW2. This contrasts when growing on glucose, where the cell-densities were the same or slightly lower after 72 hr than they were at 24 hr (Figure 30b).

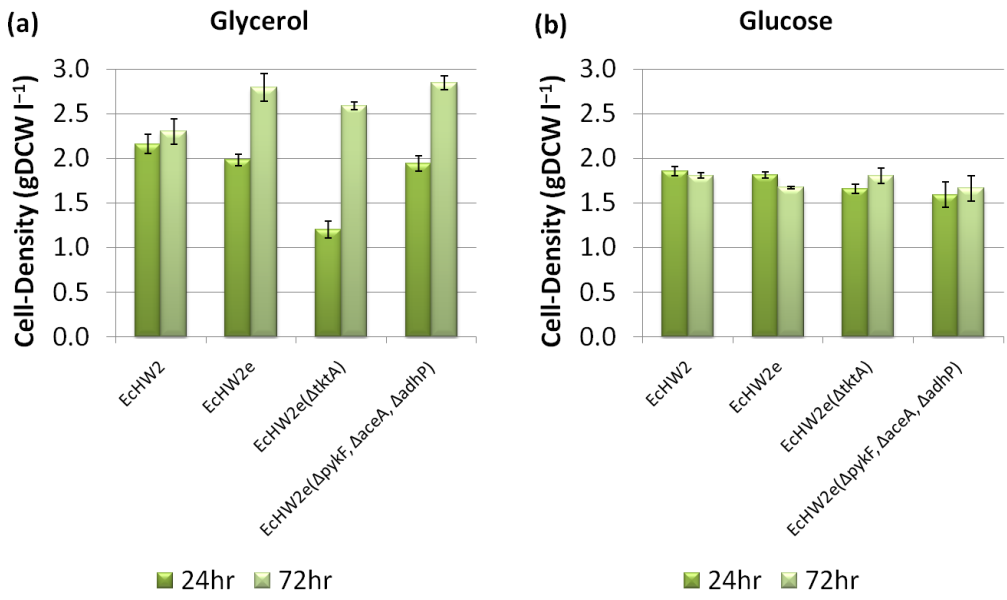

Figure 30 Cell-density of MAGE-constructed strains.
Cell-density of the four strains tested after 24hr and 72hr of culture in production medium supplemented with either (a) glycerol or (b) glucose. Error bars designate ± one standard deviation of three replicates.

Figure 31 shows the lycopene production titer after 24 and 72 hr for the four strains growing on glycerol and glucose. As expected, EcHW2e shows improved lycopene titer (as compared to EcHW2), when growing on both glycerol and glucose, due to the optimized expression of *dxs* and *idi*. Again, the lycopene titer increases significantly after 24 hr when growing on glycerol, but does not increase significantly when growing on glucose, except for EcHW2e(Δ*pykF*, Δ*aceA*, Δ*adhP*). Both of the knockout strains generated show titer improvements when growing on glycerol. After 72 hr, EcHW2e(Δ*pykF*, Δ*aceA*, Δ*adhP*) has a titer of 72.7 ± 1.6 mg l^{-1} (25.5 ± 1.0 mg $gDCW^{-1}$) while EcHW2e(Δ*tktA*) has 62.6 ± 6.2 mg l^{-1} (24.2 ± 2.2 mg $gDCW^{-1}$) as compared to 47.9 ± 2.8 mg l^{-1} (17.2 ± 0.7 mg $gDCW^{-1}$) for EcHW2e. Interestingly, while the general trends in improvement for the four strains hold somewhat constant, the raw titers are significantly lower for growth on glucose than on glycerol. Even when normalized by cell-density, the specific lycopene titers are higher for all strains when growing on glycerol than glucose (Figure 32).

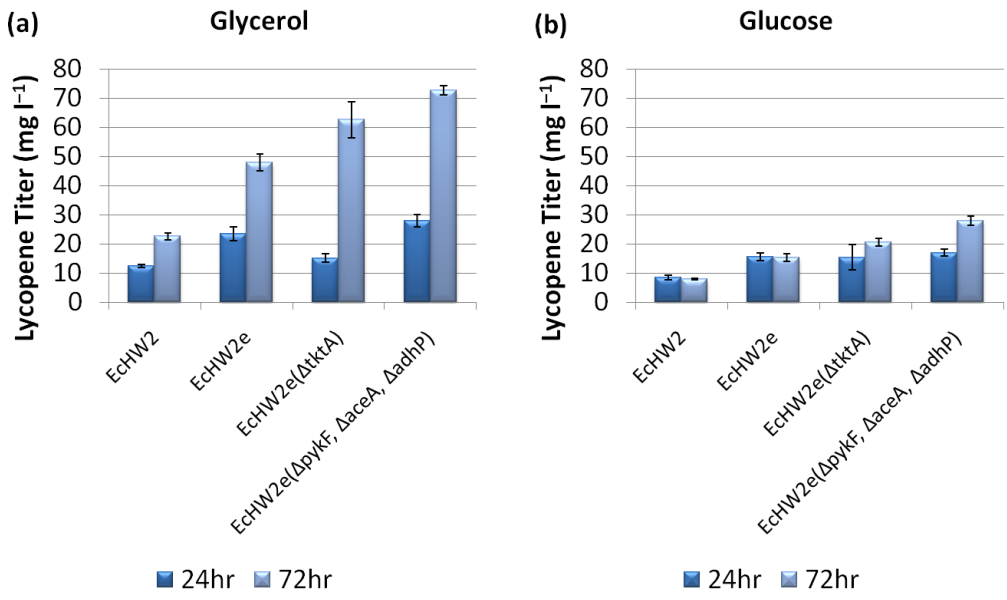

Figure 31 Lycopene titer of the MAGE-constructed strains.
Lycopene titer of the four strains tested after 24hr and 72hr of culture in production medium supplemented with either (a) glycerol or (b) glucose. Error bars designate ± one standard deviation of three replicates.

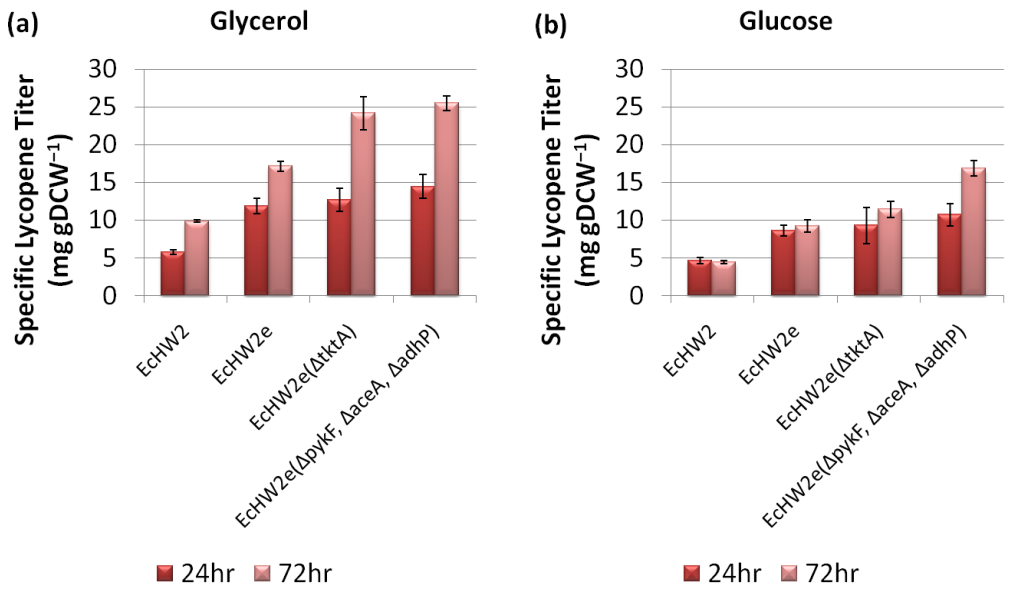

Figure 32 Specific lycopene titer of the MAGE-constructed strains.
Specific lycopene titer of the four strains tested after 24hr and 72hr of culture in production medium supplemented with either (a) glycerol or (b) glucose. Error bars designate ± one standard deviation of three replicates.

At the end of the 72 hr, the principle carbon sources had not yet been depleted (Figure 33). In general, glucose consumption was very similar between the strains, however, varied more for when glycerol was the principle carbon source. While significant amounts (<1 mM) of pyruvate, ethanol, or lactate were not detected in the culture medium after 72 hr of culture, significant amounts of acetate were observed. After 24 hr, the acetate production was between 64% and 268% higher for growth on glucose than on glycerol, while after 72 hr, it was between 68% and 305% higher (Figure 34). A scatter plot of this data (lycopene production versus acetate production) showed that after 24 hr, the ratio of lycopene produced to acetate produced was quite similar (Figure 60, in Appendix). However, after 72 hr, the trajectory for the glycerol cultures was much more favorable to lycopene production, while the trajectory for the glucose cultures was much more favorable to acetate production (Figure 60, in Appendix). In the absence of *pyk*, acetate could be produced from glucose or glycerol by the oxidation of pyruvate (through the action of the *poxB* gene product).

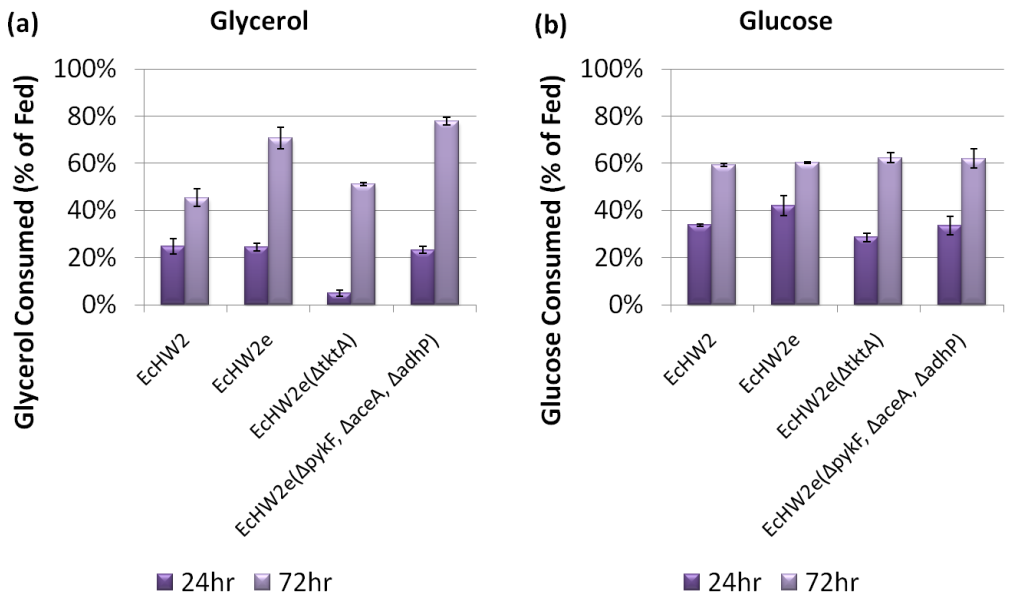

Figure 33 Carbon source consumption of the MAGE-constructed strains.
Carbon source consumption of the four strains tested after 24hr and 72hr of culture in production medium supplemented with either (a) glycerol or (b) glucose. Percentage imported of the amount fed (15 g l^{-1} in each case) is shown. Error bars designate ± one standard deviation of three replicates.

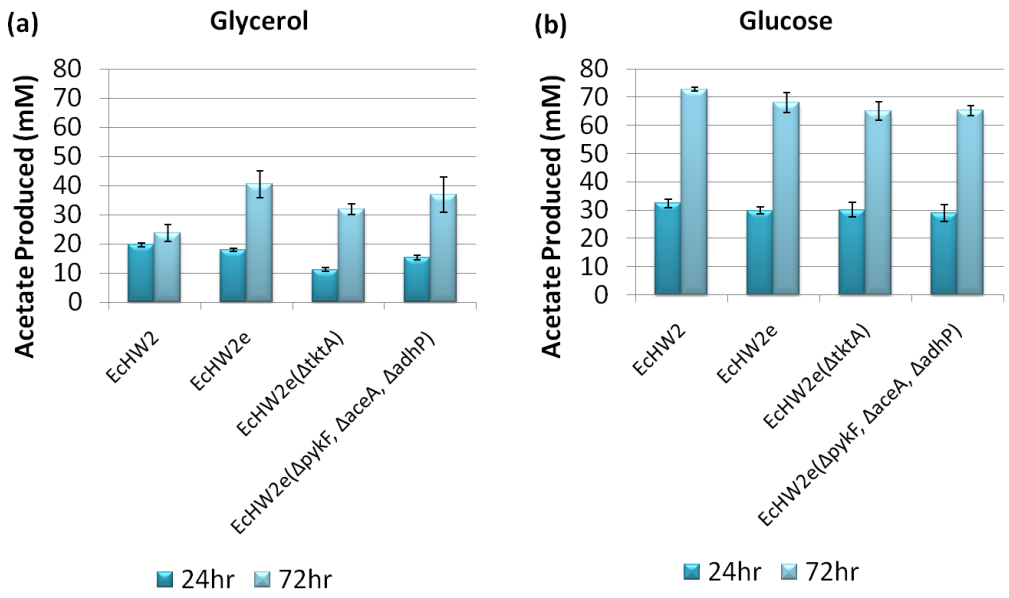

Figure 34 Acetate production of the MAGE-constructed strains.
Acetate production the four strains tested after 24hr and 72hr of culture in production medium supplemented with either (a) glycerol or (b) glucose. Error bars designate ± one standard deviation of three replicates.

Lastly, as a means to show direct comparison with the modeling data generated, the lycopene yields were measured on their respective carbon source. Although, it is clear that the minimum yields predicted (Table 11) were not met, most likely because all of the genetic targets were implemented. For example, EcHW2e(Δ*tktA*) had a yield of 18.9 mg lycopene g glycerol^{-1} (roughly 6% of the theoretical yield), while EcHW2 and EcHW2e had yields of 3.4 mg g^{-1} and 6.4 mg g^{-1}, respectively. While predicted yields were not met, this corresponded to a 2.95-fold improvement in lycopene yield over the already "optimized" EcHW2e strain, owing to the power of using this modeling strategy.

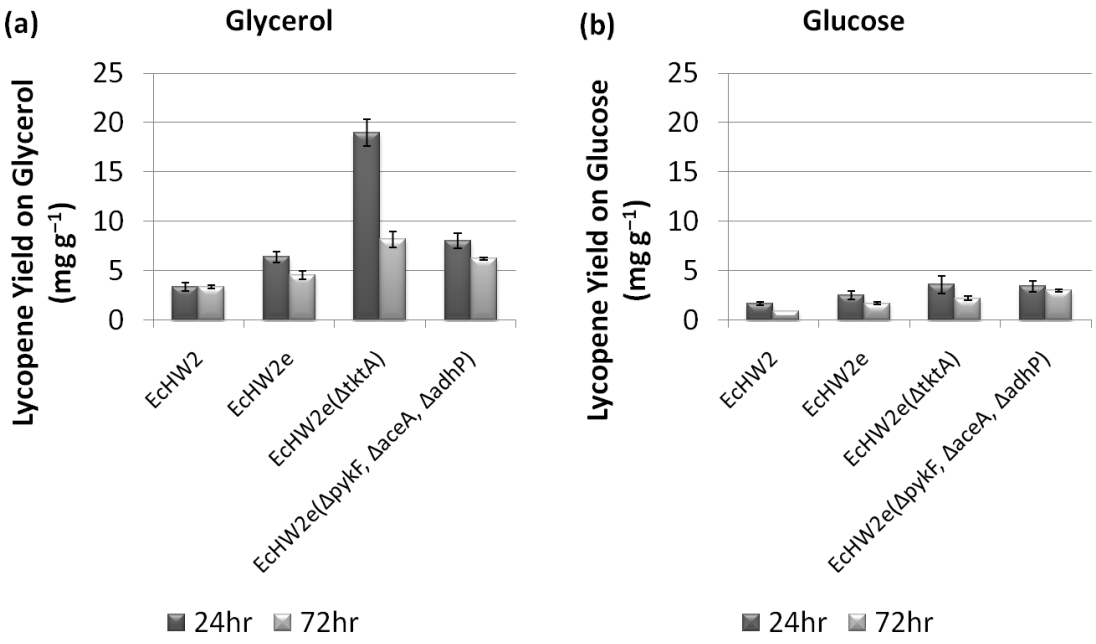

Figure 35 Lycopene yields on carbon sources of the MAGE-constructed strains. Lycopene yields on the carbon sources of the four strains tested after 24hr and 72hr of culture in production medium supplemented with either (a) glycerol or (b) glucose. Error bars designate ± one standard deviation of three replicates.

Microplate Growth Assay

While the production cultures described in the previous section yielded a significant amount of information regarding overall cellular phenotype, high resolution growth profiles could be utilized to further examine the growth phenotypes under different genotypes and environmental conditions. Five strains (EcHW1, EcHW2, EcHW2e, EcHW2e(Δ*pykF*, Δ*aceA*, Δ*adhP*), and EcHW2e(Δ*tktA*)) were grown under four medium formulations and growth parameters were determined by a non-linear least-squares regression of the OD_{600nm} vs. time data to a logistic population model equation. As a frame of comparison, Figure 36 shows the growth profiles of the parent (EcHW1) in all four medium formulations. Figure 61 (in Appendix) shows OD_{600nm} vs. time (average ± one standard devotion, $n = 4$) for the five strains on M9 + glucose. Figure 62, Figure 63, and Figure 64 show the same data for M9 + glycerol, PM + glucose, and PM + glycerol, respectively (PM is the abbreviation for the "production medium" used in previous studies). Figure 37 shows a summary of this data: the specific growth rates as a function of these strains and medium formulations. Only one strain/medium combination showed no significant growth, EcHW2e(Δ*tktA*) in M9 supplemented with glycerol. In some cases, the specific growth rate exceeded 1.0 hr^{-1}. Generally, the specific growth rates were higher for glucose than that for glycerol

and the complex medium showed faster specific growth rates than the minimal medium.

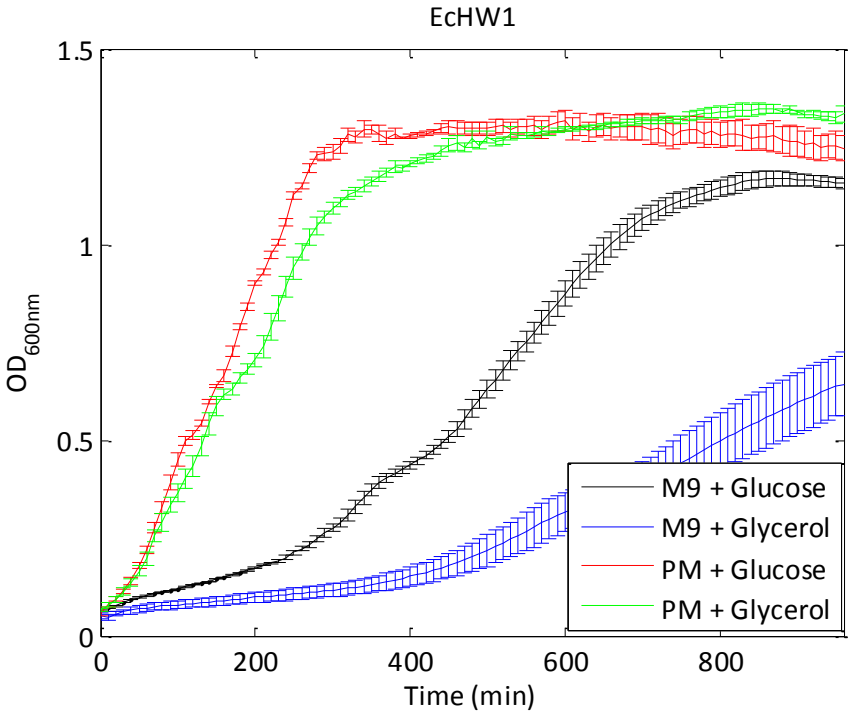

Figure 36 Microplate growth assay.
OD_{600nm} vs. time for EcHW1 growing in 96-well plate format at 30°C. Error bars represent ± one standard deviation of four replicates.

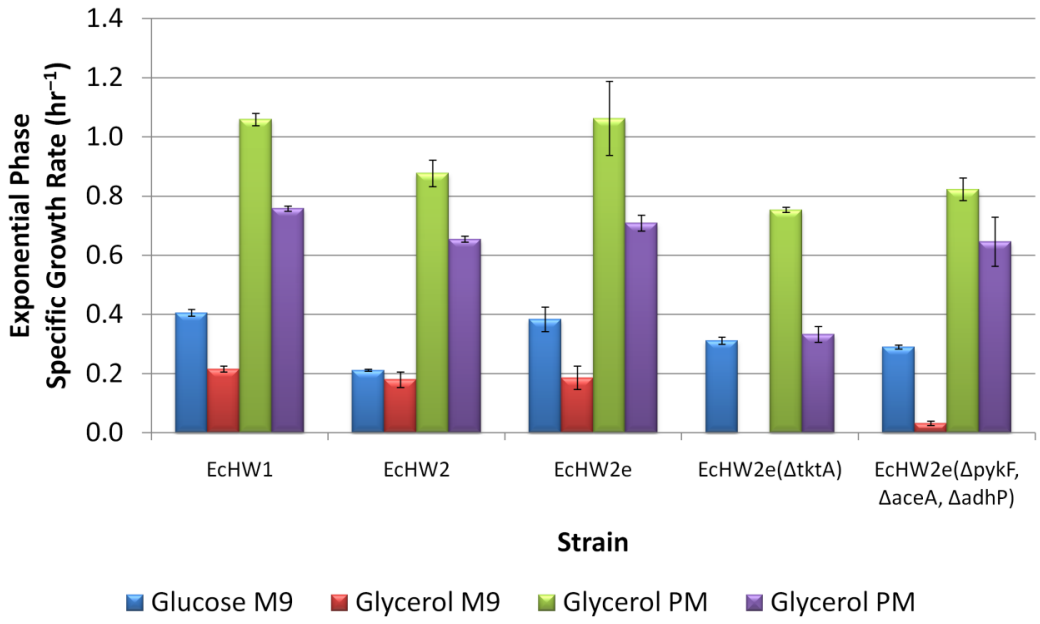

Figure 37 Specific growth rates of MAGE-constructed strains.
The exponential phase specific growth rates at 30°C for the five strains used in this study as determined by a non-linear regression of the OD_{600nm} vs. time data to Equation 11. Error bars represent ± one standard deviation of four replicates.

Discussion

Elementary mode analysis is a powerful tool for pathway decomposition in and topological studies of metabolic networks (de Figueiredo et al. 2009; Papin et al. 2004; Schilling et al. 2000; Schuster et al. 1999), and can be utilized for metabolic engineering (Trinh et al. 2009). Elementary mode analysis has been utilized to design strains of *E. coli* that are efficient at producing biomass from glucose (Trinh et al. 2006), ethanol from five- and six-carbon sugars (Trinh et al. 2008), diapolycopendioic acid from glucose (Unrean et al. 2010), ethanol from glycerol (Trinh and Srienc 2009), and succinate from glycerol (Chen et al. 2010). In two cutting-edge applications, EMA was combined with linear programming to determine flux distributions from external measurements in lysine-producing *Corynebacterium glutamicum* (Gayen and Venkatesh 2006), and to determine the metabolic fluxes of *Lactobacillus rhamnosus* growing on medium containing mixed substrates (Gayen et al. 2007). EMA has also been utilized to determine flux distributions in polyhydroxybutyrate-producing *E. coli*, mediated by a thermodynamic analysis of the EMs (Wlaschin et al. 2006). Due to the issues regarding the applicability of the optimality assumptions of FBA and MoMA in engineered systems, this goal of this chapter was to develop an algorithm for identifying knockout targets based strictly upon network topology.

Randomized mutational strategies have long been employed for "classical strain development", one notable example being penicillin production from *Penicillium chrysogeum* (Demain 2006). During each cycle of classical strain development, a mutation is introduced, screened for a phenotype, mutated again, and so on (known as asexual recursive mutagenesis) (Zhang et al. 2002). Genome shuffling emerged as a tool in which, after mutagenesis and screening, multiple top performers in a progeny would then be shuffled further mixing the genomes and screening. This was applied to *Streptomyces fradiae* for improving tylosin (a complex polyketide antibiotic) production roughly 6-fold in roughly 1 year and 24,000 assays. A similar improvement in titer was undertaken at Eli Lilly over the course of 20 years and requiring roughly 1,000,000 assays (Zhang et al. 2002), therefore demonstrating the power of this strategy (Stephanopoulos 2002). The same methodology was applied towards improving lactic acid tolerance in an unspecified industrial *Lactobacillus* species (Patnaik 2008).

While gene shuffling relies simply on mutation and protoplast fusion, the advent of advanced DNA synthesis and sequencing technologies has provided for a strong underpinning for massively parallel genome *engineering* and evolution. Protein engineering techniques such as directed evolution and gene shuffling have allowed for the randomized, but targeted improvement of a *specific* enzyme

function (such as broadened or reduced substrate specificity, or improved turnover rates). However, it is clear that to truly "maximize" or "optimize" flux through a pathway of interest, one must modulate the expression of multiple genes, both within and outside the pathway, in a time efficient manner. Global transcription machinery engineering (gTME) emerged as an extremely powerful tool to rapidly tune the expression of many genes by using error-prone PCR to mutate a transcription factor, therefore effecting the expression of a large number of genes (rather than one or a few). This technique was first applied to *S. cerevisiae*'s TATA-binding protein (SPT15 and TAF25) for improving tolerance to high glucose and ethanol concentrations in the culture medium (Alper et al. 2006b). This technique was successfully applied to *E. coli*'s *rpoD* gene (encoding for its main sigma factor, σ^{70}) for improving ethanol tolerance, SDS tolerance, and lycopene production (Alper and Stephanopoulos 2007), as well as the α subunit of its RNA polymerase for improving tolerance to short chain alcohols and production of hyaluronic acid and L-tyrosine (Klein-Marcuschamer et al. 2009).

MAGE differs from gTME in that it is: 1) automated such that multiple rounds of mutagenesis/evolution can be undertaken in operational cycles; and 2) it targets deletion or over-expression of *specific* genes. The high efficiency of allelic replacement (up to 30%

under some conditions without selection) allows for the searching of a significant fraction of genomic space. In the end, a central difference between MAGE and gTME is that, with enough rounds of MAGE, the population will converge to a specific genotype (see in the Appendix), whereas gTME should never converge to a single, specific genotype. The question now becomes, is this a good thing? The answer to this question may be reduced to the ability to actually assess a phenotype (or, more importantly, multiple phenotypes). For phenotypes that can be measured in a high-throughput manner, gTME might be a better option to better explore. The convergent trajectory of the fitness landscape generated with MAGE might allow for "low-throughput" screening of mutants generated.

This particular study decided to use elementary mode analysis coupled with a genetic algorithm as a means of more efficiently searching genomic space for targets for improving taxadiene production. As a means of comparison, an exhaustive search of all four-knockout strains would have required evaluating roughly 1.17×10^7 individuals, and assuming an average computation time of 2.7117 s (as determined earlier), would take roughly a year (367.2 days) to compute on the current platform. The genetic algorithm allowed for the identification of a four knockout mutant after evaluating 5,000 individuals, taking roughly one hour. In the end, the utilization of the

genetic algorithm allowed for roughly a five order of magnitude decrease in computation time.

Further, using a high-throughput phenotype (lycopene production, or colony red color) as a surrogate for another phenotype (taxadiene production, a similar isoprenoid molecule) combined with MAGE, these strains were created. Three sets of knockouts were identified: Glucose 1 Δ*pntAB*, Δ*pykAF*, Δ*aceA*, Δ*adhP*, Δ*adhE*, Δ*pgk*, Glucose 2: Δ*pntAB*, Δ*lpdA*, Δ*aceEF*, Δ*pgk*, Glycerol: Δ*pntAB*, Δ*adhE*, Δ*gapA*, Δ*tktAB*. A central component to MAGE is λ-Red recombination, which requires phage Gam, Exo, and Beta (Muniyappa and Radding 1986) functions but does not require *E. coli* RecA function (Ellis et al. 2001). After the utilization of λ-Red recombination to construct precise alterations of the *E. coli* chromosome with PCR products (that is, dsDNA) and a resistance marker, it was discovered that ssDNA could recombine at higher efficiency with even as small as 30 bp chromosomal homology (Ellis et al. 2001). As a result, this was used as the basis for these MAGE experiments.

The clear issue with the study as presented was that the full genotypes were not constructed in EcHW2e using MAGE. One gene in each set (*pgk* for set 1 and 2, and *gapA* for set 3) was determined to be essential in *E. coli* as identified through the process of creating the Keio collection (a genome wide single-gene knockout collection of *E.*

coli). However, these mutants were constructed by selecting on LB plates, as was MAGE cycling. It has been often cited of the conditional dependence of certain genes in many organisms such as *E. coli*, mycobacteria (Sassetti et al. 2001), *Pseudomonas putida* KT2440 (Molina-Henares et al. 2010). For example, a gene might be non-essential under aerobic growth, but might be under anaerobic growth. While these are most often identified by genome-wide single or double gene deletions, they can also be identified computationally (Joyce and Palsson 2008) with excellent accuracy in some cases (Joyce et al. 2006). For example, out of 3,888 single-deletion mutants tested, 119 mutants were unable to grow on glycerol minimal medium. These conditionally essential genes were then evaluated using a genome scale metabolic model and the correct phenotype was identified in approximately 91% of the cases (Joyce et al. 2006).

In the case of *E. coli* here, while *gapA* has been cited to be essential for growth on rich media (as in the Keio collection) (Baba et al. 2006; Yamamoto et al. 2009), it has been verified both computationally and experimentally that it is not essential when glycerol is the principle carbon source (Joyce et al. 2006). This is a simple explanation for why this knockout was not observed to have been constructed, and further exemplifies the importance of context when it comes to randomized mutagenesis and screening. The case for

why the *pgk* knockout was not observed in the glucose strains is a more difficult one to explain. When this gene was attempted to be replaced with a kanamycin resistance gene in strain *E. coli* BW25113 using λ-Red recombination, it could not be constructed. As a result, this group determined to be essential (Baba et al. 2006; Yamamoto et al. 2009). However, shortly before this study was published, another group reported the construction of this mutant in MG1655 (a close parent of BW25113) using an method based on *in vitro* transposition of a modified Tn*5* element (also containing a kanamycin resistance gene) (Kang et al. 2004). This construct was also selected on a rich medium supplemented with kanamycin. The conflicting evidence of the essentiality of this gene promoted a further analysis of this operon and its essentiality. There could be multiple explanations for why this knockout was not found to be present. First, it could be due to subtle differences between the BW25113 and MG1655 genomes. However, strain EcHW2e is a derivative of MG1655 containing the λ-Red genes in the *bioAB* locus (Wang et al. 2009). Second, it could be the directionality or position of the kanamycin resistance gene. The Keio collection aimed to replace the *pgk* gene with a FRT-site flanked kanamycin resistance gene upstream of the start codon and downstream of the stop codon (Baba et al. 2006). The strategy was also used to remove the entire gene from before the start codon to after the

stop codon. Whereas, the modified Tn*5* element will insert itself randomly *within* the gene. The expression of the downstream gene, *fbaA* could be effected by the removal of the upstream gene, *pgk*, which have been shown to be are coordinatively transcribed (Charpentier et al. 1998). Interestingly, *fbaA* was determined to be essential by the Keio group (Baba et al. 2006). As a result, removal of the entire *pgk* gene, which may contain promoter elements for the *fbaA* gene, would be lethal to the cell. To verify that these could really not be constructed, they should have been attempted to be constructed individually and not with other gene targets.

Even though the MAGE cycling in this study was not able to fully generate the computationally designed strains, the results most definitely support the utility of this computational method for designing strains for metabolite over-production. Both MAGE-generated strains improved lycopene production titers and yields in both glycerol- and glucose-based complex medium. Figure 38 shows the comparison between the modeling results and the experimental results. It can be clearly seen that, without constructing all of the knockouts, lycopene (or taxadiene) production cannot be fully coupled to biomass production. As a result, the overall yields are quite low: 18.9 mg g^{-1} and 3.4 mg g^{-1} for the glycerol and glucose designed

strains, respectively. These correspond to 6.1% and 1.1% of the theoretical yield of lycopene on the respective carbon sources.

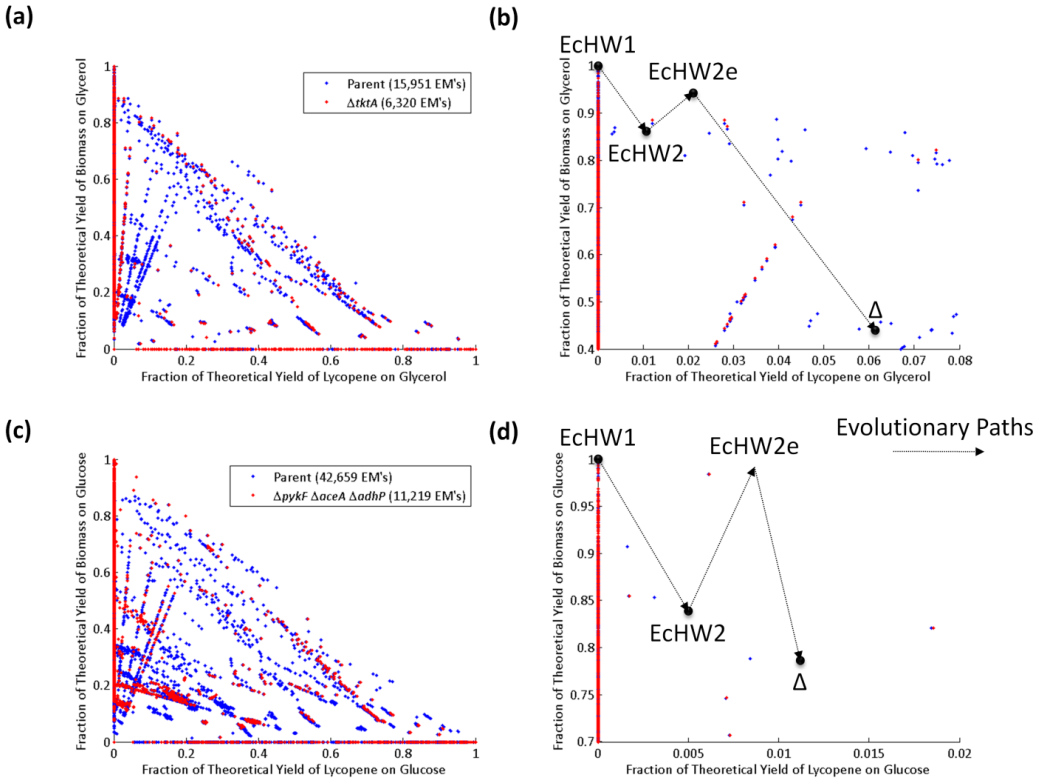

Figure 38 Comparison of the modeling results with the experimental results. (a) A scatter plot of the lycopene model and their corresponding yields of biomass on glycerol, and lycopene on glycerol (normalized to the fraction of the theoretical yield). Shown is the parent model in blue points and the constructed Δ*tktA* model in red points. (b) A zoomed-in portion of (a) and the experimental biomass and lycopene yields of the MAGE-constructed strains. (c) A scatter plot of the lycopene model and their corresponding yields of biomass on glucose, and lycopene on glucose (normalized to the fraction of the theoretical yield). Shown is the parent model in blue points and the constructed Δ*pykF* Δ*aceA* Δ*adhP* model in red points. (d) A zoomed-in portion of (c) and the experimental biomass and lycopene yields of the MAGE-constructed strains. Δ indicates the MAGE constructed strains and the dotted, arrowed lines indicates the evolutionary paths of these strains.

A few significant differences between the glycerol- and glucose-based media were observed across all strains: 1) cell-densities were consistently lower for growth on glucose, 2) lycopene production was significantly higher for growth on glycerol, and 3) acetate production was significantly higher for growth on glucose. Acetate is the main fermentative product of *E. coli*, and it's reduction has been a target for improving recombinant protein production (Wong et al. 2008) and succinate production (Jantama et al. 2008), of many others. It has been recently shown that controlling glycerol supplementation in a fed-batch fermentation of taxadiene production in *E. coli* could reduce acetate levels to less than 1 g l^{-1} and dramatically improve taxadiene production (Ajikumar et al. 2010). Previous studies have shown that exogenous feeding of pyruvate (one of the two precursors for the non-mevalonate isoprenoid pathway) in addition to glycerol, did not improve lycopene production in *E. coli* (Farmer and Liao 2001). This was also observed when feeding pyruvate in addition to glycerol for taxadiene production in *E. coli* (see Chapter 8). Moreover, over-expression of the gluconeogenic PEP carboxykinase (*pck*) improved lycopene titer, while over-expression of the glycolytic PEP carboxylase (*ppc*) had the opposite effect. These all support the utilization of a gluconeogenic carbon source for production of isoprenoid natural products, likely enabling the balancing of pyruvate and

glyceraldehydes-3-phopshate levels, and decreasing overflow metabolism and therefore acetate production.

Chapter 6 – Identification of over-expression targets using optimization, and experimental implementation for improving taxadiene production

Introduction

Two previous chapters of this dissertation have been devoted to using stoichiometric modeling to identify gene knockouts to improve product titer. This chapter presents the development and application of an algorithm for identifying over-expression targets to improve product titer. For identifying knockout targets, the binary nature of the problem (the gene and therefore reaction either exists or does not exist) simplifies its mathematical abstraction. For identifying over-expression targets, the problem is now non-binary, meaning that describing gene over-expression in this context is more difficult. Over-expressing a gene on a five-copy plasmid does not necessarily correlate to five times the amount of transcript, which does not necessarily correlate to five times the amount of soluble protein, which does not necessarily correlate to five times the amount of metabolic flux through that reaction. Nonetheless, this chapter aims to formulate a mathematical abstraction for modeling gene over-expression, as well as utilize it to simulate metabolic fluxes and identify over-expression targets for improving heterologous isoprenoid titer. Taxadiene was

selected for the over-expression study because of the low titers currently observed (see Chapter 7).

Materials & Methods

Model Construction

As was the case in Chapter 2, the *E. coli* genome-scale metabolic model *i*AF1260 was used as a base for the model used in this study (Feist et al. 2007). This model contains 2,077 reactions, 1,039 metabolites, and 1,261 genes (Feist et al. 2007). Reactions then had to be added to account for the reactions catalyzed by the two heterologous enzymes introduced to produce taxadiene through *E. coli*. These reactions are: 1) a geranylgeranyl-diphosphate synthase (to catalyze: farnesyl-diphosphate + isopentenyl-diphosphate → geranylgeranyl-diphosphate + diphosphate), 2) a cyclizing taxadiene synthase (to catalyze: geranylgeranyl-diphosphate → taxa-4,11-diene + diphosphate), and 3) a taxadiene transport reaction (taxadiene → [nothing]) (Ajikumar et al. 2008; Ajikumar et al. 2010).

Calculations were made in MATLAB® 7.4 (Mathworks Inc.; Natick, MA) utilizing the SMBL Toolbox (version 2.0.2, http://sbml.org/software/sbmltoolbox/) (Keating et al. 2006; Schmidt and Jirstrand 2006) and the COBRA Toolbox (version 1.3.3, http://gcrg.ucsd.edu/) (Becker et al. 2007). Optimization was undertaken using the CPLEX (version 11.0) algorithm of the

TOMLAB™ Optimization Environment (TOMLAB™/CPLEX) interfaced with the COBRA Toolbox and MATLAB® 7.4.

Over-Expression Target Identification

The over-expression algorithm involves: 1) imposing a taxadiene production rate (as determined experimentally), 2) solving a FBA problem, 3) imposing an amplification in individual reaction fluxes (to simulate the effect of gene over-expression, 4) solving a MoMA problem, and 5) identifying over-expressions that led to a phenotype fraction value, f_{ph}, greater than unity (an overflow of this algorithm can be seen in Figure 39). The overall algorithm is very similar to the knockout identification section in Chapter 2, however, instead of setting a reaction value to zero, individual reaction values are amplified five-fold. Steps "3" and "4" were iterated for every reaction within the network.

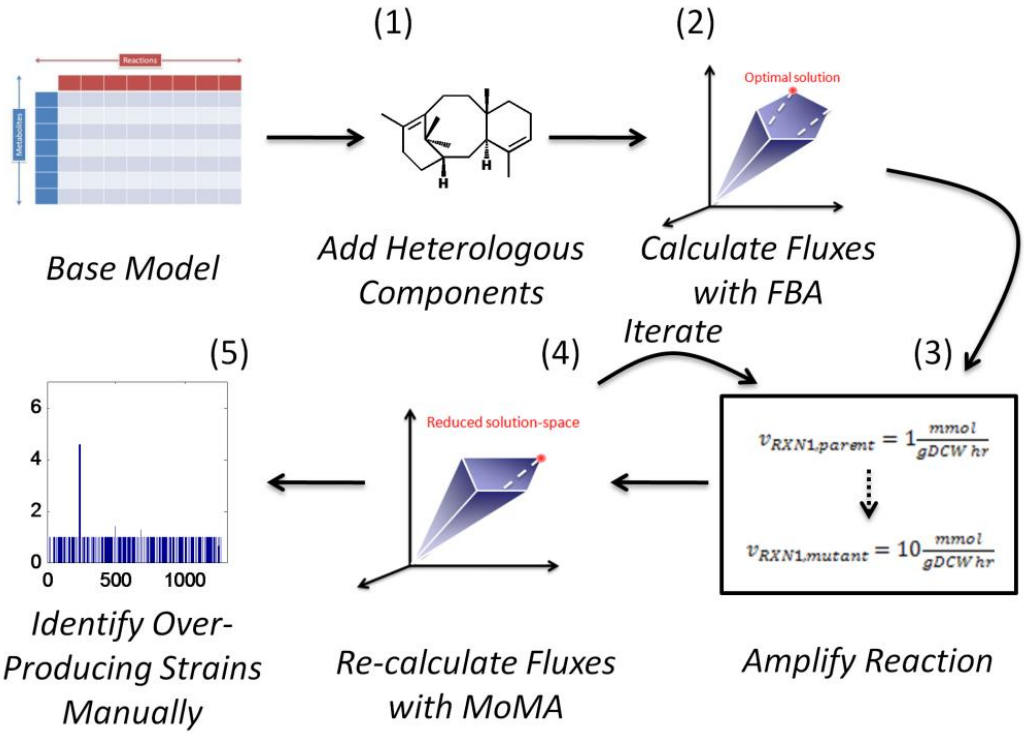

Figure 39 An overview of the proposed algorithm for identifying over-expression targets to improve product titer.

As was conducted in Chapter 3, calculations were made under conditions to simulate complex medium. Medium composition can be approximated by setting uptake rates of specific chemical components known to exist in the medium of interest. The "computational complex medium" contained all twenty naturally-occurring amino acids (L-isomers) (Oh et al. 2007). The lower bounds of the amino acid transport reactions were set to –0.1 mmol $gDCW^{-1}$ hr^{-1} (negative sign indicates metabolite uptake into the cell) and were chosen based upon previous literature values and because they satisfied the relative biomass differences experimentally observed between media (Oh et al. 2007; Selvarasu et al. 2009a). Glycerol transport rates were set to –3.0 mmol $gDCW^{-1}$ hr^{-1}, as in Chapter 3.

Strains & Plasmids

Strain YW22(pTrcHis2B-TXS-GGPPs) was used as the base strain for taxadiene production in this study (see Chapter 7). This strain is a derivative of JM109(DE3) containing a $T7_{prom}$-*dxs*-*idi*-*ispB*-*ispDF*-$T7_{term}$ operon in the *araA* of its chromosome. Plasmid pTrcHis2B-TXS-GGPPs is a carbenicillin resistant plasmid with a pBR322 origin of replication, containing synthetic *txs* and *crtE* genes under the control of an IPTG-inducible Trc promoter (also described in Chapter 7).

Plasmids containing over-expression targets were obtained from the ASKA(-) library of the National Institute of Genetics in Japan, containing each gene of interest as identified by the algorithm (and controls) (Kitagawa et al. 2005). The genes are cloned into a derivative of pQE31 (called pCA24N) under the control of an IPTG-inducible T5 promoter and with an N-terminal 6× histidine tag (Kitagawa et al. 2005). These plasmids are chloramphenicol resistant and have a ColE1 origin of replication, and are therefore compatible with YW22(pTrcHis2B-TXS-GGPPs).

Small-Scale Production Cultures

A stab of glycerol stock was inoculated into 2 ml LB medium with appropriate antibiotics and grown overnight at 37°C and 250 rpm. For production cultures, 3 ml production medium (5 g l^{-1} yeast extract, 10 g l^{-1} tryptone, 10 g l^{-1} sodium chloride, 15 g l^{-1} glycerol, 3 ml l^{-1} 50% (v v^{-1}) Antifoam B, 100 mM HEPES, and were adjusted to pH 7.60 with 5 M sodium hydroxide) was inoculated into 16 × 100 mm culture tubes with the precultures to an OD_{600nm} = 0.1. These production cultures were grown for 120 hr at 22°C and 250 rpm. At the end of the culture period, cell-density was measured spectrophotometrically at 600 nm and a single, 1 ml aliquot was stored at –20°C for subsequent analyses. When needed, antibiotics were supplemented at concentrations of 100

mg l^{-1} for carbenicillin, 34 mg l^{-1} for chloramphenicol and IPTG was supplemented at a concentration of 100 µM.

Taxadiene Quantification

For taxadiene quantification, a culture aliquot (750 µl) was supplemented with (-)-*trans*-caryophyllene (TC) at a final concentration of 1 µg l^{-1} to serve as an internal standard for quantification. The samples were then extracted with an equal volume of hexane, followed by 20 s of vortexing and centrifugation for 10 min at 10,000 × g. The hexane layer (150 µl) was removed and stored in glass vials at –20°C until analysis with gas chromatography-mass spectroscopy (GC-MS) could be conducted.

Samples were analyzed on a Shimadzu QP5050A GC-MS using splitless injection. Gas chromatography was run on a non-polar Rxi®-XLB column (30 m x 0.25 mm ID, 0.25µm). The inlet pressure for the column was set at 120 kPa and column flow velocity was 1.6 ml min^{-1}. The flow rate of the ultra high purity helium carrier gas was 20 ml min^{-1}. Temperature of the column was initially set and maintained at 100°C for 2 min and was then increased to 235°C at a rate of 15.0°C min^{-1}. The column was then maintained at this temperature for 1 min. Mass spectrometry was performed in Single Ion Monitoring (SIM) mode scanning for mass to charge ratios of 107 m z^{-1}, 122 m z^{-1}, and 272 m z^{-1}, corresponding to principle daughter ions and parent ion of

taxadiene, respectively (Koepp et al. 1995). Under these conditions, TC and taxadiene eluted at approximately 6.8 min and 11.4 min, respectively. Quantification of taxadiene was accomplished based on a calibration curve of taxadiene concentration (kindly provided by Drs. Ajikumar Parayil and Gregory Stephanopoulos) versus the peak area ratio of TC to taxadiene.

Results

Over-Expression Target Identification

The first demonstration of the over-expression algorithm was undertaken in glycerol-based complex medium (to mimic the "production medium" used in previous chapters). The experimental specific production rate of YW22(pTrcHis2B-TXS-GGPPs) was determined to be 1.53×10^{-4} mmol $gDCW^{-1}$ hr^{-1} and was set as the lower bound for taxadiene transport flux. The specific uptake rate of glycerol was 3.0 mmol $gDCW^{-1}$ hr^{-1} while the specific uptake rates of all twenty L-amino acids were set to 0.1 mmol $gDCW^{-1}$ hr^{-1}. Under these carbon limited conditions, the specific growth rate was determined by FBA to be 0.2671 hr^{-1}. Next, for all reactions that had a non-zero flux value in the FBA simulations, were over-expressed computationally. Once all reactions had been cycled, reactions that produced an f_{ph} value of greater than one were chosen as potential

over-expression targets. Figure 40 shows all of these genes and their corresponding f_{ph} values.

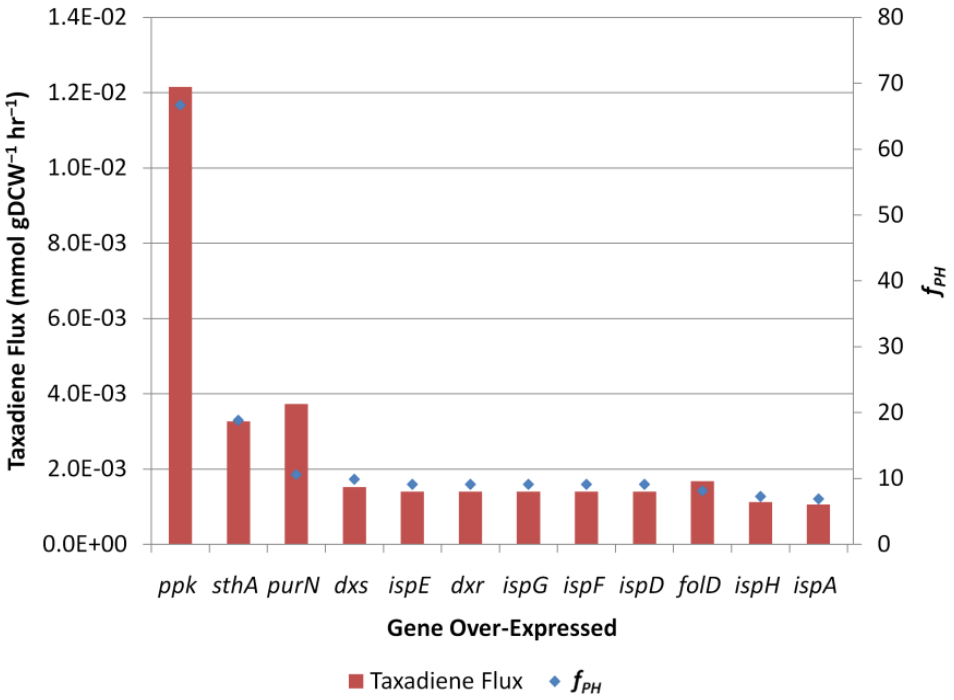

Figure 40 Calculated taxadiene production flux (left *y*-axis) and f_{PH} (right *y*-axis) as a function of gene over-expressed.

Of the twelve targets identified, four of them were outside of the isoprenoid biosynthetic pathway (*ppk*, *sthA*, *purN*, and *folD*). The other eight (*dxs*, *ispE*, *dxr*, *ispG*, *ispF*, *ispD*, *ispH*, and *ispA*) were within the biosynthetic pathway (Figure 3). While these targets produced an f_{ph} greater than one, it was expected that amplifying a reaction within the linear isoprenoid biosynthetic pathway would improve taxadiene flux, and therefore provided an internal control for verification of the algorithm itself. As a result, the four targets for experimental implementation were *ppk*, *sthA*, *purN*, and *folD*, as summarized in Table 12.

Table 12 Genetic targets identified by the over-expression algorithm to implement in the laboratory.
*** indicates that the *folD* gene product is a bifunctional enzyme. The net reaction is shown below.**

Gene	Reaction Name	Reaction	Predicted f_{ph}
ppk	Polyphosphate kinase	ATP + P_i ↔ ADP + PP_i	66.7
sthA	Pyridine nucleotide transhydrogenase	NAD^+ + NADPH ↔ NADH + $NADP^+$	18.9
purN	Phosphoribosylglycinamide formyltransferase	5-phospho-ribosyl-glycineamide + 10-formyl-tetrahydrofolate ↔ 5'-phosphoribosyl-*N*-formylglycineamide + tetrahydrofolate + 3 H^+	10.6
folD	5,10-methylene-tetrahydrofolate dehydrogenase / cyclohydrolase*	$NADP^+$ + 5,10-methylene-THF + H_2O ↔ NADPH + 10-formyl-tetrahydrofolate + H^+	8.1

As another means of implementing a control for the algorithm, the same algorithm was run with glucose as the principle carbon source instead of glycerol. Targets that were identified in the glucose case that were not identified in the glycerol case were also included in the analysis (determined to be *fumA*, *fumB*, *fumC*, and *mdh*). Experimentally, these targets should not be able to improve titer.

Experimental Implementation

The four positive targets (*ppk*, *sthA*, *purN*, and *folD*), the four algorithm controls (*fumA*, *fumB*, *fumC*, and *mdh*) were combined with two experimental controls encoding for only a subunit of a functional enzyme (*sucC* and *pntA*) to ensure that other experimental factors associated with plasmid-based over-expression inadvertently improved taxadiene titer. The genes were over-expressed using a T5 promoter from a chloramphenicol-resistant ColE1-based plasmid in YW22 also co-expressing GGPPs and TXS to produce taxadiene. At the end of the culture period (120 hr), taxadiene was extracted and quantified using GC-MS. Cell-density was also measured spectrophotometrically at 600 nm. Another aliquot of the culture was stored for SDS-PAGE analysis to verify gene-expression of TXS, GGPPs, and the other gene. Raw taxadiene titer was normalized by cell-density, creating a specific taxadiene titer (Figure 41).

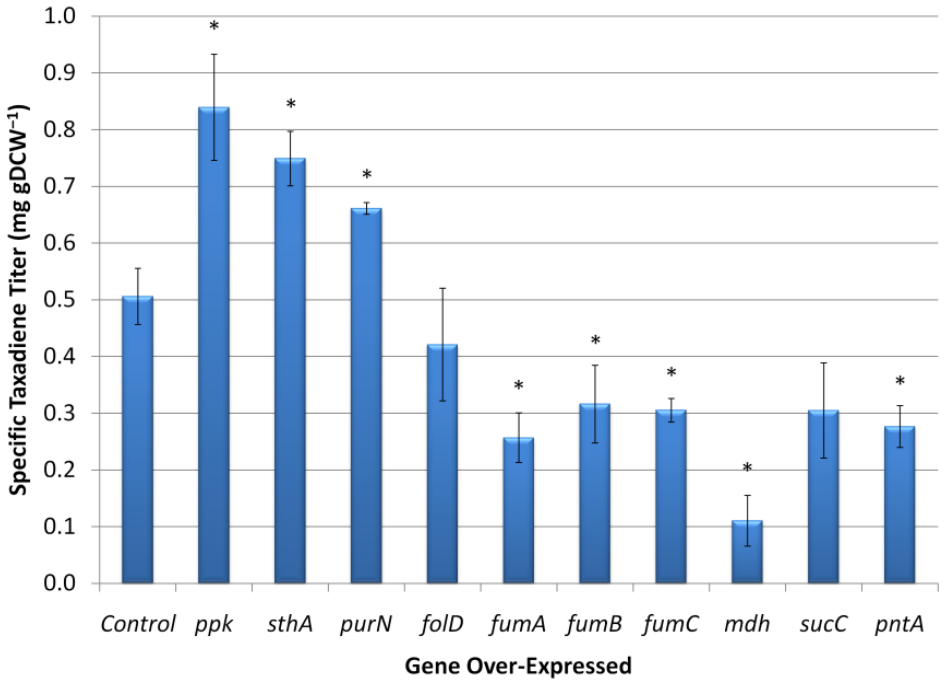

Figure 41 Specific taxadiene titer data for the computationally identified gene over-expression targets and corresponding controls.
The "Control" sample is YW22(pTrcHis2B-TXS-GGPPs). Error bars represent ± one standard deviation of three independent replicates. * indicates a statistically significant ($p < 0.05$, as determined by paired Student's *t*-test) difference from the YW22(pTrcHis2B-TXS-GGPPs) control.

Three of the four (*ppk*, *sthA*, *purN*) experimental targets improved titer ($p < 0.05$ when compared to YW22(pTrcHis2B-TXS-GGPPs) control), however, the improvements were well below what was predicted by the algorithm. The over-expression of *ppk*, *sthA*, and *purN* improved specific production of taxadiene 1.66-, 1.48-, and 1.31-fold, respectively. The fourth target, *folD*, appears to decrease titer but the difference is statistically insignificant from the control ($p = 0.400$). Except for *folD*, showed the same trend of improvement as predicted by the model (*ppk* > *sthA* > *purN*). The three fumarase isozymes (*fumA*, *fumB*, and *fumC*) all decreased titer, while *mdh* decreased titer even more significantly. For the two experimental controls, over-expression of *sucC* had no effect on specific taxadiene titer ($p = 0.119$), while over-expression of *pntA* decreased specific taxadiene titer to roughly half of the control ($p = 0.010$). As a result, all six of the chosen computational and experimental controls were verified as having no positive effect on taxadiene titer. The over-expression of the control genes likely increases metabolic burden to the host, which may partially account for the decrease in taxadiene production. While this is certainly not an exhaustive analysis of potential control targets, no previous knowledge was used to inherently bias the choice of these targets.

Gene-expression was verified in the soluble portion of the whole cell protein by SDS-PAGE (Figure 42). The GGPPs is quite strong is all

lanes, while TXS is faintly visible slightly under the 97.6 kDa molecular weight marker. While the majority of the proteins can be easily visualized, it is questionable whether FumC is visible on the gel. Although all of the genes were over-expressed on the same copy number plasmid and using the same promoter, their expression level did vary, perhaps due to codon bias or gene/protein size.

Lane	Protein	Size (kDa)
1	Control (TXS GGPPs)	91.8 32.1
2	Ppk	80.4
3	SthA	51.6
4	PurN	23.2
5	FolD	31.0
6	Mdh	32.3
7	FumC	50.5
8	SucC	41.4
9	PntA	54.6

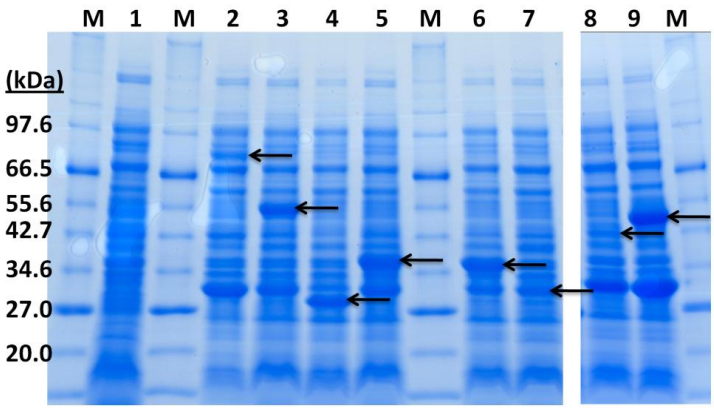

Figure 42 Gene-expression is qualitatively visualized by SDS-PAGE. The proteins and their size are shown in the table to the left.

Over-Expression of the Isoprenoid Biosynthetic Genes

The lack of dramatic improvement in specific taxadiene titer prompted us to investigate the role of a potential bottleneck in the isoprenoids biosynthetic pathway. It has been shown that there is an incredibly complex response (isoprenoids production titer) when it comes to modulating the expression of the upstream and downstream portions of the isoprenoid biosynthetic pathway (Ajikumar et al. 2010). Even though YW22 already contains over-expressed versions of the *dxs*, *idi*, *ispB*, *ispD*, and *ispF* genes from its chromosome, higher levels of some of these or other isoprenoid pathway genes might be needed to fully access potential improved precursor flux and debottleneck this pathway. As a result, all of the upstream pathway genes were over-expressed, individually, from the same plasmid. Six of these genes produced less taxadiene: *dxs*, *ispE*, *ispF*, *ispG*, *ispH*, and *ispA* ($p < 0.05$) Oddly, the *dxr* plasmid could not be stably transformed (multiple trials) into YW22(pTrcHis2B-TXS-GGPPs). Over-expression of *ispD* did not produce any taxadiene, but did produce a viable strain. One target, *idi*, improved specific production of taxadiene 3.77-fold ($p = 0.008$) (Figure 43).

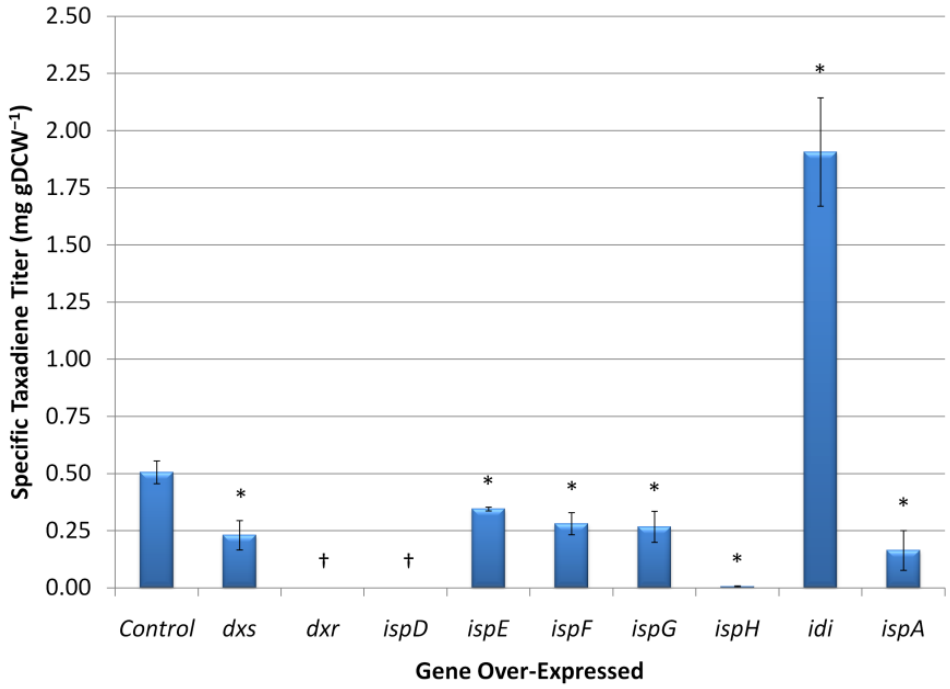

Figure 43 Specific taxadiene titer data for the isoprenoid pathway gene over-expression targets.
The "Control" sample is YW22(pTrcHis2B-TXS-GGPPs). Error bars represent ± one standard deviation of three independent replicates. * indicates a statistically significant ($p < 0.05$, as determined by paired Student's *t*-test) difference from the YW22(pTrcHis2B-TXS-GGPPs) control. † indicates that taxadiene was not detected by GC-MS.

Discussion

The use of systematic methods for identifying over-expression targets, at the start of this study, had only been explored once theoretically (Pharkya and Maranas 2006) and never been implemented experimentally. The goal of this chapter was to use a previously developed optimization strategy and modify it for usage in identifying over-expression targets, and implement it in the laboratory for improving taxadiene production. A variation of the MoMA algorithm was used as an extension to indentifying gene over-expression targets to improve a product titer. As cited in Chapter 3, MoMA was proposed as an alternative to FBA as a means of quantifying metabolic fluxes in networks that had been perturbed, originally by gene knockouts. In a similar sense, forced over-expression of a particular gene could also be considered a genetic perturbation, so this computational framework was extended for identifying genetic over-expression targets. Unfortunately, there is almost no experimental data available on global flux distributions upon over-expression of single genes in *E. coli*, so there was no means of verifying this algorithm short of testing it in our case study of improving taxadiene production.

Four targets were identified by the algorithm as candidates to implement in the laboratory. While an in depth examination of the

physiological effects of over-expression of these four targets was outside of the scope of this chapter's work, it would appear that all four targets improve cofactor availability. The over-expression of *ppk* allows for the reversible generation of ATP and P_i from ADP and the PP_i generated by the IspA and GGPPS reactions in "elongation and cyclization" pathway of isoprenoid biosynthesis. The next three targets are involved with improving NADPH supply for isoprenoid biosynthesis. Transhydrogenases are responsible for the reversible NAD^+ + NADPH ↔ $NADP^+$ NADH reaction, and can therefore theoretically be used to modulate the level of reduction within the cell. Over-expression of *sthA* led to improved production of poly(3-hydroxybutyrate) in *Escherichia coli* (requiring NADPH reducing equivalents) (Sanchez et al. 2006a). The same function appears to be the case here, as one molecule of taxadiene requires four molecules of NADPH (at the DXR reaction step). Lastly, *purN* and *folD* is a two-step linear pathway in tetrahydrofolate biosynthesis. While it appears that tetrahydrofolate has no immediate metabolic relation to the isoprenoid biosynthetic pathway, this step also produces an equivalent of NADPH. Moreover, because this is a linear pathway, perhaps both genes would need to be over-expressed in series to observe the full benefits of this pathway.

Nonetheless, of the four targets identified, three of the four improved specific production titer. All six controls implemented showed either no change or a decrease in specific taxadiene titer. However, the improvements did not exceed 2-fold, and failed to reproduce the flux improvements as predicted by the algorithm. At the same time, all of the controls failed to improve titer. It would be advantageous to further investigate some factors that might improve the predictability of this algorithm. Particularly, the amplification factor could be varied to observe whether other targets appears at different flux levels. For better validation of this algorithm, strains with various single over-expressions should be cultured under different carbon sources. Then, ^{13}C-MFA should be used to quantify fluxes in central metabolism and observe how well the predicted fluxes compare with the measured fluxes under these varying environmental and genetic conditions.

It has been shown and readily recognized that the first step in the DXP-based isoprenoid biosynthetic pathway (Figure 3) is the rate-limiting step in the pathway (Begley et al. 1999; Lawhorn et al. 2004; Matthews and Wurtzel 2000). Bacterial D-1-deoxyxylulose 5-phosphate synthase (encoded by *dxs*), catalyzing the condensation of pyruvate and D-glyceraldehyde-3-phosphate to D-1-deoxyxylulose 5-phosphate (requiring thiamine diphosphate as a cofactor and Mg^{2+} as a metal

adduct) has a turnover rate of 1.9 s^{-1} (the reported value for the *E. coli* enzyme has not been reported) (Eubanks and Poulter 2003). While that of the next step in the pathway, the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (encoded by *dxr*) has been reported to be between 29 s^{-1} and 38 s^{-1} for *E. coli* (Fox and Poulter 2005a; Fox and Poulter 2005b) (a k_{cat}/K_M value of 2.2×10^7 M^{-1} s^{-1}), providing *in vitro* evidence that DXS activity may bottleneck the pathway from the beginning. *In vivo*, it has been shown in numerous cases that over-expression of *dxs* improves isoprenoid or carotenoid titer in native and heterologous hosts (Alper et al. 2005b; Alper et al. 2005c; Brown et al. 2010; Chiang et al. 2008; Choi et al. 2009; Leonard et al. 2010; Morrone et al. 2010; Tyo et al. 2009; Yuan et al. 2006). However, the lack of effect in this case could be due to the fact that increased expression of one rate-limiting enzyme (*dxs*) may result in shifting the control of pathway flux to another enzyme in the pathway (perhaps *idi*) (Kacser and Burns 1973). While it seems odd that an isomerization reaction would be limiting the pathway, the turnover number for IDI has been reported to be 0.33 s^{-1} (also having a much smaller k_{cat}/K_M value of 4.2×10^4 M^{-1} s^{-1}) (Hahn et al. 1999). A highly complex phenotype has been observed recently with respect to the downstream and upstream isoprenoid biosynthetic pathways (Ajikumar et al. 2010), but it appears that IDI could be the rate-limiting step after even both *dxs* and *dxr*

over-expression. As a result, over-expression of the other genes in this pathway would only lead to further unbalancing of this pathway, and lead to decreased or no taxadiene production.

During the course of this dissertation research, an algorithm was developed for identifying over-expression targets and applied towards heterologous lycopene production in *E. coli* (Choi et al. 2010). This algorithm, called Flux Scanning based on Enforced Objective Flux (FSEOF), is based on simulating cellular metabolism by maximizing biomass formation (utilizing FBA), and then imposing specific production rates of a product of interest. Fluxes that increase through a reaction step as the product flux increases are considered to be over-expression targets. Interestingly, both fumarase (encoded by the homologous *fumA*, *fumB*, and *fumC* genes) and malate dehydrogenase (encoded by *mdh*) were identified as targets in this method for producing lycopene from glucose, as were identified as targets for producing taxadiene from glucose in this chapter. Numerous targets were implemented in the laboratory (*pfkA*, *pgi*, *fbaA*, *tpiA*, *icdA*, and *mdh*), however, only with plasmid-based *dxs* and *idi* over-expression as well. Three of the six targets (*pfkA*, *pgi*, and *icdA*) decreased or did not change lycopene titer, while the other three (*fbaA*, *tpiA*, *mdh*) improved titer between 3- and 4-fold. The top over-expression only (*dxs*, *idi*, and *mdh*) strain improved titer from 2.52 mg l^{-1} to 12.85 mg

l^{-1}. Upon combining with MoMA deletion targets, this titer increased to 26.77 mg l^{-1} (with a Δ*lacI*, Δ*gdhA*, and Δ*gpmB* strain) (Choi et al. 2010).

This algorithm has been patented (and therefore the code is unavailable), such that these two methods could not be directly compared, however, there is significant evidence that could be both be used to accomplish the same task. Computationally, this strategy can be much less intensive than our method described, as it allows for the identification of multiple gene targets through solving only a small number of linear optimization problems. This method also has the advantage of not imposing an artificial amplification factor that corresponds to every reaction, allowing for the actual flux values to dictate this factor, which can change for enzyme to enzyme.

Chapter 7 – Multi-scale engineering of taxadiene biosynthesis

Introduction

This chapter presents similar heuristic engineering of taxadiene biosynthesis in *E. coli*, in a similar manner as Chapter 4 was for 6-dEB biosynthesis. In the broader context of heterologous isoprenoid production, more recent work has focused on engineering and optimizing the precursor pathways that support biosynthesis (Martin et al. 2003; Yuan et al. 2006). Of particular relevance to the present study, the DXP pathway (Figure 3) native to *E. coli* was engineered to boost production of the carotenoid compound lycopene (Yuan et al. 2006). More specifically, the *dxs*, *idi*, and various *isp* genes of the DXP pathway were placed under inducible T5 promoters with concomitant improvement in lycopene biosynthesis (Alper et al. 2005c; Jin and Stephanopoulos 2007). Further improvement was observed in a Δ*gdhA*, Δ*aceE*, and Δ*fdhF* mutant genotype as identified using stoichiometric modeling (Alper et al. 2005b). High-cell density bioreactor cultivations resulted in 220 mg l^{-1} lycopene (roughly 21 mg $gDCW^{-1}$) (Alper et al. 2006a), while a replicon-free *E. coli* system harboring chromosomal over-expressions produced approximately 100 mg l^{-1} lycopene (roughly 36 mg $gDCW^{-1}$) (Chiang et al. 2008). More recently, multiplexed automated genome engineering (MAGE) was used to construct a

lycopene over-producing *E. coli* strain (almost 9 mg $gDCW^{-1}$) in only three days time (Wang et al. 2009).

The newly engineered strains then lend themselves to the heterologous production of related compounds dependent upon the same precursors (see Figure 7). However, while previous efforts identified crucial precursor pathway genes for engineered over-expression, there are still many aspects of the heterologous systems that remain to be explored. These elements include both recombinant parameters in addition to alternative cellular systems available to house heterologous production. Bioprocess engineering offers a second option to improve heterologous natural product biosynthesis from hosts like *E. coli*. As opposed to directly targeting the cellular and molecular components responsible for biosynthesis, process optimization focuses on the surrounding environment affecting both growth and heterologous biosynthesis. Just as with recombinant parameters, there are several options to be explored within process engineering towards improved heterologous production: medium composition, temperature, aeration, pH, and so forth.

In this chapter, strain backgrounds, promoter systems, and bioprocess conditions were varied to compare and improve the production of taxadiene from *E. coli*. To begin, a transcriptomic study was conducted to provide a systems-level profile of gene-expression

between the strains tested for production. The differences at the transcript level between the K and B strains were analyzed to understand expression differences that could account for the cellular differences in taxadiene production, and more generally, to provide a basis of comparison between the JM109(DE3) and BL21(DE3) strains. In addition, bioprocess improvements in the form of statistical medium optimization, an *in situ* product capture system, and a temperature modulation study was used to further boost production. The cumulative specific titer improvement exhibited in this study is 240-fold (from 0.05 mg gDCW^{-1} to 12 mg gDCW^{-1}).

Materials & Methods

Reagents & Chemicals

The reagents and chemicals used in this study were purchased from ThermoFisher Scientific (Waltham, MA, USA) or Sigma-Aldrich (St. Louis, MO, USA). PCR primers were synthesized by Eurofins MWG Operon (Ebersberg, Germany). TaKaRa LA Taq™ DNA polymerase was from Clontech Laboratories/Takara Mirus Bio (Madison, WI, USA).

Gene, Plasmid, & Strain Construction

The geranylgeranyl diphosphate synthase (*crtE*) and taxadiene synthase (*txs*) genes were synthesized using the method of Kodumal *et*

al. (Kodumal et al. 2004). The original template sequences were a *crtE* gene from *Taxus canadensis* (GenBank accession code AF081514) (Hefner et al. 1998) and a *txs* gene from *Taxus brevifolia* (GenBank accession code U48796) (Wildung and Croteau 1996). The synthesized products were optimized for codon usage in *E. coli* and sequenced to confirm gene design. Standard molecular biology techniques were then used to generate the plasmids presented in (Table 13) (Sambrook and Russell 2001) (constructed by Dr. Yong Wang).

Table 13 Plasmids constructed in this chapter.

Plasmid Name	Description
pQE-TXS-GGPPS	$T5_{prom}$-*txs*$_{syn}$-*crtE*$_{syn}$-$T5_{term}$; both *txs* and *crtE* synthetic; background plasmid pQE30 (Qiagen)
pTrc-TXS-GGPPS	Trc_{prom}-*txs*$_{syn}$-*crtE*$_{syn}$-Trc_{term}; both *txs* and *crtE* synthetic; background plasmid pTrcHis2B (Invitrogen)
pACYCDuet-TXS-GGPPS	$T7_{prom}$-*txs*$_{syn}$-*crtE*$_{syn}$-$T7_{term}$; both *txs* and *crtE* synthetic; background plasmid pACYCDuet-1 (Novagen)

Table 14 also presents the strains constructed to support taxadiene biosynthesis. *E. coli* strains MG1655, JM109(DE3), and BL21(DE3) were used as the original hosts for YW140, YW22, and YW23, respectively (constructed by Dr. Yong Wang). Strains YWS140 and YWGAF were re-constructed as described previously (Yuan et al. 2006). For the construction of YW22 and YW23, a polycistronic operon containing $T7_{prom}$-*dxs-idi-ispB-ispDF*-$T7_{term}$ was constructed in a pET21c expression plasmid. Next, λ-Red mediated homologous recombination (Datsenko and Wanner 2000) was used to insert the operon into the *araA* location of JM109(DE3) and BL21(DE3), respectively, as described previously (Wang and Pfeifer 2008). All integrants and knockouts were verified by PCR. All strains were stored at –80°C in LB medium supplemented with 10% (v v^{-1}) glycerol.

Table 14 Strains constructed in this chapter.

Strain Name	Description
YW22	JM109(DE3); *araA*::T7$_{prom}$-*dxs-idi-ispB-ispDF*-T7$_{term}$
YW23	BL21(DE3); *araA*::T7$_{prom}$-*dxs-idi-ispB-ispDF*-T7$_{term}$
YWS140	MC1061; *dxs*, *idi*, and *ispDF* genes over-expressed through T5 promoter replacement
YWGAF	YWS140; *gdhA*::FRT, *aceE*::FRT, *fdhF*::FRT

Small-Scale Production Cultures

The strains were inoculated in 2 ml of production medium in 16 × 100 mm culture tubes to an OD_{600nm} = 0.1 and cultured for five days at 22°C and 250 rpm. Initial experiments testing recombinant parameters were conducted using production medium (described below). Selection was maintained with 100 mg l^{-1} carbenicillin, 50 mg l^{-1} kanamycin, or 34 mg l^{-1} chloramphenicol when necessary. Expression of *lacI*-repressed genes was accomplished by induction with 100 µM isopropyl ß-D-1-thiogalactopyranoside (IPTG). For the two-phase *in situ* extraction, *n*-dodecane was added to the cultures in 10, 20, 30, 40, and 50% volume ratios. Control cultures of each strain without *n*-dodecane or without IPTG induction were also tested. For the temperature modulation studies, YW22 and YW23 were cultured in the same manner at 12, 17, 22, 27, 32, and 37°C.

Transcript Preparation & Analysis

YW22(pACYCDuet-TXS-GGPPs) and YW23(pACYCDuet-TXS-GGPPs) were grown as described below in the "Small-Scale Cultures" section. At 72hr, total RNA was first prepared with the RNeasy Protect Bacteria Mini Kit (QIAGEN) and then purified by spin column with the RNeasy Mini Kit (QIAGEN). Extracted total RNA (in triplicate) was sent to the Tufts University Computational Genomics Core Facility for analysis, amplification, labeling, hybridization to the

GeneChip® *E. coli* Genome 2.0 Array (Affymetrix; Santa Clara, CA, USA), and scanning. The GeneChip® *E. coli* Genome 2.0 Array contains 10,208 probe sets for *E. coli* strains K-12 MG1655, CFT073, O157:H7-EDL933, and O157:H7-Sakai. The six Affymetrix CEL files containing intensity information for each probe were exported to the MATLAB (The MathWorks™; Natick, MA, USA) Bioinformatics Toolbox. Microarray data was processed using the RMA procedure for background adjustment (Irizarry et al. 2003). Normalized intensities for each probe set and each strain were averaged and *t*-tests were conducted in the $\log_2$ values to calculate the *p*-value between the YW22 and YW23 strains. Probe data were then filtered to contain only probes specific for *E. coli* K-12 MG1655. Statistically significant differences between the two strains were considered for cases with $p < 0.01$, while differentially expressed genes were considered for cases with a greater than 2-fold difference in $\log_2$ values.

Plackett-Burman Screening

A Plackett-Burman screening methodology was used to identify media components significantly influencing taxadiene production. Briefly, the components of two media were used within the Plackett-Burman screening scheme (Plackett and Burman 1946). Defined medium consisted of 6.3 g l^{-1} KH_2PO_4, 18.1 g l^{-1} K_2HPO_4, 0.6 g l^{-1} $(NH_4)_2SO_4$, 15 g l^{-1} glycerol, 0.47 g l^{-1} $MgSO_4 \cdot 7H_2O$, 1.875 ml l^{-1} trace

metal solution (per liter deionized water: 27 g $FeCl_3 \cdot 6H_2O$, 2 g $CaCl_2 \cdot 6H_2O$, 2 g $NaMoO_4 \cdot 2H_2O$, 1.9 g $CuSO_4 \cdot 5H_2O$, 1.3 g $ZnCl_2$, 0.5 g H_3BO_3, and 1.21 mol HCl), and 1.875 ml l^{-1} vitamin solution (per liter deionized water: 6 g niacin, 5.42 g D-pantothenic acid, 1.4 g pyroxidine, 0.42 g riboflavin, 0.06 g biotin, and 0.04 g folic acid). Production medium was composed of 5 g l^{-1} yeast extract, 10 g l^{-1} tryptone, 10 g l^{-1} NaCl with 15 g l^{-1} glycerol, 3 ml l^{-1} 50% (v v^{-1}) Antifoam B Emulsion, and 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) sodium salt, adjusted to pH 7.6 with 5 M NaOH (Lau et al. 2004; Pfeifer et al. 2002). Using an experimental rationale based on balanced incomplete blocks, two-level sample matrices that utilize $4n$ experimental cases to simultaneously evaluate the effect of $4n$-1 parameters were defined (Stanbury et al. 1995). In this study, eight experimental conditions were utilized to evaluate seven medium components. As a comparison, a two-level full factorial to evaluate seven components would have required 2^n (or 128) experiments. Experiments were repeated in order to account for variability of the process (including medium formulation, cell growth, and GC-MS analysis). The effect of each component was calculated using the sum of the four positive conditions compared to the sum of the four negative conditions normalized to the net effect of each experimental condition.

The control limits (CL) were calculated as $CL = t \cdot \sigma \cdot 2\sqrt{N}$, where N is the total number of responses (16 in our case), σ is the average standard deviation derived from the day-to-day variability of the method calculated from probability tables, and t is the Student's t-distribution value at a 95% confidence interval. Any value greater than this control limit indicates a statistically significant positive effect on taxadiene production. Conversely, any value less than the negative value of the control limit indicates a significant inhibitory response.

Bioreactor Production Cultures

YW22 was inoculated in 1.5 l of production medium (as described above but containing 45 g l^{-1} glycerol) in a 3 l bioreactor (New Brunswick Scientific BioFlo 110) to an OD_{600nm} = 0.1 and cultured at 22°C for five days at 400 rpm. Air was supplied at 0.5 VVM (1.5 l min^{-1}). Selection was maintained with 34 mg l^{-1} chloramphenicol. The cultures where induced with 100 µM IPTG at the beginning of each bioreactor run. Dodecane was added to the culture at 20% volumetric ratio. Duplicate samples from the bioreactor were collected periodically over the course of the culture.

Taxadiene Quantification

The aqueous phase of a culture aliquot was extracted with equal volume of ethyl acetate, followed by vortexing for 20 s and centrifuging at 10,000 rpm for 10 min. In cultures grown with *n*-dodecane, the

samples were first centrifuged at 10,000 rpm for 10 min and then the organic phase was diluted 100× in ethyl acetate. Taxadiene analysis was performed on a Shimadzu (Kyoto, Japan) QP5050A GC-MS using splitless injection. Gas chromatography was performed on a Restek (Bellefonte, PA, USA) Rtx®-XLB column (30 m × 0.25 mm ID, 0.25 µm). The column was initially held at 50°C for 1 min, then it was heated to 320°C using a gradient of 8°C min^{-1}, and it was finally held at that temperature for 2 min. Mass spectrometry was conducted in Single Ion Monitoring (SIM) mode scanning for mass to charge ratios of 107 m z^{-1}, 122 m z^{-1} and 272 m z^{-1}, as determined previously (Jennewein et al. 2001). Taxadiene eluted at 22.21 min and quantification was accomplished by using a six-point calibration curve (0 mg l^{-1}, 5 mg l^{-1}, 10 mg l^{-1}, 25 mg l^{-1}, 50 mg l^{-1}, and 100 mg l^{-1}) created with a purified taxadiene standard (kindly provided by Drs. Ajikumar Parayil and Gregory Stephanopoulos).

Metabolite Quantification

The aqueous phase of a culture aliquot was analyzed by HPLC (Agilent 1100 Series) coupled to a Refractive Index Detector (RID). 20 µl of the clarified culture supernatant (by 10,000 rpm for 10 minutes) was applied to a Bio-Rad Aminex® HPX-87H Ion Exchange (300 mm × 7.8 mm, 9 µm) column. The isocratic analysis used a solvent of composition of 9.5 mM H_2SO_4 held at a flow rate of 0.3 ml min^{-1}. These

conditions were identified by using an iterative stochastic search HPLC optimization program based on the compounds anticipated to be present in the culture medium (Dharmadi and Gonzalez 2005). A five-point standard calibration curve was created and used for quantification of glycerol, pyruvate, acetate, ethanol, succinate, formate, and lactate. The elution order was as follows: pyruvate (16.7 min), succinate (22.7 min), lactate (24.2 min), glycerol (25.1 min), formate (26.8 min), acetate (29.1 min), and ethanol (41.3 min).

Results & Discussion

In this study, taxadiene production through *E. coli* was initiated with K and B strains designed to support isoprenoid precursor supply. In addition, the *crtE* and *txs* genes needed for taxadiene biosynthesis were both designed and synthesized to optimize codon usage within *E. coli*. Having taken these initial steps, two routes were then pursued to further improve taxadiene production. First, different promoter/strain combinations were used to test expression of the chromosomal genes supporting precursor supply and the heterologous genes needed for taxadiene biosynthesis. Strain design was based upon the genotype described previously for *E. coli* carotenoid biosynthesis (Yuan et al. 2006). Because carotenoid and isoprenoids derive from the same substrate precursors, this initial strain design was chosen as a template for subsequent variation. Hence, the same genes targeted for

T5 promoter over-expression in the original strain were also altered for expression from a stronger T7 promoter. Besides promoter variation between the genes responsible for taxadiene precursor supply, other chromosomal modifications included specific knockouts (Δ*gdhA*, Δ*aceA*, and Δ*fdhF*) predicted to improve carotenoid production (Alper et al. 2005b) and JM109(DE3) versus BL21(DE3) pertaining to T7-based gene-expression. Modifications were made to both JM109(DE3) and BL21(DE3) strains harboring the required T7 RNA polymerase. Of particular interest were any differences in taxadiene production between these two hosts given differences previously observed with these strains for heterologous polyketide biosynthesis (Wu et al. 2010). The situation therefore provided a range of strains supporting taxadiene precursor supply. Plasmid-borne expression of the codon optimized *crtE* and *txs* genes was tested across Trc, T5, and T7 promoter systems.

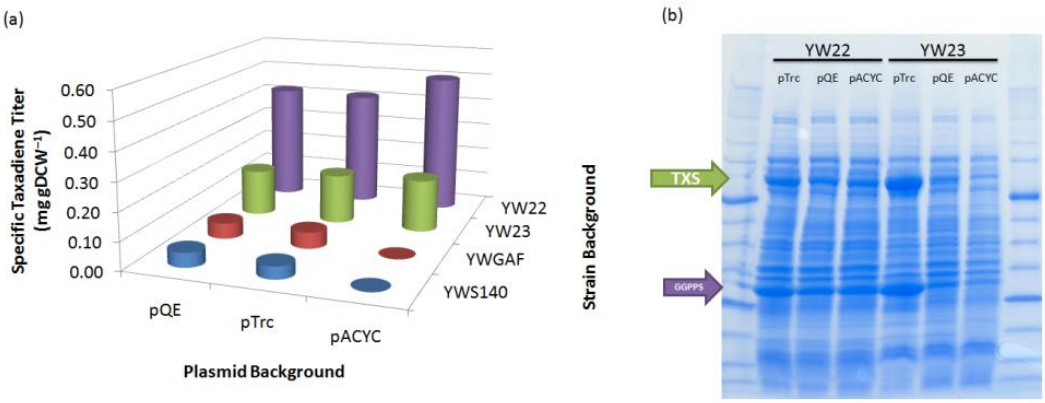

Figure 44 Modulation of heterologous gene promoter and strain background. T5, Trc, and T7 promoters were tested for biosynthetic genes expressed in strains designed for either T5 (YWS140 and YWGAF) or T7 gene expression (YW22 and YW23). Plasmids with T7 promoters were not tested in strains YWS140 and YWGAF and therefore show no taxadiene production. (b) SDS-PAGE analysis of heterologous protein levels for the three promoter systems tested in YW22 and YW23).

Figure 44a presents *E. coli* taxadiene production as a function of strain and recombinant parameter variation. Taxadiene production in YWS140 and YWGAF show similar trends for both the T5 and Trc promoter systems. Strain YWS140 was designed to replicate the original modifications used to better support carotenoid biosynthesis. As such, this was considered baseline production with respect to comparisons between the remaining strain and plasmid combinations. The YWGAF strain contained gene deletions (Δ*gdhA*, Δ*aceE*, and Δ*fdhF*) found capable of improving carotenoid production within *E. coli*. This triple knockout strain was rationally predicted from stoichiometric modeling and experimentally verified for lycopene (Alper et al. 2005b). Interestingly, when the native *crtE* gene is expressed, production is improved and matches those levels seen when strain YW23 serves as the background host. Though the reason for this is unclear, others have noticed an imbalance in heterologous expression when using synthesized genes, implying that coupling native *crtE* and synthesized *txs* genes may provide a more optimal expression profile and/or precursor supply (Menzella et al. 2006). For those synthetically derived genes, the YW23 host shows an increase in production compared to counterparts from YWS140 and YWGAF. Given the precursor network shared between carotenoid and isoprenoids compounds, it was reasonable to expect that the

Δ*gdhA*Δ*aceE*Δ*fdhF* genotype would similarly aid taxadiene production. However, in this study, only slight but non-significant improvements resulted from strain YWGAF, implying that the improvements observed were specific for lycopene production or the culture conditions used. This could have been due to the fact that the triple knockout improved titer in minimal medium with glucose; whereas, the medium used here was a complex medium with glycerol. Maximum taxadiene production was associated with YW22. Of the improvements observed, strain YW22(pACYCDuet-TXS-GGPPS) showed the best production with an 18.5-fold improvement in specific production from the lowest strain-plasmid combination (of those comparing synthetically derived biosynthetic genes). Both strains carrying the T7 RNA polymerase showed a noticeably reduced production when the native *crtE* gene was expressed, in comparison to expression from the YWS140 and YWGAF strains.

The T7 strains led to a pronounced improvement in taxadiene biosynthesis, but improvement varied across B (YW23) and K (YW22) genotypes. These differences prompted us to analyze the K and B *E. coli* hosts using DNA microarray technology. A transcriptomics study (global transcript profiling) was undertaken to better understand why the K strain produces roughly twice as much taxadiene with qualitative lower expression of the heterologous genes (as seen in

Figure 44b, particularly for the pTrc plasmid) . Total RNA was extracted from YW22 and YW23 during late-stage exponential growth (approximately 48 hr) in the production medium preciously described and analyzed using the GeneChip® *E. coli* Genome 2.0 Array. After adjustment, normalization, and filtering, 348 of the 4070 MG1655 genes were differentially expressed ($\log_2$(YW22/YW23) > ± 2) at a statistically significant level (p-value < 0.01) (as can be seen in the scatter plot in Figure 45a). Of these 348 genes, 243 were up-regulated in YW22 while 105 were up-regulated in YW23. Although a large number of the identified genes had no clearly assigned function (hypothetical proteins), a number of enzymes involved in central metabolic pathways were identified. In YW23, both pyruvate kinase I (*pykF*) and phosphoenolpyruvate carboxykinase (*pck*) were up-regulated, indicating that phosphoenolpyruvate is being produced from both pyruvate (through the action of *pykF*) and oxaloacetate (through the action of *pck*). Because pyruvate is one of the direct precursors (along with glyceraldehyde-3-phosphate) for the DXP pathway, decreasing pyruvate flux from this pathway by up-regulation of these two enzymes is a likely explanation for why the taxadiene production titer in YW23 is roughly half that of YW22. In addition, *pykF* was a main target identified through a novel modeling algorithm developed

in our laboratory (unpublished work) as one that, when deleted, predicted improved taxadiene titer.

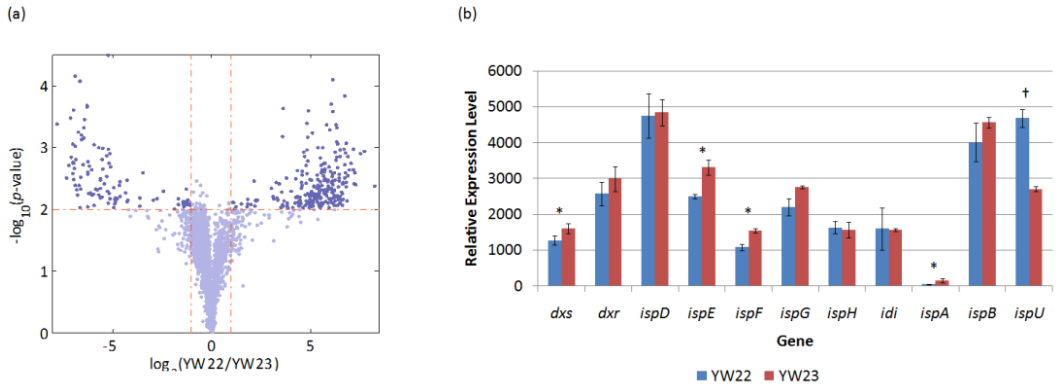

Figure 45 Microarray data comparing late-stage exponential growth gene-expression between YW22(pACYCDuet-TXS-GGPPS) and YW23(pACYCDuet-TXS-GGPPS).
(a) A volcano plot shows the relationship between *p*-value and differential expression. Points in the upper right quadrant correspond to genes up-regulated greater than two-fold in YW22(pACYCDuet-TXS-GGPPS) at a statistically significant level ($p < 0.01$) while points in the upper left quadrant correspond to genes that are up-regulated greater than two-fold in YW23(pACYCDuet-TXS-GGPPS) at $p < 0.01$ (as determined by paired Student's *t*-test). (b) A plot showing the relative expression level of the genes in the isoprenoid biosynthetic pathway for both YW22(pACYCDuet-TXS-GGPPS) and YW23(pACYCDuet-TXS-GGPPS). * indicates $p < 0.05$ between the two strains, while † indicates $p < 0.01$. All data shown are from three replicates ($n = 3$).

Another key difference between these two strains was the activity around the fructose-6-phoshate metabolite node. YW22 significantly up-regulated phosphofructokinase I (*pfkA*) which catalyzes the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate. At the same time, YW23 significantly up-regulated fructose-1,6-bisphosphatase (*fbp*), which catalyzes the opposite reaction. Although both of these enzymes are inhibited by phosphoenolpyruvate, it would appear that the expression of the glycolytic *pfkA* would be preferable over the gluconeogenic *fbp*, especially being that fructose-1,6-bisphosphate is degraded during glycolysis to yield glyceraldehyde-3-phosphate, the other direct precursor for the isoprenoid biosynthetic pathway. Over-expression of *pfkA* and/or deletion of *fbp* would likely increase cellular pool of glycerol-3-phosphate and would therefore increase flux toward the isoprenoid biosynthetic pathway and improve taxadiene production.

As an attempt to identify an overall bottleneck in the isoprenoid biosynthetic pathway and to identify differences between YW22 and YW23, the expression of the genes in the isoprenoid biosynthetic pathway (*dxs*, *dxr*, *ispDEFGH*, *idi*, *ispA*, *ispB*, and *ispU*) (Figure 45b) were directly compared. Of the eleven genes, four (*dxs*, *ispE*, *ispF*, and *ispA*) were expressed higher ($p < 0.05$) in YW23 and one (*ispU*) was expressed higher in YW22 ($p < 0.01$). It should be noted that *ispB* and

ispU occur further downstream in the isoprenoid biosynthetic pathway required for production of octaprenyl diphosphate (required for menaquinol and ubiquinol biosynthesis) and undecaprenyl diphosphate (used in peptidoglycan biosynthesis), respectively. It is clear that slight increased expression of certain genes in the isoprenoid biosynthetic pathway by YW23 does not result in improved taxadiene production, leading us to believe the main difference in taxadiene phenotype between the K and B strains lies within central carbon metabolism. More generally, it is interesting to note that *ispA* expression is approximately an order of magnitude lower than that of the next lowest expressed gene in this pathway, indicating a potential candidate for over-expression to alleviate a biosynthetic bottleneck and further improve taxadiene production.

The study also featured a second route to improved *E. coli* taxadiene production through bioprocess engineering. This involved improving three subcomponents: 1) medium composition, 2) an *in situ* extraction technique, and 3) a temperature modulation study. Because strain YW22(pACYCDuet-TXS-GGPPS) demonstrated the highest taxadiene levels from the recombinant parameter analysis, it was chosen for use during media optimization. To expeditiously identify media components having a significant impact on taxadiene biosynthesis, a Plackett-Burman screening methodology was employed.

The Plackett-Burman method aims at minimizing the number of experiments in order to determine the effects of main variables, in this case, media components. Hence, a two-dimensional analysis between these components was completed to identify maximum taxadiene production. Here, components of the production medium, used for the recombinant parameter study outlined above, were tested with other media components from a defined medium commonly used by our group (Wang et al. 2007a). The Plackett-Burman method allows a rapid screening of media components and provides quantifiable basis for identifying which components impact biosynthesis. In this case, glycerol and yeast extract showed significant positive impact on taxadiene biosynthesis. Glycerol's positive effect has been observed before for closely related carotenoids and sesquiterpenes also produced through *E. coli* (Lee et al. 2004; Martin et al. 2003). Antifoam solution was also identified as a medium component positively influencing titers; however, the remainder of the study focused on glycerol and yeast extract since these were considered more likely to influence both cellular biomass and taxadiene biosynthesis.

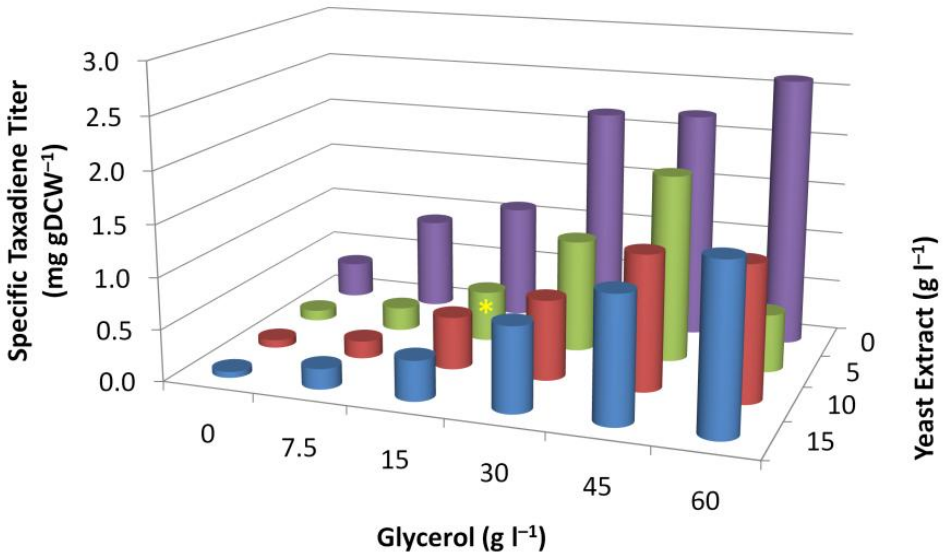

Figure 46 Medium component modulation study.
Strain YW22(pACYCDuet-TXS-GGPPS) was tested across different glycerol and yeast extract concentrations to identify production maxima as a function of media. Data are presented for specific (mg gDCW^{-1}) production and are averaged between two separate experiments. Original production medium is denoted with a yellow * at the base of the production column.

Glycerol and the yeast extract were therefore varied so as to provide a surface response curve with the *z*-axis as taxadiene production (Figure 46). Holding the remaining components of the production media constant, glycerol and yeast extract were varied to survey the production landscape as a function of these two components. Figure 46 presents this data as specific production titers (mg $gDCW^{-1}$), highlighting maxima beyond those provided by the original medium formulation. It can be seen that the specific production in the quadrant of the surface response encompassing increased glycerol and reduced yeast extract content presents new media compositions for maximum taxadiene production. Maximum production from the media screen shows an additional 10-fold improvement in specific production beyond that provided by recombinant parameter/strain combinations.

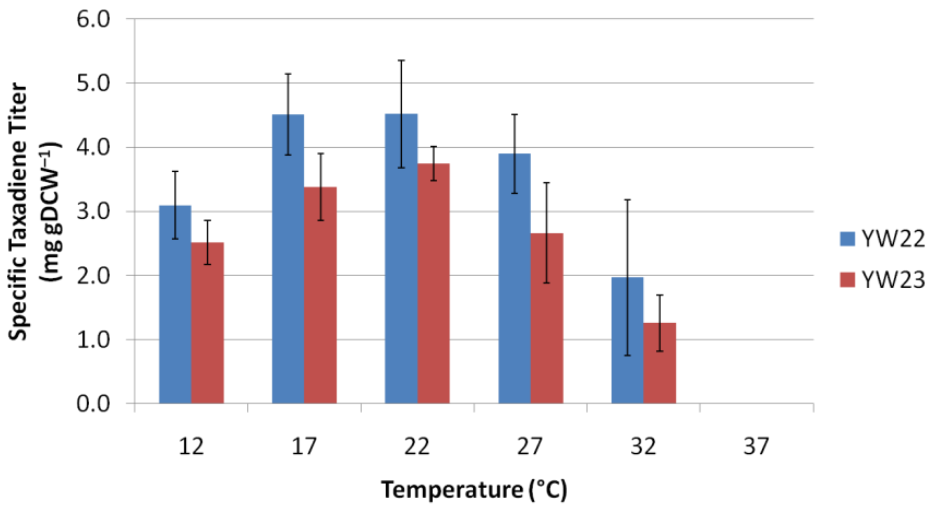

Figure 47 Temperature modulation study, specific taxadiene titer.
Specific taxadiene titer (mg $gDCW^{-1}$) is shown as a function of temperature between 12°C and 37°C at 5°C intervals for YW22(pACYCDuet-TXS-GGPPS) and YW23(pACYCDuet-TXS-GGPPS). Error bars designate ± one standard deviation of four replicates.

Next, a study was undertaken to determine the optimal temperature for taxadiene production. Here, a temperature down-shift from 37°C to between 12°C and 32°C (at 5°C intervals), concurrently with induction of gene-expression, at OD_{600nm} = 0.6 was utilized to observe the differences in cell-density attained, taxadiene titer, substrate uptake, and by-product formation for both YW22(pACYCDuet-TXS-GGPPS) and YW23(pACYCDuet-TXS-GGPPS). Taxadiene production was observed at all temperatures except 37°C for both YW22 and YW23. Interestingly, for each strain, the specific taxadiene titers did not vary significantly between 12°C and 27°C, with a slight decrease observed at 32°C.It has been shown previously that BL21(DE3) grows to a higher cell-density than JM107 (a close relative to JM109(DE3)) (Yau et al. 2008). However, there have not been many side-to-side comparisons of *E. coli* strains and their physiology (Phue et al. 2008; Phue et al. 2005; Phue and Shiloach 2004), much less in terms of metabolic engineering applications (Tseng et al. 2009; Wu et al. 2010). Here, it was also observed that YW23 (the B derivative) grew to higher cell-densities than YW22 (the K derivative); however, the cell-densities achieved were not statistically significantly different at 22°C, 27°C, and 32°C (Figure 48a). Also as expected, the specific uptake rate of glycerol was higher at higher temperatures (Figure 48b). Interestingly, YW23 showed lower specific

uptake rates of glycerol than YW22 but grew to a higher cell-density, lending itself to better utilization of the initial carbon source. Across all strains and culture temperatures, acetate was the primary by-product (Figure 48c). Ethanol was only observed in the residual culture medium at temperatures of and between 22°C and 37°C for both YW22 and YW23 (Figure 48d). Lactate, succinate, formate, or pyruvate were not observed (less than 0.5 mM) in the residual culture medium for all strains and culture temperatures.

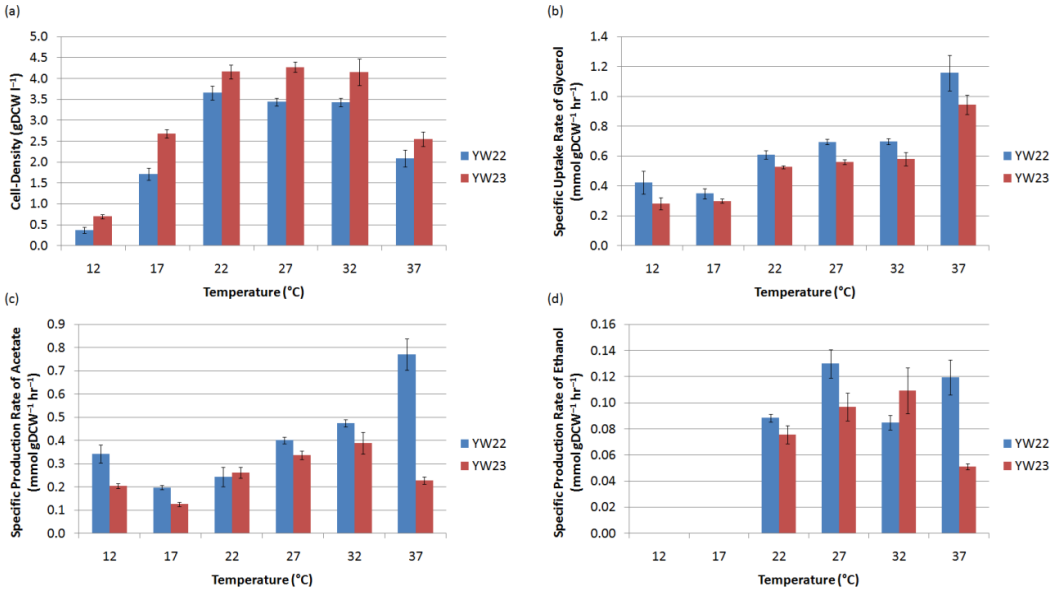

Figure 48 Temperature modulation study, cell-density and metabolite production rates.
Data shown for YW22(pACYCDuet-TXS-GGPPS) and YW23(pACYCDuet-TXS-GGPPS), as a function of temperature between 12°C and 37°C in 5°C increments, the (a) specific uptake rate of glycerol, (b) specific production rate of acetate, (c) specific production rate of ethanol, (d) and specific taxadiene titer are plotted. Error bars designate ± one standard deviation of four replicates.

Finally, bioreactor studies were conducted to test whether the optimal conditions as determined in the previous sections would hold at increased scale. Like other isoprenoid compounds recently produced heterologously (Newman et al. 2006), taxadiene is highly volatile. Preliminary studies revealed that taxadiene titer sharply decreased approximately half way through a standard bioreactor process. Whereas, in 100 ml shake-flask cultures, the maximal taxadiene titer was higher than that in the 1.5 l bioreactor, and the decrease in titer was attenuated. The inverse relationship between culture aeration and taxadiene titer emphasized taxadiene loss through evaporation. Given this condition and the fact that taxadiene has a log *P* value (an octanol:water partition coefficient, as calculated by ChemDraw Ultra using Crippen's fragmentation method) of 5.74, cultures were first conducted at 2 ml and *n*-dodecane (the non-volatile twelve-membered alkane) was overlaid to prevent taxadiene loss. The dodecane overlay was tested at volume percentages of 10% to 50% in increments of 10% (Figure 49a and Figure 49b). Figure 49c shows the specific taxadiene titer as a function of *n*-dodecane concentration, indicating that *n*-dodecane concentrations between 20% and 50% lead to similar specific titers. Overall, the specific taxadiene titer increased from 0.5 mg $gDCW^{-1}$ to 5.0 mg $gDCW^{-1}$ with the use of 50% *n*-dodecane. It is interesting to note that taxadiene volatility may have been indicated

during the Plackett-Burman media optimization study. The antifoam solution was included as a media component within the analysis and exhibited a statistically significant influence upon taxadiene levels, possibly because of air-water interface effects and influence upon taxadiene volatilization.

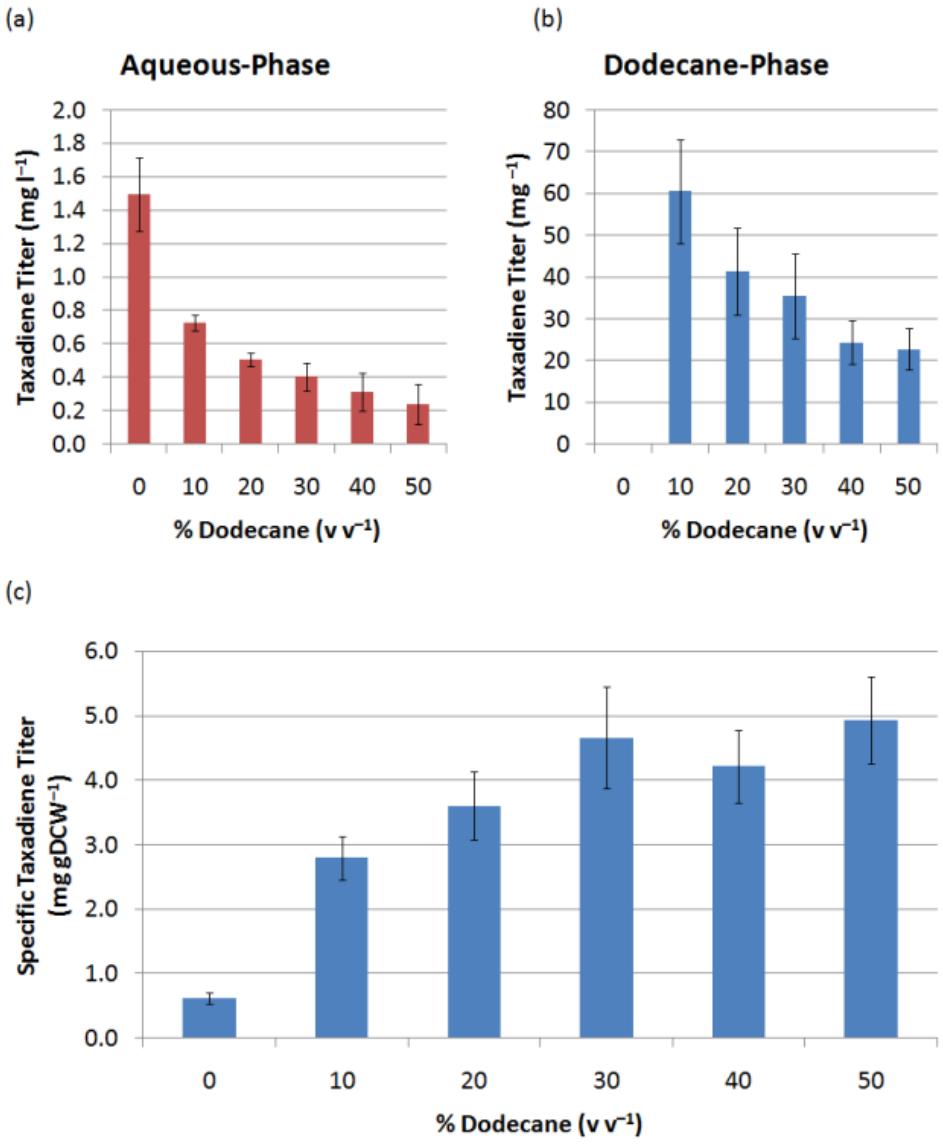

Figure 49 ***In situ*** **extraction of taxadiene with** ***n*****-dodecane.**
For YW22(pACYCDuet-TXS-GGPPS), (a) Taxadiene titer in the aqueous phase and (b) in the dodecane phase is plotted as a function of volume percentage of dodecane. (c) Overall specific taxadiene titer. Error bars designate ± one standard deviation of three or four replicates.

As a result, the batch bioreactor process was redesigned in three respects to aid in the *in situ* capture of taxadiene: 1) n-dodecane was added to the culture at 20% (v v^{-1}), 2) the air flow-rate was decreased from 1 VVM to 0.5 VVM, and 3) the agitation was decreased from 1000 RPM to 400 RPM. The latter two adjustments were to minimize mixing of the abiotic organic phase and the aqueous culture medium. The specific taxadiene titer increased ten-fold from approximately 1.2 mg $gDCW^{-1}$ to approximately 12 mg $gDCW^{-1}$ without significant differences in the cell-density achieved (approximately 5 gDCW l^{-1}) (Figure 50a). Glycerol was exhausted at the end of the culture period (127 hr), with large amounts of acetate (almost 200 mM in the later stages of the bioprocess) generated throughout the batch bioprocess (Figure 50b). As the glycerol became exhausted, the culture started utilizing the acetate generated earlier in the bioprocess as a secondary carbon source. Presumably, if taxadiene is stable in the organic phase, continual usage of the remainder of the acetate would likely further improve taxadiene production. As was the case with the small-scale cultures, pyruvate, formate, and succinate were not observed (less than 0.5 mM) in the culture medium. Expression of the heterologous genes appears to be strong throughout the entire bioprocess (*Figure 7c*).

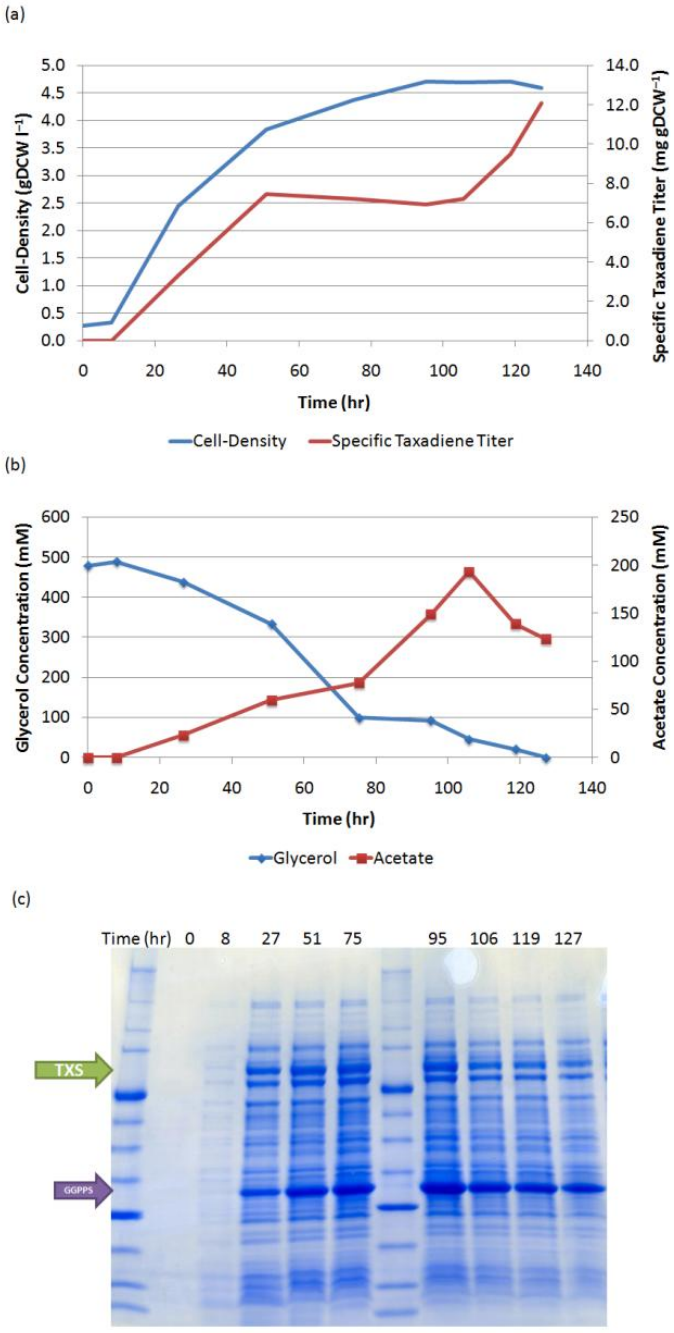

Figure 50 Bioreactor cultivation of YW22(pACYCDuet-TXS-GGPPS with 20% (v v^{-1}) dodecane.
(a) Specific taxadiene titer (in the aqueous and organic phases combined) and cell-density are plotted as a function of time. (b) Glycerol and acetate concentration in the culture medium are plotted as a function of time. (c) Gene-expression by SDS-PAGE is shown as a function of time. Bands corresponding to the taxadiene synthase (TXS) and the geranylgeranyl diphosphate synthase (GGPPS) are marked.

In summary, this study featured efforts to improve and compare taxadiene biosynthesis through a range of recombinant/process parameters and K- and B-based *E. coli* strains. Highest specific production levels were observed at 22°C for K-based *E. coli* strains utilizing the T7 expression system. A transcript analysis between the K and B strains identified several interesting variations that provided insight into potential mechanisms for the differences observed. This insight provides future metabolic engineering targets to further overproduce taxadiene, or other isoprenoids, heterologously through *E. coli*. Expression of *txs* and $crtE_{syn}$ was qualitatively verified by SDS-PAGE throughout the process (Figure 50c).

Chapter 8 – Development of a platform system for simultaneous production and partitioning of polyketide and isoprenoid natural products in a two-phase bioprocess

Introduction

Heterologous production of erythromycin and Taxol intermediates has now been established using *E. coli*. In the previous chapters of this dissertation, the engineering of *E. coli* designed to support the production of 6-dEB and taxadiene was reported, in separate systems. In this chapter, the potential of producing both compounds simultaneously was explored. *E. coli* was first genetically modified based upon the known requirements for polyketide and isoprenoid biosynthesis (Figure 51). A co-production process was then assessed, aided greatly by the ability to readily partition the two nascently formed products. The implications of this study range from a consolidated bioprocess for producing multiple therapeutic natural compounds to an *E. coli* host now capable of supporting continual drug discovery and development opportunities.

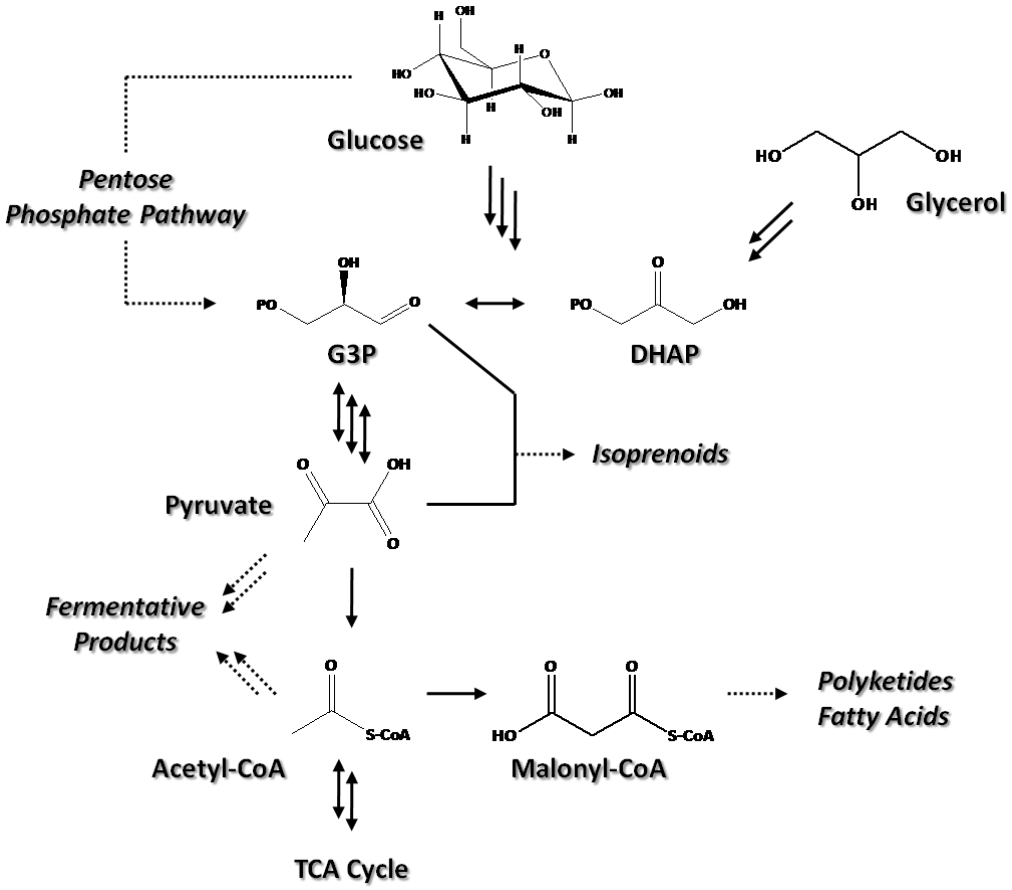

Figure 51 An overview of *E. coli* metabolism and its relation to polyketide and isoprenoid biosynthesis.
Abbreviations: G3P = glyceraldehyde-3-phosphate, DHAP = dihydroxyacetone phosphate.

Materials & Methods

Background Strains & Plasmids

E. coli, strain BAP1 has been developed previously for the heterologous production of polyketide and nonribosomal peptide natural products (Pfeifer et al. 2001). The *Bacillus subtilis* surfactin phosphopantetheinyl transferase gene (*sfp*) (Quadri et al. 1998) was inserted into the *prpRBCD* location of the BL21(DE3) chromosome, under the control of an inducible T7 promoter (Pfeifer et al. 2001). The *sfp* gene is required to post-translationally activate PKS's by transferring the 4'-phosphopantetheinyl moiety from coenzyme A to a conserved, reactive serine residue within each acyl carrier protein (ACP) domain of polyketide synthases (or the peptidyl carrier protein (PCP) domains of nonribosomal peptide synthetases) (Quadri et al. 1998). *E. coli*'s native holo-acyl carrier protein synthase (ACPS, required for activation of fatty acid synthases) does not recognize the apo-forms of many ACP and PCP domains (Quadri et al. 1998). Thus, without *sfp*, 6-dEB cannot be produced through *E. coli*. During this genetic insertion, a T7 promoter was also inserted before the native *prpE* gene to increase metabolic flux towards propionyl-CoA, a direct precursor of 6-dEB. This strain was termed BAP1 (Pfeifer et al. 2001) and is used as the base production system in this study. Strain JW1 was derived from BAP1 (Wu et al. 2010), and contains a Flippase

Recognition Target (FRT)-(Cherepanov and Wackernagel 1995) flanked *cat* gene (coding for chloramphenicol acyl transferase, derived from pKD3 (Datsenko and Wanner 2000)) in the chromosome before the *sfp* gene. Strain YW23 was derived from BL21(DE3), and contains an additional copy of the native *dxs*, *idi*, *ispB*, *ispD*, and *ispF* genes under the control of a single T7 promoter in the *araA* location of the chromosome (see Chapter 7).

The genes required for the production of 6-dEB from propionate were previously inserted into plasmids pBP130 and pBP144 (Pfeifer et al. 2001). Briefly, pBP130 (approximately 26kb and derived from pET21c (Novagen)) contains the *eryA2* and *eryA3* genes (coding for the DEBS2 and DEBS3 enzymes) under a single T7 promoter. Plasmid pBP144 (approximately 19kb and derived from pET28 (Novagen)) contains *eryA1* (coding for the DEBS1 enzyme) under a T7 promoter and genes coding for the two subunits of the *S. coelicolor* propionyl-CoA carboxylase enzyme (*accA*1 and *pccB*) (Rodriguez and Gramajo 1999) under another T7 promoter. All three *eryA* genes are from the native erythromycin producer, *S. erythraea* (Cortes et al. 1990; Donadio et al. 1991).

Two heterologous enzymes are needed to produce taxadiene through *E. coli*: 1) a geranylgeranyl-diphosphate synthase (to catalyze: farnesyl-diphosphate + isopentenyl-diphosphate → geranylgeranyl-

diphosphate + diphosphate) and 2) a cyclizing taxadiene synthase (to catalyze: geranylgeranyl-diphosphate → taxa-4,11-diene + diphosphate). The original template sequences were *crtE* (for the geranylgeranyl-diphosphate synthase) from *Erwinia herbicola* (also known as *Enterobacter agglomerans*) (Cunningham et al. 1994) and *txs* (for the taxadiene synthase) from *Taxus baccata* (Besumbes et al. 2004). The sequences were codon optimized for *E. coli*, constructed synthetically (Kodumal et al. 2004), and cloned into the MCS1 of pACYCDuet-1 (Novagen) between *Nco*I and *Sal*I sites (under the control of a single T7 promoter) (see Chapter 7). This plasmid was named pACYCDuet-TXS-GGPPS (approximately 7.5 kb).

Strain Construction

YW23^{sfp} was generated through P1 bacteriophage transduction (Masters 1977) in which JW1 was used as the donor strain and YW23 was the recipient. The resulting strain would then be able to support isoprenoid, polyketide, and nonribosomal peptide biosynthesis. The infected YW23 was plated on LB agar containing 20 mg l^{-1} chloramphenicol, incubated overnight at 37°C, and successful recombinants were verified by PCR, as described previously (Wu et al. 2010). This strain was stored as a glycerol stock, prepared electrocompetent, and transformed with pCP20 (Cherepanov and Wackernagel 1995). This plasmid contains the gene coding for the

Saccharomyces cerevisiae Flp recombinase (Cherepanov and Wackernagel 1995), used to excise the *cat* gene between the FRT sites. The ultimate result was a chloramphenicol sensitive strain, YW23^{sfp}. This strain was stored a glycerol stock, prepared electrocompetent, and transformed with 1) pBP130 and pBP144, 2) pACYCDuet-TXS-GGPPS, or 3) all three plasmids.

Shake-Flask Production Cultures

Glycerol stocks were used to inoculate 2 ml Luria-Bertani (LB) medium cultures with appropriate antibiotics for overnight incubation at 37°C and 250 rpm. Production cultures were conducted in 125 ml non-baffled Erlenmeyer flasks in a rich medium containing 5 g l^{-1} yeast extract, 10 g l^{-1} tryptone, 10 g l^{-1} sodium chloride, 15 g l^{-1} glycerol, 3 ml l^{-1} 50% (v v^{-1}) Antifoam B, 100 mM HEPES, and pH 7.60. Production cultures (15 ml) were inoculated with the precultures to an OD_{600nm} = 0.1 and supplemented with 20 mM appropriate precursors (propionate, pyruvate, both, or neither), appropriate antibiotics, and 100 µM isopropyl ß-D-1-thiogalactopyranoside (IPTG). When *n*-dodecane was used as the organic phase, 12 ml culture medium was supplemented with 3 ml *n*-dodecane (for 20% v v^{-1}). Cultures were incubated for 120 hr at 22°C and 250 rpm. At the end of the culture period, cell-density was measured spectrophotometrically at 600 nm, and 1 ml aliquots were stored at –20°C for subsequent analyses. Cell-density in gram dry

cell weight per liter (gDCW l^{-1}) was calculated using an experimentally determined correlation of 1 OD_{600nm} = 0.52 gDCW l^{-1} (data not shown). When needed, antibiotics were supplemented at concentrations of 100 mg l^{-1} for carbenicillin, 50 mg l^{-1} for kanamycin, and 34 mg l^{-1} for chloramphenicol.

Two-Phase Batch Bioprocess

Strain YW23^{sfp}(pBP130/pBP144/pACYCDuet-TXS-GGPPS) was first inoculated in 2 ml of selective LB medium and grown for approximately 8 hr at 37°C and 250 rpm. This culture was then used to inoculate a larger culture containing 100 ml of LB medium with appropriate antibiotics. After an overnight incubation at 37°C and 250 rpm, 1.2 l of medium (as described above) was inoculated with this starter culture in a 3 l bioreactor (New Brunswick Scientific BioFlo 110) to an OD_{600nm} = 0.1. The system was then charged with 300 ml of -*n*-dodecane, bringing the final volume to 1.5 l. Air was supplied at 0.5 vessel volumes per minute (VVM), temperature was controlled with a water bath at 22°C, and pH was controlled at 7.60 with the addition of 5 M NH_4OH. The bioprocess was run for 5 days at 22°C and 500 rpm. The medium also contained 100 mg l^{-1} carbenicillin, 50 mg l^{-1} kanamycin, and 34 mg l^{-1} chloramphenicol to maintain plasmid selection and 100 µM IPTG to induce gene-expression. Sample aliquots

(1 ml) were taken from the bioreactor every 6-12 hr over the course of the culture period and stored at –20°C until subsequent analysis.

6-dEB Quantification

The HPLC method for 6-dEB separation and quantification has been described previously (Wang et al. 2007a). Briefly, quantification of 6-dEB was carried out on an Agilent 1100 series HPLC coupled with an Alltech 800 series evaporative light-scattering detector (ELSD). The guard column used was an Inertsil ODS3 C_{18} 5 µm, 4.6 mm × 10 mm while the analytical column used was an Inertsil ODS3 C_{18} 5 µm, 4.6 mm × 150 mm (GL Sciences). Column temperature was maintained at 25°C while the autosampler was maintained at 4°C. Ultra-high purity grade nitrogen gas (AirGas East) was used as the mobile phase for the ELSD at a pressure of 3.00 ± 0.05 bar, while the ELSD drift tube temperature was maintained at 55°C and the gain setting was set at 16.

Culture samples were first centrifuged for 10 min at 10,000 × g to remove insolubles. A 20.0 µl supernatant injection volume was then used with a solvent system of 100% distilled, deionized water (ddH_2O) from 0-2.0 min as the mobile phase passed through the guard column and to a waste collection. A six-port switching valve then directed the mobile phase to the analytical column as a gradient to 100% HPLC-grade acetonitrile (Sigma-Aldrich) began from 2.0-5.0 min followed by

100% acetonitrile from 5.0-8.0 min, a quick gradient back to 100% ddH_2O from 8.0-8.1 min, and finally 100% ddH_2O maintained from 8.1-9.5 min. The mobile-phase flow-rate was 1 ml min^{-1}. Under these conditions, 6-dEB eluted at 7.92 ± 0.05 min. Quantification was carried out against a five-point calibration curve of purified 6-dEB (kindly provided by Kosan Biosciences).

Taxadiene Quantification

For taxadiene quantification from the aqueous phase, samples were first centrifuged for 10 min at 10,000 × g to remove insolubles. Aqueous supernatant (750 µl) was supplemented with (-)-*trans*-caryophyllene (TC) at a final concentration of 1 µg l^{-1} to serve as an internal standard for quantification. The samples were then extracted with an equal volume of hexane, followed by 20 s of vortexing and centrifugation for 10 min at 10,000 × g. The hexane layer (150 µl) was removed and stored in glass vials at –20°C until analysis with gas chromatography-mass spectroscopy (GC-MS) could be conducted. For quantification of taxadiene in the organic phase, an aliquot of the n-dodecane was diluted 100-fold in hexane containing 1 µg l^{-1} TC. Samples were placed in glass vials stored at –20°C until GC-MS analysis was conducted.

Samples were analyzed on a Shimadzu QP5050A GC-MS using splitless injection. Gas chromatography was run on a non-polar Rxi®-

XLB column (30 m x 0.25 mm ID, 0.25µm). The inlet pressure for the column was set at 120 kPa and column flow velocity was 1.6 ml min^{-1}. The flow rate of the ultra high purity helium carrier gas was 20 ml min^{-1}. Temperature of the column was initially set and maintained at 100°C for 2 min and was then increased to 235°C at a rate of 15.0°C min^{-1}. The column was then maintained at this temperature for 1 min. Mass spectrometry was performed in Single Ion Monitoring (SIM) mode scanning for mass to charge ratios of 107 m z^{-1}, 122 m z^{-1}, and 272 m z^{-1}, corresponding to principle daughter ions and parent ion of taxadiene, respectively (Koepp et al. 1995). Under these conditions, TC and taxadiene eluted at approximately 6.8 min and 11.4 min, respectively. Quantification of taxadiene was accomplished based on a calibration curve of taxadiene (kindly provided by Drs. Ajikumar Parayil and Gregory Stephanopoulos) concentration versus the peak area ratio of TC to taxadiene.

Metabolite Quantification

The clarified aqueous phase was also analyzed for quantities of precursor and byproduct metabolites. This analysis was conducted with HPLC (Agilent 1100 Series) coupled with a Refractive Index Detector (RID). Clarified culture supernatant (20 µl) was applied to a Bio-Rad Aminex® HPX-87H Ion Exchange (300 mm × 7.8 mm, 9 µm) column, preceded by a 30 mm guard column of the same resin. The

isocratic analysis used a 9.5 mM H_2SO_4 solvent held at a flow rate of 0.3 ml min^{-1}. Under these conditions, the elution order of the analyzed compounds was pyruvate (16.7 min), glycerol (25.1 min), acetate (29.1 min), and propionate (36.6 min). Quantification of these compounds was conducted against a five-point calibration curve of purchased standards (Sigma-Aldrich).

Results

Strain Construction

Production of 6-dEB and Taxadiene Separately and Together

The production of 6-dEB and taxadiene, both separately and together, was first analyzed in the newly constructed YW23^{sfp}. Both 6-dEB and taxadiene were quantified in the culture medium after 120 hr of shake-flask cultivation at 22°C and 250 rpm. As can be seen in Figure 52a, YW23^{sfp}(pBP130/pBP144) produced 48.8 ± 7.1 mg l^{-1} 6-dEB while the titer decreased to 17.9 ± 7.3 mg l^{-1} ($p < 0.001$) with the addition of the pACYCDuet-TXS-GGPPS plasmid. The taxadiene titers were approximately an order of magnitude lower than that of 6-dEB (Figure 52b); however, the titers are similar to those previously reported in the parent to YW23^{sfp} (Chapter 7). The YW23^{sfp}(pACYCDuet-TXS-GGPPS) strain produced taxadiene at 1.53 ±

0.27 mg l^{-1}, and there was no statistically significant difference when pBP130 and pBP144 was added to the system ($p = 0.772$).

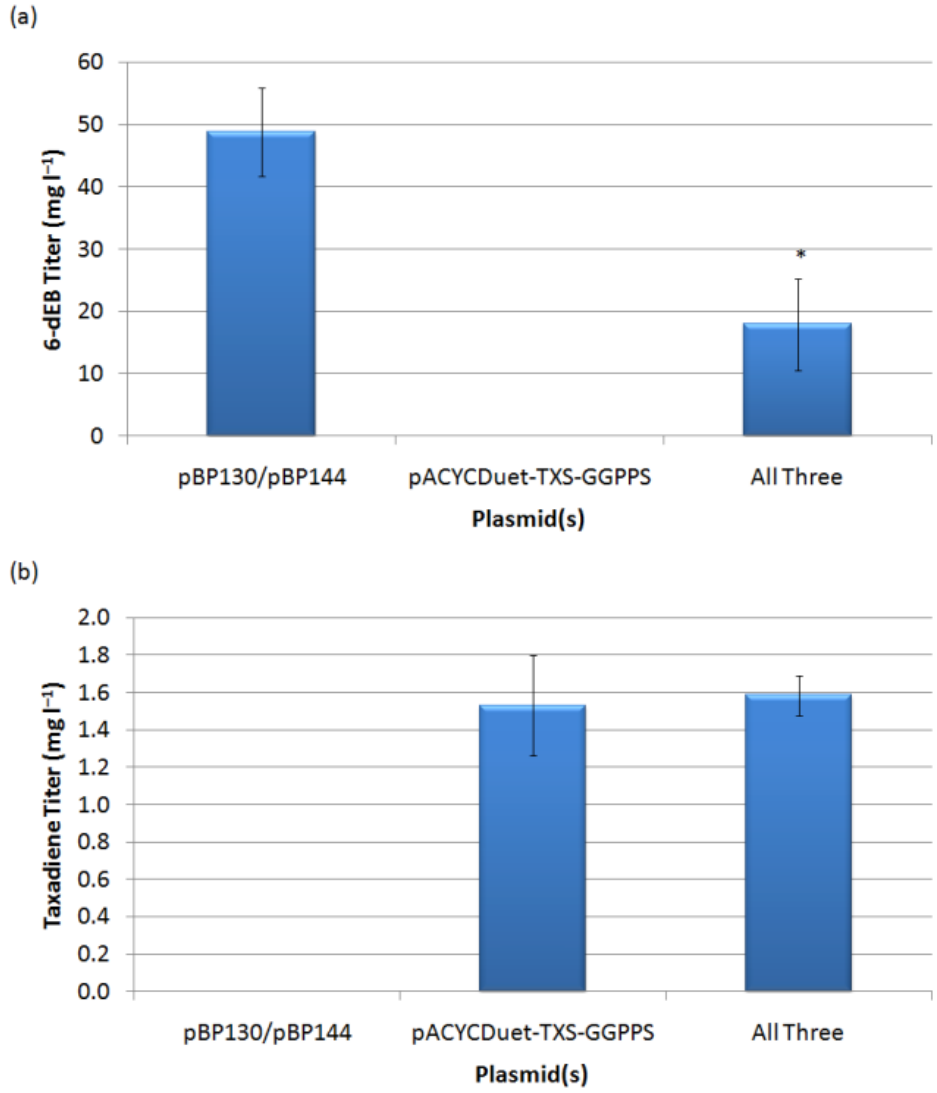

Figure 52 Co-production of 6-dEB and taxadiene in a single-phase system. Production of (a) 6-dEB and (b) taxadiene in YW23sfp containing either the plasmids required for 6-dEB biosynthesis (pBP130/pBP144), the plasmid required for taxadiene biosynthesis (pACYCDuet-TXS-GGPPS), or all three plasmids together. Error bars represent ± one standard deviation of three replicates. * indicates statistically significant results at $p < 0.05$.

Co-Production of 6-dEB and Taxadiene in a Two-Phase System

It was previously reported that taxadiene has a strong preference for partitioning into organic environments *in situ*, improving titer dramatically, as can be seen in Chapter 7 and (Ajikumar et al. 2010). In these studies, *n*-dodecane was chosen as an organic phase based upon its successful use in sequestering *E. coli*-produced amorpha-4,11-diene (a sesquiterpene) (Newman et al. 2006). A concentration of 20% *n*-dodecane (v v^{-1}) was previously determined to be the optimal concentration for improving specific titer (see previoius work). As a result, the two-phase systems reported here contain an aqueous medium for cell growth and 20% *n*-dodecane (v v^{-1}) to capture secreted taxadiene.

Figure 53a presents 6-dEB titer from YW23^{sfp}(pBP130/pBP144/pACYCDuet-TXS-GGPPS) with the inclusion of *n*-dodecane in the production medium. Of note, 6-dEB was not found in the *n*-dodecane phase (<1 mg l^{-1}) though this compound also exhibits non-polar properties and is readily extracted by ethyl acetate. The two-phase 6-dEB titers for YW23^{sfp}(pBP130/pBP144) and YW23^{sfp}(pBP130/pBP144/pACYCDuet-TXS-GGPPS) were similar to single-phase production, indicating that the addition of *n*-dodecane had no effect on polyketide levels. As in the single phase system, 6-dEB

production decreased approximately two-fold with the addition of the pACYCDuet-TXS-GGPPS plasmid (p = 0.023).

Figure 53b presents taxadiene titer in the same three scenarios. As expected, the addition of *n*-dodecane to the shake-flask cultures increased the titer of taxadiene significantly to 40.8 ± 6.9 mg l^{-1}. There was residual taxadiene found in the aqueous phase (0.40 ± 0.21 mg l^{-1}). As before, there was no statistically significant difference (p = 0.588 for the dodecane phase and p = 0.374 in the aqueous phase) in taxadiene titer when the 6-dEB plasmids were included. The organic:aqueous partitioning coefficient for taxadiene ($K_P^{taxadiene}$) is roughly 10^2, while the same partitioning coefficient for 6-dEB ($K_P^{6\text{-}dEB}$) is at most $10^{-1.65}$. This separation factor ($K_P^{taxadiene}/K_P^{6\text{-}dEB}$) of over 5000 is a conservative estimate due to the roughly 1 mg l^{-1} detection limit for 6-dEB in the HPLC-ELSD system used here.

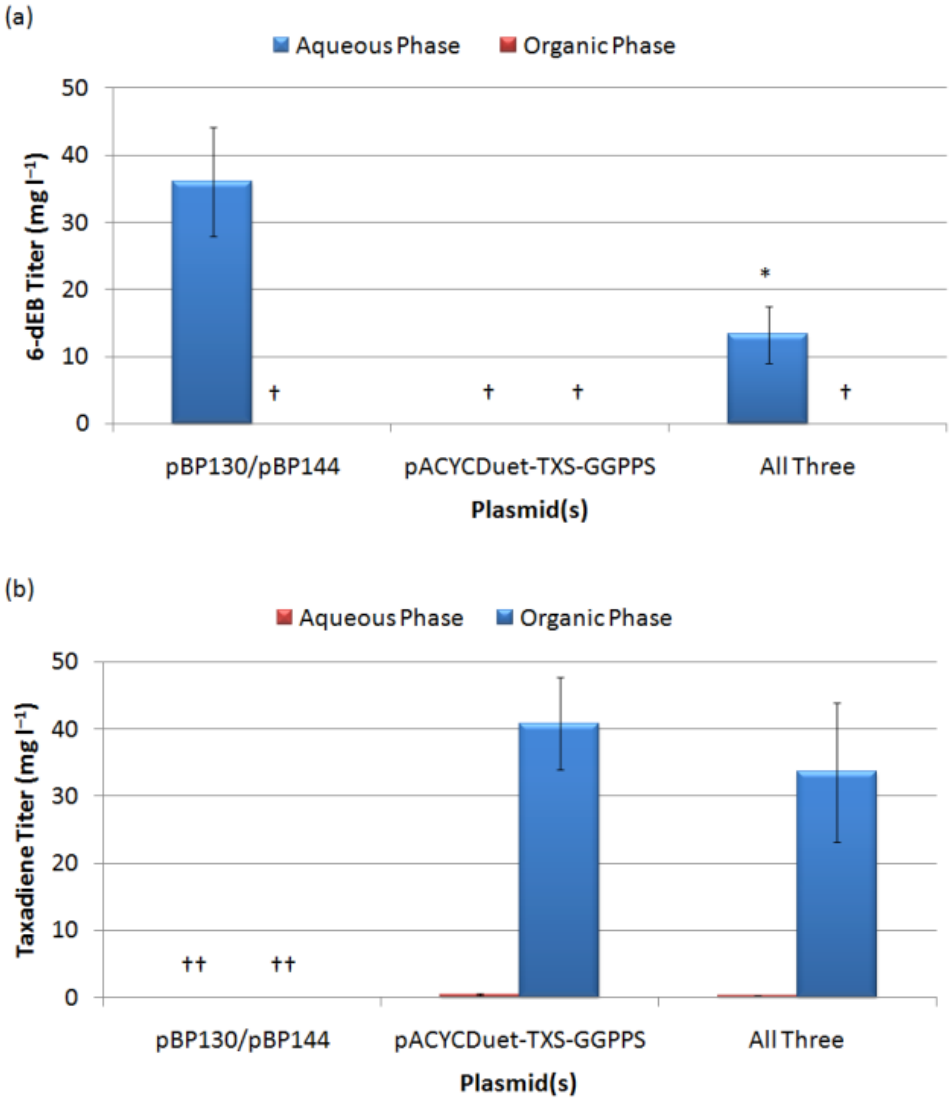

Figure 53 Co-production of 6-dEB and taxadiene in a dual-phase system. Production of (a) 6-dEB and (b) taxadiene in YW23^{sfp} in the two-phase shake flask cultures under the same plasmid combinations as before. The concentration in both the aqueous and the organic phases are shown. Error bars represent ± one standard deviation of three replicates. * indicates statistically significant results at $p < 0.05$ (as determined by paired Student's *t*-tests. † indicates not detectable at a limit of detection of 1 mg l^{-1}. †† indicates not detectable at a limit of detection of 0.1 mg l^{-1}.

Effect of Precursor Supplementation on 6-dEB and Taxadiene Production

The exogenous feeding of propionate is required for >1 mg l^{-1} production of 6-dEB. It was reasoned that this was one of the reasons why the 6-dEB titer was much higher than that of taxadiene in the single phase system (Figure 52). As a result, the effect of the addition of 20 mM propionate, 20 mM pyruvate, or 20 mM propionate and pyruvate was tested on 6-dEB and taxadiene titers. Pyruvate, as illustrated in Figure 51, can be considered a precursor of the DXP-based isoprenoid biosynthetic pathway. The dual producer YW23^{sfp}(pBP130/pBP144/pACYCDuet-TXS-GGPPS) was solely used in this study. As shown in Figure 54a, the addition of pyruvate had no effect on 6-dEB titer, while the sole feeding of pyruvate was not able to produce 6-dEB >1 mg l^{-1}. The addition of the precursors in the medium in any combination had no effect on taxadiene titer (ANOVA $p = 0.782$, Figure 54b).

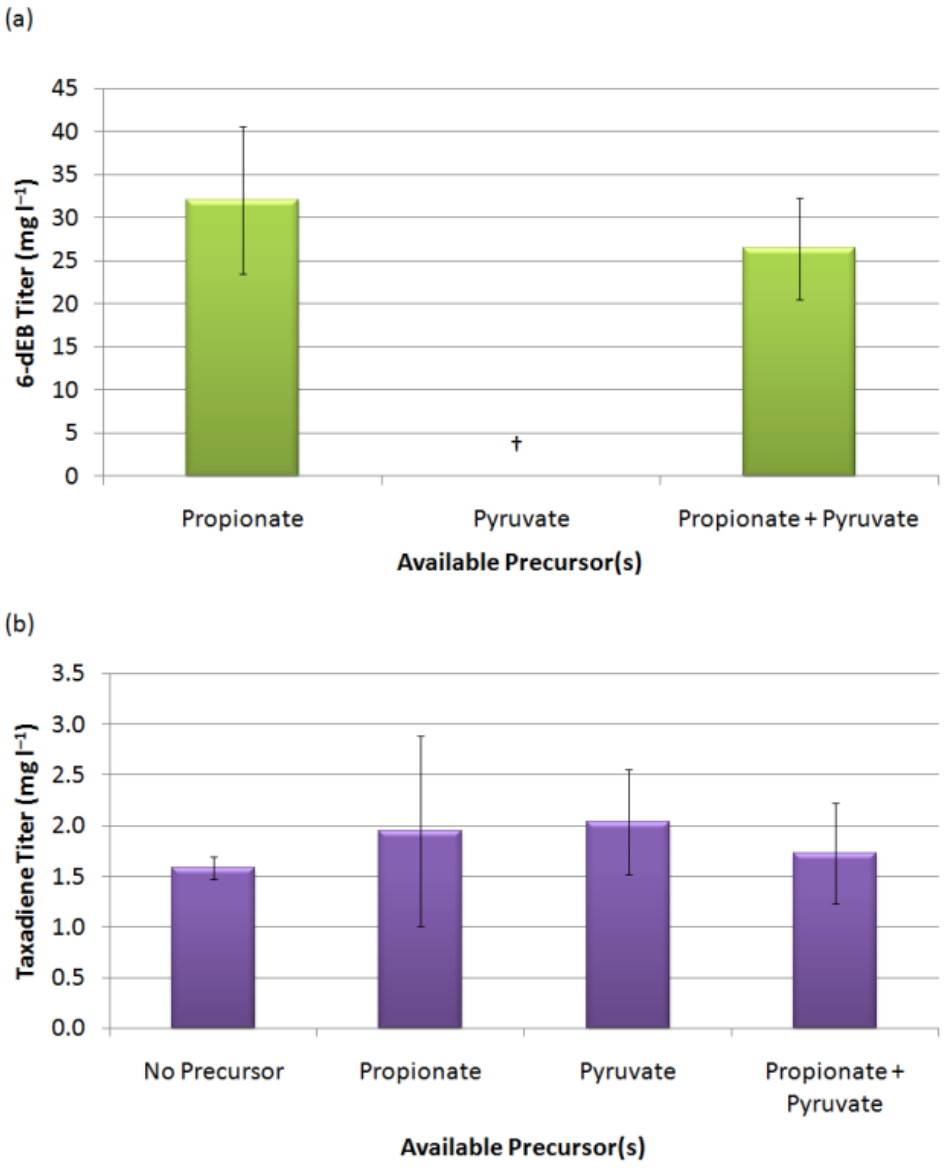

Figure 54 Precursor supplementation effect on 6-dEB and taxadiene production. (a) Production of 6-dEB with the supplementation of 20 mM propionate, 20 mM pyruvate, or 20 mM propionate and 20 mM pyruvate. (b) Production of taxadiene with the supplementation of nothing, 20 mM propionate, 20 mM pyruvate, or 20 mM propionate and 20 mM pyruvate. Error bars represent ± one standard deviation of three replicates. † indicates not detectable at a limit of detection of 1 mg l^{-1}.

Propionate consumption was next analyzed between the single- and two-phase production systems across the three different plasmid combinations (Figure 55a). Interestingly, regardless of whether the 6-dEB production plasmids were used or not, propionate uptake was unaffected. This indicates that the production of 6-dEB has no influence on propionate uptake, at least in the production scenarios here. Across all three single-phase systems, propionate consumption did not vary (ANOVA $p = 0.139$). This same result was observed in the dual phase system as well (ANOVA $p = 0.688$). Pyruvate was completely consumed at the end of the culture period (Figure 55b). Neither propionate nor pyruvate consumption was affected by the presence of the other compound in the culture medium.

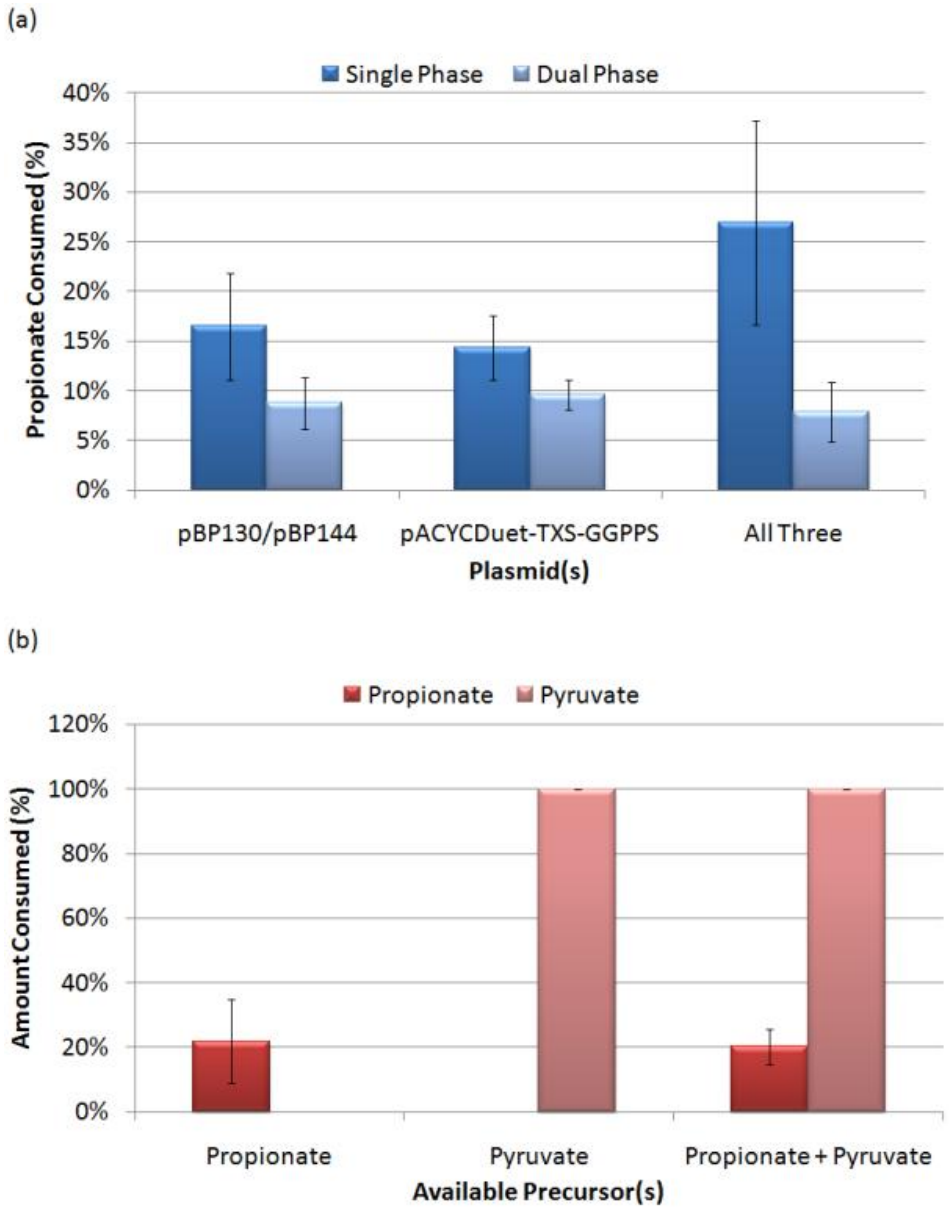

Figure 55 Precursor consumption.
(a) Consumption of propionate in YW23^{sfp} across the three plasmid systems described previously and in the single- vs. dual-phase culture systems. (b) Consumption of propionate and pyruvate by themselves or together in YW23^{sfp}(pBP130/pBP144/pACYCDuet-TXS-GGPPS). Error bars represent ± one standard deviation of three replicates.

Co-Production of 6-dEB and Taxadiene in a Two-Phase Bioreactor

To further test our polyketide and isoprenoid co-production platform, the culture was scaled to a 3 l bioreactor system. In this setting, aeration, temperature, and pH were all controlled, and 1.2 l of culture medium was again charged with 20% *n*-dodecane. A batch run was conducted over the course of 120 hr and sampled every 6-12 hr.

Figure 56a presents cell density and specific product titers as a function of time. Not surprisingly, with better aeration (0.5 VVM) and a higher charge of principle carbon source (45 g l^{-1} glycerol), the cell-density increased from approximately 4 gDCW l^{-1} in the shake-flask (data not shown) to approximately 20 gDCW l^{-1} in stationary phase growth. The specific titers of 6-dEB and taxadiene reached 2.73 and 4.90 mg $gDCW^{-1}$, resulting in raw titers of 32.8 and 94.2 mg l^{-1}, respectively. Figure 56b presents the consumption of glycerol and propionate, as well as the production of acetate, as a function of time. By 100 hr, the glycerol had been completely consumed, resulting in an averaged specific uptake rate of 0.28 mmol $gDCW^{-1}$ hr^{-1}. Interestingly, propionate consumption stalled around 50 hr at approximately 5 mM. This corresponded to an accompanying plateau in 6-dEB production titer. Oddly, this occurred during the middle of exponential growth phase and no changes were observed in glycerol consumption,

taxadiene production, or acetate production. However, in general, the acetate by product titers were quite low increasing to approximately 22 mM when glycerol was still present in the medium and then consumed by the culture after glycerol was exhausted.

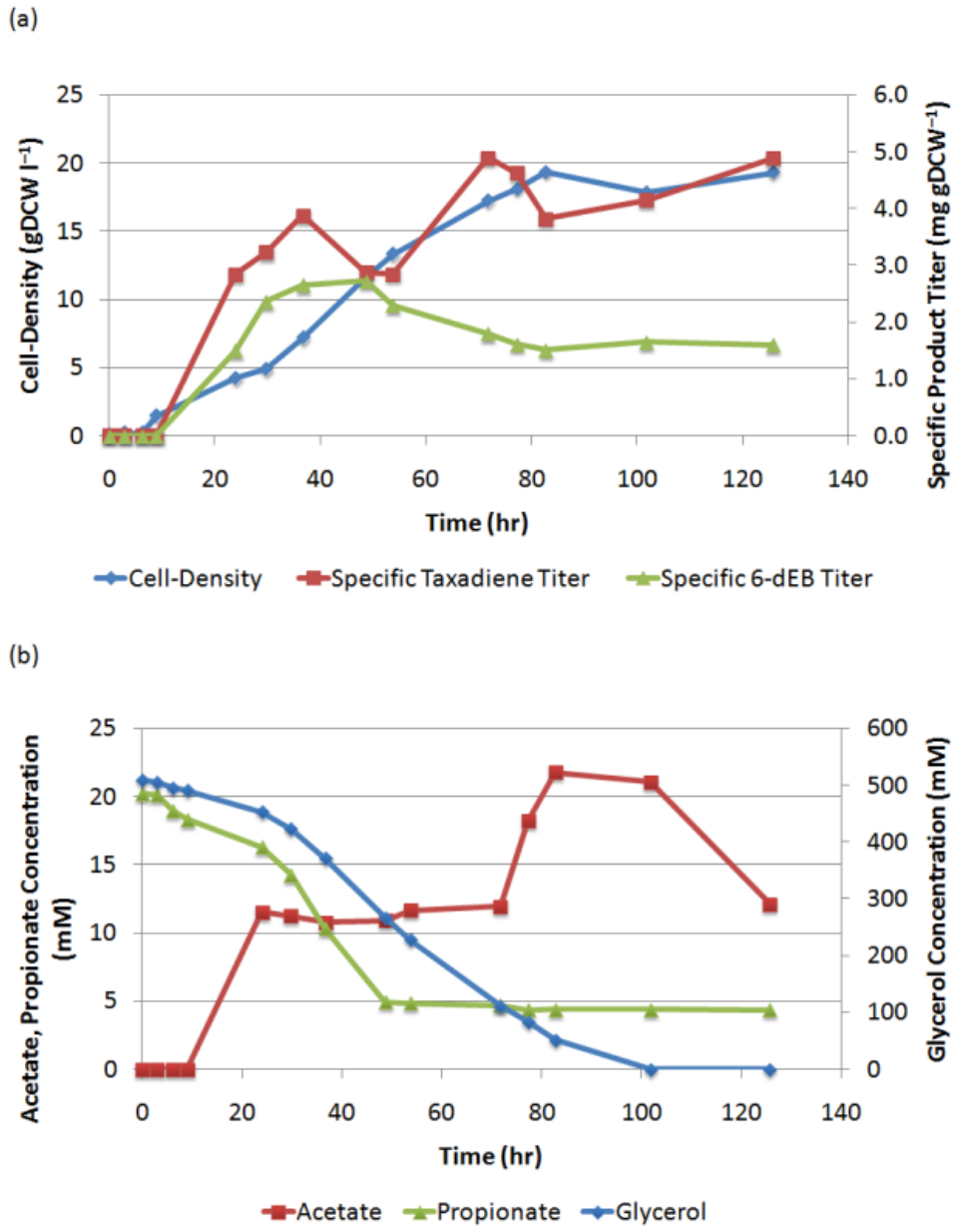

Figure 56 Bioreactor cultivation.
Two-phase bioreactor culture of YW23^{sfp}(pBP130/pBP144/pACYCDuet-TXS-GGPPS) supplemented with 20 mM propionate. (a) Cell-density, specific taxadiene titer, and specific 6-dEB titer are shown as a function of time. (b) Propionate, acetate, and glycerol concentrations in the medium are shown as a function of time)

Discussion

Due to the immense therapeutic potential of natural products, there has been a growing interest in developing efficient and scalable heterologous production processes. Here, it was reported the co-production of a polyketide and an isoprenoid in mg l^{-1} quantities with a two-phase bioreactor. First, a strain of *E. coli* was constructed with a chromosomally over-expressed isoprenoid biosynthetic pathway and a separate chromosomal modification to allow PKS (and NRPS) posttranslational modification and precursor supply. This strain, YW23^{sfp}, was tested for the production of a polyketide (6-dEB) and an isoprenoid (taxadiene) separately and together. While the 6-dEB titer decreased during co-production, the taxadiene titer was unchanged. Then, the production capabilities of this strain in a two-phase culture system was tested. It was previously reported that amorpha-4,11-diene titers were reduced due to compound volatilization in a highly aerated bioreactor (Newman et al. 2006). After confirming the same result for taxadiene, a similar two-phase culture strategy was used here. The 6-dEB titer was unchanged with the addition of the second phase, while the taxadiene titer was greatly improved. The substantial difference in the organic:aqueous partitioning coefficients for taxadiene and 6-dEB resulted in a separation factor ($K_P^{taxadiene}/K_P^{6\text{-}dEB}$) of greater than 5,000

and allowed the simultaneously production and separation of these two complex natural products.

Next, the effects of exogenous precursors on the simultaneous production of 6-dEB and taxadiene in *E. coli* were tested. Results showed that addition of either pyruvate or propionate, alone or together, had no significant effect on either 6-dEB or taxadiene production. Results did, however, verify that an exogenous supply of propionate is essential for 6-dEB production. Equal amounts of propionate and pyruvate were consumed regardless of the product(s) produced or the quantities in which they were being generated. This phenomenon is particularly noteworthy because it suggests other metabolic pathways for propionate consumption. Scenarios characterized by less production of 6-dEB but equal consumption of propionate clearly suggest that the bacteria are utilizing the substrate for purposes other than 6-dEB production.

Finally, a scaled production scenario was conducted in a bioreactor with a working volume of 1.5 l and consisting of 20% *n*-dodecane. Results verified that production trends observed in the smaller-scale 15-ml shake flasks. Final specific product titers in the 15 ml shake flasks were nearly equivalent to those achieved in the bioreactor for 6-dEB and slightly higher for taxadiene. Drawing parallels to other dual-purpose bioprocesses, the polyketide/isoprenoid

co-production system could be considered a type of consolidated production process. Similar terminology has been coined to describe the efficient conversion of biomass to useful biofuels. In this example, the goal is to simultaneously combine cellulosic biomass degradation with subsequent biofuel fermentation using *S. cerevisiae* as a host (Lynd et al. 2005; van Zyl et al. 2007). While co-production of two compounds of interests may not be suitable in all cases, an economic analysis of the compounds of interest would have to be conducted and assessed for co-production based on drug cost and annual demand.

The concept of co-production could also be extended to other heterologous systems. For example, a genome-minimized strain of *S. avermitilis* was recently constructed and reported for the heterologous production of streptomycin (from *S. griseus*) and cephamycin C (from *S. clavuligerus*) (Komatsu et al. 2010). These titers reached nearly 200 mg l^{-1} and 120 mg l^{-1} for streptomycin and cephamycin C, respectively. The strain was then used to express a codon optimized amorpha-4,11-diene synthase gene (*ads*, from *Artemisia annua*) under the control of the *rpsJ* promoter to produce amorpha-4,11-diene at an estimated titer of 10-30 mg l^{-1} (Komatsu et al. 2010). While this genome-minimized strain was able to produce approximately 10^1-10^2 mg l^{-1} of an aminoglycoside, a ß-lactam, and a sesquiterpene, the co-production of

these compounds was not explored, though presumably this would be feasible.

In addition to a bioprocess production platform, a strain capable of supporting multiple classes of natural products could also be of great utility in drug discovery efforts. It has long been known that actinomycetes (particularly from the soil-dwelling *Streptomyces* genus) produce a vast number of secondary metabolites and that drug discovery through isolating and culturing actinomycetes has proven successful. Daptomycin (marketed as Cubicin® by Cubist Pharmaceuticals) (Baltz 2008) was isolated from an extended search of almost 10^7 actinomycetes. Once isolated, advances in genome-sequencing technologies and bioinformatic approaches have aided the identification of secondary metabolite gene clusters from those organisms characterized by genome sequencing. For example, the native erythromycin producer, *S. erythraea,* contains sequence information for a predicted 25 polyketide, nonribosomal peptide, and terpene compounds (Oliynyk et al. 2007), even though no products derived from these clusters (besides erythromycin) were indentified on 50 different types of solid and liquid media (Boakes et al. 2004). Similar phenomena have been observed for *S. avermitilis* (Ikeda et al. 2003) and *S. griseus* (Ohnishi et al. 2008). Accessing this potential may involve innovative culturing techniques, which have been used as a

means to stimulate expression of dormant secondary metabolite pathways (Pettit 2009), and this is not necessarily a high-throughput, process-friendly approach. Furthermore, it has been cited that 99% of all microorganisms cannot be cultured with conventional laboratory methods (Li and Vederas 2009)), thus severely limiting current drug screening approaches through these methods. To address these concerns, the concept of metagenomics was introduced (Handelsman et al. 1998). This approach generally consists of the following steps: 1) collection of environmental samples, 2) isolation of genomic DNA from these samples, 3) cloning of the environmental DNA into bacterial artificial chromosome (BAC) vectors, 4) transformation into a heterologous host, and 5) screening for biological activity. This workflow, which can be designed in a high-throughput manner, has been successfully utilized to screen soil DNA for new antimicrobial natural products (MacNeil et al. 2001). Even though *E. coli* is a primary choice for the heterologous host used in metagenomic efforts, a significant shortcoming is the lack of native metabolism needed to support many complex secondary metabolites. For example, even if an entire PKS cluster was captured in a single BAC, a host lacking a functional phosphopantetheinyl transferase (PPTase) would be unable to support subsequent biosynthesis. In addition, wild-type *E. coli* hosts will typically not provide the needed acyl-CoA precursors, such as the

(2*S*)-ethylmalonyl-CoA, and chloroethylmalonyl-CoA extender units described previously, to support complex natural product biosynthesis.

Although not the particular focus of this study, the strain developed here could be used for metagenomic studies, having the ability to now support the production of three classes of natural products (polyketides, non-ribosomal peptides, and isoprenoids). The over-expression of a promiscuous PPTase gene and numerous genes to support PKS and isoprenoid precursor supply should facilitate the production of a wide range of natural products. From a development standpoint, a single strain capable of discovery and production has the potential to greatly decrease strain and process development time.

Chapter 9 – Conclusions & Future Directions

Summary

This dissertation presents the use of systematic and heuristic methods for the engineering of heterologous polyketide (6-dEB) and isoprenoid (taxadiene, lycopene) natural products in *Escherichia coli*. Upon reconstruction of the pathways to produce these molecules, the development and application of these methods were used to probe and understand both basic and applied aspects of the development of heterologous natural product systems. These methods and molecules were investigated in parallel. For polyketide production, initially, genome-scale stoichiometric models were used to evaluate three commonly used heterologous hosts and their ability to support heterologous polyketide biosynthesis under varying environmental conditions (Chapter 3). Further, targets were identified for gene-deletion to improve specific polyketide production, in some cases over 25-fold. Throughout this process, examination of the relevant metabolism surrounding polyketide biosynthesis in *E. coli* revealed that multiple short-chain organic acids could be utilized to support polyketide production. The varying of substrate feeding strategies, the incorporation of pathway over-expressions, heterologous pathways, and/or pathway deletions, revealed an extremely complex phenotype.

In the end, some cases led to improvement in titer around two-fold (Chapter 4).

For isoprenoid production, another type of stoichiometric modeling was used (elementary mode analysis) to identify knockout targets to improve taxadiene titer (Chapter 5). A central limitation identified in Chapter 3 was the inability to thoroughly search genotypic space in a time efficient manner. To address this problem, a genetic algorithm was used as an efficient means to arriving at a global phenotypic optima, which reduced computation time roughly five orders of magnitude. Coupling the genetic algorithm to the elementary mode analysis algorithm allowed for the identification multiple strains capable of producing a high yield of isoprenoid, which were then partially constructed in a high-throughput manner utilizing multiplex automated genome engineering. The two strains constructed showed significant improvements in titer to nearly 80 mg l^{-1} in shake-flask cultures, providing partial experimental verification of the modeling algorithm. The original *E. coli* stoichiometric model and MoMA was reformulated in an effort to identify over-expression targets. Four over-expressions were identified computationally and implemented in the laboratory in a previously engineered strain. While three of the over-expressions showed improvements in titer, the titer improvements were lower than predicted. However, the trend of the

improvement is the same with prediction (*ppk* > *sthA* > *purN*), showing the potential of this modeling methodology for predicating the effect of gene over-expression. Over-expressing individual genes in the isoprenoid biosynthetic pathway revealed that, under tested conditions, the rate-limiting step towards improving titer may lie within the pathway itself, and not upstream of the pathway. While this was being examined, similar heuristic approaches for polyketide production were used for isoprenoid production. By varying promoters of the heterologous genes, strain background, medium formulation, and other bioprocess related conditions (temperature, aeration, *in situ* extraction), heterologous isoprenoid specific titer was improved over 240-fold. Transcriptomic profiling of two high-producing strains identified differences between these two strains, and identified more metabolic engineering targets that could be exploited for further improvement in titer. Figure 57 shows a summary of all of the improvements observed for taxadiene production. In general, the heuristic methods have been more successful at improving titer than the systematic method applied.

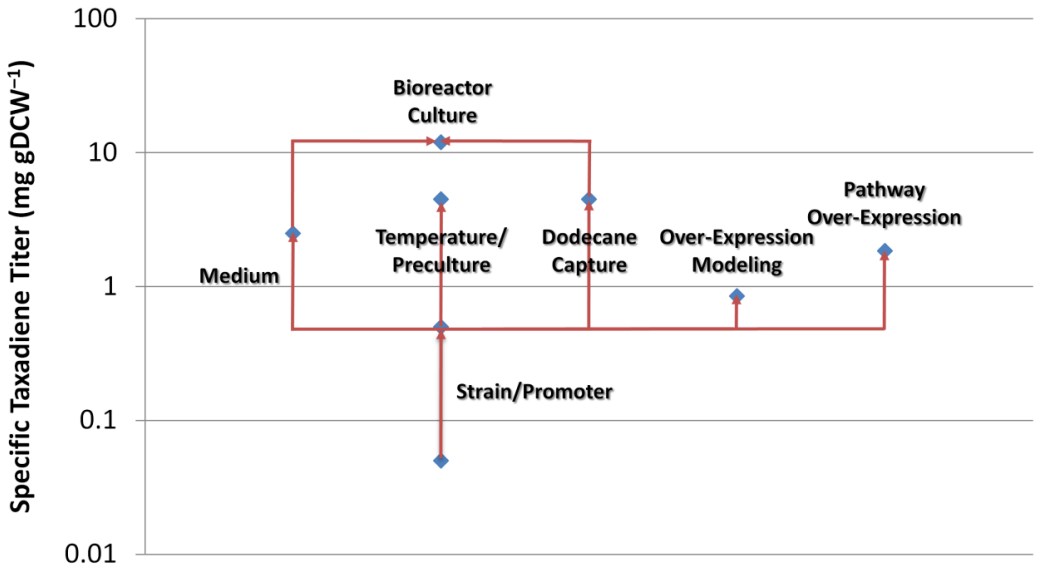

Figure 57 A summary of the methods applied to improving specific taxadiene titer in *E. coli* and their respective means of improvement.

The last chapter of this dissertation presents the integration of these two natural products. A strain capable of simultaneously producing > 10 mg l^{-1} of both polyketide and isoprenoid natural products was developed. In a two-phase bioreactor, the products strongly partitioned in the two phases, theoretically allowing for simplification of downstream purification processes. Given the metabolic and process engineering approaches available to *E. coli*, it is reasonable to believe that these techniques could be utilized to further improve the production titers. This strain presents the first attempt to create a platform strain for producing both polyketide and isoprenoid natural products through a single heterologous host, and could be used for natural product discovery efforts.

Conclusions

The optimization of biological systems for pharmaceutical production is an extremely daunting task. Metabolic networks can exhibit non-linear phenotypic landscape which can produce multiple local optima. The sheer number of variables when it comes to optimizing even simple metabolic networks is computationally intensive, much less experimentally. While using systematic approaches for improving cellular phenotype has been shown to work in some cases, the improvements observed experimentally are significantly smaller than the improvements predicted computationally

(from this dissertation and other cases). The many genome-scale stoichiometric models developed in recent years have been able to predict growth phenotypes fairly well, however, their accuracy tends to decrease the further the strain becomes from the wild-type (as in engineered strains in this dissertation). These models are also strictly stoichiometric in nature, failing to account for genetic, transcriptional, or metabolic regulation. Additionally, to solve these models, a pseudo-steady state assumption has been applied. Most industrial fermentations are conducted in batch (or fed-batch) schemes, which do exhibit steady-states but are transient. As such, these models only describe a portion of an industrial fermentation. The development of genome-scale kinetic models would require an exorbitant number of parameters, and would be extremely difficult to solve. In the end, the *directed* engineering of biological systems based *strictly* on systematic methods is significantly hampered by the current inability to formalize or accurately describe some cellular phenomena that can directly control metabolic flux.

In general, the use of systematic and heuristic methods can both be used to improve heterologous natural product biosynthesis. In this particular dissertation, heuristic methods outperformed systematic methods. However, it is clear that a combination of systematic, heuristic, *and* combinatorial methods could be utilized for improving

cellular phenotype. While ingenious methods such as gTME and directed evolution have been utilized to great results at improving cellular phenotypes such as growth inhibitor tolerance and product formation, they come at a cost of needing a high-throughput screen. At the same time, combinatorial methods such as MAGE still required inputs that were both systematic and heuristic in nature (that is, the genes to be knocked out and over-expressed). Although the first MAGE-consturcted strain was identified also using a phenotypic screen, by decreasing this search space and using statistical calculations, these methods could be applied towards systems that would require a chromatography-based quantification method, such as 6-dEB and taxadiene reported in this dissertation. While MAGE is still considered to be a method that can create combinatorial diversity, as cited previously, a central difference between MAGE and gTME is that, with enough rounds of MAGE, the population will converge to a specific genotype, whereas gTME should never converge to a single, specific genotype (as can be seen pictorially in Figure 58). MAGE can be used for future genome evolution, whereas gTME is suitable for randomized evolution to achieve a wider array of genotypic changes. The method of choice should be used depending on the purpose of specific studies, however, could most certainly be utilized in series for strain development. One might envision using MAGE to remove a

number of genes known to decrease product synthesis, and then use gTME to further tune expression levels across a broader range of genes and improve product titer further.

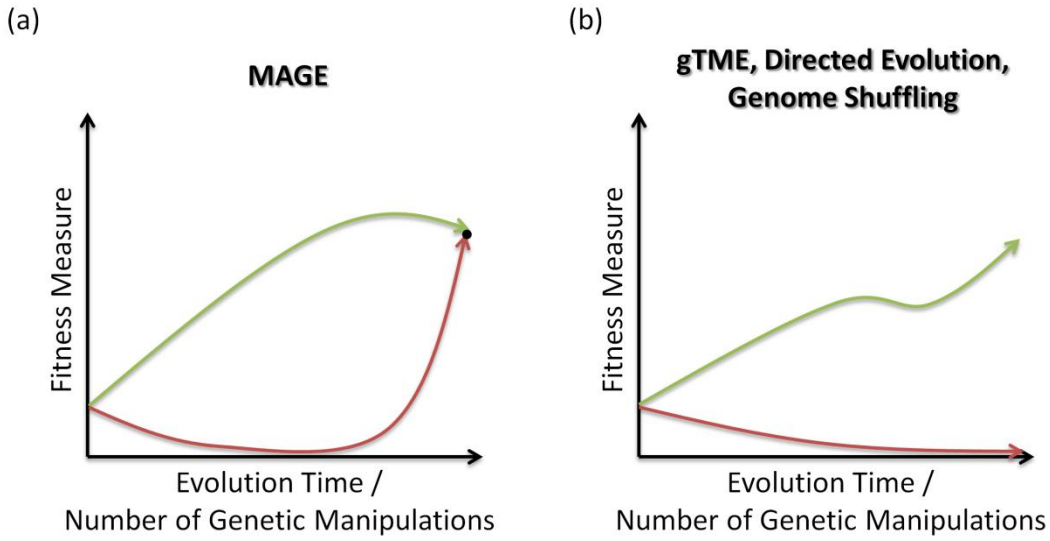

Figure 58 Pictorial fitness trajectories for combinatorial methods.
A pictorial representation of fitness trajectories with respect to evolution time for (a) MAGE and (b) other combinatorial methods such as gTME, directed evolution, or genome shuffling. The green line is meant to signify the trajectory of the individual with the highest fitness value in the population, while the red line is meant to signify that of the lowest fitness value in the population.

Summary of findings:

- Genome-scale metabolic models can be used as a means of investigating cellular phenotype, as well as give insights into genetic deletions as for candidates to removal with respect to heterologous natural product biosynthesis
- For polyketide biosynthesis, many of the identified targets revolved around the same metabolic nodes and were consistent across hosts, principle carbon source, and medium formulation
- Engineering substrate pathways provided as a feasible means to improve titer, however, the effect of engineered substrate pathways is extremely context-specific
- Coupling a genetic algorithm for efficient searching of cellular genotype allows for the high-throughput identification of knockout targets in central metabolic pathways
- MAGE cycling is also highly context-dependent and failed to produce numerous knockout targets, however, partial strain constructions improved isoprenoid titer and yield
- Utilizing MoMA for over-expression target identification appears to be feasible, however, the effect of the over-expressed targets is diminished by pathway bottlenecking

- Culture temperature has a significant effect on cellular phenotype as it relates to heterologous polyketide and isoprenoid production
- Model heterologous polyketides and isoprenoids can be simultaneously produced and partitioned in a two-phase bioreactor
- Heuristic methods are generally more powerful at improving heterologous natural product biosynthesis. However, utilizing systematic methods can provide insight into genetic targets to be implemented for engineering.

Recommendations for Future Work

There are several arenas of future work that could be extended from the results and findings presented in this dissertation. While 6-dEB and DEBS have served as the model polyketide and PKS, respectively, the prospects for *E. coli*-derived erythromycin A production remains limited. The true potential for harnessing *E. coli*-derived production is more likely in the rapid production of erythromycin derivatives for potential next-generation antibiotics. The genetic strategies available to *E. coli* have allowed for production of the full erythromycin A molecule, as well as glyco-randomized versions of this molecule (Zhang et al. 2010b). Of course, there are many other polyketide natural products which have prospects for industrialized

production through *E. coli*. Particularly, the emerging area of marine natural products have produced some interesting evidence that are well suited for heterologous production in *E. coli*. An example of this is salinosporamide, a potent proteasome inhibitor used as an anticancer agent, originally produced by *Salinispora tropica* (Gulder and Moore 2009). While *S. tropica* can produce salinosporamide on the order of 70-100 mg l^{-1}, its potential for usage as a drug has been highly touted (it is in phase I clinical trials only three years after discovery) (Eustaquio et al. 2009; Liu et al. 2009). The investigation of the scalability of bioreactor processes based on marine actinomycetes has been fairly limited as compared to more conventional microbes, and with complete genetic information available on salinosporamide biosynthesis, it seems ideally suited for heterologous production.

In terms of utilizing FBA and MoMA based stoichiometric modeling, there are certainly questions to be addressed in terms of its applicability to batch cultures and perhaps growth in complex media. The knockout and over-expression targets identified in this study are growth-associated, meaning that they were identified as improving product titer while the cells are growing. However, most industrial batch fermentations run beyond the exponential growth phase into a stationary phase where the net specific growth rate is zero. How these results are impacted by the change in steady state is yet to be known,

and may be complicated by the optimality assumptions used by FBA and MoMA. The development of genome-scale dynamic models has been addressed in a largely theoretical manner (Jamshidi and Palsson 2008), while smaller-scale dynamic models have been developed and applied to industrial antibody production from CHO cells (Nolan and Lee 2010). However, they have not been used for the identification of genetic modifications for improving cellular phenotype.

Extension of these methods to other heterologous host systems is also entirely plausible. While *E. coli* shows many advantages when serving as a heterologous host, it is likely not a universal solution to heterologous biosynthesis (for natural products or otherwise). Cellular transport of certain substrate or product metabolites may not be a rate-limiting step for a native producer (due to specific enzymatic transporters), however, could be for *E. coli*. Functional expression of cytochrome P450 enzymes is also particularly challenging in *E. coli*, often requiring significant engineering. First, the absence of cytochrome P450 reductases (required for electron transfer in these reactions) in *E. coli*, and secondly, the absence of an endoplasmic reticulum for these membrane-bound cytochrome P450's, both make this a daunting task. While they have been used to produce roughly 100 mg l^{-1} titers of functionalized isoprenoids in *E. coli* (Ajikumar et al. 2010; Chang et al. 2007), it has been shown that (after GGPP) eight of

the roughly twenty enzymatic steps needed to produce Taxol are catalyzed by cytochrome P450 enzymes. To overcome this limitation, some groups have utilized yeast systems due to the fact that they possess the aforementioned requirements for successful P450 expression and functionality (Ro et al. 2006). Otherwise, engineering an oxygen transport system, such as expression of the *Vitreoscilla* sp. hemoglobin gene, has been shown to improve the titer of a variety of products in both native and heterologous systems (Andersson et al. 2000; Bollinger et al. 2001; Chen et al. 2007; Farres et al. 2005; Frey et al. 2000; Frey and Kallio 2003; Hart et al. 1994; Hofmann et al. 2009; Kallio et al. 2008; Khosla and Bailey 1988; Khosla and Bailey 1989a; Khosla and Bailey 1989b; Khosla et al. 1990; Koskenkorva et al. 2006; Villarreal et al. 2008), and has been sometimes attributed to improved cytochrome P450 activity.

A central tenet to a chemical engineer, or more specifically a chemical reaction engineer, is that an overall process (or overall reaction rate) is dictated by the slowest step (or single reaction) in the overall process. While biological systems are amazing chemical reactors, their inherent impressive robustness can often lead to problems in accurate prediction of phenotype with regards to genetic and environmental perturbations. With the numerous cellular components and spatial and temporal scales, identifying a rate-

limiting step with respect to product formation in a biological system is an extremely daunting task. For example, the knockout target simulations inherently assume that product flux is substrate limited, such that any improvement in flux to this pathway would improve titer. While the modeling results implemented saw small, incremental improvements in titer, the systems in question could perhaps be kinetically-limited, such that protein engineering strategies might improve product titer to a greater extent. Speaking more generally, future work in this field should be dedicated to more thorough systems biological analyses (integrated transcriptomics, proteomics, and metabolomics studies) to generate more information to either integrate into cellular models or serve as a means for further engineering.

Appendix

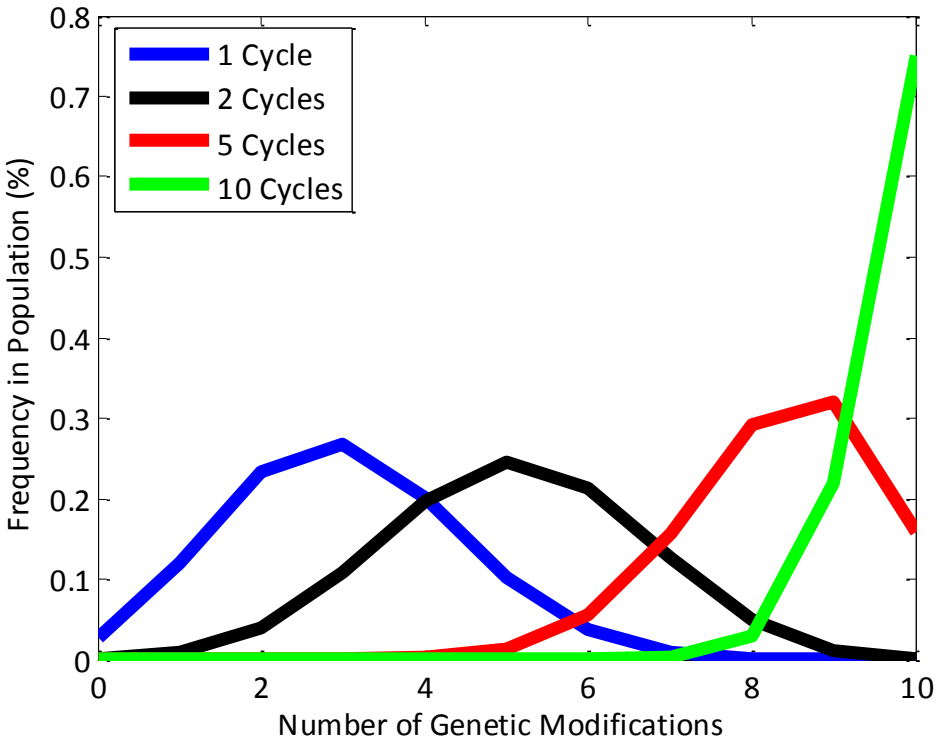

Figure 59 Predicted genetic distribution of a population of cells that have undergone 10 simultaneous genetic manipulations by MAGE.
This assumes a 30% replacement efficiency and follows Equation 12.

Equation 12

$$P(K,N) = \frac{K!}{j!\,(K-j)!}[(1-M)^N]^{K-j}[1-(1-M)^N]^j$$

Equation 12 describes the probability (P) of number of genetic modifications (j) as a function of the total number of modifications (K), replacement efficiency (M = 0.3), and number of MAGE cycles (N). Described in (Wang et al. 2009).

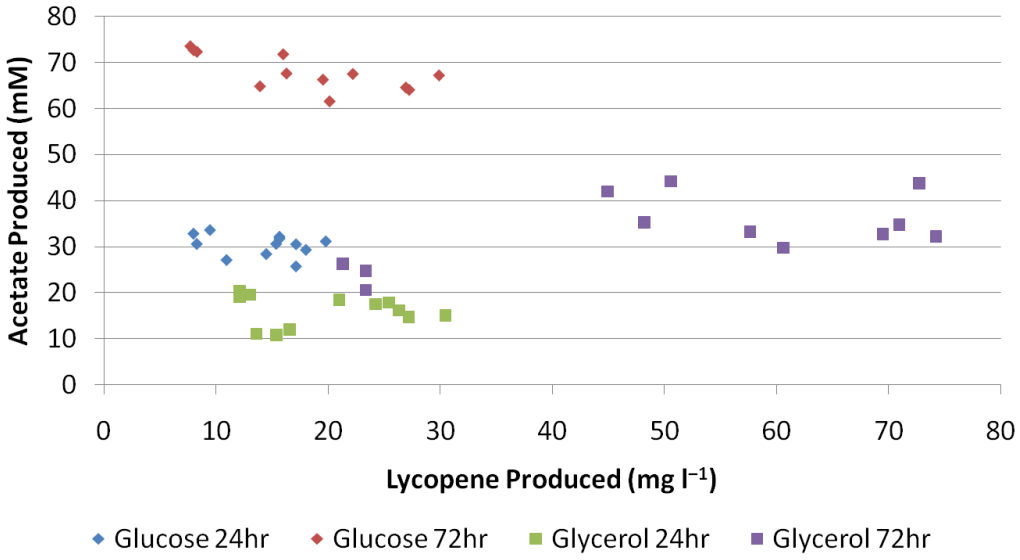

Figure 60 Scatter plot of the lycopene produced versus acetate produced for all cultures after 24hr and 72hr.

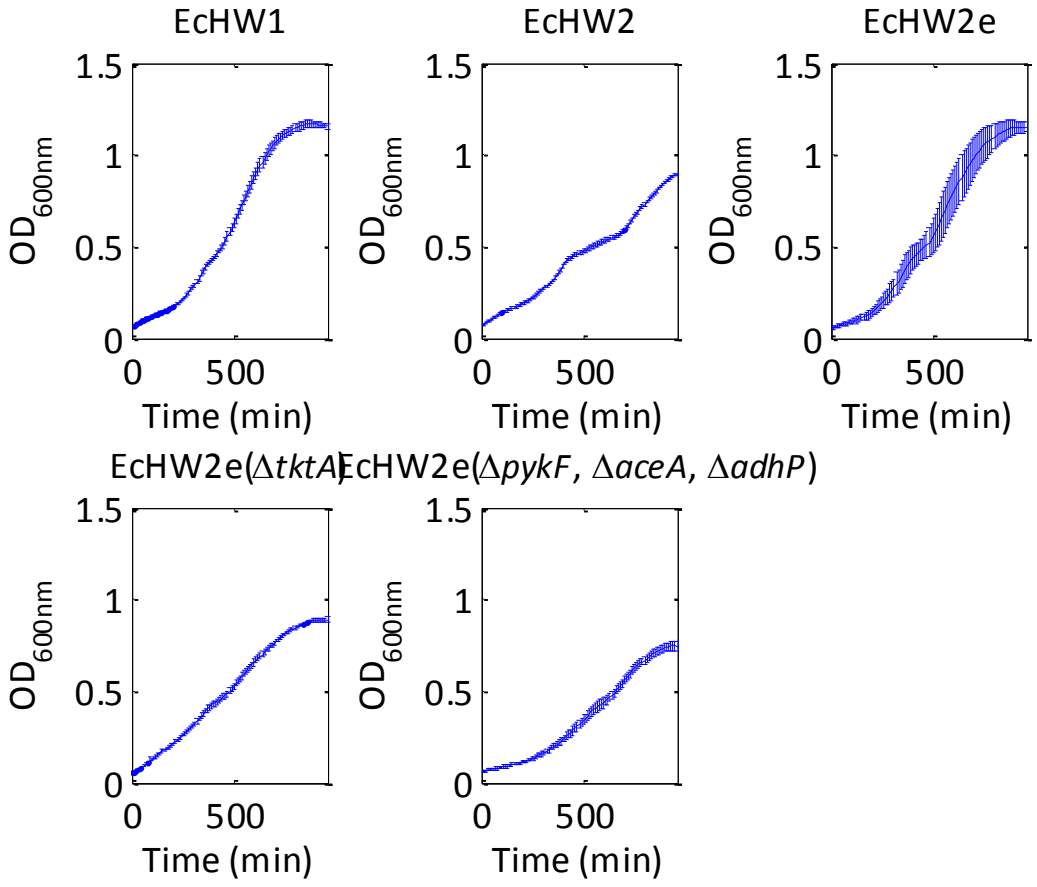

Figure 61 Microplate growth data.
Cell-density (OD_{600nm}) vs. time plots for the MAGE-cycled strains grown in 300 µl microplate format in M9 medium with 15 g l^{-1} glucose. Error bars represent ± one standard deviation of four replicates.

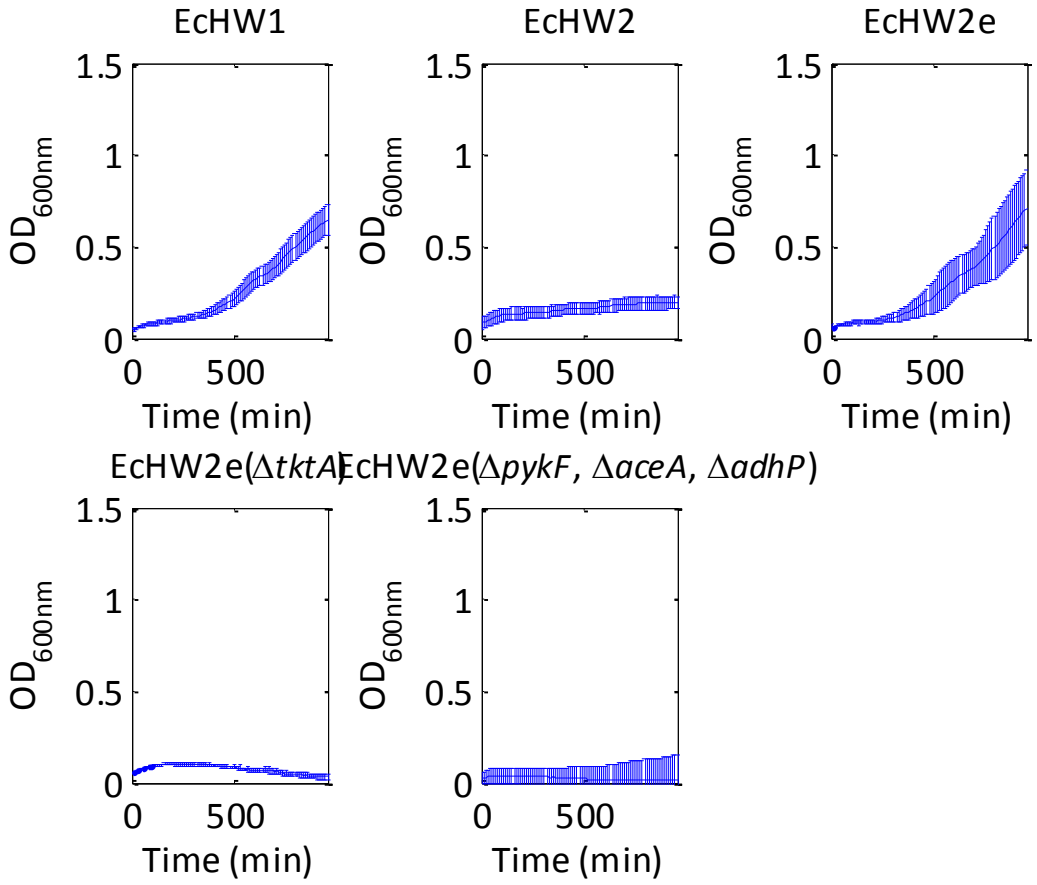

Figure 62 Microplate growth data.
Cell-density (OD_{600nm}) vs. time plots for the MAGE-cycled strains grown in 300 µl microplate format in M9 medium with 15 g l^{-1} glycerol. Error bars represent ± one standard deviation of four replicates.

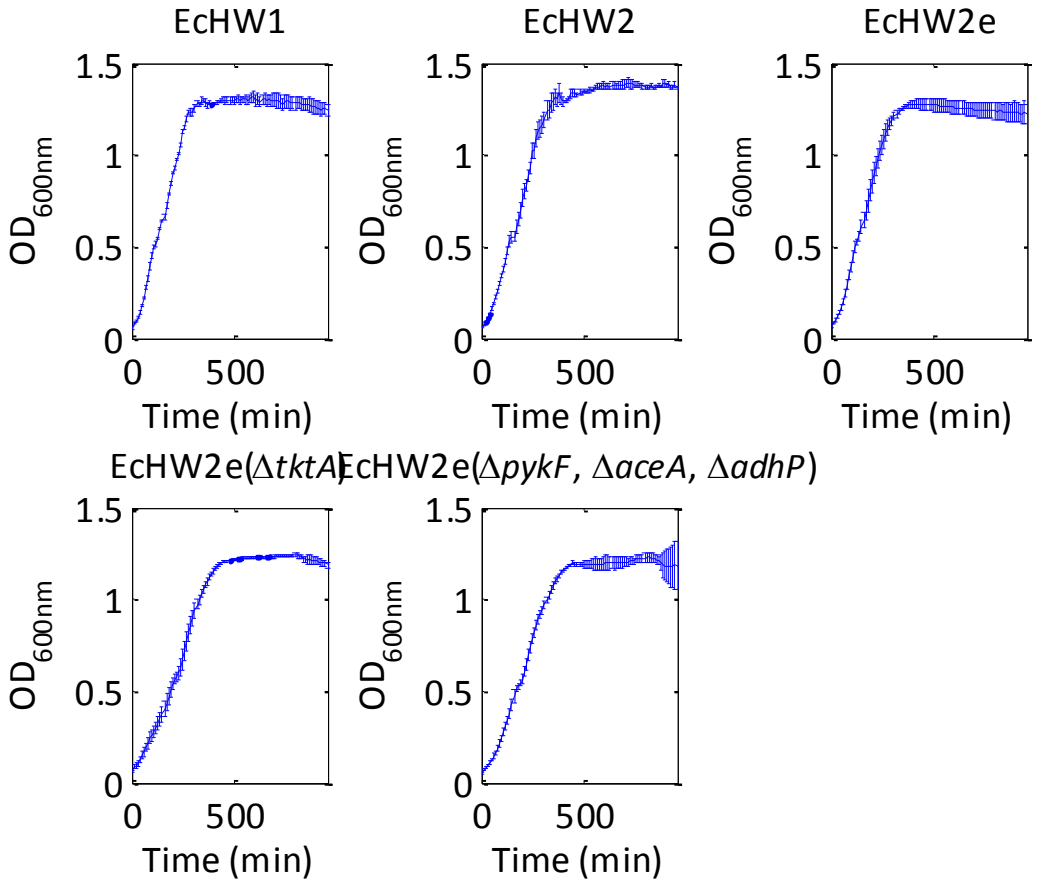

Figure 63 Microplate growth data.
Cell-density (OD_{600nm}) vs. time plots for the MAGE-cycled strains grown in 300 µl microplate format in production medium with 15 g l^{-1} glucose. Error bars represent ± one standard deviation of four replicates.

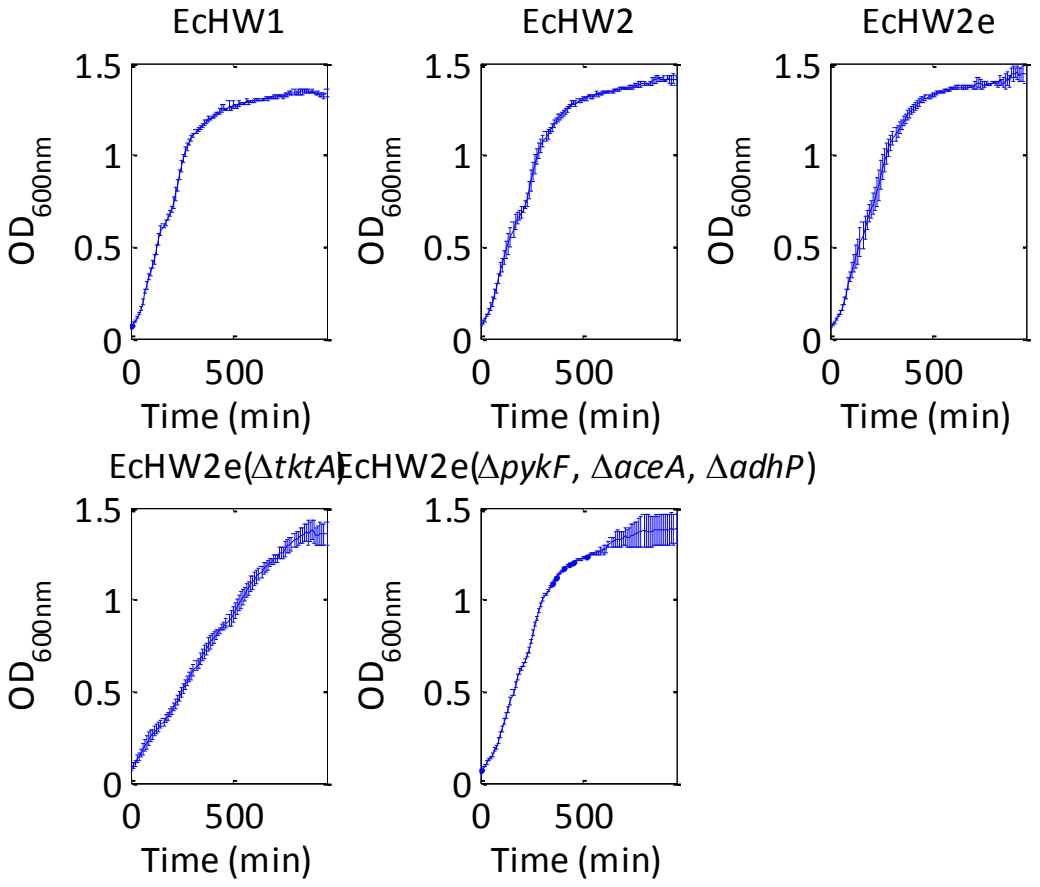

Figure 64 Microplate growth data.
Cell-density (OD_{600nm}) vs. time plots for the MAGE-cycled strains grown in 300 µl microplate format in production medium with 15 g l^{-1} glycerol. Error bars represent ± one standard deviation of four replicates.

Works Cited

AbuOun M, Suthers PF, Jones GI, Carter BR, Saunders MP, Maranas CD, Woodward MJ, Anjun MF. 2009. Genome scale reconstruction of a Salmonella metabolic model: comparison of similarity and differences with a commensal Escherichia coli strain. J Biol Chem.

Adrio JL, Demain AL. 2006. Genetic improvement of processes yielding microbial products. FEMS Microbiol Rev 30(2):187-214.

Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. Nature 422(6928):198-207.

Aggarwal K, Lee KH. 2003. Functional genomics and proteomics as a foundation for systems biology. Brief Funct Genomic Proteomic 2(3):175-84.

Ahsanul Islam M, Edwards EA, Mahadevan R. 2010. Characterizing the metabolism of dehalococcoides with a constraint-based model. PLoS Comput Biol 6(8).

Ajikumar PK, Tyo K, Carlsen S, Mucha O, Phon TH, Stephanopoulos G. 2008. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. Mol Pharm 5(2):167-90.

Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330(6000):70-4.

Aldor IS, Krawitz DC, Forrest W, Chen C, Nishihara JC, Joly JC, Champion KM. 2005. Proteomic profiling of recombinant Escherichia coli in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. Appl Environ Microbiol 71(4):1717-28.

Alm E, Arkin AP. 2003. Biological networks. Curr Opin Struct Biol 13(2):193-202.

Almaas E. 2007. Biological impacts and context of network theory. J Exp Biol 210(Pt 9):1548-58.

Alper H. 2006. Development of systematic and combinatorical approaches for the metabolic engineering of microorganisms. Cambridge, MA: Massachusetts Institute of Technology. 262 p.

Alper H, Fischer C, Nevoigt E, Stephanopoulos G. 2005a. Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102(36):12678-83.

Alper H, Jin YS, Moxley JF, Stephanopoulos G. 2005b. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. Metab Eng 7(3):155-64.

Alper H, Miyaoku K, Stephanopoulos G. 2005c. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23(5):612-6.

Alper H, Miyaoku K, Stephanopoulos G. 2006a. Characterization of lycopene-overproducing E. coli strains in high cell density fermentations. Appl Microbiol Biotechnol 72(5):968-74.

Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. 2006b. Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314(5805):1565-8.

Alper H, Stephanopoulos G. 2007. Global transcription machinery engineering: A new approach for improving cellular phenotype. Metab Eng 9(3):258-267.

Alper H, Stephanopoulos G. 2008. Uncovering the gene knockout landscape for improved lycopene production in E. coli. Appl Microbiol Biotechnol 78(5):801-10.

Alper H, Stephanopoulos G. 2009. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat Rev Microbiol 7(10):715-23.

An JH, Kim YS. 1998. A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in Rhizobium trifolii--cloning, sequencing, and expression of the enzymes in Escherichia coli. Eur J Biochem 257(2):395-402.

An JH, Lee GY, Jung JW, Lee W, Kim YS. 1999. Identification of residues essential for a two-step reaction by malonyl-CoA synthetase from Rhizobium trifolii. Biochem J 344 Pt 1:159-66.

Andersen MR, Nielsen ML, Nielsen J. 2008. Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. Mol Syst Biol 4:178.

Andersson CI, Holmberg N, Farres J, Bailey JE, Bulow L, Kallio PT. 2000. Error-prone PCR of Vitreoscilla hemoglobin (VHb) to support the growth of microaerobic Escherichia coli. Biotechnol Bioeng 70(4):446-55.

Anesiadis N, Cluett WR, Mahadevan R. 2008. Dynamic metabolic engineering for increasing bioprocess productivity. Metab Eng 10(5):255-66.

Arikawa Y, Kuroyanagi T, Shimosaka M, Muratsubaki H, Enomoto K, Kodaira R, Okazaki M. 1999. Effect of gene disruptions of the TCA cycle on production of succinic acid in Saccharomyces cerevisiae. J Biosci Bioeng 87(1):28-36.

Arita M. 2009. What can metabolomics learn from genomics and proteomics? Curr Opin Biotechnol 20(6):610-5.

Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, Nielsen J. 2009. Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. Metab Eng.

Askenazi M, Driggers EM, Holtzman DA, Norman TC, Iverson S, Zimmer DP, Boers ME, Blomquist PR, Martinez EJ, Monreal AW and others. 2003. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. Nat Biotechnol 21(2):150-6.

Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, Chou KJ, Hanai T, Liao JC. 2008a. Metabolic engineering of Escherichia coli for 1-butanol production. Metab Eng 10(6):305-11.

Atsumi S, Hanai T, Liao JC. 2008b. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451(7174):86-9.

Atsumi S, Liao JC. 2008a. Directed evolution of Methanococcus jannaschii citramalate synthase for biosynthesis of 1-propanol and 1-butanol by Escherichia coli. Appl Environ Microbiol 74(24):7802-8.

Atsumi S, Liao JC. 2008b. Metabolic engineering for advanced biofuels production from Escherichia coli. Curr Opin Biotechnol 19(5):414-9.

Baart GJ, Zomer B, de Haan A, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2007. Modeling Neisseria meningitidis metabolism: from genome to metabolic fluxes. Genome Biol 8(7):R136.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006 0008.

Baerga-Ortiz A, Popovic B, Siskos AP, O'Hare HM, Spiteller D, Williams MG, Campillo N, Spencer JB, Leadlay PF. 2006. Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. Chem Biol 13(3):277-85.

Bailey JE. 1991. Toward a science of metabolic engineering. Science 252(5013):1668-1675.

Bailey JE. 2001. Reflections on the scope and the future of metabolic engineering and its connections to functional genomics and drug discovery. Metab Eng 3(2):111-4.

Bailey JE, Ollis DF. 1986. Biochemical engineering fundamentals. New York: McGraw-Hill. xxi, 984 p. p.

Baltz RH. 2008. Renaissance in antibacterial discovery from actinomycetes. Curr Opin Pharmacol 8(5):557-63.

Baltz RH. 2010. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol.

Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY. 2006. Systems biology as a foundation for genome-scale synthetic biology. Curr Opin Biotechnol 17(5):488-92.

Barrett JF. 2005. Can biotech deliver new antibiotics? Curr Opin Microbiol 8(5):498-503.

Beard DA, Liang SD, Qian H. 2002. Energy balance for analysis of complex metabolic networks. Biophys J 83(1):79-86.

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc 2(3):727-38.

Becker SA, Palsson BO. 2005. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. BMC Microbiol 5:8.

Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon AP, Taylor S, Campobasso N, Chiu HJ, Kinsland C, Reddick JJ and others. 1999. Thiamin biosynthesis in prokaryotes. Arch Microbiol 171(5):293-300.

Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S. 2008. Structural robustness of metabolic networks with respect to multiple knockouts. J Theor Biol 252(3):433-41.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D and others. 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417(6885):141-7.

Beste DJ, Hooper T, Stewart G, Bonde B, Avignone-Rossa C, Bushell ME, Wheeler P, Klamt S, Kierzek AM, McFadden J. 2007. GSMN-TB: a web-based genome-scale network model of Mycobacterium tuberculosis metabolism. Genome Biol 8(5):R89.

Besumbes O, Sauret-Gueto S, Phillips MA, Imperial S, Rodriguez-Concepcion M, Boronat A. 2004. Metabolic engineering of isoprenoid biosynthesis in Arabidopsis for the production of taxadiene, the first committed precursor of Taxol. Biotechnol Bioeng 88(2):168-75.

Binz TM, Wenzel SC, Schnell HJ, Bechthold A, Muller R. 2008. Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster by using red/ET recombineering. Chembiochem 9(3):447-54.

Blanchard S, Thorson JS. 2006. Enzymatic tools for engineering natural product glycosylation. Curr Opin Chem Biol 10(3):263-71.

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF and others. 1997. The complete genome sequence of Escherichia coli K-12. Science 277(5331):1453-74.

Boakes S, Oliynyk M, Cortes J, Bohm I, Rudd BA, Revill WP, Staunton J, Leadlay PF. 2004. A new modular polyketide synthase in the erythromycin producer Saccharopolyspora erythraea. J Mol Microbiol Biotechnol 8(2):73-80.

Boghigian BA, Lee K, Pfeifer BA. 2010. Computational analysis of phenotypic space in heterologous polyketide biosynthesis--applications to Escherichia coli, Bacillus subtilis, and Saccharomyces cerevisiae. J Theor Biol 262(2):197-207.

Boghigian BA, Pfeifer BA. 2008. Current status, strategies, and potential for the metabolic engineering of heterologous polyketides in Escherichia coli. Biotechnol Lett 30(8):1323-30.

Bollinger CJ, Bailey JE, Kallio PT. 2001. Novel hemoglobins to enhance microaerobic growth and substrate utilization in Escherichia coli. Biotechnol Prog 17(5):798-808.

Borodina I, Krabben P, Nielsen J. 2005. Genome-scale analysis of Streptomyces coelicolor A3(2) metabolism. Genome Res 15(6):820-9.

Borodina I, Siebring J, Zhang J, Smith CP, van Keulen G, Dijkhuizen L, Nielsen J. 2008. Antibiotic overproduction in Streptomyces coelicolor A3 2 mediated by phosphofructokinase deletion. J Biol Chem 283(37):25186-99.

Boyle NR, Morgan JA. 2009. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. BMC Syst Biol 3:4.

Brotz-Oesterhelt H, Bandow JE, Labischinski H. 2005. Bacterial proteomics and its role in antibacterial drug discovery. Mass Spectrom Rev 24(4):549-65.

Brown AC, Eberl M, Crick DC, Jomaa H, Parish T. 2010. The nonmevalonate pathway of isoprenoid biosynthesis in Mycobacterium tuberculosis is essential and transcriptionally regulated by Dxs. J Bacteriol 192(9):2424-33.

Burgard AP, Maranas CD. 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng 82(6):670-7.

Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84(6):647-57.

Cane DE, Walsh CT, Khosla C. 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. Science 282(5386):63-8.

Cann AF, Liao JC. 2008. Production of 2-methyl-1-butanol in engineered Escherichia coli. Appl Microbiol Biotechnol 81(1):89-98.

Castellanos M, Kushiro K, Lai SK, Shuler ML. 2007. A genomically/chemically complete module for synthesis of lipid membrane in a minimal cell. Biotechnol Bioeng 97(2):397-409.

Castellanos M, Wilson DB, Shuler ML. 2004. A modular minimal cell model: purine and pyrimidine transport and metabolism. Proc Natl Acad Sci U S A 101(17):6681-6.

Chan YA, Podevels AM, Kevany BM, Thomas MG. 2009. Biosynthesis of polyketide synthase extender units. J Nat Prod 26:90-114.

Chang MC, Eachus RA, Trieu W, Ro DK, Keasling JD. 2007. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. Nat Chem Biol 3(5):274-7.

Chang Y, Suthers PF, Maranas CD. 2008. Identification of optimal measurement sets for complete flux elucidation in metabolic flux analysis experiments. Biotechnol Bioeng 100(6):1039-49.

Charpentier B, Bardey V, Robas N, Branlant C. 1998. The EIIGlc protein is involved in glucose-mediated activation of Escherichia coli gapA and gapB-pgk transcription. J Bacteriol 180(24):6476-83.

Chemler JA, Fowler ZL, McHugh KP, Koffas MA. 2010. Improving NADPH availability for natural product biosynthesis in Escherichia coli by metabolic engineering. Metab Eng 12(2):96-104.

Chemler JA, Koffas MA. 2008. Metabolic engineering for plant natural product biosynthesis in microbes. Curr Opin Biotechnol 19(6):597-605.

Chen H, Chu J, Zhang S, Zhuang Y, Qian J, Wang Y, Hu X. 2007. Intracellular expression of Vitreoscilla hemoglobin improves S-adenosylmethionine production in a recombinant Pichia pastoris. Appl Microbiol Biotechnol 74(6):1205-12.

Chen K, Baran PS. 2009. Total synthesis of eudesmane terpenes by site-selective C-H oxidations. Nature 459(7248):824-8.

Chen Y, Deng W, Wu J, Qian J, Chu J, Zhuang Y, Zhang S, Liu W. 2008. Genetic modulation of the overexpression of tailoring genes eryK and eryG leading to the improvement of erythromycin A purity and production in Saccharopolyspora erythraea fermentation. Appl Environ Microbiol 74(6):1820-8.

Chen Z, Liu H, Zhang J, Liu D. 2010. Elementary mode analysis for the rational design of efficient succinate conversion from glycerol by Escherichia coli. J Biomed Biotechnol 2010:518743.

Cherepanov PP, Wackernagel W. 1995. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158(1):9-14.

Chiang CJ, Chen PT, Chao YP. 2008. Replicon-free and markerless methods for genomic insertion of DNAs in phage attachment sites and controlled expression of chromosomal genes in Escherichia coli. Biotechnol Bioeng 101(5):985-95.

Chng C, Lum AM, Vroom JA, Kao CM. 2008. A key developmental regulator controls the synthesis of the antibiotic erythromycin in Saccharopolyspora erythraea. Proc Natl Acad Sci U S A 105(32):11346-51.

Choi HS, Lee SY, Kim TY, Woo HM. 2010. In silico identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76(10):3097-105.

Choi JH, Ryu YW, Park YC, Seo JH. 2009. Synergistic effects of chromosomal ispB deletion and dxs overexpression on coenzyme Q(10) production in recombinant Escherichia coli expressing Agrobacterium tumefaciens dps gene. J Biotechnol.

Chou CP. 2007. Engineering cell physiology to enhance recombinant protein production in Escherichia coli. Appl Microbiol Biotechnol 76(3):521-32.

Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee DY. 2010. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact 9:50.

Church GM. 2005. From systems biology to synthetic biology. Mol Syst Biol 1:2005 0032.

Clardy J, Fischbach MA, Walsh CT. 2006. New antibiotics from bacterial natural products. Nat Biotechnol 24(12):1541-50.

Clardy J, Walsh C. 2004. Lessons from natural molecules. Nature 432(7019):829-37.

Connor MR, Cann AF, Liao JC. 2010. 3-Methyl-1-butanol production in Escherichia coli: random mutagenesis and two-phase fermentation. Appl Microbiol Biotechnol 86(4):1155-64.

Connor MR, Liao JC. 2008. Engineering of an Escherichia coli strain for the production of 3-methyl-1-butanol. Appl Environ Microbiol 74(18):5769-75.

Connor MR, Liao JC. 2009. Microbial production of advanced transportation fuels in non-natural hosts. Curr Opin Biotechnol 20(3):307-15.

Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF. 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. Nature 348(6297):176-8.

Covert MW, Schilling CH, Palsson B. 2001. Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213(1):73-88.

Cox RS, 3rd, Surette MG, Elowitz MB. 2007. Programming gene expression with combinatorial promoters. Mol Syst Biol 3:145.

Cox SJ, Shalel Levanon S, Bennett GN, San KY. 2005. Genetically constrained metabolic flux analysis. Metab Eng 7(5-6):445-56.

Cragg GM, Newman DJ, Snader KM. 1997. Natural products in drug discovery and development. J Nat Prod 60(1):52-60.

Cragg GM, Schepartz SA, Suffness M, Grever MR. 1993. The taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. J Nat Prod 56(10):1657-68.

Cragg GM, Snader KM. 1991. Taxol: the supply issue. Cancer Cells 3(6):233-5.

Cunningham FX, Jr., Sun 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(8):1107-21.

Daniels BC, Chen YJ, Sethna JP, Gutenkunst RN, Myers CR. 2008. Sloppiness, robustness, and evolvability in systems biology. Curr Opin Biotechnol 19(4):389-95.

Danishefsky SJ, Masters JJ, Young WB, Link JT, Snyder LB, Magee TV, Jung DK, Isaacs RCA, Bornmann WG, Alaimo CA and others.

1996. Total synthesis of baccatin III and taxol. Journal of the American Chemical Society 118(12):2843-2859.

Das A, Yoon SH, Lee SH, Kim JY, Oh DK, Kim SW. 2007. An update on microbial carotenoid production: application of recent metabolic engineering tools. Appl Microbiol Biotechnol 77(3):505-12.

Datamonitor. 2008. Monoclonal antibodies update - 2008.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97(12):6640-5.

David H, Ozcelik IS, Hofmann G, Nielsen J. 2008. Analysis of Aspergillus nidulans metabolism at the genome-scale. BMC Genomics 9:163.

Dayem LC, Carney JR, Santi DV, Pfeifer BA, Khosla C, Kealey JT. 2002. Metabolic engineering of a methylmalonyl-CoA mutase-epimerase pathway for complex polyketide biosynthesis in Escherichia coli. Biochemistry 41(16):5193-201.

de Figueiredo LF, Schuster S, Kaleta C, Fell DA. 2009. Can sugars be produced from fatty acids? A test case for pathway analysis tools. Bioinformatics 25(1):152-8.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol 152(2):579-89.

Demain AL. 2006. From natural products discovery to commercialization: a success story. J Ind Microbiol Biotechnol 33(7):486-95.

Denis JN, Greene AE, Guenard D, Gueritte-Voegelein F, Mangatal L, Potier P. 1988. A highly efficient, practical approach to natural taxol. Journal of the American Chemical Society 110(17):5917-5919.

Desai RP, Nielsen LK, Papoutsakis ET. 1999. Stoichiometric modeling of Clostridium acetobutylicum fermentations with non-linear constraints. J Biotechnol 71(1-3):191-205.

Dharmadi Y, Gonzalez R. 2005. A better global resolution function and a novel iterative stochastic search method for optimization of high-

performance liquid chromatographic separation. J Chromatogr A 1070(1-2):89-101.

Dhurjati P, Mahadevan R. 2008. Systems biology: The synergistic interplay between biology and mathematics. Canadian Journal of Chemical Engineering 86(2):127-141.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L. 1991. Modular organization of genes required for complex polyketide biosynthesis. Science 252(5006):675-9.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A 104(6):1777-82.

Durfee T, Nelson R, Baldwin S, Plunkett G, 3rd, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M and others. 2008. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. J Bacteriol 190(7):2597-606.

Durot M, Bourguignon PY, Schachter V. 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev 33(1):164-90.

Edwards JS, Covert M, Palsson B. 2002. Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol 4(3):133-40.

Edwards JS, Ibarra RU, Palsson BO. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19(2):125-30.

Edwards JS, Palsson BO. 2000. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97(10):5528-33.

Eisenreich W, Bacher A, Arigoni D, Rohdich F. 2004. Biosynthesis of isoprenoids via the non-mevalonate pathway. Cell Mol Life Sci 61(12):1401-26.

Ellis HM, Yu D, DiTizio T, Court DL. 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98(12):6742-6.

Engels B, Dahm P, Jennewein S. 2008. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 10(3-4):201-6.

Eppelmann K, Doekel S, Marahiel MA. 2001. Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host Bacillus subtilis. J Biol Chem 276(37):34824-31.

Eubanks LM, Poulter CD. 2003. Rhodobacter capsulatus 1-deoxy-D-xylulose 5-phosphate synthase: steady-state kinetics and substrate binding. Biochemistry 42(4):1140-9.

Eustaquio AS, McGlinchey RP, Liu Y, Hazzard C, Beer LL, Florova G, Alhamadsheh MM, Lechner A, Kale AJ, Kobayashi Y and others. 2009. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. Proc Natl Acad Sci U S A 106(30):12295-300.

Farmer WR, Liao JC. 2001. Precursor balancing for metabolic engineering of lycopene production in Escherichia coli. Biotechnol Prog 17(1):57-61.

Farres J, Rechsteiner MP, Herold S, Frey AD, Kallio PT. 2005. Ligand binding properties of bacterial hemoglobins and flavohemoglobins. Biochemistry 44(10):4125-34.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121.

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. 2009. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7(2):129-43.

Feist AM, Palsson BO. 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat Biotechnol 26(6):659-67.

Feist AM, Scholten JC, Palsson BO, Brockman FJ, Ideker T. 2006. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. Mol Syst Biol 2:2006 0004.

Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN, Timmis KN. 2003. Chaperonins govern growth of Escherichia coli at low temperatures. Nat Biotechnol 21(11):1266-7.

Fischbach MA, Walsh CT, Clardy J. 2008. The evolution of gene collectives: How natural selection drives chemical innovation. Proc Natl Acad Sci U S A 105(12):4601-8.

Fong SS, Marciniak JY, Palsson BO. 2003. Description and interpretation of adaptive evolution of Escherichia coli K-12 MG1655 by using a genome-scale in silico metabolic model. J Bacteriol 185(21):6400-8.

Fong SS, Palsson BO. 2004. Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat Genet 36(10):1056-8.

Fowler ZL, Gikandi WW, Koffas MA. 2009. Increased malonyl coenzyme A biosynthesis by tuning the Escherichia coli metabolic network and its application to flavanone production. Appl Environ Microbiol 75(18):5831-9.

Fox DT, Poulter CD. 2005a. Mechanistic studies with 2-C-methyl-D-erythritol 4-phosphate synthase from Escherichia coli. Biochemistry 44(23):8360-8.

Fox DT, Poulter CD. 2005b. Synthesis and evaluation of 1-deoxy-D-xylulose 5-phosphoric acid analogues as alternate substrates for methylerythritol phosphate synthase. J Org Chem 70(6):1978-85.

Frense D. 2007. Taxanes: perspectives for biotechnological production. Appl Microbiol Biotechnol 73(6):1233-40.

Frey AD, Bailey JE, Kallio PT. 2000. Expression of Alcaligenes eutrophus flavohemoprotein and engineered Vitreoscilla hemoglobin-reductase fusion protein for improved hypoxic growth of Escherichia coli. Appl Environ Microbiol 66(1):98-104.

Frey AD, Kallio PT. 2003. Bacterial hemoglobins and flavohemoglobins: versatile proteins and their impact on microbiology and biotechnology. FEMS Microbiol Rev 27(4):525-45.

Fu J, Wenzel SC, Perlova O, Wang J, Gross F, Tang Z, Yin Y, Stewart AF, Muller R, Zhang Y. 2008. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. Nucleic Acids Res 36(17):e113.

Garcia DE, Baidoo EE, Benke PI, Pingitore F, Tang YJ, Villa S, Keasling JD. 2008. Separation and mass spectrometry in microbial metabolomics. Curr Opin Microbiol.

Gayen K, Gupta M, Venkatesh KV. 2007. Elementary mode analysis to study the preculturing effect on the metabolic state of Lactobacillus rhamnosus during growth on mixed substrates. In Silico Biol 7(2):123-39.

Gayen K, Venkatesh KV. 2006. Analysis of optimal phenotypic space using elementary modes as applied to Corynebacterium glutamicum. BMC Bioinformatics 7:445.

Gershenzon J, Dudareva N. 2007. The function of terpene natural products in the natural world. Nat Chem Biol 3(7):408-14.

Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA. 2008. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformatics 9:43.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M and others. 1996. Life with 6000 genes. Science 274(5287):546, 563-7.

Gokhale RS, Tsuji SY, Cane DE, Khosla C. 1999. Dissecting and exploiting intermodular communication in polyketide synthases. Science 284(5413):482-5.

Gomes de Oliveira Dal'molin C, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. C4GEM - Genome-Scale Metabolic Model to study C4 plant metabolism. Plant Physiol.

Gonzalez-Lergier J, Broadbelt LJ, Hatzimanikatis V. 2005. Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. J Am Chem Soc 127(27):9930-8.

Gonzalez-Lergier J, Broadbelt LJ, Hatzimanikatis V. 2006. Analysis of the maximum theoretical yield for the synthesis of erythromycin precursors in Escherichia coli. Biotechnol Bioeng 95(4):638-44.

Gonzalez O, Oberwinkler T, Mansueto L, Pfeiffer F, Mendoza E, Zimmer R, Oesterhelt D. 2010. Characterization of growth and metabolism of the haloalkaliphile Natronomonas pharaonis. PLoS Comput Biol 6(6):e1000799.

Gruschow S, Chang LC, Mao Y, Sherman DH. 2007. Hydroxyquinone O-methylation in mitomycin biosynthesis. J Am Chem Soc 129(20):6470-6.

Gulder TA, Moore BS. 2009. Chasing the treasures of the sea - bacterial marine natural products. Curr Opin Microbiol 12(3):252-60.

Hahn FM, Hurlburt AP, Poulter CD. 1999. Escherichia coli open reading frame 696 is idi, a nonessential gene encoding isopentenyl diphosphate isomerase. J Bacteriol 181(15):4499-504.

Haller T, Buckel T, Retey J, Gerlt JA. 2000. Discovering new enzymes and metabolic pathways: conversion of succinate to propionate by Escherichia coli. Biochemistry 39(16):4622-9.

Han MJ, Yoon SS, Lee SY. 2001. Proteome analysis of metabolically engineered Escherichia coli producing Poly(3-hydroxybutyrate). J Bacteriol 183(1):301-8.

Hanai T, Atsumi S, Liao JC. 2007. Engineered synthetic pathway for isopropanol production in Escherichia coli. Appl Environ Microbiol 73(24):7814-8.

Hancock WS, Wu SL, Shieh P. 2002. The challenges of developing a sound proteomics strategy. Proteomics 2(4):352-9.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5(10):R245-9.

Hart RA, Kallio PT, Bailey JE. 1994. Effect of biosynthetic manipulation of heme on insolubility of Vitreoscilla hemoglobin in Escherichia coli. Appl Environ Microbiol 60(7):2431-7.

Harvey AL. 2008. Natural products in drug discovery. Drug Discov Today 13(19-20):894-901.

Haus UU, Klamt S, Stephen T. 2008. Computing knock-out strategies in metabolic networks. J Comput Biol 15(3):259-68.

Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H and others. 2006. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. Mol Syst Biol 2:2006 0007.

Hefner J, Ketchum RE, Croteau R. 1998. Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from Taxus canadensis and assessment of the role of this prenyltransferase in cells induced for taxol production. Arch Biochem Biophys 360(1):62-74.

Heilman FR, Herrell WE, Wellman WE, Geraci JE. 1952. Some laboratory and clinical observations on a new antibiotic, erythromycin (ilotycin). Proc Staff Meet Mayo Clin 27(15):285-304.

Heinemann M, Kummel A, Ruinatscha R, Panke S. 2005. In silico genome-scale reconstruction and validation of the Staphylococcus aureus metabolic network. Biotechnol Bioeng 92(7):850-64.

Henry CS, Broadbelt LJ, Hatzimanikatis V. 2007. Thermodynamics-based metabolic flux analysis. Biophys J 92(5):1792-805.

Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28(9):977-82.

Henry CS, Zinner JF, Cohoon MP, Stevens RL. 2009. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. Genome Biol 10(6):R69.

Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M and others. 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26(10):1155-60.

Hofmann G, Diano A, Nielsen J. 2009. Recombinant bacterial hemoglobin alters metabolism of Aspergillus niger. Metab Eng 11(1):8-12.

Holton RA, Kim HB, Somoza C, Liang F, Biediger RJ, Boatman PD, Shindo M, Smith CC, Kim SC, Nadizadeh H and others. 1994a. First Total Synthesis of Taxol .2. Completion of the C-Ring and D-Ring. Journal of the American Chemical Society 116(4):1599-1600.

Holton RA, Somoza C, Kim HB, Liang F, Biediger RJ, Boatman PD, Shindo M, Smith CC, Kim SC, Nadizadeh H and others. 1994b. First Total Synthesis of Taxol .1. Functionalization of the B-Ring. Journal of the American Chemical Society 116(4):1597-1598.

Horwitz SB. 1994. How to make taxol from scratch. Nature 367(6464):593-4.

Hu Z, Pfeifer BA, Chao E, Murli S, Kealey J, Carney JR, Ashley G, Khosla C, Hutchinson CR. 2003. A specific role of the Saccharopolyspora erythraea thioesterase II gene in the function of modular polyketide synthases. Microbiology 149(Pt 8):2213-25.

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S. 2003. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat Biotechnol 21(5):526-31.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2):249-64.

Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A and others. 2007. Multiple high-throughput analyses monitor the response of E. coli to perturbations. Science 316(5824):593-7.

Jamshidi N, Palsson BO. 2007. Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. BMC Syst Biol 1:26.

Jamshidi N, Palsson BO. 2008. Formulating genome-scale kinetic models in the post-genome era. Mol Syst Biol 4:171.

Jantama K, Zhang X, Moore JC, Shanmugam KT, Svoronos SA, Ingram LO. 2008. Eliminating side products and increasing succinate yields in engineered strains of Escherichia coli C. Biotechnol Bioeng 101(5):881-93.

Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, Ingram LO. 2010. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. J Biomed Biotechnol 2010:761042.

Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H and others. 2007. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis. Nat Biotechnol 25(3):319-26.

Jenkins LS, Nunn WD. 1987. Regulation of the ato operon by the atoC gene in Escherichia coli. J Bacteriol 169(5):2096-102.

Jennewein S, Rithner CD, Williams RM, Croteau RB. 2001. Taxol biosynthesis: taxane 13 alpha-hydroxylase is a cytochrome P450-dependent monooxygenase. Proc Natl Acad Sci U S A 98(24):13595-600.

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. Nature 407(6804):651-4.

Jevremovic D, Trinh CT, Srienc F, Boley D. 2010. On algebraic properties of extreme pathways in metabolic networks. J Comput Biol 17(2):107-19.

Jewett MC, Hofmann G, Nielsen J. 2006. Fungal metabolite analysis in genomics and phenomics. Curr Opin Biotechnol 17(2):191-7.

Jin YS, Stephanopoulos G. 2007. Multi-dimensional gene target search for improving lycopene biosynthesis in Escherichia coli. Metab Eng.

Jones AC, Gu L, Sorrels CM, Sherman DH, Gerwick WH. 2009. New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. Curr Opin Chem Biol 13(2):216-23.

Jones KL, Kim SW, Keasling JD. 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab Eng 2(4):328-38.

Joshi A, Palsson BO. 1989a. Metabolic dynamics in the human red cell. Part I--A comprehensive kinetic model. J Theor Biol 141(4):515-28.

Joshi A, Palsson BO. 1989b. Metabolic dynamics in the human red cell. Part II--Interactions with the environment. J Theor Biol 141(4):529-45.

Joshi A, Palsson BO. 1990a. Metabolic dynamics in the human red cell. Part III--Metabolic reaction rates. J Theor Biol 142(1):41-68.

Joshi A, Palsson BO. 1990b. Metabolic dynamics in the human red cell. Part IV--Data prediction and some model computations. J Theor Biol 142(1):69-85.

Joyce AR, Palsson BO. 2008. Predicting gene essentiality using genome-scale in silico models. Methods Mol Biol 416:433-57.

Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S. 2006. Experimental and computational assessment of conditionally essential genes in Escherichia coli. J Bacteriol 188(23):8259-71.

Kacser H, Burns JA. 1973. The control of flux. Symp Soc Exp Biol 27:65-104.

Kallio PT, Bollinger CJ, Koskenkorva T, Frey AD. 2008. Assessment of biotechnologically relevant characteristics of heterologous hemoglobins in E. coli. Methods Enzymol 436:255-72.

Kang MJ, Lee YM, Yoon SH, Kim JH, Ock SW, Jung KH, Shin YC, Keasling JD, Kim SW. 2005. Identification of genes affecting lycopene accumulation in Escherichia coli using a shot-gun method. Biotechnol Bioeng 91(5):636-42.

Kang Y, Durfee T, Glasner JD, Qiu Y, Frisch D, Winterberg KM, Blattner FR. 2004. Systematic mutagenesis of the Escherichia coli genome. J Bacteriol 186(15):4921-30.

Kao CM, Katz L, Khosla C. 1994. Engineered biosynthesis of a complete macrolactone in a heterologous host. Science 265(5171):509-12.

Kao CM, Pieper R, Cane DE, Khosla C. 1996. Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. Biochemistry 35(38):12363-8.

Kealey JT, Liu L, Santi DV, Betlach MC, Barr PJ. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc Natl Acad Sci U S A 95(2):505-9.

Keasling JD. 2008. Synthetic biology for synthetic chemistry. ACS Chem Biol 3(1):64-76.

Keasling JD, Chou H. 2008. Metabolic engineering delivers next-generation biofuels. Nat Biotechnol 26(3):298-9.

Keating SM, Bornstein BJ, Finney A, Hucka M. 2006. SBMLToolbox: an SBML toolbox for MATLAB users. Bioinformatics 22(10):1275-7.

Kell DB. 2004. Metabolomics and systems biology: making sense of the soup. Curr Opin Microbiol 7(3):296-307.

Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR. 1999. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. Science 284(5418):1368-72.

Ketchum RE, Gibson DM, Croteau RB, Shuler ML. 1999. The kinetics of taxoid accumulation in cell suspension cultures of Taxus following elicitation with methyl jasmonate. Biotechnol Bioeng 62(1):97-105.

Khosla C, Bailey JE. 1988. The Vitreoscilla hemoglobin gene: molecular cloning, nucleotide sequence and genetic expression in Escherichia coli. Mol Gen Genet 214(1):158-61.

Khosla C, Bailey JE. 1989a. Characterization of the oxygen-dependent promoter of the Vitreoscilla hemoglobin gene in Escherichia coli. J Bacteriol 171(11):5995-6004.

Khosla C, Bailey JE. 1989b. Evidence for partial export of Vitreoscilla hemoglobin into the periplasmic space in Escherichia coli. Implications for protein function. J Mol Biol 210(1):79-89.

Khosla C, Curtis JE, DeModena J, Rinas U, Bailey JE. 1990. Expression of intracellular hemoglobin improves protein synthesis in oxygen-limited Escherichia coli. Biotechnology (N Y) 8(9):849-53.

Khosla C, Keasling JD. 2003. Metabolic engineering for drug discovery and development. Nat Rev Drug Discov 2(12):1019-25.

Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE. 2007. Structure and mechanism of the 6-deoxyerythronolide B synthase. Annu Rev Biochem 76:195-221.

Kim HU, Kim TY, Lee SY. 2008. Metabolic flux analysis and metabolic engineering of microorganisms. Mol Biosyst 4(2):113-20.

Kim HU, Kim TY, Lee SY. 2010. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen Acinetobacter baumannii AYE. Mol Biosyst 6(2):339-48.

Kim PJ, Lee DY, Kim TY, Lee KH, Jeong H, Lee SY, Park S. 2007a. Metabolite essentiality elucidates robustness of Escherichia coli metabolism. Proc Natl Acad Sci U S A 104(34):13638-42.

Kim TY, Kim HU, Park JM, Song H, Kim JS, Lee SY. 2007b. Genome-scale analysis of Mannheimia succiniciproducens metabolism. Biotechnol Bioeng 97(4):657-71.

Kingston DGI. 2001. Taxol, a molecule for all seasons. Chem Commun (Camb)(10):867-938.

Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H. 2005. Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res 12(5):291-9.

Kitano H. 2004. Biological robustness. Nat Rev Genet 5(11):826-37.

Kitano H. 2007. Towards a theory of biological robustness. Mol Syst Biol 3:137.

Kjeldsen KR, Nielsen J. 2009. In silico genome-scale reconstruction and validation of the Corynebacterium glutamicum metabolic network. Biotechnol Bioeng 102(2):583-97.

Klein-Marcuschamer D, Ajikumar PK, Stephanopoulos G. 2007. Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene. Trends Biotechnol 25(9):417-24.

Klein-Marcuschamer D, Santos CN, Yu H, Stephanopoulos G. 2009. Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. Appl Environ Microbiol 75(9):2705-11.

Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV. 2004. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. Proc Natl Acad Sci U S A 101(44):15573-8.

Koepp AE, Hezari M, Zajicek J, Vogel BS, LaFever RE, Lewis NG, Croteau R. 1995. Cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene is the committed step of taxol biosynthesis in Pacific yew. J Biol Chem 270(15):8686-90.

Koffas M, Stephanopoulos G. 2005. Strain improvement by metabolic engineering: lysine production as a case study for systems biology. Curr Opin Biotechnol 16(3):361-6.

Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. 2010. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A 107(6):2646-51.

Koskenkorva T, Frey AD, Kallio PT. 2006. Characterization of heterologous hemoglobin and flavohemoglobin promoter regulation in Escherichia coli. J Biotechnol 122(2):161-75.

Kovacs K, Zhang L, Linforth RS, Whittaker B, Hayes CJ, Fray RG. 2007. Redirection of carotenoid metabolism for the efficient production of taxadiene [taxa-4(5),11(12)-diene] in transgenic tomato fruit. Transgenic Res 16(1):121-6.

Kumar VS, Maranas CD. 2009. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. PLoS Comput Biol 5(3):e1000308.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S and others. 1997. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 390(6657):249-56.

Lam KS. 2007. New aspects of natural products in drug discovery. Trends Microbiol 15(6):279-89.

Lau J, Tran C, Licari P, Galazzo J. 2004. Development of a high cell-density fed-batch bioprocess for the heterologous production of 6-deoxyerythronolide B in Escherichia coli. J Biotechnol 110(1):95-103.

Lawhorn BG, Gerdes SY, Begley TP. 2004. A genetic screen for the identification of thiamin metabolic genes. J Biol Chem 279(42):43555-9.

Lee J, Yun H, Feist AM, Palsson BO, Lee SY. 2008. Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. Appl Microbiol Biotechnol 80(5):849-62.

Lee JM, Gianchandani EP, Papin JA. 2006. Flux balance analysis in the era of metabolomics. Brief Bioinform 7(2):140-50.

Lee KH, Park JH, Kim TY, Kim HU, Lee SY. 2007. Systems metabolic engineering of Escherichia coli for L-threonine production. Mol Syst Biol 3:149.

Lee KH, Reardon KF. 2003. Proteomics: An exciting new science, but where are the chemical engineers? Aiche Journal 49(11):2682-2686.

Lee PC, Mijts BN, Schmidt-Dannert C. 2004. Investigation of factors influencing production of the monocyclic carotenoid torulene in

metabolically engineered Escherichia coli. Appl Microbiol Biotechnol 65(5):538-46.

Lee PS, Lee KH. 2003. Escherichia coli - A model system that benefits from and contributes to the evolution of proteomics. Biotechnology and Bioengineering 84(7):801-814.

Lee S, Phalakornkule C, Domach MM, Grossmann IE. 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. Computers & Chemical Engineering 24(2-7):711-716.

Lee SY, Lee DY, Kim TY. 2005. Systems biotechnology for strain improvement. Trends Biotechnol 23(7):349-58.

Leeds JA, Schmitt EK, Krastel P. 2006. Recent developments in antibacterial drug discovery: microbe-derived natural products--from collection to the clinic. Expert Opin Investig Drugs 15(3):211-26.

Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KL. 2010. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proc Natl Acad Sci U S A 107(31):13654-9.

Leonard E, Runguphan W, O'Connor S, Prather KJ. 2009. Opportunities in metabolic engineering to facilitate scalable alkaloid production. Nat Chem Biol 5(5):292-300.

Leonard E, Yan Y, Fowler ZL, Li Z, Lim CG, Lim KH, Koffas MA. 2008. Strain improvement of recombinant Escherichia coli for efficient production of plant flavonoids. Mol Pharm 5(2):257-65.

Li JW, Vederas JC. 2009. Drug discovery and natural products: end of an era or an endless frontier? Science 325(5937):161-5.

Liu Y, Hazzard C, Eustaquio AS, Reynolds KA, Moore BS. 2009. Biosynthesis of Salinosporamides from alpha,beta-Unsaturated Fatty Acids: Implications for Extending Polyketide Synthase Diversity. J Am Chem Soc.

Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM. 2009. Large-scale identification of genetic design strategies using local search. Mol Syst Biol 5:296.

Luo G, Pieper R, Rosa A, Khosla C, Cane DE. 1996. Erythromycin biosynthesis: exploiting the catalytic versatility of the modular polyketide synthase. Bioorg Med Chem 4(7):995-9.

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J and others. 2008. How biotech can transform biofuels. Nat Biotechnol 26(2):169-72.

Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16(5):577-83.

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R and others. 2001. Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 3(2):301-8.

Mahadevan R, Bond DR, Butler JE, Esteve-Nunez A, Coppi MV, Palsson BO, Schilling CH, Lovley DR. 2006. Characterization of metabolism in the Fe(III)-reducing organism Geobacter sulfurreducens by constraint-based modeling. Appl Environ Microbiol 72(2):1558-68.

Mahadevan R, Edwards JS, Doyle FJ, 3rd. 2002. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83(3):1331-40.

Maillet I, Berndt P, Malo C, Rodriguez S, Brunisholz RA, Pragai Z, Arnold S, Langen H, Wyss M. 2007. From the genome sequence to the proteome and back: evaluation of E. coli genome annotation with a 2-D gel-based proteomics approach. Proteomics 7(7):1097-106.

Majewski RA, Domach MM. 1990. Simple constrained-optimization view of acetate overflow in E. coli. Biotechnol Bioeng 35(7):732-8.

Man WJ, Li Y, O'Connor CD, Wilton DC. 1995. The binding of propionyl-CoA and carboxymethyl-CoA to Escherichia coli citrate synthase. Biochim Biophys Acta 1250(1):69-75.

Martin JF, Demain AL. 1980. Control of antibiotic biosynthesis. Microbiol Rev 44(2):230-51.

Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. 2003. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotechnol 21(7):796-802.

Masters M. 1977. The frequency of P1 transduction of the genes of Escherichia coli as a function of chromosomal position: preferential

transduction of the origin of replication. Mol Gen Genet 155(2):197-202.

Matthews PD, Wurtzel ET. 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(4):396-400.

Mavrovouniotis ML. 1991. Estimation of standard Gibbs energy changes of biotransformations. J Biol Chem 266(22):14440-5.

McGuire JM, Bunch RL, Anderson RC, Boaz HE, Flynn EH, Powell HM, Smith JW. 1952. [Ilotycin, a new antibiotic.]. Schweiz Med Wochenschr 82(41):1064-5.

Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R. 2008. Microbiology in the post-genomic era. Nat Rev Microbiol 6(6):419-30.

Menzella HG, Reeves CD. 2007. Combinatorial biosynthesis for drug development. Curr Opin Microbiol.

Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV. 2005. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat Biotechnol 23(9):1171-6.

Menzella HG, Reisinger SJ, Welch M, Kealey JT, Kennedy J, Reid R, Tran CQ, Santi DV. 2006. Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster. J Ind Microbiol Biotechnol 33(1):22-8.

Minas W, Brunker P, Kallio PT, Bailey JE. 1998. Improved erythromycin production in a genetically engineered industrial strain of Saccharopolyspora erythraea. Biotechnol Prog 14(4):561-6.

Molina-Henares MA, de la Torre J, Garcia-Salamanca A, Molina-Henares AJ, Herrera MC, Ramos JL, Duque E. 2010. Identification of conditionally essential genes for growth of Pseudomonas putida KT2440 on minimal medium through the screening of a genome-wide mutant library. Environ Microbiol 12(6):1468-85.

Mollapour M, Shepherd A, Piper PW. 2008. Novel stress responses facilitate Saccharomyces cerevisiae growth in the presence of the monocarboxylate preservatives. Yeast 25(3):169-77.

Morihira K, Hara R, Kawahara S, Nishimori T, Nakamura N, Kusama H, Kuwajima I. 1998. Enantioselective total synthesis of taxol. Journal of the American Chemical Society 120(49):12980-12981.

Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ. 2010. Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. Appl Microbiol Biotechnol 85(6):1893-906.

Mukaiyama T, Shiina I, Iwadare H, Saitoh M, Nishimura T, Ohkawa N, Sakoh H, Nishimura K, Tani Y, Hasegawa M and others. 1999. Asymmetric total synthesis of Taxol (R). Chemistry-a European Journal 5(1):121-161.

Mukhopadhyay A, Redding AM, Rutherford BJ, Keasling JD. 2008. Importance of systems biology in engineering microbes for biofuel production. Curr Opin Biotechnol 19(3):228-34.

Muniyappa K, Radding CM. 1986. The homologous recombination system of phage lambda. Pairing activities of beta protein. J Biol Chem 261(16):7472-8.

Murli S, Kennedy J, Dayem LC, Carney JR, Kealey JT. 2003. Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production. J Ind Microbiol Biotechnol 30(8):500-9.

Mutka SC, Bondi SM, Carney JR, Da Silva NA, Kealey JT. 2006a. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. FEMS Yeast Res 6(1):40-7.

Mutka SC, Carney JR, Liu Y, Kennedy J. 2006b. Heterologous production of Epothilone C and D in Escherichia coli. Biochemistry 45(4):1321-30.

Nakamura CE, Whited GM. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol 14(5):454-9.

Neumann CS, Fujimori DG, Walsh CT. 2008. Halogenation strategies in natural product biosynthesis. Chem Biol 15(2):99-109.

Newman DJ. 2008. Natural products as leads to potential drugs: an old process or the new hope for drug discovery? J Med Chem 51(9):2589-99.

Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last 25 years. J Nat Prod 70(3):461-77.

Newman DJ, Cragg GM, Snader KM. 2003. Natural products as sources of new drugs over the period 1981-2002. J Nat Prod 66(7):1022-37.

Newman JD, Marshall J, Chang M, Nowroozi F, Paradise E, Pitera D, Newman KL, Keasling JD. 2006. High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered Escherichia coli. Biotechnol Bioeng 95(4):684-91.

Nguyen KT, Ritz D, Gu JQ, Alexander D, Chu M, Miao V, Brian P, Baltz RH. 2006. Combinatorial biosynthesis of novel antibiotics related to daptomycin. Proc Natl Acad Sci U S A 103(46):17462-7.

Nicolaou KC, Liu JJ, Yang Z, Ueno H, Sorensen EJ, Claiborne CF, Guy RK, Hwang CK, Nakada M, Nantermet PG. 1995a. Total Synthesis of Taxol .2. Construction of a-Ring and C-Ring Intermediates and Initial Attempts to Construct the Abc Ring-System. Journal of the American Chemical Society 117(2):634-644.

Nicolaou KC, Nantermet PG, Ueno H, Guy RK, Couladouros EA, Sorensen EJ. 1995b. Total Synthesis of Taxol .1. Retrosynthesis, Degradation, and Reconstitution. Journal of the American Chemical Society 117(2):624-633.

Nicolaou KC, Ueno H, Liu JJ, Nantermet PG, Yang Z, Renaud J, Paulvannan K, Chadha R. 1995c. Total Synthesis of Taxol .4. The Final Stages and Completion of the Synthesis. Journal of the American Chemical Society 117(2):653-659.

Nicolaou KC, Yang Z, Liu JJ, Nantermet PG, Claiborne CF, Renaud J, Guy RK, Shibayama K. 1995d. Total Synthesis of Taxol .3. Formation of Taxols Abc Ring Skeleton. Journal of the American Chemical Society 117(2):645-652.

Nicolaou KC, Yang Z, Liu JJ, Ueno H, Nantermet PG, Guy RK, Claiborne CF, Renaud J, Couladouros EA, Paulvannan K and others. 1994. Total synthesis of taxol. Nature 367(6464):630-4.

Nielsen J, Jewett MC. 2008. Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. FEMS Yeast Res 8(1):122-31.

Nielsen J, Oliver S. 2005. The next wave in metabolome analysis. Trends Biotechnol 23(11):544-6.

Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. Appl Environ Microbiol 64(5):1694-9.

Nishihara K, Kanemori M, Yanagi H, Yura T. 2000. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. Appl Environ Microbiol 66(3):884-9.

Nolan RP, Fenley AP, Lee K. 2006. Identification of distributed metabolic objectives in the hypermetabolic liver by flux and energy balance analysis. Metab Eng 8(1):30-45.

Nolan RP, Lee K. 2010. Dynamic model of CHO cell metabolism. Metab Eng.

Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J and others. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum. J Bacteriol 183(16):4823-38.

Novere NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM and others. 2009. The systems biology graphical notation. Nat Biotechnol 27(8):735-41.

Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA. 2008. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. J Bacteriol 190(8):2790-803.

Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007. Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282(39):28791-9.

Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S. 2008. Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350. J Bacteriol 190(11):4050-60.

Oliveira AP, Nielsen J, Forster J. 2005. Modeling Lactococcus lactis using a genome-scale flux model. BMC Microbiol 5:39.

Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF. 2007. Complete genome sequence of the

erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. Nat Biotechnol.

Otero JM, Nielsen J. 2010. Industrial systems biology. Biotechnol Bioeng 105(3):439-60.

Panagiotou G, Andersen MR, Grotkjaer T, Regueira TB, Nielsen J, Olsson L. 2009. Studies of the production of fungal polyketides in Aspergillus nidulans by using systems biology tools. Appl Environ Microbiol 75(7):2212-20.

Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO. 2004. Comparison of network-based pathway analysis methods. Trends Biotechnol 22(8):400-5.

Papini M, Salazar M, Nielsen J. 2010. Systems biology of industrial microorganisms. Adv Biochem Eng Biotechnol 120:51-99.

Papoutsakis ET. 1984. Equations and calculations for fermentations of butyric acid bacteria. Biotechnol Bioeng 26(2):174-87.

Park JH, Lee KH, Kim TY, Lee SY. 2007. Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci U S A 104(19):7797-802.

Park JH, Lee SY, Kim TY, Kim HU. 2008. Application of systems biology for bioprocess development. Trends Biotechnol 26(8):404-12.

Pastink MI, Teusink B, Hols P, Visser S, de Vos WM, Hugenholtz J. 2009. Genome-scale model of Streptococcus thermophilus LMG18311 for metabolic comparison of lactic acid bacteria. Appl Environ Microbiol 75(11):3627-33.

Paterson I, Anderson EA. 2005. Chemistry. The renaissance of natural products as drug candidates. Science 310(5747):451-3.

Patil KR, Rocha I, Forster J, Nielsen J. 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6:308.

Patnaik R. 2008. Engineering complex phenotypes in industrial strains. Biotechnol Prog 24(1):38-47.

Pettit RK. 2009. Mixed fermentation for natural product drug discovery. Appl Microbiol Biotechnol 83(1):19-25.

Pfeifer B, Hu Z, Licari P, Khosla C. 2002. Process and metabolic strategies for improved production of Escherichia coli-derived 6-deoxyerythronolide B. Appl Environ Microbiol 68(7):3287-92.

Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291(5509):1790-2.

Pfeifer BA, Khosla C. 2001. Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev 65(1):106-18.

Pfeifer BA, Wang CC, Walsh CT, Khosla C. 2003. Biosynthesis of Yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host. Appl Environ Microbiol 69(11):6698-702.

Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S. 1999. METATOOL: for studying metabolic networks. Bioinformatics 15(3):251-7.

Pharkya P, Burgard AP, Maranas CD. 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol Bioeng 84(7):887-99.

Pharkya P, Burgard AP, Maranas CD. 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14(11):2367-76.

Pharkya P, Maranas CD. 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8(1):1-13.

Phue JN, Lee SJ, Trinh L, Shiloach J. 2008. Modified Escherichia coli B (BL21), a superior producer of plasmid DNA compared with Escherichia coli K (DH5alpha). Biotechnol Bioeng 101(4):831-6.

Phue JN, Noronha SB, Hattacharyya R, Wolfe AJ, Shiloach J. 2005. Glucose metabolism at high density growth of E. coli B and E. coli K: differences in metabolic pathways are responsible for efficient glucose utilization in E. coli B as determined by microarrays and Northern blot analyses. Biotechnol Bioeng 90(7):805-20.

Phue JN, Shiloach J. 2004. Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in E. coli B (BL21) and E. coli K (JM109). J Biotechnol 109(1-2):21-30.

Pieper R, Ebert-Khosla S, Cane D, Khosla C. 1996. Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates. Biochemistry 35(7):2054-60.

Pieper R, Luo G, Cane DE, Khosla C. 1995. Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. Nature 378(6554):263-6.

Pitera DJ, Paddon CJ, Newman JD, Keasling JD. 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metab Eng 9(2):193-207.

Plackett RL, Burman JP. 1946. The design of optimum multifactorial experiments. Biometrika 33(4):305-325.

Plata G, Hsiao TL, Olszewski KL, Llinas M, Vitkup D. 2010. Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Mol Syst Biol 6:408.

Poolman MG, Miguet L, Sweetlove LJ, Fell DA. 2009. A genome-scale metabolic model of Arabidopsis thaliana and some of its properties. Plant Physiol.

Pramanik J, Keasling JD. 1997. Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng 56(4):398-421.

Pramanik J, Keasling JD. 1998. Effect of Escherichia coli biomass composition on central metabolic fluxes predicted by a stoichiometric model. Biotechnol Bioeng 60(2):230-8.

Puchalka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VA. 2008. Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol 4(10):e1000210.

Quadri LE, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT. 1998. Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry 37(6):1585-95.

Reed JL, Famili I, Thiele I, Palsson BO. 2006a. Towards multidimensional genome annotation. Nat Rev Genet 7(2):130-41.

Reed JL, Palsson BO. 2004. Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 14(9):1797-805.

Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO. 2006b. Systems approach to refining genome annotation. Proc Natl Acad Sci U S A 103(46):17480-4.

Reed JL, Vo TD, Schilling CH, Palsson BO. 2003. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol 4(9):R54.

Reeves AR, Brikun IA, Cernota WH, Leach BI, Gonzalez MC, Weber JM. 2007. Engineering of the methylmalonyl-CoA metabolite node of Saccharopolyspora erythraea for increased erythromycin production. Metab Eng 9(3):293-303.

Renneberg R. 2007. Biotech History: Yew trees, paclitaxel synthesis and fungi. Biotechnol J 2(10):1207-9.

Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BO. 2007. Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli. PLoS Comput Biol 3(10):1887-95.

Reszko AE, Kasumov T, Pierce BA, David F, Hoppel CL, Stanley WC, Des Rosiers C, Brunengraber H. 2003. Assessing the reversibility of the anaplerotic reactions of the propionyl-CoA pathway in heart and liver. J Biol Chem 278(37):34959-65.

Rhie HG, Dennis D. 1995. Role of fadR and atoC(Con) mutations in poly(3-hydroxybutyrate-co-3-hydroxyvalerate) synthesis in recombinant pha+ Escherichia coli. Appl Environ Microbiol 61(7):2487-92.

Risso C, Sun J, Zhuang K, Mahadevan R, DeBoy R, Ismail W, Shrivastava S, Huot H, Kothari S, Daugherty S and others. 2009. Genome-scale comparison and constraint-based metabolic reconstruction of the facultative anaerobic Fe(III)-reducer Rhodoferax ferrireducens. BMC Genomics 10:447.

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J and others. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440(7086):940-3.

Roberts SB, Gowen CM, Brooks JP, Fong SS. 2010. Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. BMC Syst Biol 4:31.

Rocha M, Maia P, Mendes R, Pinto JP, Ferreira EC, Nielsen J, Patil KR, Rocha I. 2008. Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinformatics 9:499.

Rochfort S. 2005. Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. J Nat Prod 68(12):1813-20.

Rodriguez E, Gramajo H. 1999. Genetic and biochemical characterization of the alpha and beta components of a propionyl-CoA carboxylase complex of Streptomyces coelicolor A3(2). Microbiology 145 (Pt 11):3109-19.

Rokem JS, Lantz AE, Nielsen J. 2007. Systems biology of antibiotic production by microorganisms. Nat Prod Rep 24(6):1262-87.

Salas JA, Mendez C. 2007. Engineering the glycosylation of natural products in actinomycetes. Trends Microbiol 15(5):219-32.

Salimi F, Zhuang K, Mahadevan R. 2010. Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. Biotechnol J 5(7):726-38.

Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sanchez AM, Andrews J, Hussein I, Bennett GN, San KY. 2006a. Effect of overexpression of a soluble pyridine nucleotide transhydrogenase (UdhA) on the production of poly(3-hydroxybutyrate) in Escherichia coli. Biotechnol Prog 22(2):420-5.

Sanchez AM, Bennett GN, San KY. 2006b. Batch culture characterization and metabolic flux analysis of succinate-producing Escherichia coli strains. Metab Eng 8(3):209-26.

Sassetti CM, Boyd DH, Rubin EJ. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A 98(22):12712-7.

Satish Kumar V, Dasika MS, Maranas CD. 2007. Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics 8:212.

Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wuthrich K. 1997. Metabolic fluxes in riboflavin-producing Bacillus subtilis. Nat Biotechnol 15(5):448-52.

Sauer U, Hatzimanikatis V, Hohmann HP, Manneberg M, van Loon AP, Bailey JE. 1996. Physiology and metabolic fluxes of wild-type and riboflavin-producing Bacillus subtilis. Appl Environ Microbiol 62(10):3687-96.

Savage DF, Way J, Silver PA. 2008. Defossiling fuel: how synthetic biology can transform biofuel production. ACS Chem Biol 3(1):13-6.

Schilling CH, Edwards JS, Letscher D, Palsson BO. 2000. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol Bioeng 71(4):286-306.

Schilling CH, Palsson BO. 2000. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. J Theor Biol 203(3):249-83.

Schmidt H, Jirstrand M. 2006. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics 22(4):514-5.

Schuetz R, Kuepfer L, Sauer U. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol 3:119.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17(2):53-60.

Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99(23):15112-7.

Selvarasu S, Ow DS, Lee SY, Lee MM, Oh SK, Karimi IA, Lee DY. 2009a. Characterizing Escherichia coli DH5alpha growth and metabolism in a complex medium using genome-scale flux analysis. Biotechnol Bioeng 102(3):923-34.

Selvarasu S, Wong VV, Karimi IA, Lee DY. 2009b. Elucidation of metabolism in hybridoma cells grown in fed-batch culture by genome-scale modeling. Biotechnol Bioeng 102(5):1494-504.

Senger RS, Papoutsakis ET. 2008. Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis. Biotechnol Bioeng 101(5):1036-52.

Sharan SK, Thomason LC, Kuznetsov SG, Court DL. 2009. Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc 4(2):206-23.

Sharma KK, Boddy CN. 2007. The thioesterase domain from the pimaricin and erythromycin biosynthetic pathways can catalyze hydrolysis of simple thioester substrates. Bioorg Med Chem Lett 17(11):3034-7.

Sheikh K, Forster J, Nielsen LK. 2005. Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus. Biotechnol Prog 21(1):112-21.

Shen CR, Liao JC. 2008. Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. Metab Eng 10(6):312-20.

Shinfuku Y, Sorpitiporn N, Sono M, Furusawa C, Hirasawa T, Shimizu H. 2009. Development and experimental verification of a genome-scale metabolic model for Corynebacterium glutamicum. Microb Cell Fact 8:43.

Shlomi T, Berkman O, Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci U S A 102(21):7695-700.

Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BO. 2010. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. BMC Syst Biol 4(1):140.

Skeel RT. 1999. Handbook of cancer chemotherapy. Philadelphia: Lippincott Williams & Wilkins. xiv, 733 p. p.

Smith RL, Bungay HR, Pittenger RC. 1962. Growth-biosynthesis relationships in erythromycin fermentation. Appl Microbiol 10(4):293-6.

Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY. 2010. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnol J 5(7):705-15.

Sorensen HP, Mortensen KK. 2005. Advanced genetic strategies for recombinant protein expression in Escherichia coli. J Biotechnol 115(2):113-28.

Souchelnytskyi S. 2005. Bridging proteomics and systems biology: what are the roads to be traveled? Proteomics 5(16):4123-37.

Stafford DE, Yanagimachi KS, Lessard PA, Rijhwani SK, Sinskey AJ, Stephanopoulos G. 2002. Optimizing bioconversion pathways through systems analysis and metabolic engineering. Proc Natl Acad Sci U S A 99(4):1801-6.

Stanbury PF, Whitaker A, Hall SJ. 1995. Principles of fermentation technology. Oxford, U.K. ; Tarrytown, N.Y., U.S.A.: Pergamon. xviii, 357 p. p.

Stephanopoulos G. 1999. Metabolic fluxes and metabolic engineering. Metab Eng 1(1):1-11.

Stephanopoulos G. 2002. Metabolic engineering by genome shuffling. Nat Biotechnol 20(7):666-8.

Stephanopoulos G. 2007. Challenges in engineering microbes for biofuels production. Science 315(5813):801-4.

Stephanopoulos G, Alper H, Moxley J. 2004. Exploiting biological complexity for strain improvement through systems biology. Nat Biotechnol 22(10):1261-7.

Stephanopoulos G, Vallino JJ. 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science 252(5013):1675-81.

Strange K. 2005. The end of "naive reductionism": rise of systems biology or renaissance of physiology? Am J Physiol Cell Physiol 288(5):C968-74.

Strocchi M, Ferrer M, Timmis KN, Golyshin PN. 2006. Low temperature-induced systems failure in Escherichia coli: insights from rescue by cold-adapted chaperones. Proteomics 6(1):193-206.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189(1):113-30.

Sun J, Sayyar B, Butler JE, Pharkya P, Fahland TR, Famili I, Schilling CH, Lovley DR, Mahadevan R. 2009. Genome-scale constraint-based modeling of Geobacter metallireducens. BMC Syst Biol 3:15.

Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, Maranas CD. 2009. A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. PLoS Comput Biol 5(2):e1000285.

Swingle B, Markel E, Costantino N, Bubunenko MG, Cartinhour S, Court DL. 2010. Oligonucleotide recombination in Gram-negative bacteria. Mol Microbiol 75(1):138-48.

Tang Y, Kim CY, Mathews, II, Cane DE, Khosla C. 2006. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc Natl Acad Sci U S A 103(30):11124-9.

Terzer M, Stelling J. 2008. Large scale computation of elementary flux modes with bit pattern trees. Bioinformatics.

Teusink B, Wiersma A, Molenaar D, Francke C, de Vos WM, Siezen RJ, Smid EJ. 2006. Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model. J Biol Chem 281(52):40041-8.

Textor S, Wendisch VF, De Graaf AA, Muller U, Linder MI, Linder D, Buckel W. 1997. Propionate oxidation in Escherichia coli: evidence for operation of a methylcitrate cycle in bacteria. Arch Microbiol 168(5):428-36.

Thiele I, Vo TD, Price ND, Palsson BO. 2005. Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. J Bacteriol 187(16):5818-30.

Trinh CT, Carlson R, Wlaschin A, Srienc F. 2006. Design, construction and performance of the most efficient biomass producing E. coli bacterium. Metab Eng 8(6):628-38.

Trinh CT, Srienc F. 2009. Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol. Appl Environ Microbiol 75(21):6696-705.

Trinh CT, Unrean P, Srienc F. 2008. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74(12):3634-43.

Trinh CT, Wlaschin A, Srienc F. 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81(5):813-26.

Tsai SP, Lee YH. 1988. Application of metabolic pathway stoichiometry to statistical analysis of bioreactor measurement data. Biotechnol Bioeng 32(5):713-5.

Tseng HC, Martin CH, Nielsen DR, Prather KL. 2009. Metabolic engineering of Escherichia coli for enhanced production of (R)- and (S)-3-hydroxybutyrate. Appl Environ Microbiol 75(10):3137-45.

Tsoka S, Simon D, Ouzounis CA. 2004. Automated metabolic reconstruction for Methanococcus jannaschii. Archaea 1(4):223-9.

Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, Regentin R, Keasling JD, Renninger NS, Newman JD. 2009. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli. PLoS ONE 4(2):e4489.

Tyo KE, Ajikumar PK, Stephanopoulos G. 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol 27(8):760-5.

Tyo KE, Alper HS, Stephanopoulos GN. 2007. Expanding the metabolic engineering toolbox: more options to engineer cells. Trends Biotechnol 25(3):132-7.

Unrean P, Trinh CT, Srienc F. 2010. Rational design and construction of an efficient E. coli for production of diapolycopendioic acid. Metab Eng 12(2):112-22.

Urbanczik R, Wagner C. 2005. An improved algorithm for stoichiometric network analysis: theory and applications. Bioinformatics 21(7):1203-10.

Valenzano CR, Lawson RJ, Chen AY, Khosla C, Cane DE. 2009. The biochemical basis for stereochemical control in polyketide biosynthesis. J Am Chem Soc 131(51):18501-11.

Vallino JJ, Stephanopoulos G. 1993. Metabolic flux distributions in Corynebacterium glutamicum during growth and lysine overproduction. Biotechnol Bioeng 41(6):633-46.

van der Werf MJ, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T. 2007. Microbial metabolomics: toward a platform with full metabolome coverage. Anal Biochem 370(1):17-25.

Van Dien SJ, Lidstrom ME. 2002. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph Methylobacterium extorquens AM1, with application to reconstruction of C(3) and C(4) metabolism. Biotechnol Bioeng 78(3):296-312.

Van Hoek P, Van Dijken JP, Pronk JT. 1998. Effect of specific growth rate on fermentative capacity of baker's yeast. Appl Environ Microbiol 64(11):4226-33.

van Zyl WH, Lynd LR, den Haan R, McBride JE. 2007. Consolidated bioprocessing for bioethanol production using Saccharomyces cerevisiae. Adv Biochem Eng Biotechnol 108:205-35.

Vanee N, Roberts SB, Fong SS, Manque P, Buck GA. 2010. A genome-scale metabolic model of Cryptosporidium hominis. Chem Biodivers 7(5):1026-39.

Varma A, Boesch BW, Palsson BO. 1993a. Biochemical production capabilities of Escherichia coli. Biotechnol Bioeng 42(1):59-73.

Varma A, Boesch BW, Palsson BO. 1993b. Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59(8):2465-73.

Varma A, Palsson BO. 1994a. Predictions for oxygen supply control to enhance population stability of engineered production strains. Biotechnol Bioeng 43(4):275-85.

Varma A, Palsson BO. 1994b. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60(10):3724-31.

Varma A, Palsson BO. 1995. Parametric sensitivity of stoichiometric flux balance models applied to wild-type Escherichia coli metabolism. Biotechnol Bioeng 45(1):69-79.

Villarreal DM, Phillips CL, Kelley AM, Villarreal S, Villaloboz A, Hernandez P, Olson JS, Henderson DP. 2008. Enhancement of recombinant hemoglobin production in Escherichia coli BL21(DE3) containing the Plesiomonas shigelloides heme transport system. Appl Environ Microbiol 74(18):5854-6.

von Kamp A, Schuster S. 2006. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22(15):1930-1.

von Nussbaum F, Brands M, Hinzen B, Weigand S, Habich D. 2006. Antibacterial natural products in medicinal chemistry--exodus or revival? Angew Chem Int Ed Engl 45(31):5072-129.

Vongsangnak W, Olsen P, Hansen K, Krogsgaard S, Nielsen J. 2008. Improved annotation through genome-scale metabolic modeling of Aspergillus oryzae. BMC Genomics 9:245.

Wackett LP. 2008. Biomass to fuels via microbial transformations. Curr Opin Chem Biol 12(2):187-93.

Walsh CT. 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science 303(5665):1805-10.

Walsh CT. 2008. The chemical versatility of natural-product assembly lines. Acc Chem Res 41(1):4-10.

Walsh CT, Fischbach MA. 2010. Natural products version 2.0: connecting genes to molecules. J Am Chem Soc 132(8):2469-93.

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature.

Wang QZ, Wu CY, Chen T, Chen X, Zhao XM. 2006. Integrating metabolomics into a systems biology framework to exploit metabolic complexity: strategies and applications in microorganisms. Appl Microbiol Biotechnol 70(2):151-61.

Wang Y, Boghigian BA, Pfeifer BA. 2007a. Improving heterologous polyketide production in Escherichia coli by overexpression of an S-adenosylmethionine synthetase gene. Appl Microbiol Biotechnol 77(2):367-73.

Wang Y, Pfeifer BA. 2008. 6-deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in E. coli. Metab Eng 10(1):33-8.

Wang Y, Wang Y, Zhang S. 2007b. High frequency transformation of the industrial erythromycin-producing bacterium Saccharopolyspora erythraea. Biotechnol Lett.

Wang Y, Wu SL, Hancock WS, Trala R, Kessler M, Taylor AH, Patel PS, Aon JC. 2005. Proteomic profiling of Escherichia coli proteins under high cell density fed-batch cultivation with overexpression of phosphogluconolactonase. Biotechnol Prog 21(5):1401-11.

Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. 1971. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 93(9):2325-7.

Wattanachaisaereekul S, Lantz AE, Nielsen ML, Andresson OS, Nielsen J. 2007. Optimization of heterologous production of the polyketide 6-MSA in Saccharomyces cerevisiae. Biotechnol Bioeng 97(4):893-900.

Weissman KJ, Leadlay PF. 2005. Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol 3(12):925-36.

Wender PA, Badham NF, Conway SP, Floreancig PE, Glass TE, Granicher C, Houze JB, Janichen J, Lee DS, Marquess DG and others. 1997a. The pinene path to taxanes .5. Stereocontrolled synthesis of a versatile taxane precursor. Journal of the American Chemical Society 119(11):2755-2756.

Wender PA, Badham NF, Conway SP, Floreancig PE, Glass TE, Houze JB, Krauss NE, Lee DS, Marquess DG, McGrane PL and others. 1997b. The pinene path to taxanes .6. A concise stereocontrolled synthesis of taxol. Journal of the American Chemical Society 119(11):2757-2758.

Wenzel SC, Gross F, Zhang Y, Fu J, Stewart AF, Muller R. 2005. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. Chem Biol 12(3):349-56.

Westergaard SL, Oliveira AP, Bro C, Olsson L, Nielsen J. 2007. A systems biology approach to study glucose repression in the yeast Saccharomyces cerevisiae. Biotechnol Bioeng 96(1):134-45.

Widiastuti H, Kim JY, Selvarasu S, Karimi IA, Kim H, Seo JS, Lee DY. 2010. Genome-scale modeling and in silico analysis of ethanologenic bacteria Zymomonas mobilis. Biotechnol Bioeng.

Wildung MR, Croteau R. 1996. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J Biol Chem 271(16):9201-4.

Wink M. 2010. Biochemistry of plant secondary metabolism. Chichester, West Sussex, U.K. ; Ames, Iowa: Wiley-Blackwell. xv, 445 p. p.

Wishart DS. 2007. Current progress in computational metabolomics. Brief Bioinform.

Withers ST, Keasling JD. 2007. Biosynthesis and engineering of isoprenoid small molecules. Appl Microbiol Biotechnol 73(5):980-90.

Witherup KM, Look SA, Stasko MW, Ghiorzi TJ, Muschik GM, Cragg GM. 1990. Taxus spp. needles contain amounts of taxol comparable to the bark of Taxus brevifolia: analysis and isolation. J Nat Prod 53(5):1249-55.

Wlaschin AP, Trinh CT, Carlson R, Srienc F. 2006. The fractional contributions of elementary modes to the metabolism of Escherichia coli and their estimation from reaction entropies. Metab Eng 8(4):338-52.

Wong MS, Wu S, Causey TB, Bennett GN, San KY. 2008. Reduction of acetate accumulation in Escherichia coli cultures for increased recombinant protein production. Metab Eng 10(2):97-108.

Woodward RB, Logusch E, Nambiar KP, Sakan K, Ward DE, Auyeung BW, Balaram P, Browne LJ, Card PJ, Chen CH and others. 1981a. Asymmetric Total Synthesis of Erythromycin .1. Synthesis of an Erythronolide a Seco Acid-Derivative Via Asymmetric Induction. Journal of the American Chemical Society 103(11):3210-3213.

Woodward RB, Logusch E, Nambiar KP, Sakan K, Ward DE, Auyeung BW, Balaram P, Browne LJ, Card PJ, Chen CH and others. 1981b. Asymmetric Total Synthesis of Erythromycin .2. Synthesis of an Erythronolide a Lactone System. Journal of the American Chemical Society 103(11):3213-3215.

Woodward RB, Logusch E, Nambiar KP, Sakan K, Ward DE, Auyeung BW, Balaram P, Browne LJ, Card PJ, Chen CH and others. 1981c.

Asymmetric Total Synthesis of Erythromycin .3. Total Synthesis of Erythromycin. Journal of the American Chemical Society 103(11):3215-3217.

Wu J, Boghigian BA, Myint M, Zhang H, Zhang S, Pfeifer BA. 2010. Construction and performance of heterologous polyketide-producing K-12- and B-derived Escherichia coli. Lett Appl Microbiol 51(2):196-204.

Xue Q, Ashley G, Hutchinson CR, Santi DV. 1999. A multiplasmid approach to preparing large libraries of polyketides. Proc Natl Acad Sci U S A 96(21):11740-5.

Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, Furubayashi A, Kinjyo S, Dose H, Hasegawa M and others. 2009. Update on the Keio collection of Escherichia coli single-gene deletion mutants. Mol Syst Biol 5:335.

Yau SY, Keshavarz-Moore E, Ward J. 2008. Host strain influences on supercoiled plasmid DNA production in Escherichia coli: Implications for efficient design of large-scale processes. Biotechnol Bioeng 101(3):529-44.

Yoon KW, Doo EH, Kim SW, Park JB. 2008. In situ recovery of lycopene during biosynthesis with recombinant Escherichia coli. J Biotechnol 135(3):291-4.

Yoon SH, Kim JE, Lee SH, Park HM, Choi MS, Kim JY, Lee SH, Shin YC, Keasling JD, Kim SW. 2007. Engineering the lycopene synthetic pathway in E. coli by comparison of the carotenoid genes of Pantoea agglomerans and Pantoea ananatis. Appl Microbiol Biotechnol 74(1):131-9.

Yoon SH, Lee YM, Kim JE, Lee SH, Lee JH, Kim JY, Jung KH, Shin YC, Keasling JD, Kim SW. 2006. Enhanced lycopene production in Escherichia coli engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate. Biotechnol Bioeng 94(6):1025-32.

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. Proc Natl Acad Sci U S A 97(11):5978-83.

Yuan LZ, Rouviere PE, Larossa RA, Suh W. 2006. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in E. coli. Metab Eng 8(1):79-90.

Zamboni N, Sauer U. 2009. Novel biological insights through metabolomics and 13C-flux analysis. Curr Opin Microbiol 12(5):553-8.

Zerikly M, Challis GL. 2009. Strategies for the discovery of new natural products by genome mining. Chembiochem 10(4):625-33.

Zhang H, Boghigian BA, Pfeifer BA. 2010a. Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. Biotechnol Bioeng 105(3):567-73.

Zhang H, Wang Y, Boghigian B, Pfeifer BA. 2009. Probing the heterologous metabolism supporting 6-deoxyerythronolide B biosynthesis in Escherichia coli. Microbial Biotechnology 2(3):390-394.

Zhang H, Wang Y, Pfeifer BA. 2008a. Bacterial hosts for natural product production. Mol Pharm 5(2):212-25.

Zhang H, Wang Y, Wu J, Pfeifer BA. 2010b. Complete biosynthesis of erythromycin A and designed analogs using E. coli as a heterologous host. Chemistry & Biology(In press.).

Zhang K, Sawaya MR, Eisenberg DS, Liao JC. 2008b. Expanding metabolism for biosynthesis of nonnatural alcohols. Proc Natl Acad Sci U S A 105(52):20653-8.

Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WP, del Cardayre SB. 2002. Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature 415(6872):644-6.

Zhou H, Xie X, Tang Y. 2008. Engineering natural products using combinatorial biosynthesis and biocatalysis. Curr Opin Biotechnol 19(6):590-6.