

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Construction and Optimization of Mevalonate Pathway for Production of Isoprenoids in Escherichia coli

Permalink

<https://escholarship.org/uc/item/54p5k63q>

Author

Nowroozi, Farnaz F.B.

Publication Date

2010

Peer reviewed|Thesis/dissertation

**Construction and Optimization of Mevalonate Pathway for Production of
Isoprenoids in *Escherichia coli***

by

Farnaz Nowroozi

**A dissertation submitted in partial satisfaction of the
requirements for the degree of**

**Joint Doctor of Philosophy
with University of California, San Francisco**

in

Engineering-Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D. Keasling, Chair

Professor Adam P. Arkin

Professor Francis C. Szoka

Professor Marc K. Hellerstein

Fall 2009

The dissertation of Farnaz Foroughi-Boroujeni Nowroozi, titled Construction and Optimization of Mevalonate Pathway for production of Isoprenoids in *Escherichia coli*, is approved:

Chair _____

Date _____
Date _____
Date _____
Date _____

University of California, Berkeley

Abstract

Construction and Optimization of Mevalonate Pathway for production of Isoprenoids in Escherichia coli

by

Farnaz Foroughi-Boroujeni Nowroozi

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Jay D. Keasling, Chair

The isoprenoid family, containing over 50,000 members, constitutes one of the most structurally diverse groups of natural products. They range from essential and relatively universal primary metabolites, such as sterols, carotenoids, and hormones, to more unique secondary metabolites that serve roles in plant defense and communication and cellular and organismal development. Although these molecules have vast potential in medicine and industry their production is limited by two factors:

- 1- The yields from harvest and extraction of these compounds from their native sources are low
- 2- Due to their complex structure, synthetic routes to most isoprenoids are difficult and inefficient

Therefore engineering metabolic pathways for production of large quantities of isoprenoids in a microbial host is an attractive approach.

A major obstacle to efficient microbial biosynthesis of isoprenoids is the production of the universal isoprenoid precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). The biosynthesis of these precursors is highly regulated in all organisms, and previous research on increasing the in vivo supply has focused on engineering a host's native isoprenoid biosynthesis pathway to avoid known regulation.

To better address this limitation, we have taken the novel approach of engineering *Escherichia coli* to over-produce IPP and DMAPP by cloning and expressing the heterologous mevalonate isoprenoid pathway from *Saccharomyces cerevisiae*. When co-expressed with a codon-optimized amorphaadiene synthase, this system successfully demonstrated high-level production of terpenes such as amorpha-4, 11-diene the sesquiterpene olefin precursor to the antimalarial drug artemisinin.

Expression of the heterologous mevalonate pathway circumvented native regulation of isoprenoid biosynthesis in *E. coli*, by providing a second, unregulated route to the isoprenoid precursors. However it was shown that unregulated flux through the bottom part of mevalonate pathway is detrimental to both production and cell growth. Accumulation of some of the mevalonate pathway intermediates isoprenyl pyrophosphates, IPP, DMAPP, and farnesyl pyrophosphate causes severe growth inhibition and affects production of amorphaadiene. Further studies of growth inhibited cells through metabolite and protein analysis suggested that accumulation of these intermediates results in down-regulation of parts of the pathway that are responsible for production of these toxic intermediates. After the cells successfully decrease protein expression and metabolite production the growth is restored.

By engineering *E. coli* for high level production of isoprenoids, we demonstrate that balancing carbon flux through the engineered biosynthetic pathway is the key factor in optimization efforts towards high level production of isoprenoids.

Table of Contents

Table of Contents

TABLE OF CONTENTS	I
LIST OF FIGURES	II
LIST OF TABELS	V
ACKNOWLEDGEMENTS	VI
CHAPTER 1: INTRODUCTION, MOTIVATION, OUTLINE	1
REFERENCES	3
CHAPTER 2: LITERATURE REVIEW	5
INTRODUCTION	5
BIOSYNTHESIS OF ISOPRENOIDS	5
BIOSYNTHESIS OF ISOPENTENYL PYROPHOSPHATE	5
Mevalonate Pathway	6
DeoxyXylulose-5-Phosphate (DXP) Pathway	7
Biosynthesis and Functionalization of Terpene Olfins from IPP.....	7
ENGINEERING MICROORGANISMS FOR THE PRODUCTION OF ISOPRENOIDS.....	8
Engineering of Escherichia coli for production of isoprenoids	8
Engineering of yeasts for the production of isoprenoids	11
REFERENCES	12
CHAPTER 3: INVESTIGATING THE IMPACT OF HETEROLOGOUS AMORPHADIENE PRODUCTION IN <i>ESCHERICHIA COLI</i>; FLUX ANALYSIS	15
INTRODUCTION	15
ISOTOPOMER ANALYSIS	16
MATERIALS AND METHOD	19
Chemostat Cultivation	19
Amorphadiene Analysis	20
Mass Isotopomer Measurements via GC-MS	20
Experimental Isotopomer Measurements	21
Construction of Large-scale Isotopomer Mapping Model	22
Mathematical Analysis of Flux Elucidation	24
RESULTS	25
DISCUSSION AND FUTURE DIRECTION	31
REFERENCES	32
CHAPTER 4: OPTIMIZATION OF THE MEVALONATE-BASED ISOPRENOID BIOSYNTHETIC PATHWAY IN <i>ESCHERICHIA COLI</i> FOR PRODUCTION OF THE ANTI-MALARIAL DRUG PRECURSOR AMORPHA-4, 11-DIENE	34

INTRODUCTION	34
MATERIALS AND METHOD	35
Strains and Plasmids	35
Growth Condition	36
Gas Chromatograph-Mass Selective (GC-MS) Analysis of Amorphadiene	36
Construction of a Vector for Rapid Cloning of Promoters and Amorphadiene Operons	37
Construction of Amorphadiene Production Strains	37
Construction of Plasmids Expressing Members of the MBIS Operon	38
Plasmid Stability Assays	38
RESULTS	38
Increasing Expression of the MevT Pathway	38
Use of Stronger Promoter and Codon Optimization of MevT Increases Amorphadiene Production	39
Identification of ADS as a Limiting Enzyme	39
Identification of Limiting Enzymes in the MBIS Operon	40
Combination of Modular Vectors with Increased Mevalonate Kinase Expression.....	41
DISCUSSION AND FUTURE DIRECTIONS	43
REFERENCES	44
CHAPTER 5: STUDY AND OPTIMIZATION OF THE LOWER PART OF HETEROLOGOUS MEVALONATE PATHWAY IN <i>ESCHERICHIA COLI</i>; THE CELLS' RESPONSE	47
INTRODUCTION	47
MATERIALS AND METHODS	49
Operon and Plasmid Construction	50
Growth Condition	53
Gas Chromatograph-Mass Selective (GC-MS) Analysis of Amorphadiene	54
Proteomics Analysis: Protein Sample Preparation	54
Proteomics Analysis: LC-MS Analysis	55
Metabolomics Analysis: The Analysis of FPP by Liquid Chromatography and Mass Spectrometry (LC-MS)	56
Metabolomics Analysis: The Analysis of Mevalonate by Liquid Chromatography and Mass Spectrometry (LC-MS)	56
RESULTS	57
Initial Steps of the Lower Mevalonate Pathway Limit Carbon Flux to Amorphadiene	57
Use of Stronger RBS in front of ADS	59
Proteomics Analysis: Measuring Protein Concentrations	65
Metabolomics Analysis: Measuring Metabolite Concentrations	67
Growth and Intermediate Metabolite Concentration Profiles: DH1: pMevT/ pTrc-MBIS2	68
Growth and Intermediate Metabolite Concentration Profiles: DH1: pMevT/ pTrc-O.ADS-MBIS1 and DH1: pMevT/ pTrcO.ADS-MBIS2	70
DISCUSSION AND FUTURE DIRECTION	74
REFERENCES	77

APPENDICES	79
APPENDIX 1: MATHEMATICAL ANALYSIS OF FLUX ELUCIDATION AND RESULTS	79
Mathematical Analysis of Flux Elucidation and Results	79
Flux Distribution Results	84
APPENDIX 2: PROTOCOLS	86
Fermentation Protocol	86
GC-MS Isotopomer Analysis of Amino Acids	87

List of Figures

Fig. 2-1. Biosynthesis of IPP and DMAPP.

6

Fig. 2-2. The C5 isoprene units combine in condensation reactions catalyzed by the prenyl transferases to form GPP (C10), FPP (C15), GGPP (C20), and longer prenyl pyrophosphate precursors. The isoprenyl pyrophosphates are then cyclized by the terpene cyclases (synthases) to form the various terpene classes.

7

Fig. 2-3. Heterologous mevalonate pathway and native DXP (non-mevalonate) pathway for the production of amorphadiene.

10

Fig. 3-1 Sample metabolic network and corresponding flux balance equation

15

Fig. 3-2 Feeding a substrate (A) with known label pattern and measuring the fraction of label in a resulting metabolite (F) allows determination of the relative flux through branching pathways (v_1 and v_2)

16

Fig. 3-3 Feeding a known label distribution of glucose results in a variety of label patterns in the resulting amino acids such as threonine, shown above. The label is scrambled as the network branches and comes back together. Amino acids are often used in isotopomer analysis because they originate from many points in central metabolism.

17

Fig. 3-4 TBDMS-derivatized amino acid and fragmentation sites.

22

Fig. 3-5 Pictorial representation of the variable I_{ik} and the parameter MDV for a sample molecule. For a molecule, i , that contains three carbons there are 23, i.e., 8, different labeling patterns that make up the k members of the isotope fraction I_{ik} . The relative fraction of each is contained in the I_{ik} . During GC-MS, the molecule is derivitized and can be fragmented into different-sized species, f , that are then analyzed. The mass distribution vector, MDV, contains the information about the relative fractions of each fragment that contains m labeled carbons. Because only the total mass of each instance of a fragment is determined, the mass distribution vector is made by collecting various isotopomers by mass, as shown for two different fragments.

23

Fig. 3-6 The isotope labeling of species is governed by an isotopomer distribution vector balance and the isotope mapping matrix. Two reactions, j_1 and j_2 , form the species E ; the first is unimolecular and the second is bi-molecular. E is also consumed by two reactions j_3 and j_4 ; again, the first is unimolecular and the second is bi-molecular. The isotope labeling of E is given by I_{Ek} . Each carbon-containing compound has an associated isotope fraction for each labeling

pattern, I_{ik} . The isotope labeling balance of the system is at the top, with the first term of the right-hand side of Eq. (3) on the left and the second

24

Fig. 3-8. Central metabolism flux maps. 29-30

Fig. 4-1. Mevalonate pathway and vector constructs used in this study. 35

Fig. 4-2. Decreasing plasmid number and additional copies of ADS increase amorphaadiene production. 41

Fig. 4-3. Mevalonate kinase limits amorphaadiene production. 42

Fig. 4-4. Amorphaadiene production increases when additional copies of mevalonate kinase are present. 43

Fig. 5-1. Growth inhibition by high flux through the mevalonate bottom pathway. 48

Fig. 5-2. Biobrick cloning method. 51

Fig. 5-3. Using PCR the five genes of the lower mevalonate pathway were amplified. By designing PCR primers that contain the selected RBS sequences, three RBSs with three different strengths were put in front of the first three genes of MBIS operon. Only one RBS of medium activity was used with the last two genes, *idi*, and *ispA*. 52

Fig. 5-4. The 9 constructed operons containing amorphaadiene synthase (ADS) and MBIS operon cloned into pTrc99a backbone. 53

Fig. 5-5. The 10 constructs containing ADS with ribosome binding sites of different activity levels and two of the MBIS operons (MBIS1 and MBIS2). 54

Fig. 5-6. Cell growth of *E. coli* DH1 strains expressing pMevT and pTrc-ADS.MBIS operons 24 hour and 48 hour after induction. Some of the strains show growth inhibition at 24 hour that is relieved at 48 hour. 58

Fig. 5-7. Amorphaadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-ADS.MBIS operons 24 hour and 48 hour after induction. The strains with growth inhibition at 24 hour also exhibit poor amorphaadiene production. 59

Fig. 5-8. Cell growth of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. Strains with different strength of RBS in front of ADS exhibit growth differences. 61

Fig. 5-9. Amorphaadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. 62

- Fig. 5-10.** Cell growth of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. 63
- Fig. 5-11.** Amorphadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. 64
- Fig. 5-12.** Specific amorphadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. 65
- Fig. 5-13.** Protein levels of all of the enzymes in the mevalonate pathway at 24 hr and 48 hr after induction. 66
- Fig. 5-14.** Protein levels of all of the enzymes in the mevalonate pathway at 24 hr and 48 hr after induction. 67
- Fig. 5-15.** Cell growth of *E. coli* DH1 strain expressing pMevT and pTrc-MBIS2 operons. 69
- Fig. 5-16.** Concentration of the intermediate metabolites in *E. coli* harboring pMevT and pTrc-MBIS2 plasmids. 70
- Fig. 5-17.** Cell growth of *E. coli* DH1 strains expressing pMevT / pTrcO.ADS-MBIS1 and pMevT / pTrc-O.ADS-MBIS2 plasmids. 71
- Fig. 5-18.** Mevalonate concentration profile in DH1: pMevT / pTrc-O.ADS-MBIS1 (A) and DH1: pMevT / pTrc-O.ADS-MBIS2 (B) At final time point strain expressing MBIS2 operon produces higher mevalonate titers. 72
- Fig. 5-19.** FPP concentration profile in DH1: pMevT/ pTrc-O.ADS-MBIS1 (A) and DH1: pMevT / pTrc-O.ADS-MBIS2 (B). 73

List of tables

Table 1 Fragments of the amino acids experimentally measured by GC–MS and simulated using the IMM models.	26
Table 2 Experimentally measured and stimulated mass distribution (mol%) of amino acid fragments.	28
Table 3 Summary of strains and plasmids used in this study .	36
Table 4 Strains and plasmids used in this study.	49
Table 5 Plasmids with different strength RBSs in front of ADS.	60
Table 6 Plasmids tested with higher glycerol concentration.	62
Table 7 Plasmid sets used in this experiment.	65

Acknowledgments

I would first like to acknowledge my advisor, Professor Jay D. Keasling. Jay works tirelessly to advise the students and scientists in his lab despite his incredibly busy schedule. Over the course of my graduate career, Jay has coordinated and supported important and fascinating projects that I am fortunate to have been a part of. He allows his students to pursue their own research directions and explore their ideas within the broader project objectives. He has created a friendly but challenging environment to give each individual the opportunity to grow and reach their potential. I thank you Jay for always being a supportive and enthusiastic mentor and for always finding the time whenever I needed your help and advice.

Second, I would like to thank my parents for their commitment to my education and their continual encouragement. I thank them for supporting me throughout my time in school, from early years to now. And of course, thanks to my husband, Mehrzad. Without his full support I would not have been able to accomplish as much and be here.

Third, I would like to acknowledge the many scientists I have been lucky enough to work with over the past years. Although I owe everyone in the Keasling lab thanks, I specifically like to thank Douglas Pitera for mentoring me and helping me with many of my questions during my early years in graduate school. I would like to thank Dr. Steve Van Dien for his assistance with flux analysis studies and Dr. Edward Baidoo for his help with metabolomics studies and insightful discussions. Dr. Wes Marner, thank you for being a wonderful office-mate and for all of the scientific discussions and general help in the lab. My wonderful undergraduate assistant Simon Ermakov, thank you for all your help.

Finally, I must thank all of my friends that have encouraged me throughout graduate school.

Chapter 1: Introduction, Motivation, outline

Engineering microorganisms for the production of industrial products has become increasingly attractive in the past decades due to the advantages over traditional synthetic methods. Creating new biosynthetic capabilities in microorganisms allows for production and purification of previously limited products, such as therapeutic proteins and complex natural chemicals at high levels while reducing the use of environmentally destructive chemical processes. In this effort, research has shifted focus from engineering the production of a single recombinant protein to the production of small molecule (non-protein) products, both natural and novel. The goal of this thesis work is to develop and apply methodologies for engineering multi-gene heterologous pathways in bacteria to achieve high-level production of isoprenoids, a large class of small molecule natural products.

Increasing production levels by re-routing cellular metabolism essentially converts the bacterial host into an efficient catalyst for production of these small molecules from simple and inexpensive substrates. By engineering *Escherichia coli* for high level production of isoprenoids, we have highlighted design rules that apply not only to this project, but also to most efforts to engineer microorganisms for high level biosynthesis of small molecules. Specifically, this work demonstrates that balancing carbon flux through the engineered biosynthetic pathway is a key factor in optimizing isoprenoid biosynthesis. Isoprenoids are a highly diverse class of natural products from which numerous commercial flavors, fragrances, chemicals, and medicines are derived^{1, 2}. These valuable compounds are commonly isolated from plants, microbes, and marine organisms where they are naturally produced in small quantities. As such, purification from native sources suffers from low yields, impurities, and excessive consumption of natural resources. Furthermore, most of these compounds are chemically complex, resulting in chemical synthesis routes that are difficult, expensive, and suffer from low yields³⁻⁵. For these reasons, the engineering of metabolic pathways to produce large quantities of complex isoprenoids in a well-studied biological host presents an attractive alternative to extractions from environmental sources or chemical syntheses. Production consistency, scalability, and efficiency of substrate-to-product conversion of microbial fermentation are of particular importance to producing isoprenoid products on the scale and cost of commodity chemicals. Decreasing the cost of isoprenoid products is an important overall goal of the Keasling Lab for the inexpensive production of different isoprenoid molecules, such as isoprenoid-based *biofuels*, potential HIV treatment *prostratin*, and the anti-malarial drug *artemisinin*. Malaria is a disease common to tropical and subtropical regions throughout the world that kills approximately 1 million people every year, mostly children⁶. Malaria has been traditionally treated with chloroquine based drugs; however, resistant strains have emerged⁶. Artemisinin is an isoprenoid produced by the plant, *Artemisia annua*, commonly known as sweet wormwood. The compound has been acclaimed as the next generation anti-malarial drug because it acts very effectively against the parasite^{7, 8} and shows little or no cross-resistance to existing anti-malarial drugs⁹⁻¹¹. However, treatment with artemisinin based therapies is up to 10 to 30 times more expensive than those based on traditional drugs, making use of this treatment impractical in poorer, underdeveloped countries that are most affected by the disease¹². Commercial production of artemisinin currently relies on its extraction and purification from plant material in an environmentally unfriendly process involving petrochemicals¹³ and yields from plant extracts are low⁶. As a crop plant, production is greatly affected by the seasonal variation in Southeast Asia, where *Artemisia* is grown, and current supplies are low, driving raw material prices^{14, 15}. Developing a successful,

large-scale fermentation process for the production of artemisinin may reduce the price of artemisinin-based treatments by an order of magnitude if high yields are efficiently produced.

Artemisinin like all isoprenoids is synthesized from the 5-carbon universal precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). DMAPP the IPP isomer, primes the head to tail condensation of IPP molecules by prenyltransferases to form the linear longer chained farnesyl pyrophosphate (FPP). Cyclization and re-arrangement of FPP is done by amorphaadiene synthase (ADS) to produce the sesquiterpene amorphaadiene. Further tailoring and modifications are done by p450 class of enzymes to convert Amorphaadiene to artemisinic acid. The final conversion to artemisinin is performed by an unknown enzyme, but can alternatively be performed synthetically with a reasonably high yield¹⁸.

The focus of this thesis is engineering *E. coli* to increase carbon flux to isoprenoid products, by increasing the in vivo supply of the universal precursors, IPP and DMAPP. These two precursors are produced naturally by either of the two biosynthetic pathways: the mevalonate pathway or the most recently discovered DXP pathway. In *E. coli*, IPP and DMAPP are synthesized via DXP isoprenoid pathway and are used for the prenylation of tRNA molecules and for the production of the sesquiterpene precursor FPP, which is necessary for quinone and cell wall biosynthesis¹⁶. *E. coli*'s native DXP pathway is tightly regulated and attempts to engineer DXP isoprenoid pathway in *E. coli* to increase IPP and DMAPP production, can only increase slightly the amount of heterologous isoprenoids produced by *E. coli*^{17, 18}. Research in our group has shown that significant increases in isoprenoid production can be achieved using *Saccharomyces cerevisiae*'s mevalonate pathway to supply isoprenoid precursors¹⁶. Engineering of this non-native enzymatic pathway in the bacterium successfully increased isoprenoid production by providing a second, un-regulated route to isoprenoid precursors IPP and DMAPP.

The following chapters describe many relevant aspects of this project. Chapter 2 provides background information to isoprenoid biosynthesis, different classes of isoprenoids and their constituents, and a review on engineering metabolic pathways in *E. coli*.

In chapter 3 we investigate the changes to *E. coli*'s metabolism wrought by overexpressing a non-native pathway that draws on *E. coli*'s supply of acetyl-CoA. We use the approach of metabolic flux analysis: combining the knowledge of biochemical pathways, measurements of metabolite concentrations, carbon-labeling studies and mathematical calculations to determine how the cells' metabolism is affected by being forced to produce a secondary metabolite that they do not benefit from. This approach requires:

1. The compilation of a central metabolic pathway network for *E. coli*
2. The development of continuous culture fermentation techniques to produce cultures growing in steady-state.
3. The development of methods for analysis of major biomass components and secreted metabolites to account for all of the carbon consumed by the cell.
4. The use of ¹³C-labeled glucose, determining the appropriate label position and the ratio of labeled to non-labeled glucose, and measuring the label pattern of the resulting amino acids.
5. The creation of a computer program that solves for the cells intracellular fluxes using the data gathered.

Given the fact that this work was aimed to develop a tool to facilitate engineering of any *E. coli* strain for production of isoprenoids, this chapter is concerned not only with the results obtained

using flux analysis, but also with the development of each of the components of a flux analysis system.

The focus of chapter 4 is optimization and study of lower part of mevalonate pathway. This chapter begins by identification of the bottlenecks that limit the flux through mevalonate pathway. Initial optimization efforts and subsequent gene titration studies demonstrated that the first two genes of the lower mevalonate pathway, mevalonate kinase and phosphomevalonate kinase are limiting. Over-expression of these two genes resulted in increased production of the pathway specific isoprenoid precursor, amorphaadiene. We also employed common methods to improve production from engineered biosynthetic pathways such as optimizing codon usage, enhancing production of rate-limiting enzymes, and eliminating the accumulation of toxic intermediates or byproducts to improve cell growth. By optimizing promoter strength to balance expression of the encoding genes, the pathway bottlenecks were alleviated and production was improved.

While chapter 4 demonstrates application of common metabolic engineering methods in optimization of the mevalonate biosynthetic pathway, chapter 5 investigates the impact of our optimization attempts on cellular growth and amorphaadiene production. Different strength ribosome binding sites were cloned in front of the genes that were suspected to be limiting the flux towards production of amorphaadiene to build many different constructs. Growth, production, metabolite, and protein analysis of some of these constructs reveal many unknown features of isoprenoid biosynthetic pathway and its expression in *E. coli*. The analytical methods developed along the way are also fundamental to establishing the success of metabolic engineering efforts.

Chapter 6 finishes the thesis by summarizing the main results and conclusions of this work and describes the future experiments that will improve upon this bacterial isoprenoid production system.

Together, the following chapters of this thesis discuss the design, development and optimization of a novel microbial production platform for high-level biosynthesis of isoprenoid products, as well as, new findings obtained by engineering this system.

References

1. McCaskill, D. & Croteau, R. Prospects for the bioengineering of isoprenoid biosynthesis. *Adv. Biochem. Eng./Biotechnol.* 55, 107-146 (1997).
2. McCaskill, D. & Croteau, R. Some caveats for bioengineering terpenoid metabolism in plants. *Trends in Biotechnology* 16, 349-355 (1998).
3. Nicolaou, K.C. et al. Total synthesis of eleutherobin. *Angew Chem Int Ed* V36, 2520-2524 (1997).
4. Danishefsky, S.J. et al. Total synthesis of baccatin III and taxol. *J Amer Chem Soc* V118, 2843-2859 (1996).
5. Avery, M.A., Chong, W.K.M. & Jenningswhite, C. Stereoselective total synthesis of (+)-artemisinin, the antimalarial constituent of *Artemisia-annua* L. *J Amer Chem Soc* V114, 974-979 (1992).
6. Kenneth, A.J., Claire, P.B. & Hellen, G. Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance. The National Academies Press, Washington, D.C.; 2004.
7. Jung, M. et al. Antimalarial activities of (+)-deoxoartemisinin and its novel C-11, 13 derivatives. *Bioorg. Med. Chem. Lett* 14, 2001-2003 (2004).

8. Sriram, D., Rao, V.S., Chandrasekhara, K.V. & Yogeewari, P. Progress in the research of artemisinin and its analogues as antimalarials: an update. *Nat. Prod. Res.* 18, 503-527 (2004).
9. White, N.J. Artemisinin-Current status. *Trans. R. Soc. Trop. Med. Hyg.* (88 Suppl). 53-54 (1994).
10. Ridley, R.G. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature (London)* **415**, 686-693 (2002).
11. Haynes, R.K. Artemisinin and derivatives: The future for malaria treatment? *Current Opinion in Infectious Diseases* **14**, 719-726 (2001).
12. Medicines for Malaria Venture - MMV Annual Report 2004. <http://www.mmv.org/IMG/pdf/2004AnnualReport.pdf>.
13. Abdin, M.Z., Israr, M., Rehman, R.U. & Jain, S.K. Artemisinin, a novel antimalarial drug: biochemical and molecular approaches for enhanced production. *Planta. Med.* 69, 289-299 (2003).
14. Wallaart, T.E., Pras, N., Beekman, A.C. & Quax, W.J. Seasonal variation of artemisinin and its biosynthetic precursors in plants of *Artemisia annua* of different geographical origin: Proof for the existence of chemotypes. *Planta Medica.* 66, 57-62 (2000).
15. McNeil, D.G. Plant shortage leaves campaigns against malaria at risk. *New York Times*. New York; 2004.
16. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., Keasling, J. D. Engineering the mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796-802 (2003).
17. Harker M., Bramley P.M. Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. *FEBS Lett.* 448, 115-119 (1999).
18. Matthews P.D., Wurtzel E.T. Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl Microbiol Biotechnol.* 53, 396-400 (2000)

Chapter 2: Literature Review

Introduction

The objective of this chapter is to provide the scientific background needed to understand the principles behind this work. Specifically, this chapter will present a general review of the biosynthesis of isoprenoids and the engineering of microorganisms for the production of isoprenoid products. Relevant literature and examples will be covered.

Biosynthesis of isoprenoids

Isoprenoids are the largest class of natural products from which various products such as flavors, fragrances, and medicines are derived. Isoprenoids comprise highly diverse class of natural compounds, encompassing >50,000 known compounds with an extreme diversity in chemical structure¹. Isoprenoids have various biological functions, such as hormones (steroids, gibberellins and abscisic acid), their role in maintaining membrane fluidity (steroids), electron transport, protein targeting and regulation (prenylation and glycosylation), respiration (quinines), cellular and organismal development, plant defense and signaling, and light-harvesting pigments (carotenoids)^{2, 3}.

Isoprenoid biosynthesis occurs in 4 steps:

- 1- Production of the 5 carbon monomers isopentenyl pyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP)
- 2- Polymerization to form the linear prenyl pyrophosphate precursors
- 3- Cyclization of the linear precursor to form terpene backbone
- 4- Tailoring and finctionalization of the terpene skeleton

The terpenes are grouped according to their carbon chain length: monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), and triterpenes (C30). Monoterpenes, such as menthol and camphor, and sesquiterpenes, such as zingiberene (ginger), are the major constituents of herbs and spices. Other sesquiterpenes and diterpenes are pheromones, defensive agents, and signal transduction agents^{4, 5}. Triterpenes, such as saponins, provide plant disease resistance⁶. Higher molecular weight isoprenoids, such as sterols, stabilize membranes or serve as photoreceptive agents, as in the case of carotenoids and other C40 compounds. Isoprenoids or their intermediates are also often bound to proteins by a process named prenylation to alter protein function or allow insertion of proteins into membranes.

Aside from their biological functions, these compounds have also found application in medicine and industry. Many effective and promising pharmaceuticals such as Taxol, vinblastine, artemisinin and prostratin belong to this family of compounds. Isoprenoids and Isoprenoid derived compounds have also been speculated to be a potential source of biofuels⁷.

Despite the vast diversity in structure and function, all isoprenoids are derived from the common 5 carbon precursor IPP and DMAPP.

Biosynthesis of isopentenyl pyrophosphate

Biosynthesis of the IPP and DMAPP building blocks takes place via two distinct pathways, the mevalonate pathway⁸ (Figure 1-a) and the more recently discovered 1-deoxyulose-5-phosphate (DXP) pathway^{9, 10} (Figure 1-b).

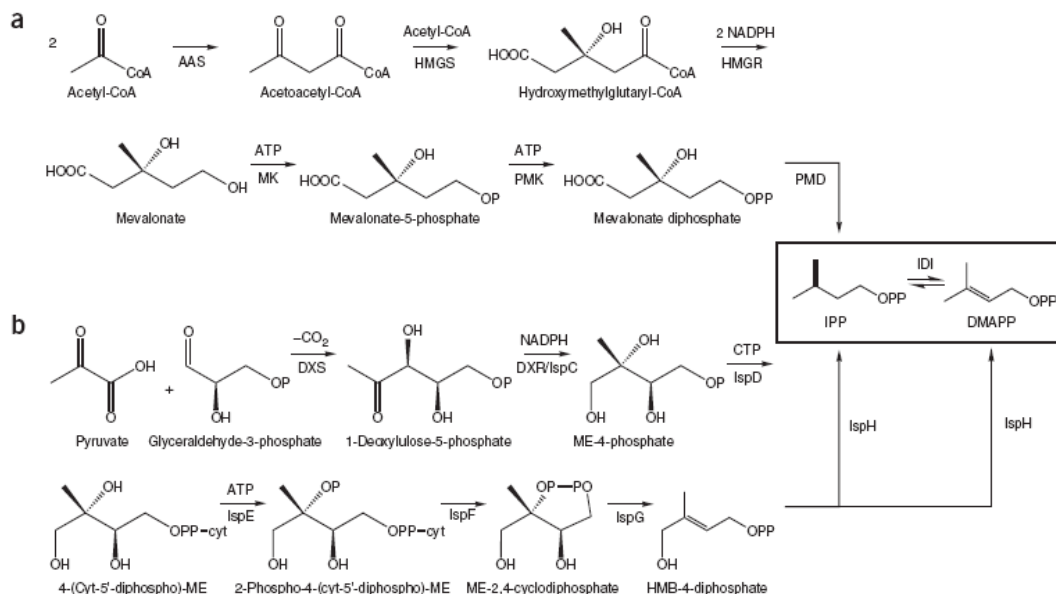


Fig. 2-1. Biosynthesis of IPP and DMAPP. (a) Via the mevalonate pathway (AAS, acetoacetyl-CoA synthase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; PMD, mevalonate diphosphate decarboxylase). (b) Via the DXP pathway (ME, 2-methylerythritol; cyt, cytidine; HMB, 1-hydroxy-2-methyl-2-butenyl; IspD, ME-4-phosphate cytidyltransferase; IspE, 4-(cyt-5'-diphospho)-ME kinase; IspF, ME-2,4-cyclodiphosphate synthase; IspG, HMB-4-diphosphate synthase; IspH, HMB-4-diphosphate reductase).

The mevalonate pathway is found primarily in eukaryotes but has also been found in *Streptomyces*¹¹, grampositive cocci¹², *Borrelia burgdorferi*¹³, and *Lactobacillus*¹⁴ whereas the DXP pathway is found primarily in prokaryotes and in the plastids of photosynthetic organisms¹⁵.

Mevalonate Pathway

The mevalonate pathway produces IPP from three molecules of acetyl-CoA (Figure 1-a). The first enzyme is a thiolase (or acetyl transferase) that catalyzes the formation of acetoacetyl-CoA from two molecules of acetyl-CoA. The second enzyme, HMG-CoA synthase, catalyzes the formation of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) from acetoacetyl-CoA and acetyl-CoA. The equilibrium of the acetyl transferase reaction is heavily directed toward the formation of acetyl-CoA¹⁶; however, the high catalytic activity of HMG-CoA synthase compensates and directs the two reaction sequence to the production of HMG-CoA¹⁶. HMG-CoA is reduced to mevalonate by HMG-CoA reductase. This enzyme is highly regulated in most eukaryotes, and is heavily studied due to its importance in cholesterol metabolism¹⁷⁻¹⁹. After production of mevalonate, two

phosphates are added sequentially to the molecule by mevalonate kinase and phosphomevalonate kinase. As one of the last proposed regulatory points prior to IPP formation, mevalonate kinase has been shown to be feedback inhibited by FPP²⁰. Finally, mevalonate-5-diphosphate is decarboxylated by mevalonate diphosphate decarboxylase to form IPP. Once produced, IPP is converted reversibly to dimethylallyl diphosphate (DMAPP) by isopentenyl diphosphate isomerase. The genes for the enzymes in the mevalonate pathway have been cloned from a number of organisms, and some of these have been expressed in functional form in *E. coli*²¹⁻²³. In addition, entire mevalonate pathway gene clusters from two eubacteria strains of the genus *Streptomyces* have been functionally expressed in *E. coli*, though expression levels appear to be very low^{11, 24}.

Deoxyxylulose-5-phosphate (DXP) pathway

The recently elucidated DXP pathway, native to *Escherichia coli*, produces IPP and DMAPP from pyruvate and glyceraldehyde-3-phosphate (Figure 1-b). This pathway is not the focus of this research and therefore is not discussed in details in this chapter.

Biosynthesis and functionalization of terpene olfins from IPP

DMAPP primes the head to tail condensation of IPP molecules by prenyl-transferases to form prenyl-diphosphate precursors geranyl diphosphate (GPP, C₁₀, monoterpenoids), farnesyl diphosphate (FPP, C₁₅, sesquiterpenoids) and geranylgeranyl diphosphate (GGPP, C₂₀, diterpenoids), as well as larger units (Figure 2)

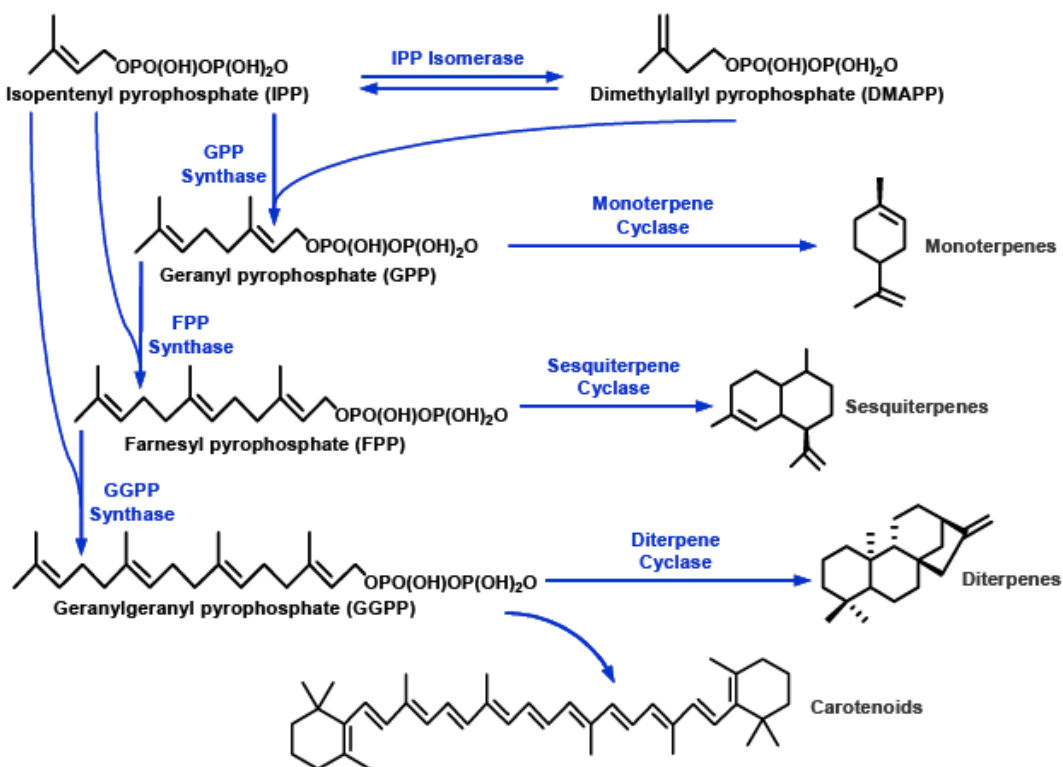


Fig. 2-2. The C5 isoprene units combine in condensation reactions catalyzed by the prenyl transferases to form GPP (C₁₀), FPP (C₁₅), GGPP (C₂₀), and longer prenyl pyrophosphate precursors. The isoprenyl pyrophosphates are then cyclized by the

terpene cyclases (synthases) to form the various terpene classes. Carotenoids are synthesized by a series of enzymatic reactions beginning with the condensation of two GGPP molecules.

Terpene cyclases (or synthases) catalyze the cyclization and rearrangement of these linear precursors to form the diverse carbon backbones found in all terpenoid molecules (Figure 2-2). The terpene cyclases are similar in reaction mechanism and structure to prenyl-transferases. However different terpene cyclases can bind to the same substrate and catalyze the formation of different products²⁵.

To form the final bioactive chemical structure, terpenoids are functionalized in at least one position on the hydrocarbon backbone. This functionalization can be in the form of glycosylation, acetylation, hydroxylation, and benzylation and other possible modifications which even further expand the diversity of these compounds.

Cytochrome P450 enzyme family catalyzes the majority of these modifications and is required for the subsequent activities of acetyl transferases and benzoyl transferases by providing them reactive oxygen attachment sites²⁶. . Cytochrome P450s are heme-containing proteins that catalyze the monooxygenation of a wide range of aromatic and aliphatic compounds and are widely distributed in nature, occurring in almost all organisms²⁶.

Engineering microorganisms for the production of isoprenoids

As mentioned earlier, with this diversity in structure and function, isoprenoids have a vast commercial application and potential. Unfortunately, most of these compounds are harvested from natural sources where they are found in very small quantities, which often results in non-efficient and expensive extraction and purification. Although some of these molecules can be produced by total or partial chemical synthesis, the processes are normally complex and generate low yields^{27, 28}. Engineering metabolic pathways to produce large quantities of isoprenoids in a well-studied microbial host presents an attractive alternative to extraction from environmental sources or chemical syntheses.

Because a number of important isoprenoids of medical and industrial value are derived from plants, researchers have begun to manipulate metabolic pathways to these products in their native hosts. However, for large scale production of isoprenoid products, the production consistency, scalability, and efficiency of substrate-to-product conversion of microbial fermentation make engineering isoprenoid production in tractable microbial hosts an increasingly attractive alternative to plant harvesting. This is particularly important for producing isoprenoid products on the scale and cost of commodity chemicals. The remainder of this chapter will discuss the ongoing research efforts to engineer bacteria and yeast for the production of isoprenoid products.

Engineering of *Escherichia coli* for production of isoprenoids

The initial step toward building a platform in *E. coli* for high-level isoprenoid production requires engineering a strain with high potential for generating the universal precursors IPP and DMAPP.

E. coli employs the DXP or non-mevalonate isoprenoid pathway for the production of precursors IPP and DMAPP. In *E. coli*, these precursors are required for the prenylation of tRNA's²⁹ and the synthesis of FPP by FPP synthase (IspA). FPP is the precursor to octaprenyl diphosphate³⁰ and undecaprenyl diphosphate³¹ used for ubiquinone and cell wall synthesis, respectively.

Mevalonate pathway, primarily found in eukaryotes and non-native to *E. coli*, is the alternative pathway for production of IPP and DMAPP. The existence of these two pathways for IPP biosynthesis offers two possible approaches to increase the production of these precursors in *E. coli*. The first approach is to manipulation of the DXP pathway by changing the metabolix flux and the native regulation; the other approach is to import a heterologous pathway to enhancement IPP production by supplementing the native pathway.

Although different steps in DXP pathway of *E. coli* have only recently been revealed, significant progress has been made in discovering rate limiting steps. These steps were found to be 1-deoxyxylulose-5-phosphate synthase (DXS)³³⁻³⁵, 1-deoxyxylulose-5-phosphate reductoisomerase (DXR, or IspC)^{33,35}, IPP-DMAPP isomerase (IDI)^{32,35} and the respective prenyltransferase (FPP or GGPP synthase)^{32,36}. Over-expression of these rate limiting steps has been shown to increase the production of isoprenoids.

As an alternative to manipulation of native pathways, native regulation of isoprenoid production can be avoided by introducing mevalonate pathway for the production of IPP and DMAPP from metabolites of central metabolism into *E. coli*.

Both of these approaches have their advantages and disadvantages. By expressing an entire heterologous pathway, the metabolites produced are foreign to the cell and may be easier to accumulate in the cell due to lack of feedback regulation and competing pathways. On the other hand there are also possibilities for toxicity and cellular regulation that is difficult to predict or understand. Also the host may not have the native machinery that is required for production of required precursors or managing the key enzymes in the pathway. But it can be very difficult to achieve high fluxes through the native pathway when the intermediate metabolites intersect multiple pathways in the host biochemical network and are subject to strong regulatory checkpoints. Toxicity and growth inhibition can be a result of high concentration of certain metabolites or depletion of native intracellular metabolites.

Although considerable improvements in isoprenoid production were achieved in the prior studies of the native *E. coli* DXP pathway, it seems that flux through this pathway remains limited by intracellular regulatory control. For this reason we bypassed the entire *E. coli*'s DXP pathway by expression and optimization of the entire mevalonate pathway of *S. cerevisiae* in *E. coli* for the synthesis of isoprenoid precursors.

Engineering and modifying the bacterial cell's metabolic and regulatory networks, outside of the DXP pathway, has also resulted in improvements in heterologous isoprenoid production. Flux to carotenoid products were enhanced by balancing the intracellular pools of glyceraldehyde-3-phosphate (G3P) and pyruvate, both substrates of DXP³⁷. By overexpressing or inactivating genes in glycolysis and the tricarboxylic acid cycle, cellular levels of G3P were increased in comparison to pyruvate. The genetic modifications were combined with over-expression of the three rate limiting enzymes discussed above (*idi*, *ispA* and a heterologous GGPP synthase) to increase the production of lycopene³⁷.

Metabolic flux experiments using labeled carbon sources are also useful for visualizing the relationships between major metabolic pathways and for identifying new opportunities to increase flux through the target pathway. Using sequential application of genome-wide stoichiometric flux balance analysis (FBA) with maximize cell growth chosen as the objective function, Alper *et al.* have identified single-gene and multiple-gene knockout targets predicted to increase lycopene yields³⁸. Choosing the predicted best knockout mutant from one round as the cellular genotype of the sequential round identified two different three-gene knockout combinations that increased lycopene production by up to 37% in a strain already overexpressing genes *dxs*, *idi*, & *ispFD*. The combinations identified were: glutamate dehydrogenase, pyruvate dehydrogenase, formate dehydrogenase H ($\Delta gdhA \Delta aceE \Delta fdhF$) and glutamate dehydrogenase, pyruvate dehydrogenase, phosphoglucosmutase II ($\Delta gdhA \Delta aceE \Delta talB$). Based on the resulting flux balances, the authors presume that the enhancement is due to increased availability of precursors and cofactors needed for lycopene biosynthesis. In optimized media, substrate, and growth conditions, this improved to 2-fold over the parental strain³⁹. However in case of expressing a heterologous pathway in *E. coli*, achieving the optimal expression of the enzymes in the pathway is more challenging than redirecting the flux through the native biochemical pathways towards the imported heterologous pathway. Specifically, the imported mevalonate pathway is not restricted by availability of acetyl-CoA but rather by the efficient and balanced expression of the enzymes involved in the mevalonate pathway. Our attempts in directing flux towards acetyl-CoA production by knocking out the entire acetate-biosynthetic pathway only resulted in accumulation of pyruvate instead of acetate and did not improve IPP and DMAPP production (Figure 3).

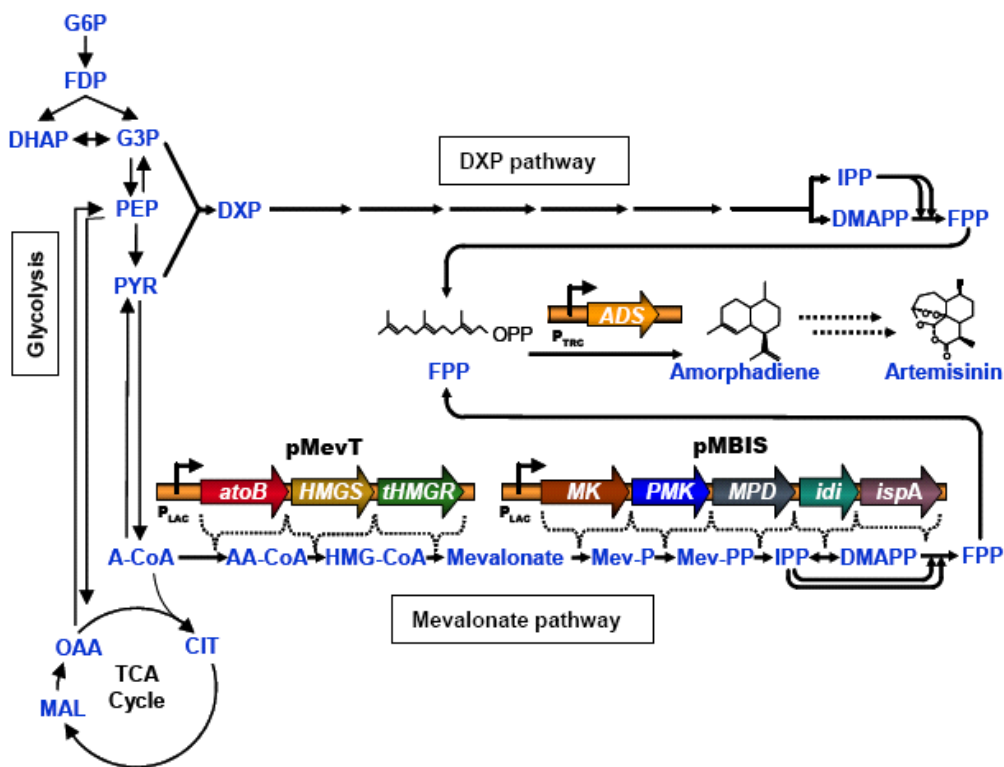


Fig. 2-3. Heterologous mevalonate pathway and native DXP (non-mevalonate) pathway for the production of amorphaadiene.

Therefore it is critical to optimize mevalonate pathway to utilize the full potential of the cell in achieving high yields of IPP and DMAPP.

The mevalonate pathway was divided into two synthetic operons, referred to as the upper (MevT) and lower (MBIS) sections of the pathway. MevT is responsible for the three-step conversion of acetyl coenzyme A (acetyl-CoA) to mevalonate, whereas MBIS takes mevalonate to FPP (Figure 3). Initial work showed that expression of the mevalonate pathway in *E. coli* leads to severe growth inhibition as an outcome of toxicity resulting from accumulation of the prenyl diphosphates (FPP, IPP and/or DMAPP)⁴⁰. Coexpression of a codon-optimized terpene synthase from the artemisinin pathway, amorpho-4, 11-diene (amorphaadiene) synthase (ADS)^{41, 42}, alleviates this toxicity and can be used to evaluate IPP and DMAPP production as the downstream amorphaadiene product. Production of amorphaadiene from the imported mevalonate pathway was found to be substantially higher than production with a supplemental DXP pathway⁴⁰. Closer examination of the mevalonate pathway suggests that availability of mevalonate may limit amorphaadiene production⁴⁶. However, when MevT expression is increased, small molecule-dependent growth inhibition of *E. coli* is observed; this result has been attributed to hydroxymethylglutaryl-CoA (HMG-CoA) toxicity⁴³. Expression of an additional copy of HMGR alleviates buildup of HMG-CoA and increases productivity of mevalonate by almost two-fold by relieving growth inhibition⁴³. Following up on these results, library-based engineering of the intergenic regions of the polycistronic message for the MevT operon was carried out to balance expression of the individual genes in the MevT operon, resulting in a seven-fold overall increase in mevalonate titers due to improvements in specific production and cell growth⁴⁴. Because of this toxicity, increasing gene stability by high-level chromosomal expression of the mevalonate pathway may also greatly increase production, as it has with the DXP pathway⁴⁵.

In an attempt to increase the effective concentration of each enzyme of the MevT operon,

Synthetic protein scaffolds were built to recruit metabolic enzymes and localize them in close proximity. Pathway enzymes are colocalized to synthetic complexes in a designable manner using engineered interactions between well-characterized protein-protein domains and their specific ligands. This increases the effective concentrations of metabolic intermediates, while preventing their accumulation to toxic levels. This approach led to a 77-fold improvement in mevalonate titer with low enzyme expression and reduced metabolic load⁴⁶. These efforts show that carbon can be effectively channeled for production of exogenous isoprenoids in *E. coli* even though the flux through native isoprenoid biosynthesis is relatively low. Despite all the attempts in optimizing the upper mevalonate pathway (MevT), there has been little done to improve the flux through the lower mevalonate pathway (MBIS) even though this part of the pathway appears to significantly restrict production of FPP.

Although the above examples have all succeeded in producing various non-native isoprenoids in *E. coli*, further work is required to increase production to levels necessary for industrial application.

Engineering of yeasts for the production of isoprenoids

The majority of research to date on the production of isoprenoids in yeast has focused on the production of endogenous compounds such as sterol products. More recently, researchers have begun to engineer the production of heterologous isoprenoids. The production of native and non-native isoprenoid products rely on IPP precursor derived from the endogenous mevalonate pathway.

From the extensive research into sterol biosynthesis in eukaryotes it was determined that HMG-CoA reductase is the main regulatory enzyme of the eukaryotic mevalonate pathway^{18-19, 47}. Because *S. cerevisiae* Hmgp1 is a transmembrane protein that localizes to the cell's ER membrane, overexpression of the full enzyme results in induction of ER membrane stacks (karmellae), but no increase in sterol production⁴⁸. Instead, truncating the N-terminal transmembrane domain of *S. cerevisiae* Hmgp1 creates an active cytosolic truncated protein. Overexpression of Hmgp1 with various N-terminal truncations in yeast produces soluble enzyme, significantly increases the accumulation of the sterol precursor squalene (derived from FPP by the action of squalene synthase)^{49, 50}, and increases the accumulation of several sterols⁵⁰, but to a lesser extent.

Non-native sterols, heterologous sesquiterpenes and diterpenes have also been produced in yeast. *S. cerevisiae* is often considered to be the logical choice for production of plant natural products as it has the capacity for posttranslational modification of enzymes and has less natural codon bias.

References

1. Hall, C.a. Dictionary of Natural Products. (2004 (2)).
2. McCaskill, D. & Croteau, R. Prospects for the bioengineering of isoprenoid biosynthesis. *Adv. Biochem. Eng./Biotechnol.* 55, 107-146 (1997).
3. McCaskill, D. & Croteau, R. Some caveats for bioengineering terpenoid metabolism in plants. *Trends in Biotechnology* 16, 349-355 (1998).
4. McGarvey, D.J. & Croteau, R. Terpenoid metabolism. *The Plant Cell* 7, 1015-1026 (1995).
5. Fraga, B.M. in Methods in Plant Biochemistry, Vol. 7. (eds. B.V. Charlwood & D.V. Banthorpe) 145-185 (Academic Press, San Diego, CA; 1991).
6. Haralampidis, K., Trojanowska, M. & Osbourn, A.E. Biosynthesis of triterpenoid saponins in plants. *Adv Biochem Eng Biotechnol* 75, 31-49 (2002).
7. Wake L.V. and Hillen L.W. Study of a 'bloom' of the oil-rich alga *Botryococcus braunii* in the Darwin river reservoir. *Biotechnol. Bioeng.* 22, 1637–1656 (1980).
8. Spurgeon, S.L. & Porter, J.W. (eds.) Biosynthesis of Isoprenoid Compounds (Wiley, New York, 1981).
9. Rohmer, M., Knani, M., Simonin, P., Sutter, B. & Sahm, H. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.* 295, 517–524 (1993).

10. Rohdich, F., Hecht, S., Bacher, A. & Eisenreich, W. Deoxyxylulose phosphate pathway of isoprenoid biosynthesis. Discovery and function of ispDEFGH genes and their cognate enzymes. *Pure Appl. Chem.* 75, 393–405 (2003).
11. Takagi, M., Kuzuyama, T., Takahashi, S. & Seto, H. A gene cluster for the mevalonate pathway from *Streptomyces* sp. Strain CL190. *J Bacteriol* 182, 4153-4157. (2000).
12. Wilding, E.I. et al. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol* 182, 4319-4327. (2000).
13. Fraser, C.M. et al. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390, 580-586 (1997).
14. Smeds, A., Joutsjoki, T. & Palva, A. Identification of a gene cluster for the mevalonate pathway in *Lactobacillus helveticus*. *DNA Seq* 12, 187-190 (2001).
15. Lange, B.M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A* 97, 13172-13177. (2000).
16. Middleton, B. The kinetic mechanism of 3-hydroxy-3-methylglutaryl-coenzyme A synthase from baker's yeast. *Biochem J* 126, 35-47 (1972).
17. Steussy, C.N. et al. X-ray crystal structures of HMG-CoA synthase from *Enterococcus faecalis* and a complex with its second substrate/inhibitor acetoacetyl-CoA. *Biochemistry* 44, 14256-14267 (2005).
18. Hampton, R.Y. & Bhakta, H. Ubiquitin-mediated regulation of 3-hydroxy-3-methylglutaryl-CoA reductase. *Proc Natl Acad Sci U S A* 94, 12944-12948. (1997).
19. Meigs, T.E., Roseman, D.S. & Simoni, R.D. Regulation of 3-hydroxy-3-methylglutaryl-coenzyme A reductase degradation by the nonsterol mevalonate metabolite farnesol in vivo. *J Biol Chem* 271, 7916-7922. (1996).
20. Oulmouden, A. & Karst, F. Nucleotide sequence of the ERG12 gene of *Saccharomyces cerevisiae* encoding mevalonate kinase. *Curr Genet* 19, 9-14 (1991).
21. Campos, N. et al. *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *Biochem J* 353, 59-67. (2001).
22. Bochar, D.A. et al. 3-hydroxy-3-methylglutaryl coenzyme A reductase of *Sulfolobus solfataricus*: DNA sequence, phylogeny, expression in *Escherichia coli* of the hmgA gene, and purification and kinetic characterization of the gene product. *J Bacteriol* 179, 3632-3638. (1997).
23. Altincicek, B. et al. GcpE is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway of isoprenoid biosynthesis in *Escherichia coli*. *J Bacteriol* 183, 2411-2416. (2001).
24. Hamano, Y. et al. Cloning of a gene cluster encoding enzymes responsible for the mevalonate pathway from a terpenoid-antibiotic-producing *Streptomyces* strain. *Biosci Biotechnol Biochem* 65, 1627-1635. (2001).
25. Chappell, J. The biochemistry and molecular biology of isoprenoid metabolism. *Plant Physiol.* 107, 1-6 (1995).
26. Nelson, D.R. et al. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6, 1-42 (1996).

27. Danishefsky, S.J. et al. Total synthesis of baccatin III and taxol. *Journal of the American Chemical Society* 118, 2843-2859 (1996).
28. Nicolaou, K.C. et al. Total synthesis of eleutherobin. *Angew Chem Int Ed* V36, 2520-2524 (1997).
29. Connolly, D.M. & Winkler, M.E. Genetic and physiological relationships among the miaA gene, 2- methylthio-N⁶-(delta 2-isopentenyl)-adenosine tRNA modification, and spontaneous mutagenesis in Escherichia coli K-12. *J Bacteriol* 171, 3233-3246. (1989).
30. Asai, K. et al. The identification of Escherichia coli ispB (cel) gene encoding the octaprenyl diphosphate synthase. *Biochem Biophys Res Commun* 202, 340-345. (1994).
31. Apfel, C.M., Takacs, B., Fountoulakis, M., Stieger, M. & Keck, W. Use of genomics to identify bacterial undecaprenyl pyrophosphate synthetase: cloning, expression, and characterization of the essential uppS gene. *J Bacteriol* 181, 483-492. (1999).
32. Wang, C.W., Oh, M.K. & Liao, J.C. Engineered isoprenoid pathway enhances astaxanthin production in Escherichia coli. *Biotechnol Bioeng* 62, 235-241 (1999).
33. Albrecht, M., Misawa, N. & Sandmann, G. Metabolic engineering of the terpenoid biosynthetic pathway of Escherichia coli for production of the carotenoids b-carotene and zeaxanthin. *Biotechnol. Lett.* 21, 791–795 (1999).
34. Matthews, P.D. & Wurtzel, E.T. Metabolic engineering of carotenoid accumulation in Escherichia coli by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl. Environ. Microbiol.* 53, 396–400 (2000).
35. Kim, S.-W. & Keasling, J.D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in Escherichia coli enhances lycopene production. *Biotechnol. Bioeng.* 72, 408–415 (2001).
36. Neudert, U., Martinez-Ferez, I.M., Fraser, P.D. & Sandmann, G. Expression of an active phytoene synthase from Erwinia uredovora and biochemical properties of the enzyme. *Biochim. Biophys. Acta* 1392, 51–58 (1998).
37. Farmer, W.R. & Liao, J.C. Precursor balancing for metabolic engineering of lycopene production in Escherichia coli. *Biotechnol Prog* 17, 57-61 (2001).
38. Alper, H., Jin, Y.S., Moxley, J.F. & Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metab Eng* 7, 155-164 (2005).
39. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 23, 612-616 (2005).
40. Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D. & Keasling, J.D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat. Biotechnol.* 21, 796–802 (2003).
41. Mercke, P., Bengtsson, M., Bouwmeester, H.J., Posthumus, M.A. & Brodelius, P.E. Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in Artemisia annua L. *Arch. Biochem. Biophys.* 381, 173–180 (2000).
42. Chang, Y.J., Song, S.H., Park, S.H. & Kim, S.U. Amorpha-4,11-diene synthase of Artemisia annua: cDNA isolation and bacterial expression of a terpene synthase involved in artemisinin biosynthesis. *Arch. Biochem. Biophys.* 383, 178–184 (2000).

43. Pitera, D.J., Paddon, C., Newman, J.D. & Keasling, J.D. Rebuilding a balanced heterologous mevalonate pathway for isoprenoid production in *Escherichia coli*. *Metab. Eng.* **2**, 193-207 (2007)
44. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24**, 1027–1032 (2006).
45. Yuan, L.Z., Rouviere, P.E., Larossa, R.A. & Suh, W. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* **8**, 79–90 (2006).
46. Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T., Petzold, C.J., Ullal, A.V., Prather, K.L., & Keasling, J.D. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753-759 (2009)
47. Hampton, R., Dimster-Denk, D. & Rine, J. The biology of HMG-CoA reductase: the pros of contra-regulation. *Trends Biochem Sci* **21**, 140-145. (1996).
48. Wright, R., Basson, M., D'Ari, L. & Rine, J. Increased amounts of HMG-CoA reductase induce "karmellae": a proliferation of stacked membrane pairs surrounding the yeast nucleus. *J Cell Biol* **107**, 101-114 (1988).
49. Donald, K., Hampton, R. & Fritz, I. Effects of overproduction of the catalytic domain of 3-hydroxy-3- methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **63**, 3341- 3344 (1997).
50. Polakowski, T., Stahl, U. & Lang, C. Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. *Appl Microbiol Biotechnol* **49**, 66-71. (1998).

CHAPTER 3: Investigating the Impact of Heterologous Amorphadiene Production in *Escherichia Coli*; Flux Analysis.

Introduction

Changes to pathways or culture conditions can affect an organism's entire metabolism. Metabolic flux analysis allows insight into such changes using a mathematical representation of the cell's metabolism and measurements of secreted metabolite and biomass component concentrations. A metabolic flux is the number of moles of a particular metabolite that are consumed or produced via a particular reaction per unit cell mass (typically in dry cell weight [g] or OD₆₀₀) per unit time. The basis of the mathematical representation is a series of simultaneous steady-state mole balances on cellular metabolites. This set of balances can be represented in matrix notation:

$$S \cdot v = b \quad (1)$$

where S is the stoichiometric matrix containing the stoichiometric coefficients for reactions in the network. The rows of this matrix correspond to reactions, and the columns correspond to metabolites. The vector v contains the unknown intracellular fluxes for each reaction (mol vol⁻¹ time⁻¹), and b is the vector of net rates of formation of metabolites (mol vol⁻¹ time⁻¹). The system

is assumed to be at steady state, implying that there is no net accumulation of intracellular metabolites, and therefore the entries of b are zero for intracellular metabolites. A small example network and the resulting matrix equation is shown in Figure 3-1.

In any branched metabolic network, the number of reactions is greater than the number of metabolites, so the stoichiometric matrix is not square and cannot be inverted. Such a system is called underdetermined, because there are not enough constraints to find a unique solution. The number of degrees of freedom of the system is the number of reactions (rows of S) minus the number of metabolites (columns of S), and this difference is the number of additional constraints that are necessary to render S square and allow it to be inverted. These additional constraints can be provided by measuring fluxes to extracellular metabolites (A, C, E, G, and I in the example network), which are not zero at steady state, and appending these measurements and their corresponding reactions to b and S , respectively. The flux vector v can then be determined. In metabolic networks, 'extracellular' metabolites are secreted metabolites such as organic acids and carbon dioxide, and biomass components such as protein, lipid, and carbohydrate.

Because the stoichiometric matrix is sparse, the solution obtained by inverting an exactly determined system is potentially sensitive to small changes in measured extracellular measurements. Measuring more extracellular species than necessary to determine the system builds some redundancy into the measurements, and allows improved flux calculations to be obtained using an iterative solution that minimizes the square of the difference between the calculation and the measured fluxes. However, it is often not possible to measure enough extracellular metabolites to determine the system. Also, regardless of the number of measurements made, this flux-balancing approach alone can not determine the relative forward and reverse fluxes in reversible reactions, the flux through branching pathways that converge later (such as glycolysis and the pentose phosphate pathway), or the flux through metabolic cycles (such as the TCA cycle). For determination of these fluxes, another approach is necessary.

Isotopomer analysis

Isotopomer analysis is a technique whereby a known distribution of ^{13}C -labeled carbon source such as glucose is fed to the cells, and the labeling pattern in resulting metabolites, often the amino acids, is measured. Because all of the atom transitions between metabolites in the biochemical network of *E. coli* are known, the path that each atom takes through the network can

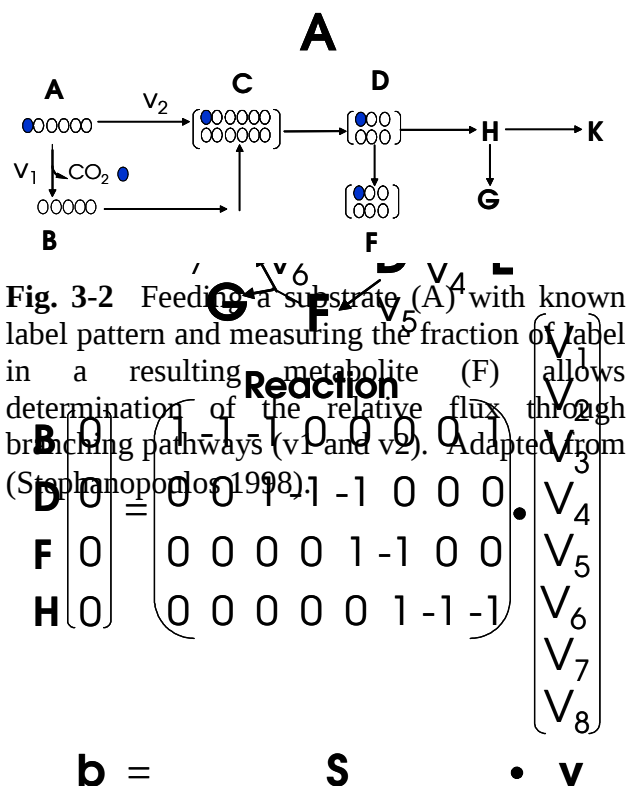
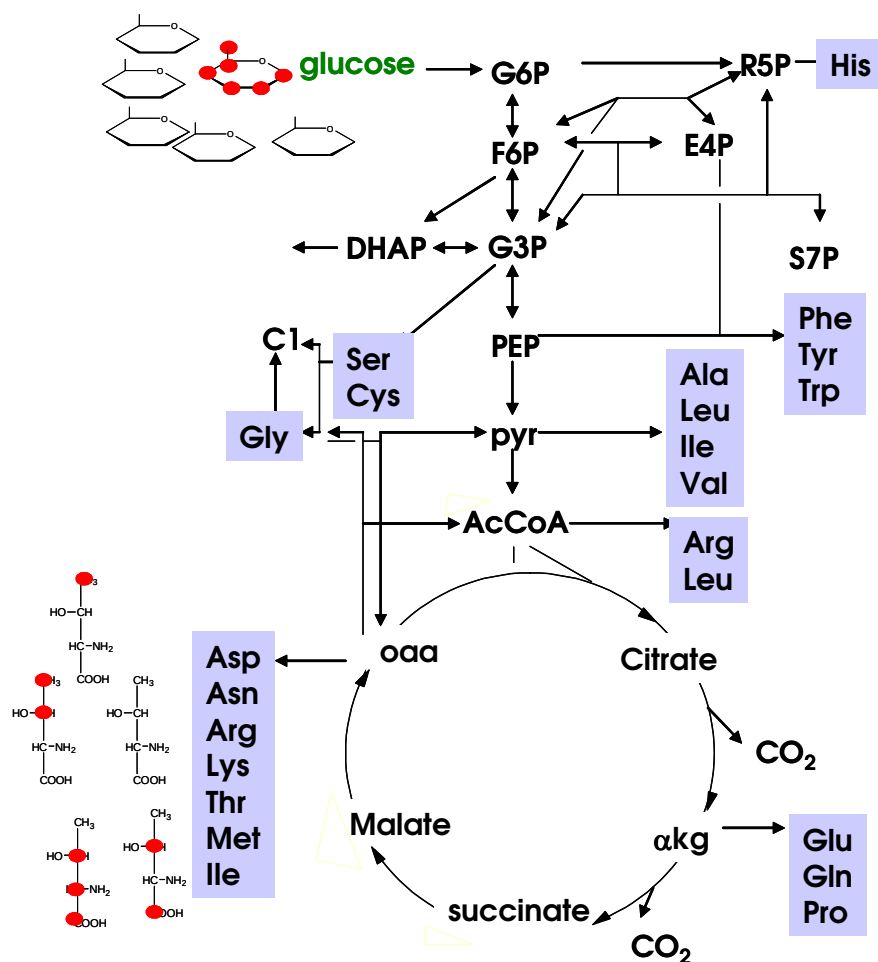


Fig. 3-2 Feeding a substrate (A) with known label pattern and measuring the fraction of label in a resulting metabolite (F) allows determination of the relative flux through branching pathways (v_1 and v_2). Adapted from (Stephanopoulos 1998).

$$\begin{bmatrix} B \\ C \\ D \\ E \\ F \\ G \\ H \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{bmatrix}$$

be traced. Isotopomer analysis allows branching pathway fluxes to be determined, as illustrated in Figure 3-2. It also can be used to determine the flux through metabolic cycles such as the TCA cycle, since the label pattern of metabolites in the TCA cycle changes with each turn through the cycle. This is represented in Fig. 3. Amino acids are often used for isotopomer analysis, because they are synthesized from many different precursors in central metabolism. Measuring their isotopomer distributions provides flux constraints on many parts of the metabolic network.

Each possible label pattern corresponds to an isotopomer. A molecule with n carbons has 2^n possible isotopomers. These are often represented using binary numbers from 0 to $2^n - 1$, where the 0s in the binary representation correspond to unlabeled carbons, and the 1s in the binary representation correspond to labeled carbons. For example, the label pattern for isotopomer #6 in a three-carbon molecule is the binary representation of 6: 1 1 0, which corresponds to a molecule with its first two carbons labeled. There are several methods that can be used to trace the transitions of metabolites through the metabolic network



The method of atom-mapping matrices (AMMs) is concise and general and was used in our work. In this method, matrices are made that map the carbon transitions for each reactant-product pair in each reaction. For example, the pentose-phosphate reaction:

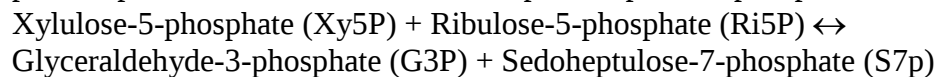


Fig. 3-3 Feeding a known label distribution of glucose results in a variety of label patterns in the resulting amino acids such as threonine, shown above. The label is scrambled as the network branches and comes back together. Amino acids are often used in isotopomer analysis because they originate from many points in central metabolism.

requires four atom-mapping matrices: Xy5P→G3P, Ri5P→G3P, Xy5P→S7P, and Ri5P→S7P. The atom mapping matrix for the first pair, Xy5P→G3P is:

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where the columns correspond to the carbons in the reactant, Xy5P, and the rows correspond to the carbons in the product, G3P. This matrix demonstrates that the third carbon in Xy5P becomes the first carbon in G3P.

A simple algorithm can be used to convert these AMMs into isotopomer mapping matrices (IMMs), which are larger in dimension¹. The columns of IMMs correspond to each of the 2ⁿ possible reactant isotopomers, and the rows correspond to each of the 2ⁿ possible product isotopomers. The binary representations are used. The left half of the 8x32 isotopomer mapping matrix for Xy5P→G3P is:

$$\begin{array}{l} \begin{matrix} 00000 \\ 00001 \\ 00010 \\ 00011 \\ 00100 \\ 00101 \\ 00110 \\ 00111 \\ 01000 \\ 01001 \\ 01010 \\ 01011 \\ 01100 \\ 01101 \\ 01110 \\ 01111 \\ \dots \end{matrix} \\ \begin{pmatrix} \begin{matrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{matrix} & \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \end{matrix} \end{pmatrix} \end{array}$$

With isotopomer mapping matrices for each reactant-product pair, the isotopomer distribution in the product molecules G3P and S7P can be calculated by first multiplying each IMM by the isotopomer distribution of the corresponding reactant molecule. The resulting vectors, of dimension equal to the number of product carbons, represent the contribution from each reactant to the isotopomer distribution of the product. The complete isotopomer distribution of the product is subsequently obtained by creating a vector containing the products of element-wise multiplication of the reactant distribution vectors: the first element of the product isotopomer distribution is equal to the first element of one contributing vector times the first element of another, and so on.

For the example reaction, this operation corresponds to:

$$\mathbf{I}_{\text{G3P}} = (\mathbf{IMM}_{\text{Xy5P} \rightarrow \text{G3P}} \cdot \mathbf{I}_{\text{Xy5P}}) \otimes (\mathbf{IMM}_{\text{Ri5P} \rightarrow \text{G3P}} \cdot \mathbf{I}_{\text{Ri5P}}) \text{ and} \\ \mathbf{I}_{\text{S7P}} = (\mathbf{IMM}_{\text{Xy5P} \rightarrow \text{S7P}} \cdot \mathbf{I}_{\text{Xy5P}}) \otimes (\mathbf{IMM}_{\text{Ri5P} \rightarrow \text{S7P}} \cdot \mathbf{I}_{\text{Ri5P}})$$

where $\mathbf{I}$ is the distribution of isotopomers for a given species, and $\otimes$ represents element-wise multiplication¹.

This calculation forms the basis of an isotopomer flux calculation. Multiplying the isotopomer distribution vectors by the flux through the corresponding reaction gives the flux to/from each isotopomer. With this method, isotopomer path tracing can be carried out. By starting with a known label distribution in a substrate, and with an initial guess for a set of fluxes, the resulting product isotopomer distribution can be determined by marching forward with the described calculation and tracing the label through successive pathways. The resulting label distribution can be compared with the observed label distribution and new guesses can be made.

Materials and Method

Studies used *Escherichia coli* DH1 containing one of two sets of three plasmids: either the three plasmids that confer amorphaadiene production via the mevalonate pathway or the corresponding three empty vectors². Strains were cultivated at 37°C in modified M9 minimal medium³ containing the following, per liter: 12.8 g NaH₂PO₄·7H₂O, 3 g KH₂PO₄, 0.5 g NaCl, 0.35 g NH₄Cl, 0.002 mol MgSO₄, 0.001 mol CaCl₂, supplemented with thiamine, iron and micronutrients. The carbon source used was glucose, supplied at 20% uniformly ¹³C-labeled (U-¹³C-glucose, Spectra Stable Isotopes, Columbia, Maryland), with an overall concentration of 10.0 g/L. The appropriate antibiotics were used at the following concentrations 25 µg ml⁻¹ chloramphenicol, 5 µg ml⁻¹ tetracycline, and 100 µg ml⁻¹ ampicillin. Cultures were incubated in the presence of 1/10 the culture volume of dodecane to sequester amorphaadiene and prevent its evaporation. Amorphaadiene-producing and control cultures were induced with 0.5 mM IPTG at an OD of 0.25-0.3 unless otherwise noted.

Chemostat cultivation

Six-hundred-milliliter cultures were grown in a 1 L Sartorius (Göttingen, Germany) fermentor. M9 medium was used, as described above, with 20% of the carbon supplied as U-¹³C glucose and with the addition of a 1/10 volume dodecane phase. Antifoam A (Sigma) was added at 100 µl per liter. pH was controlled at 7.0 by the addition of NaOH, the temperature was set at 37 °C,

and the vessel was aerated with sterile air. The agitation rate was set at 500 rpm and the dissolved oxygen level was at 450% of saturation at all times. The composition of the fermentation off-gas was monitored using a VG Prima δ B mass spectrometer. The vessel was inoculated with 30mL of culture already grown for several generations at the appropriate label concentration. The fermentor was initially maintained in batch mode. After CO₂ evolution peaked, continuous culture was started. Two inlet streams were operated, one feeding modified M9 medium containing antifoam, the appropriate antibiotics, and IPTG, operating at a dilution rate of 0.19; and the other feeding dodecane at 1/10 the flow rate of the medium inlet stream. A constant culture volume was maintained by positioning a harvest tube at the top surface of the working volume and operating the harvest line pump at a higher flow rate than the feed pump, so that the liquid level remained constant at the height of the harvest tube position. The dodecane and M9 medium proved to be well-emulsified so that a single harvest tube removed cell-containing medium and dodecane in their volumetric proportions. Optical density at a wavelength of 600 nm (OD₆₀₀) and amorphadiene measurements were taken periodically over the course of the experiment. Steady state was indicated by a constant concentration of CO₂ in the off-gas and a constant cell density. Once stable optical density was reached, steady state was insured by waiting an additional three vessel volume changes before sampling was begun. Flowrate was maintained to give a dilution rate of 0.19 h⁻¹, and the average OD₆₀₀ of the culture during the sampling period was 3.9. After taking each 50mL sample, the vessel was immediately filled to the original volume using fresh media, and the system was allowed to return to steady state (three volume changes) before the next sample.

Amorphadiene analysis

Amorphadiene was measured by first diluting 10 μ l of the dodecane phase in 690 μ l of ethyl acetate containing approximate 5 μ g/L of caryophyllene as an internal standard. Samples were subsequently analyzed using a gas chromatograph-mass spectrometer (GC-MS, Agilent) equipped with a DB-5 column (Agilent). Unsplit injections of 1 μ l were performed. The following temperature program was used: hold at 80°C for 2 minutes, increase at 30°C per minute to 160°C, increase at 3°C per minute to 170°C, increase at 50°C per minute to 300°C and hold for 2 minutes. The mass spectrometer was operated in single ion mode, monitoring the abundance of ions 189 and 204. Amorphadiene levels were determined using an amorphadiene standard curve and adjusting for the relative abundance of ions 189 and 204 in the mass spectra of caryophyllene and amorphadiene.

When cultures were grown on labeled carbon, the mass spectrometer was operated in scan mode, scanning ions 189-202, which spans the range of all possible label states for the 14-carbon 189 ion (loss of a methyl group from the molecular ion of weight 204). The sum of all ions in this range was used to determine the total abundance of amorphadiene. The 204 ion could not be used in this case, because there is a relatively large background peak at 207.

Mass isotopomer measurements via GC-MS

Pellets from 5 ml of culture at OD₆₀₀ between 0.5-1.8 (depending on the culture and time point) were used for isotopomer analysis. Pellets were washed once with 0.9% NaCl, flash frozen in liquid nitrogen and stored at -87 °C for later analysis. Washed pellets were resuspended in 1 ml water and lysed by sonication for 3 minutes (Vir Sonic S-300-200). Proteins were precipitated from the resulting samples using TCA, and the protein pellet was hydrolyzed for 24 hours at 110 °C in a glass vial containing 1.8 ml 6 M HCl. Samples were dried under a stream of air, and

derivatized by adding 100 μ l of THF and 100 μ l of N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (TBDMS-FA) ⁴ and incubating the samples at 74°C for 1 hour. Samples were subsequently injected (1 μ l with 1:10 split) into a GC-MS (Agilent) equipped with a DB-5 column. The following oven conditions were used: hold at 150°C for 2 minutes; increase at 3°C per minute to 280°C; increase at 20°C per minute to 300°C; hold at 300°C for 5 minutes. Amino acid peaks were identified using derivatized pure amino acid standards and the NIST mass spectral library. Ion fragments were identified by molecular weight analysis and comparison of results with other published reports⁵. Each derivatized sample was injected four times, thus providing at least eight total data points for the calculation of means and variances. Raw mass isotopomer data were corrected for naturally occurring ¹³C in the derivatization reagents and non-carbon isotopes in the entire fragment using an infinite dimensional matrix calculus method⁶. In addition, the aqueous phase of culture supernatants was assayed for residual glucose (2300 STAT Plus glucose analyzer, YSI Life Sciences, Yellow Springs, Ohio) and acetate (kit from R-Biopharm AG, Darmstadt, Germany), while the concentration of amorphadiene was determined in the organic phase.

Experimental Isotopomer Measurements

The experimental measurements of labeled carbon in amino acids can be done either by nuclear magnetic resonance (NMR) spectroscopy or by gas chromatography-mass spectrometry (GC-MS). NMR spectroscopy offers the advantage that individual isotopomers can be distinguished, i.e. all singly labeled isotopomers of an asymmetric three-carbon molecule will have distinct chemical shifts. However, not all isotopomers can be detected using this technique, since carbon atoms separated by more than one bond do not influence each other's NMR spectra. Also, the sensitivity of this technique is less than that of GC-MS. GC-MS, on the other hand, detects the mass distributions of labeled molecules, rather than individual isotopomers. That is, GC-MS determines what fraction of the total population of a particular molecule or molecular fragment is unlabeled, singly labeled, doubly labeled, etc., rather than exactly which carbons are labeled. GC-MS provides less information than is obtained by measuring individual isotopomers via NMR spectroscopy, but all carbon atoms in larger molecules can be considered using GC-MS, and the detection limit of GC-MS is very low. Also, because molecules fragment in the mass spectrometer detector, the mass distribution for pieces of the molecules can also be determined, this provides additional constraints.

Analysis of amino acids via gas chromatography requires that the amino acids be derivatized, commonly with silylation reagents. This not only makes the molecules volatile enough to enter the GC column, but it also creates large enough molecules that the amino acid carbon skeleton fragments in the mass-spectrometer detector.

N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (TBDMS) ⁴ is commonly used and was used in our work. The molecule that results from derivatization of amino acids is shown in Figure 4 with possible fragmentation sites marked.

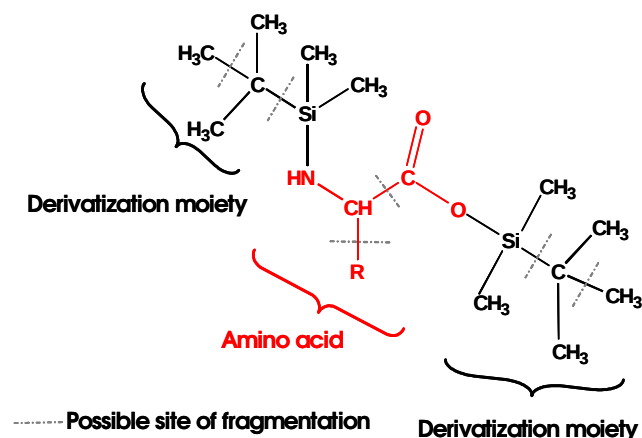


Fig. 3-4 TBDMS-derivatized amino acid and fragmentation sites.

Construction of large-scale isotopomer mapping model

The isotopomer model was constructed using as a starting point the large-scale stoichiometric model of *E. coli* metabolism, iJR904⁷, which has been successfully applied to predict the phenotypes of various *E. coli* strains, both wild-type and mutant, under certain conditions^{8, 9}. The iJR904 model contains 931 intracellular reactions although many of these reactions do not contribute to the labeling patterns of the examined amino acids. Accordingly, in this study, we developed a smaller, though still biologically comprehensive, metabolic model for the flux elucidation. The model reduction procedure first involved removing from iJR904 all blocked reactions defined as reactions that cannot carry flux during aerobic growth on glucose due to stoichiometric limitations¹⁰. Examples of blocked reactions include transporter reactions for components absent from the media such as xylose, glycerol, or fructose, and reactions involved in their utilization. This enabled the removal of nearly one third of the reactions from iJR904. In addition, reactions in the cell envelope, membrane lipid, nucleotide, and cofactor biosynthetic pathways were not explicitly included in the model enabling the removal of approximately another one third of the reactions from iJR904. However, because these biosynthetic pathways can affect amino acid labeling patterns by draining precursor metabolites away from central metabolism, the biomass equation, which was based largely on the one described in Edwards and Palsson¹¹, was modified to account for these additional drains on central metabolism. Lastly, we assumed that the catabolic reactions in the nucleotide salvage pathways do not contribute significantly to the labeling patterns of the amino acids. The model constructed for this study contained 238 reactions (including biomass) and 184 metabolites, as well as 32 exchange fluxes. This total includes a set of reactions absent from iJR904 that enable amorphaadiene production via the non-native mevalonate pathway². Reversible reactions were broken into their forward and backward components to enable the investigation of how reaction reversibility affects the labeling patterns of the amino acids. The network contained 80 such reversible reactions, bringing the total number of flux describing variables to 350. The model included all reactions of Embden–Meyerhoff–Parnas (EMP) and Entner–Doudoroff (ED) glycolysis, the tricarboxylic acid (TCA) cycle, and the PPP. In addition, all of the anaplerotic reactions and amino acid biosynthesis and degradation pathways were included. Finally, the model enforced the explicit balancing of all metabolic cofactors (e.g., ATP, NADH, NADPH) by including reactions for energy generation via substrate-level and oxidative phosphorylation as well as transhydrogenase

activity. The AMMs, which describe the transfer of carbon atoms from the reactants to products¹², were calculated using Pipeline PilotTM (SciTegic Inc., San Diego, CA), a high-throughput data analysis and mining system for chemo-informatics applications, using a structural matching algorithm. AMMs are matrices where the columns represent carbon atoms in the reactant and the rows represent carbon atoms in the product. The numbering scheme is consistent with the order of atoms represented in the molecule (.mol) file. There is one AMM per reactant–product pair. The input to the program was a list of reactions with associated reactants and products and their KEGG ID numbers (<http://www.genome.jp/kegg/ligand.html>). The program then extracted the appropriate molecule files (.mol format) and calculated the predicted AMM as well as a score indicating the quality of the match. The above procedure was applied for the reactions in the E. coli model described above, and the results checked manually based on known biochemistry. Approximately 80% of the auto-generated AMMs were found to be correct while the remaining were subsequently corrected manually. IMMs were then calculated directly from the AMMs using the algorithm introduced by Van Dien et al.¹³. IMMs indicate the possible product isotopomers that can be created from each reactant isotopomer¹. IMMs were used in the isotopomer balance equations to determine the IDVs for each metabolite. The numbering scheme shown in Fig. 5 is used for the isotopomers. The final step of the initial analysis was to calculate the mass distribution vector (MDV) for all observed products from the mass spectrometry data¹⁴. For each amino acid that was detected, elements of the IDV with the same number of labeled carbons were summed to yield an element of the MDV, as shown in Fig. 1. For amino acid fragments (for example, a common fragment is missing the carboxy-terminal carbon), the procedure was modified to include only the relevant carbon atoms. The process was facilitated by a Matlab program that makes use of matrix algebra^{13, 14}.

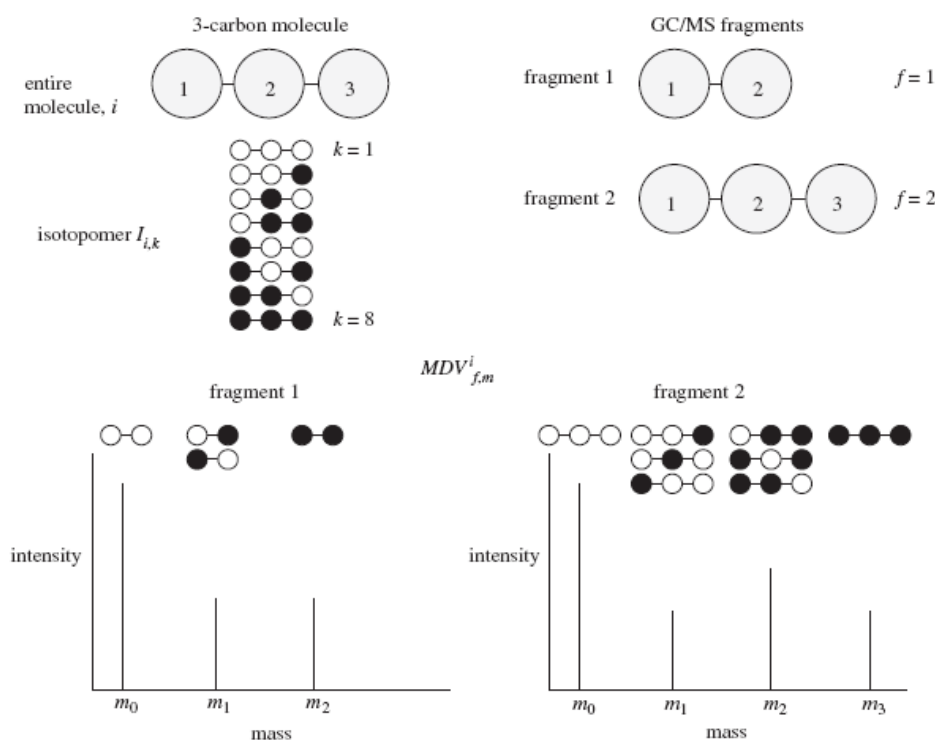


Fig. 3-5 Pictorial representation of the variable I_{ik} and the parameter MDV for a sample molecule. For a molecule, i , that contains three carbons there are 23, i.e., 8, different labeling patterns that make up the k members of the isotope fraction I_{ik} . The relative fraction of each is contained in the I_{ik} . During GC–MS, the molecule is derivitized and can be fragmented into different-sized species, f , that are then analyzed. The mass distribution vector, MDV, contains the information about the relative fractions of each fragment that contains m labeled carbons. Because only the total mass of each instance of a fragment is determined, the mass distribution vector is made by collecting various isotopomers by mass, as shown for two different fragments.

Mathematical analysis of flux elucidation

The problem of calculating the fluxes and labeling patterns for a given set of GC–MS data is formulated as a nonlinear optimization problem (FluxCalc). This problem minimizes the sum of the variance-weighted differences between the experimental data and the calculated values for the MDVs of the measured amino acid fragments by solving isotope balances (see Fig. 6). (FluxCalc) is solved multiple times using CONOPT version 3 accessed within the GAMS modeling environment. The uniformly labeled glucose was modeled to be 99% isotopically pure, with the ^{12}C impurities being equally distributed among each of the carbons. Only isotope forms having five labeled carbons were considered for the impurities; each of the six possibilities was equally likely. The initial flux distributions are provided by solving (FluxInit) that generates a set of random feasible flux distributions; using CPLEX version 10, also accessed within GAMS. Given to the non-convex nature of (FluxCalc), the size of the resulting formulation and the experimental variability in the experimental MDV, we pursue the identification of as many local optima solutions as possible.

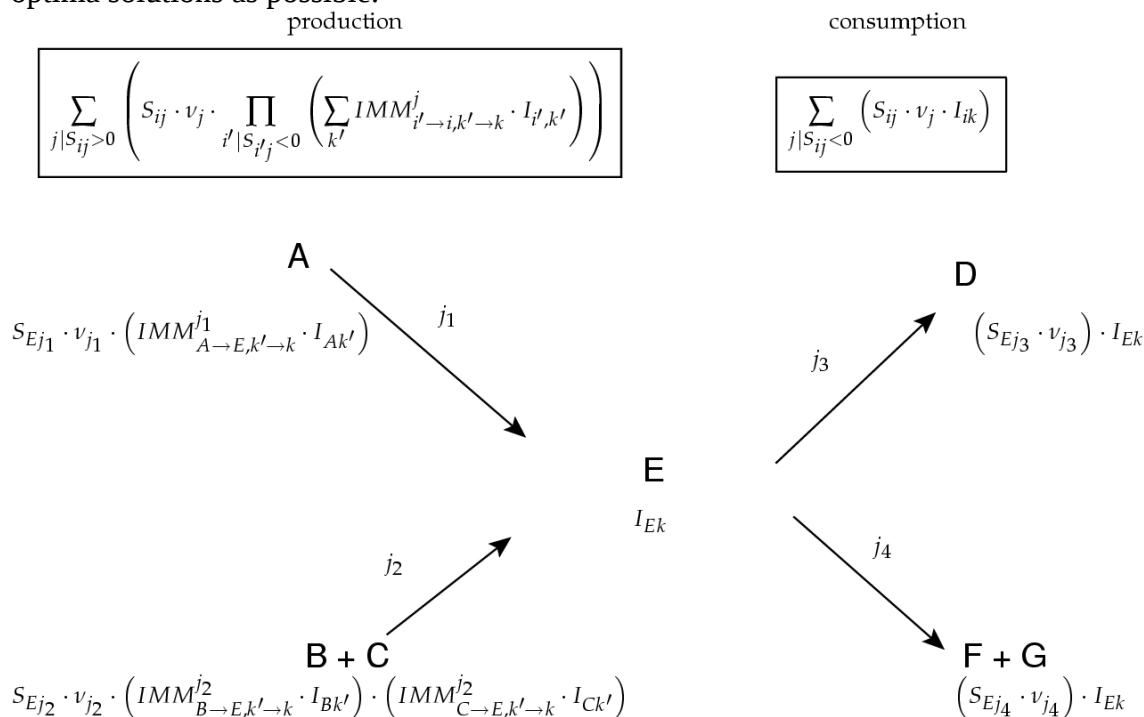


Fig. 3-6 The isotope labeling of species is governed by an isotopomer distribution vector balance and the isotope mapping matrix. Two reactions, j_1 and j_2 , form the species E; the first is unimolecular and the second is bi-molecular. E is also consumed by two reactions j_3 and j_4 ; again, the first is unimolecular and the second is bi-molecular. The isotope labeling of E is given by I_{Ek} . Each carbon-containing compound has an associated isotope fraction for each labeling pattern, I_{ik} . The isotope labeling balance of the system is at the top, with the first term of the right-hand side of Eq. (3) on the left and the second term on the right. The product symbol $[]$ is used for term by term multiplication. The isotope labeling, I_{ik} , and the isotope mapping matrix, $IMM_{i' \rightarrow i, k' \rightarrow k}^j$, that describes the contributions from the k' isotopes forms of the reactant i' on the k isotopes forms of product E are shown for the two production reactions. Note that the unimolecular reaction j_1 creates bilinear terms, and the bi-molecular reaction j_2 creates trilinear terms, as indicated. For the consumption terms, on the right, only the overall labeling patterns of E are considered.

Results

Three macroscopic measurements from the chemostat experiment were used to constrain the system. These measurements are comprised the initial glucose concentration minus the residual glucose concentration (6.1 ± 2 g/L), the cell density (1.6 ± 2 g/L), and the final amorphadiene concentration (0.077 ± 0.005 g/L) all of which were taken in replicate. The dilution rate was held constant at $0.19 \text{ h}^{-1} \pm 0.00 \text{ h}^{-1}$. The glucose entering the system was assumed to be converted to biomass, amorphadiene, CO_2 , or acetate. Although the fluxes towards biomass and amorphadiene were fixed based on the experimental measurements, the employed metabolic model was free to appropriately partition the fluxes towards CO_2 and acetate while attempting to match the observed labeling patterns of the amino acids. The measured acetate flux (10.2–12.8 mM/h) was not used as an explicit constraint on the system, but rather was left as a final consistency check. All computational results are reported using a basis glucose uptake rate of 10mM/h. Using GC–MS, we observed the mass distributions (MDV) for fragments of 13 amino acids collected from two samples taken at different time points. Each derivatized sample was injected four times and the standard deviations for each sample and an overall standard deviation were calculated. In general, the overall standard deviations of the mass distributions among replicate measurements were less than the instrument error of 0.4%, which is consistent with

Table 1

Fragments of the amino acids experimentally measured by GC–MS and simulated using the IMM models

Amino acid ^a	Monitored ions	Amino acid carbon atoms ^b	Fragment
Val-159	186–192	1-2-4-5	M-C ₇ H ₁₅ O ₂ Si
Val-85	260–266	1-2-4-5	M-C ₅ H ₉ O
Val-57	288–295	1-2-3-4-5	M-C ₄ H ₉
Gly-85	218–221	2	M-C ₅ H ₉ O
Gly-57	246–249	1-2	M-C ₄ H ₉
Ala-85	232–235	1-3	M-C ₅ H ₉ O
Ala-57	260–264	1-2-3	M-C ₄ H ₉
Glu-159 R	330–335	1-2-4-5	M-C ₇ H ₁₅ O ₂ Si
Glu-85 R	404–409	1-2-4-5	M-C ₅ H ₉ O
Glu-57 R	432–438	1-2-3-4-5	M-C ₄ H ₉
Asp-159 R	316–320	1-2-4	M-C ₇ H ₁₅ O ₂ Si
Asp-57 R	418–423	1-2-3-4	M-C ₄ H ₉
Asp-15 R	460–465	1-2-3-4	M-CH ₃
Met-159	218–223	1-3-4-5	M-C ₇ H ₁₅ O ₂ Si
Met-85	292–297	1-3-4-5	M-C ₅ H ₉ O
Met-57	320–325	1-2-3-4-5	M-C ₄ H ₉
Leu-159	200–205	1-2-4-5-6	M-C ₇ H ₁₅ O ₂ Si
Leu-85	274–280	1-2-4-5-6	M-C ₅ H ₉ O
Ile-159	200–205	1-2-4-5-6	M-C ₇ H ₁₅ O ₂ Si
Ile-85	274–280	1-2-4-5-6	M-C ₅ H ₉ O
Pro-159	184–189	1-3-4-5	M-C ₇ H ₁₅ O ₂ Si
Ser-159 R	288–291	1-3	M-C ₇ H ₁₅ O ₂ Si
Ser-85 R	362–365	1-3	M-C ₅ H ₉ O
Ser-57 R	390–394	1-2-3	M-C ₄ H ₉
Thr-85 R	376–379	1-2-4	M-C ₅ H ₉ O
Thr-57	290–294	1-2-3-4	M-C ₄ H ₉
Thr-57 R	404–408	1-2-3-4	M-C ₄ H ₉
Phe-159	234–243	1-2-3-4-5-6-7-9	M-C ₇ H ₁₅ O ₂ Si
Phe-85	308–317	1-2-3-4-5-6-7-9	M-C ₅ H ₉ O
Phe-57	336–345	1-2-3-4-5-6-7-8-9	M-C ₄ H ₉
Lys-159 R	329–335	1-2-4-5-6	M-C ₇ H ₁₅ O ₂ Si

^aThe number after the amino acid is the size of the fragment removed from the molecule with both N and C groups derivatized. An R following a fragment denotes that the R-group is also derivatized.

^bThe carbon numbering follows that in the .mol file from the KEGG database¹⁵. For those fragments that have a removed C, it corresponds to that of the carboxyl.

previous estimates of instrument error¹⁶. Next, the measurement error was estimated as the quadrature addition of the instrument error and the standard deviation¹⁷. After correcting for naturally occurring ¹³C, we validated the internal consistency of the MDV distributions of the fragments for each amino acid.

The 32 fragments used for MFA are listed in Table 1. These fragments contain 162 mass entries giving $162 - 32 = 130$ independent mass measurements for use in the MFA. The mean values of the experimental mass measurements and corresponding estimates of the measurement error are given in Table 2. The best computationally derived MDV values (weighted residual sum of squares = 56.4) are also given for each fragment. Although the un-weighted differences are quite small, we performed a goodness-of-fit analysis. The isotope model contains 130 free fluxes. Using direct experimental flux measurements, three internal fluxes (glucose uptake, amorphadiene production, and biomass production) and 22 exchange fluxes were set, giving a total of 105 independent flux variables for (FluxCalc) to determine. The flux ranges for the entire network are available as supplementary material.

However, as expected, nearly all central metabolic fluxes span wide ranges of values motivating the need for ¹³C-based MFA. Next the isotopomer balancing constraints were applied to determine (i) which fluxes (if any) were fully resolved; (ii) how many of them were only partially resolved; and (iii) if there were any remaining fluxes in the model that are completely unaffected by the isotopomer data. The calculated flux ranges are included as supplementary material.

Table 2
Experimentally measured and simulated mass distributions (mol %) of amino acid fragments

Fragment		M + 0	M + 1	M + 2	M + 3	M + 4	M + 5	M + 6	M + 7	M + 8	M + 9
Val-159	Meas ^a	61.9 ± 0.6	5.3 ± 0.6	27.6 ± 0.5	2.1 ± 0.4	3.1 ± 0.4					
	Sim ^b	62.2	5.6	27.9	1.2	3.1					
Val-85	Meas	61.7 ± 0.5	6.0 ± 0.6	27.8 ± 0.5	1.4 ± 0.4	3.1 ± 0.4					
	Sim	62.2	5.6	27.9	1.2	3.1					
Val-57	Meas	59.6 ± 0.6	7.5 ± 0.5	16.4 ± 0.7	12.9 ± 0.6	1.1 ± 0.4	2.6 ± 0.4				
	Sim	59.5	7.5	16.1	13.1	1.1	2.7				
Gly-85	Meas	80.9 ± 0.5	19.2 ± 0.5								
	Sim	80.6	19.4								
Gly-57	Meas	77.8 ± 0.5	5.6 ± 0.5	16.7 ± 0.4							
	Sim	77.9	5.6	16.5							
Ala-85	Meas	78.7 ± 0.5	3.6 ± 0.6	17.8 ± 0.5							
	Sim	78.9	3.5	17.6							
Ala-57	Meas	75.0 ± 0.8	6.2 ± 0.8	3.7 ± 0.8	15.1 ± 0.7						
	Sim	75.5	6.1	3.4	15.1						
Glu-159 R	Meas	58.1 ± 0.7	13.5 ± 1.1	23.3 ± 0.7	2.9 ± 0.5	2.2 ± 0.4					
	Sim	58.2	13.4	23.4	2.8	2.2					
Glu-85 R	Meas	58.1 ± 0.8	13.6 ± 1.2	23.4 ± 0.7	2.7 ± 0.5	2.1 ± 0.4					
	Sim	58.2	13.4	23.4	2.8	2.2					
Glu-57 R	Meas	51.9 ± 0.7	15.3 ± 0.8	21.4 ± 0.4	8.1 ± 0.4	2.3 ± 0.4	1.1 ± 0.4				
	Sim	52.1	15.1	21.5	7.9	2.3	1.0				
Asp-159 R	Meas	65.7 ± 0.6	16.1 ± 0.7	11.9 ± 0.4	6.2 ± 0.5						
	Sim	66.1	16.2	11.8	5.9						
Asp-57 R	Meas	60.7 ± 0.5	16.5 ± 0.5	10.6 ± 0.6	9.6 ± 0.5	2.7 ± 0.4					
	Sim	60.9	16.5	10.4	9.7	2.5					
Asp-15 R	Meas	60.8 ± 0.6	16.3 ± 0.5	10.6 ± 0.6	9.6 ± 0.5	2.8 ± 0.5					
	Sim	60.9	16.5	10.4	9.7	2.5					
Met-159	Meas	53.6 ± 0.5	25.8 ± 0.5	12.7 ± 0.5	6.9 ± 0.5	1.1 ± 0.4					
	Sim	53.3	25.9	12.6	7.0	1.1					
Met-85	Meas	53.2 ± 1.0	25.8 ± 0.6	12.8 ± 0.5	7.0 ± 0.6	1.2 ± 0.7					
	Sim	53.3	25.9	12.6	7.0	1.1					
Met-57	Meas	49.2 ± 0.6	24.9 ± 0.6	11.6 ± 0.5	9.8 ± 0.5	4.0 ± 0.5	0.5 ± 0.4				
	Sim	49.1	25.1	11.6	9.8	3.9	0.5				
Leu-159	Meas	50.5 ± 0.5	15.9 ± 0.6	23.9 ± 0.5	6.3 ± 0.4	2.9 ± 0.4	0.6 ± 0.4				
	Sim	50.1	16.6	23.6	6.4	2.7	0.6				
Leu-85	Meas	50.3 ± 0.5	16.4 ± 0.5	23.7 ± 0.5	6.3 ± 0.4	2.8 ± 0.4	0.6 ± 0.4				
	Sim	50.1	16.6	23.6	6.4	2.7	0.6				
Ile-159	Meas	52.0 ± 0.5	15.1 ± 0.7	21.7 ± 0.4	7.9 ± 0.5	2.3 ± 0.4	1.0 ± 0.4				
	Sim	52.1	15.3	21.5	7.9	2.3	1.0				
Ile-85	Meas	51.5 ± 0.5	15.5 ± 0.7	21.7 ± 0.4	8.0 ± 0.4	2.3 ± 0.4	1.0 ± 0.4				
	Sim	52.1	15.3	21.5	7.9	2.3	1.0				
Pro-159	Meas	58.8 ± 0.7	13.3 ± 1.1	23.1 ± 0.7	2.6 ± 0.5	2.1 ± 0.5					
	Sim	58.2	13.4	23.4	2.8	2.2					
Ser-159 R	Meas	77.5 ± 0.7	6.2 ± 0.8	16.3 ± 0.7							
	Sim	78.3	4.7	17.0							
Ser-85 R	Meas	77.4 ± 0.8	6.2 ± 0.9	16.4 ± 0.7							
	Sim	78.3	4.7	17.0							
Ser-57 R	Meas	75.0 ± 0.5	6.1 ± 0.5	3.7 ± 0.5	15.2 ± 0.5						
	Sim	75.7	5.7	3.4	15.2						
Thr-85 R	Meas	66.0 ± 0.5	16.4 ± 0.7	11.7 ± 0.5	5.9 ± 0.5						
	Sim	66.0	16.4	11.8	5.8						
Thr-57	Meas	61.0 ± 0.5	16.5 ± 0.6	10.3 ± 0.5	9.5 ± 0.5	2.7 ± 0.4					
	Sim	60.8	16.6	10.5	9.6	2.5					
Thr-57 R	Meas	60.6 ± 0.5	16.7 ± 0.5	10.5 ± 0.6	9.6 ± 0.5	2.6 ± 0.4					
	Sim	60.8	16.6	10.5	9.6	2.5					
Phe-159	Meas	43.6 ± 0.7	10.7 ± 0.7	20.9 ± 0.7	9.5 ± 0.7	8.8 ± 0.7	3.1 ± 0.7	2.7 ± 0.7	0.4 ± 0.7	0.3 ± 0.7	
	Sim	43.4	11.3	19.9	9.3	9.1	3.4	2.9	0.4	0.3	
Phe-85	Meas	43.4 ± 0.7	10.8 ± 0.7	21.0 ± 0.7	9.4 ± 0.7	8.7 ± 0.7	3.1 ± 0.7	2.6 ± 0.7	0.3 ± 0.7	0.6 ± 0.7	
	Sim	43.4	11.3	19.9	9.3	9.1	3.4	2.9	0.4	0.3	
Phe-57	Meas	42.4 ± 0.5	11.3 ± 0.4	12.3 ± 0.4	16.7 ± 0.4	8.5 ± 0.4	4.1 ± 0.4	2.9 ± 0.4	1.3 ± 0.4	0.3 ± 0.4	0.3 ± 0.4
	Sim	42.0	11.6	12.1	16.1	9.0	4.2	3.0	1.4	0.4	0.3
Lys-159 R	Meas	51.9 ± 0.8	15.5 ± 1.1	21.6 ± 0.7	7.9 ± 0.7	2.2 ± 0.7	1.1 ± 0.7				
	Sim	50.1	17.8	21.8	7.2	2.3	0.7				

^aMeas are the values measured experimentally, reported as mean ± experimental error. The mass distributions are corrected for naturally occurring isotopes.

^bSim are the values simulated by the isotope model, reported for the lowest objective (56.4).

A

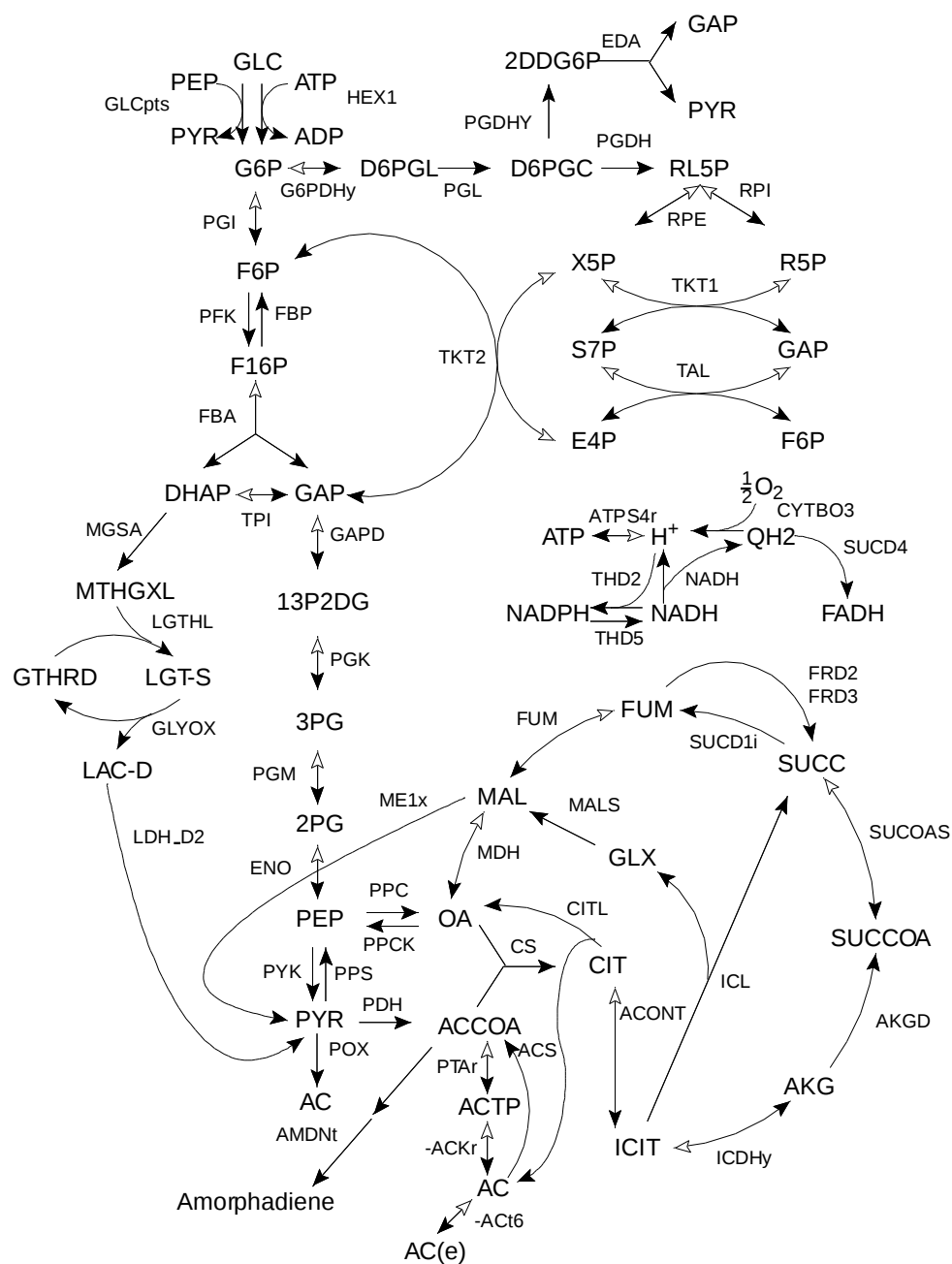
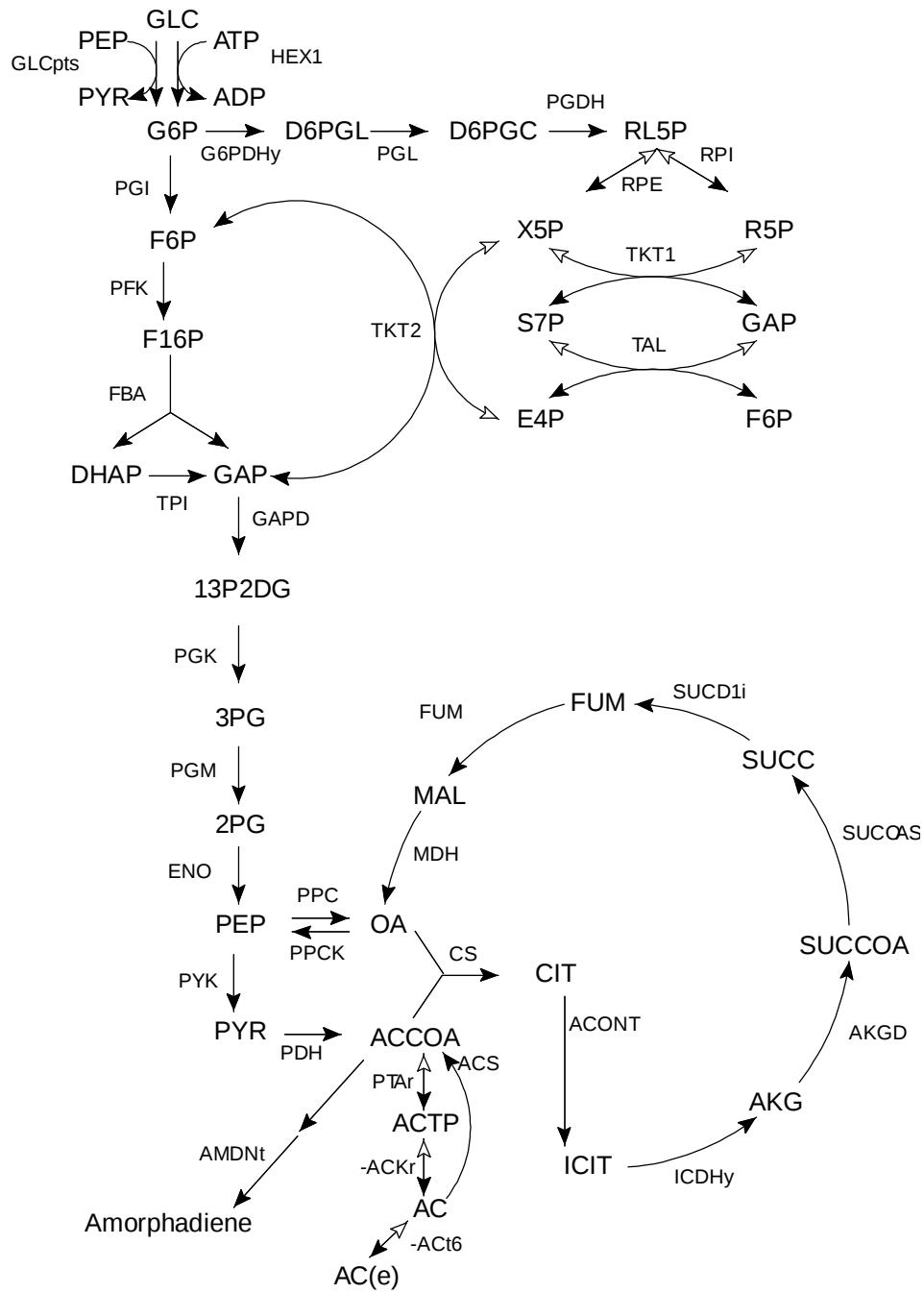


Fig. 3-8. Central metabolism flux maps. The flux map of central metabolism for the full model(panel A) shows the metabolite names (sans-serif typeface) and reaction names (smaller, serif typeface). Irreversible reactions have single, filled, arrows. Reversible reactions have two arrows: the forward direction is filled and the backward direction is unfilled. In comparison, most of the reactions are irreversible for the smaller basic model (panel B).

B



Discussion and Future Direction

Innovative approaches are needed to utilize the information generated from genome research in an integrated fashion to analyze, interpret, and predict the function of biological systems. This chapter addresses these needs with novel engineering approaches for studying the systemic capabilities of metabolism in completely sequenced bacterial genomes. We have focused on the development of tools for the analysis of experimental data and the integration of such data into the modeling framework.

This has a direct impact in metabolic engineering of bacteria for the production of drugs and pharmaceutical intermediates. Specifically we studied an amorphadiene-producing strain of *Escherichia coli*. By comparing this strain with a non-producing control strain we hoped to better understand what an impact over-expression of a non-native pathway has on *E.coli*'s central metabolic pathways. Measurements of fluxes to biomass and extracellular metabolites during continuous culture showed no significant variation in the amorphadiene strain relative to the control. Even though high energy demand of the amorphadiene pathway led us to expect some flux differences between the two strains but the production level of amorphadiene specifically at exponential phase was not high enough to have a significant impact on cellular metabolism. While this work is expected to serve as a valuable platform for studying engineered high-level-production strains, there are some limitations that need to be addressed. Complementary experiments revealed that only a very small amount of amorphadiene is produced at exponential phase. In continuous culture fermentation the cells are forced to remain in exponential phase and not enter stationary phase, however the majority of amorphadiene production occurs when the cells are at stationary phase. Therefore determining the changes to *E. coli*'s metabolism wrought by over-expressing of amorphadiene pathway and production of amorphadiene cannot be achieved through flux balance analysis at exponential phase. Also the difference in amorphadiene production among different engineered *E. coli* strains producing varying levels of amorphadiene cannot be detected until later in growth when the cells enter stationary phase. This would limit the use of this model in further engineering *E. coli* strains for increased amorphadiene production and determining the restrictions in directing the flux towards production of amorphadiene. However, this model holds great promise in designing *E. coli* for productions of many valuable secondary metabolites and prediction of mutations and gene knock-outs that would redirect the flux toward our desired metabolite.

Keeping in mind that microbial production of many commercially important secondary metabolites occurs during stationary phase, methods to measure metabolic flux during this growth phase are of great value^{18, 19}. To perform such flux analysis, one cannot simply use the proteinogenic amino acids from cells harvested during the stationary phase because their labeling patterns mainly reflect the metabolism when the amino acid was synthesized, the exponential phase. In addition, the main assumption behind most flux balance analysis models is that the cells are at steady state. While there have been attempts to study stationary phase metabolism via isotopomer analysis (Reference), it is not yet as standard and reliable as steady state metabolic flux analysis but methods to measure metabolic flux during this growth phase would be valuable and revolutionize the field of metabolic engineering.

With this in mind, the major objective of this chapter remains to utilize the power of experimental data in combination with predictive models of cell metabolism to guide strain engineering for applications in both basic and applied microbial research. There are direct commercial applications to modeling metabolism including metabolic engineering efforts in

bioprocessing, bioremediation, and protein production, as well as in such diverse areas as human therapeutics and antimicrobial drug development.

References

- 1- Schmidt, K., Carlsen, M, Nielsen, J, and Villadsen, J. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnology and Bioengineering* 55: 831-840 (1997).
- 2- Martin, V.J., et al., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796–802 (2003)
- 3- Neidhardt, F.C., et al. Culture medium for enterobacteria. *J. Bacteriol.* 119, 736–747 (1974).
- 4- Siddiquee, K.A., et al. Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on ¹³C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Appl. Microbiol. Biotechnol.* 63, 407–417 (2004).
- 5- Dauner, M. and U. Sauer. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnol Prog* 16(4): 642-9 (2000).
- 6- Wahl, S.A., et al. New tools for mass isotopomer data evaluation in (¹³)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnol. Bioeng.* 85, 259–268 (2004).
- 7- Reed, J.L., et al. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4, R54 (2003).
- 8- Fong, S.S., Palsson, B.O. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.* 36, 1056–1058 (2004).
- 9- Ibarra, R.U., et al. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420, 186–189 (2002).
- 10- Burgard, A.P., et al. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 14, 301–312 (2004).
- 11- Edwards, J.S., Palsson, B.O., The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* 97, 5528–5533 (2000).
- 12- Zupke, C., Stephanopoulos, G., Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnol. Prog.* 10, 489–498 (1994).
- 13- Van Dien, S.J., et al. Quantification of central metabolic fluxes in the facultative methylotroph *Methylobacterium extorquens* AM1 using ¹³C-label tracing and mass spectrometry. *Biotechnol. Bioeng.* 84, 45–55 (2003).
- 14- Wittmann, C., Heinzle, E. Mass spectrometry for metabolic flux analysis. *Biotechnol. Bioeng.* 62, 739–750 (1999).
- 15- Kanehisa, M., et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354–D357 (2006).
- 16- Wittmann, C., et al. In vivo analysis of intracellular amino acid labelings by GC/MS. *Anal. Biochem.* 307, 379–382 (2002).
- 17- Taylor, J.C. Statistical Techniques for Data Analysis. Lewis Publishers, Chelsea, MI (1990).
- 18- Antoniewicz M.R., Kraynie D.F., Laffend L.A., González-Lergier J., Kelleher J.K., Stephanopoulos G. Metabolic flux analysis in a nonstationary system: fed-batch

- fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metab. Eng.* 9, 277-292 (2007).
- 19- Shaikh A.S., Tang Y.J., Mukhopadhyay A., Martín H.G., Gin J., Benke P.I., Keasling J.D. Study of stationary phase metabolism via isotopomer analysis of amino acids from an isolated protein. *Biotechnol. Prog.* (2009)

Chapter 4: Optimization of the Mevalonate-based Isoprenoid Biosynthetic Pathway in *Escherichia coli* for Production of the Anti-malarial Drug Precursor Amorpha-4, 11-diene

Introduction

In recent years, microorganisms have been engineered to produce compounds of commercial and pharmacological importance. Many of these compounds are either too difficult to synthesize chemically or are produced in limited quantities by their native hosts. Isoprenoids are an example of a class of natural compounds that are of particular interest for microbial production. A number of flavors, fragrances, and medicines are isoprenoids, and most of these compounds are produced in low quantities by their native hosts. In an attempt to produce large quantities of amorpha-4,11-diene (amorphadiene), a precursor to the anti-malarial compound artemisinin, a multi-gene biosynthetic pathway (Fig. 4-1A) was engineered into *Escherichia coli*^{1,2,3}. While the engineered *E. coli* produced relatively high titers of amorphadiene, imbalances in the pathway prevented further improvements in production.

Although engineering microorganisms to produce natural compounds has shown great promise, optimization of these heterologous pathways and host strains is required to obtain maximal titers. Imbalances in gene expression can lead to over- or under-production of enzymes in the pathway, accumulation of toxic metabolic intermediates, and metabolic burden on the host, all of which result in suboptimal product titers³⁻⁷. Previously, codon optimization, appropriate promoter and ribosome binding site choice, and directed evolution of enzymes have been used to improve production in non-native host strains^{1,3,6,8}.

This work describes a complimentary method for pathway optimization that focuses on balancing gene expression, identifying rate-limiting enzymes, and decreasing the metabolic burden of the pathway on the host.

We sought to improve production of amorphadiene in a previously engineered strain of *E. coli*¹. This strain contains plasmid-borne copies of the mevalonate pathway from *Saccharomyces cerevisiae*, as well as the amorphadiene synthase (ADS) gene from *Artemisia annua*. The pathway genes were originally constructed by assembly of the operons onto three different plasmids, each with a different copy number and *E. coli* promoter (Fig. 1A). While this *E. coli* strain proved successful in producing amorphadiene in a two-phase fermentation (~25 mg/L/OD600 over 72 h), it did not achieve the desired titers necessary for providing inexpensive artemisinin to people in the Developing World^{2,9}. To further optimize expression from this strain, we developed a plasmid system where combinations of various promoters and operon constructs could be easily substituted and tested. Using this system, we identified two rate-limiting enzymes in the pathway and improved production seven-fold over that from the original strain. The strategy presented in this chapter of a cloning system with compatible and unique restriction sites for the rapid testing of multiple promoters and operons could be applied readily to any system to identify problems in gene expression and improve production from a heterologous pathway.

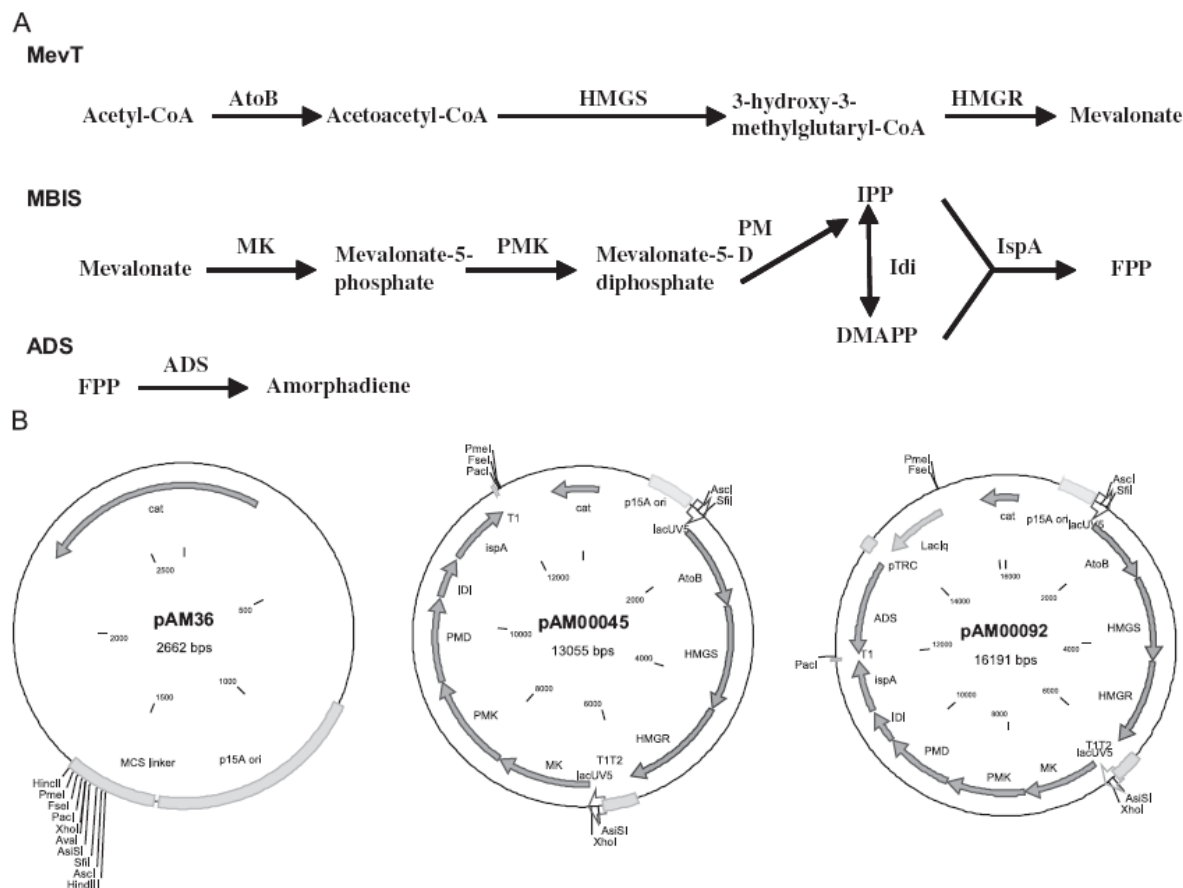


Fig. 4-1. Mevalonate pathway and vector constructs used in this study. (A) Production of amorphaadiene via the mevalonate pathway. The abbreviations for enzymes and pathway intermediates are as follows: AtoB, acetoacetyl-CoA thiolase; HMGS, HMG-CoA synthase; tHMGR, truncated HMG-CoA reductase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; PMD, mevalonate pyrophosphate decarboxylase; Idi, IPP isomerase; IspA, FPP synthase; ADS, amorphaadiene synthase; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate. (B) Plasmid maps for vectors with multiple cloning site only (pJA4), MevT and MBIS operons (pAM45), or MevT, MBIS, and ADS operons (pAM92). Restriction sites used in cloning are shown, open arrows indicate P_{LacUV5} promoters, and shaded boxes indicate the p15A origin and rmBT1T2 terminators. The following abbreviations are used: cat, chloramphenicol resistance and LacI^q , lac repressor.

Materials and Method

Strains and plasmids

E. coli DH10B was used for cloning. *E. coli* DH1 was used for expression and production studies with the amorphaadiene modular vectors. The MevT pathway (operon) contains the following

enzymes (genes): acetoacetyl-CoA synthase (*ackA*), HMG-CoA synthases (*HMG*), and an N-terminally truncated version of HMG-CoA reductase (*tHMG*R). The MBIS pathway (operon) contains the following enzymes (genes): mevalonate kinase (*MK*), phosphomevalonate kinase (*PMK*), phosphomevalonate decarboxylase (*PMD*), isopentenyl diphosphate isomerase (*idi*), and farnesyl diphosphate synthase (*ispA*). Plasmids used in this study are listed in Table 3. Media were supplemented with 100 µg/ml carbenicillin 35 µg/ml chloramphenicol, 50 µg/ml kanamycin, or 10 µg/ml tetracycline to select for plasmid maintenance. All strains were grown at 30 °C.

Table 3: Summary of strains and plasmids used in this study

	Description	Reference
Strain		
DH1	F- supE44 hsdR17 recA1 gyrA96 relA1 endA1 thi-1 λ-	Hanahan (1983)
DH10B	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu) 7697 galU galK λ- rpsL nupG	Invitrogen
Plasmids		
pMevT	pLac33 derivative containing MevT operon; Cm ^R	Martin et al. (2003)
pMBIS	pBBR1MCS-3 derivative containing MBIS operon; Tc ^R	Martin et al. (2003)
pADS	pTrc99A derivative containing ADS; Ap ^R	Martin et al. (2003)
pBlueScript IIS	Cloning vector; Ap ^R	Stratagene
pBlueScript-MCS	pBlueScript IIS derivative containing engineered MCS; Ap ^R	This study
pACYC184	Low copy number cloning vector; Cm ^R	NEB
pJA4	pACYC184 derivative containing engineered MCS; Cm ^R	This study
pAM36-MevT	pJA4 derivative containing codon-optimized MevT operon; Cm ^R	This study
pAM43	pAM36-MevT derivative containing MBIS operon; Cm ^R	This study
pAM45	pAM43 derivative containing lacUV5 promoters 50 of the MevT and MBIS operon; Cm ^R	This study
pAM45-MevT	pAM45 derivative containing MevT operon; Cm ^R	This study
pAM92	pAM45 derivative containing lacI ^q -P _{trc} -ADS; Cm ^R	This study
pAM94	pADS derivative containing codon-optimized MK	This study
pAM29	Cloning vector; p15A origin; lacUV5 promoter; Kn ^R	This study
pAM29-MK	pAM29 derivative containing MK	This study
pAM29-PMK	pAM29 derivative containing PMK	This study
pAM29-PMD	pAM29 derivative containing PMD	This study
pAM29-idi	pAM29 derivative containing idi	This study
pAM29-ispA	pAM29 derivative containing ispA	This study

Growth Condition

For time-course experiments, overnight cultures were started from fresh transformations into 5 ml of Luria Broth (Difco) with antibiotics. Overnight cultures were diluted to an OD₆₀₀ (optical density at a wavelength of 600 nm) of 0.05 in 40 ml of Terrific Broth (Difco) with 1% glycerol and overlaid with 20% (v/v) dodecane (Sigma-Aldrich). Experiments were initiated when cultures reached an OD₆₀₀ of 0.2–0.3 (~2–3 h after subculture) by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and for the MBIS gene dosage experiments, 20 mM mevalonate (Sigma-Aldrich). Cultures were incubated for 68–75 h, and at each time point the OD₆₀₀ was read and samples from the dodecane phase were taken to determine amorphadiene production. Previous studies have shown that over 97% of the amorphadiene produced partitions into the organic dodecane overlay of cultures².

Gas chromatograph-mass selective (GC-MS) analysis of amorphadiene

Samples (100 μ l) were removed from the dodecane layer of each culture, and centrifuged at 13,300g for 3 min at room temperature. After centrifugation, 10 μ l of the organic phase was added to 990 μ l of ethyl acetate (Sigma-Aldrich). To measure amorphadiene levels, 1 μ l of the sample was injected into an Agilent 6890 series gas chromatograph (GC) equipped with an Agilent 5973 mass selective (MS) detector and a cyclosyl-B (chiral) capillary column (30 m $\times$ 250 μ m $\times$ 0.25 μ m thickness, Agilent). The GC oven temperature program was as follows: 100 $^{\circ}$ C for 45 s, a ramp of 35 $^{\circ}$ C/min to 200 $^{\circ}$ C, a ramp of 100 $^{\circ}$ C/min to 250 $^{\circ}$ C, and a hold at 250 $^{\circ}$ C for 1 min. The MS was operated in selected ion monitoring (SIM) mode scanning for the molecular ions at 189 and 204 m/z. Purified amorphadiene was used to generate standard curves for obtaining production titers.

Construction of a vector for rapid cloning of promoters and amorphadiene operons

An insert containing restriction sites necessary for cloning the amorphadiene pathway was constructed in two pieces by annealing and extending two pairs of oligonucleotides (pACYCfor1, 5'-CAGGAAAGCTTCCGATGGCGCGCCGTAAAGGCCATCCTGGCCTACGCGCGATCGCG GTAGCTCGAGAACGGGTTGACATACG-3' with

pACYCrev1, 5'-CGTATGTCAACCCGTTCTCGAG-3' and
pACYCfor2, 5'-GGTAGCTCGAGAACGGTTAATTAAGTCCAGGCCGGCCTACGCGTTTAACTTCCGGT TGACATACG-3' with pACYCrev2, 5'-CGTATGTCAACCGGAAGTTTAAAG-3').

Each pair of oligonucleotides (150 pmol) was incubated at 95 $^{\circ}$ C for 5 min in TE buffer with 50mM NaCl and 10mM MgCl₂. Reactions were cooled to 25 $^{\circ}$ C over 20min. Both products were elongated by adding 10mM dNTP mix and Klenow DNAP in Buffer 3 (New England Biolabs) and incubating the reactions at 25 $^{\circ}$ C for 20min. Fragments were purified via a phenol/chloroform/isoamyl alcohol extraction followed by an isopropanol precipitation. The final products were resuspended in 20 ml TE and digested with XhoI. Fragments were ligated to form the multiple cloning site (MCS) insert. Purified ligations were cloned into pBluescript IKS (digested with SmaI, Stratagene) creating pBluescript-MCS. Clones containing the insert were identified via a blue-white screen, and plasmid was purified using Qiagen's Mini-prep kit. The backbone modular vector (pJA4) was constructed by digesting pBluescript-MCS with HincII and HindIII and ligating this insert into a 2.6-kb HincII/HindII pACYC184 backbone (Fig. 4-1B).

Construction of amorphadiene production strains

To increase translation and expression levels in *E. coli*, a codon-optimized MevT operon (MevT66) was synthesized using the *E. coli* codon usage tables. The codon-optimized operon was PCR-amplified with SfiI and AsiSI sites and subcloned into the pJA4 vector, resulting in plasmid pAM36-MevT66. The MBIS operon was PCR-amplified from pMBIS and subcloned into the XhoI and PacI sites of pAM36-MevT66, creating pAM43. LacUV5 promoters were

synthesized from oligonucleotides and subcloned into the AscI/SfiI and AsiSI/XhoI sites in pAM43, resulting in pAM45 (Fig. 1B). In addition, a non-codon-optimized MevT operon was amplified from pMevT and placed into the SfiI and AsiSI sites of pAM45 to create pAM45-MevT. In order to combine the entire amorphaadiene pathway into one vector, the lacIq-P_{trc}-ADS region was PCR-amplified from pADS with PacI and FseI restriction sites, and subcloned into pAM45 to create pAM92 (Fig. 4-1B).

Construction of plasmids expressing members of the MBIS operon

To conduct gene dosage experiments, a series of expression vectors containing individual MBIS genes were constructed using the low-copy pAM29 plasmid as a backbone. Plasmid pAM29 was created by assembly of the p15A origin of replication and the kanamycin resistance gene from pZS24-MCS1¹⁰ with an oligonucleotide-generated lacUV5 promoter. MK, PMK, PMD, idi, and ispA were PCR-amplified from pMBIS using primers that added 5' EcoRI 3' SalI restriction sites. The amplified genes were subcloned individually into pAM29 to create the single-gene expression vectors. Plasmid pMBISopt was created synthetically using DNA synthesis (Codon Devices, Cambridge, MA). It comprises a nucleotide sequence encoding the *S. cerevisiae* mevalonate kinase (MK) codon-optimized for expression in *E. coli* (Supplementary Table 4-1). The nucleotide sequence encoding mevalonate kinase (MK) was PCR-amplified from pMBISopt with BamHI and HindIII restriction sites, and subcloned into pTrc-ADS to create pAM94.

Plasmid stability assays

Overnight cultures of DHI pAM92 or DH1 pAM92 pADS were inoculated into 250ml baffled flasks containing 40ml of TB+1% glycerol with appropriate antibiotics at a OD₆₀₀ of 0.1. Cultures were incubated with shaking at 220 rpm for a total of 96 h. At 24-, 48-, 72-, and 96-h time points, culture samples were removed and adjusted to OD₆₀₀ of 1.0 with fresh LB medium. Serial dilutions of each culture were plated onto LB and LB with chloramphenicol (for pAM92) and incubated overnight. The number of colonies on each plate were scored and used to calculate viable counts and individual plasmid stabilities.

Results

Increasing expression of the MevT pathway

The design of metabolic pathways can be troublesome because of the large diversity of elements involved in controlling gene expression at both the transcriptional and translational levels: promoter strength, translation efficiency, codon choice, and operon structure. One of the initial modifications made to the original amorphaadiene-producing *E. coli* strain was to replace the MevT operon (pMevT¹) with the *E. coli* codon-optimized operon MevT66. In addition, the wild-type lac promoter of pMevT was also replaced with the lacUV5 promoter (pAM25), a mutated

version of the lac promoter whose basal activity is dramatically less sensitive to intracellular levels of cyclic AMP and is ~2-fold stronger than the lac promoter¹¹. These two modifications in DH1 pAM25 pMBIS pADS resulted in an approximate 1.5-fold increase in amorphaadiene production at 75 h compared with DH1 pMevT pMBIS pADS (Fig. 4-2). The improvements resulting from optimization of the MevT operon motivated us to examine the use of alternative promoters and gene dosages with the remaining segments of the mevalonate pathway.

Use of stronger promoters and codon optimization of MevT increases amorphaadiene production

To facilitate construction and analysis of mevalonate pathway improvements, we constructed a plasmid expression vector with unique and buffer-compatible restriction sites to efficiently analyze combinations of bacterial promoters and operon constructs. pAM43 served as the initial scaffold for testing potential improvements and contained restriction sites for inserting promoters upstream of the two pathway operons. Due to the previous increase in titers with the lacUV5 promoter in pAM25, two lacUV5 promoters were inserted into pAM43 to control expression of the mevalonate pathway operons (pAM45-MevT). *E. coli* DH1 harboring pAM45MevT and pADS exhibited a significant increase in amorphaadiene production (121mg/L/OD600 at 75 h), which was 3-fold higher than the original base strain, DH1 pMevT pMBIS pADS (Fig. 4-2). We also tested the codon-optimized MevT66 operon in the context of the vector (pAM45), because of the increase in titer observed previously with pAM25. DH1 pAM45 pADS produced approximately 2-fold more amorphaadiene (221mg/L/OD600) than the non-codon-optimized MevT strain, and a 5-fold increase over the original strain, DH1 pMevT pMBIS pADS. The design of pAM43 with unique and compatible restriction sites enabled the rapid sub-cloning and testing of other combinations of promoters and operons including the wild-type lac promoter for MevT expression, a constitutive promoter driving MBIS expression, and the use of an *E. coli* codon-optimized MBIS operon. However, none of these combinations resulted in a significant increase in amorphaadiene production.

Identification of ADS as a limiting enzyme

The presence of multiple plasmids within a bacterial cell can have a profound impact on cellular physiology and often imposes a metabolic burden on the cell^{12,13}. This burden can be amplified when multiple plasmids are used for high-level gene expression. It is also thought that expression of plasmid-borne, antibiotic resistance genes necessary for plasmid maintenance also account for a significant portion of the metabolic burden¹⁴. The original amorphaadiene production strain and the derivatives described above contain multiple plasmids and antibiotic resistance genes. This may have created a metabolic burden and compromised product titers. To address this, we placed ADS on pAM45 to create a single plasmid with one antibiotic resistance gene. This p15A-replicon plasmid, pAM92, was sufficient for amorphaadiene production as it contains the entire mevalonate pathway and ADS gene. Although DH1 harboring pAM92 generated more amorphaadiene than the original strain, DH1 pMevT pMBIS pADS, it produced significantly less

amorphadiene than the parent strain, DH1 pAM45 pADS (Fig. 4-2). Because the mevalonate pathway is identical on both pAM45 and pAM92, we hypothesized that ADS levels may limit production. Placing ADS on pAM45 lowered the ADS copy number from 20–30 copies/cell on pADS (pMB1 origin) to 10–15 copies/cell on pAM92 (p15A origin). Previous studies had also shown that amorphadiene production and enzyme activity dropped to 12% when ADS was present on a pBeloBac-11 vector (1 copy/cell) in E.coli (C. Paddon, personal communication). In addition, a decrease in cell viability was seen in these strains, presumably due to the accumulation of the intermediate FPP (data not shown). To increase ADS expression, we transformed DH1 pAM92 with pADS, resulting in a combined ADS gene dosage of 30–45 copies/cell. This strain exhibited an increase in amorphadiene production over DH1 pAM92, but still did not approach the peak titers obtained with DH1 pAM45 pADS (Fig. 4-2). Experiments to measure pAM92 stability indicated that when this plasmid was expressed alone in DH1 (DH1 pAM92) it was stable (<10% loss) over a 96 h time course. However, pAM92 in the context of DH1 pAM92 pADS exhibited a decrease in plasmid retention, with an approximate 30% loss over 96 h. These results suggest that the decreased amorphadiene production by DH1 pAM92 pADS was the result of fewer amorphadiene biosynthetic genes due to the loss of pAM92 when these two plasmids were harbored in the same cell.

Identification of limiting enzymes in the MBIS operon

The primary differences between DH1 pAM25 pMBIS pADS and DH1 pAM45 pADS are the regulation and gene dosage of the MBIS operon (Fig. 4-1A). In the original pMBIS plasmid, the plasmid backbone was derived from the broad-host range pBBR1-MCS3 plasmid^{15,16} which is maintained at approximately 6–8 copies per cell. In addition, the MBIS operon is controlled by the catabolite-repressed lac promoter. In pAM45, we utilized the lacUV5 promoter to control MBIS expression and elevated the copy number to 10–15 copies/cell by using the p15A origin of replication rather than the pBBR1 origin. The significant increase in amorphadiene production observed with DH1 pAM45 pADS suggests that one or more enzymes in the MBIS operon is limiting in the original strain. The results observed with increasing the copy number and promoter strength of the MBIS operon in pAM45 prompted us to perform gene dosage experiments to identify which of the five enzymes in the pathway were limiting. Each of the individual genes in the MBIS operon was PCR-amplified from pMBIS and sub-cloned into the p15A-based expression vector pAM29. These single-gene expression vectors were transformed into DH1 pMBIS pADS for the gene dosage experiments. Cultures were induced with IPTG and fed 20mM mevalonate to obtain flux through the pathway. We hypothesized that supplementation of any limiting MBIS gene via the pAM29-derived expression vectors would increase amorphadiene production. This experiment identified MK, which encodes mevalonate kinase, the first enzyme in the pathway, as a bottleneck and that additional copies of MK increased amorphadiene production 3-fold over the empty vector control (pAM29; Fig. 4-3).

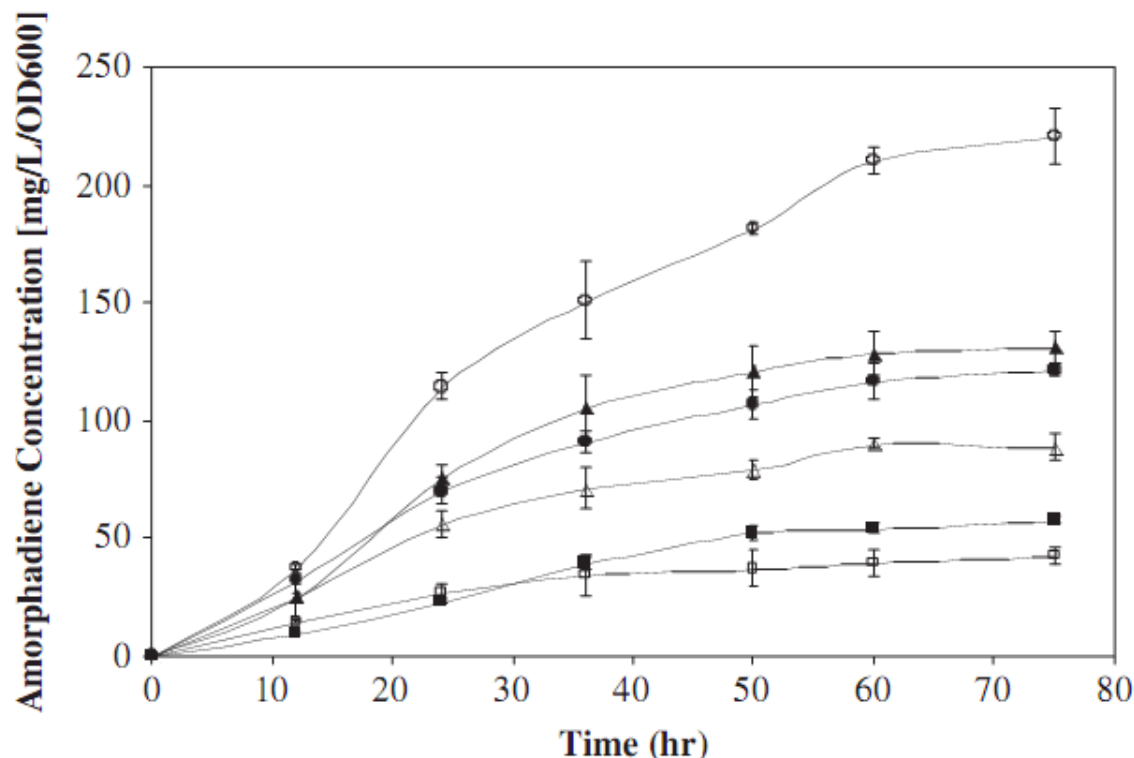


Fig. 4-2. Decreasing plasmid number and additional copies of ADS increase amorphadiene production. Comparison of amorphadiene production from shake flask cultures of *E. coli* DH1 harboring pMevT pMBIS pADS (□), pAM25 pMBIS pADS (■), pAM45-MevT pADS (●), pAM45 pADS (○), pAM92 (△), and pAM92 pADS (▲). Samples were taken every 12 h for a 72 h period for OD600 measurements and amorphadiene titers. Results are the average from three biological replicates with error bars indicating the average standard deviation from the mean.

Over-expression of the remaining four genes in the MBIS operon had no significant effect on amorphadiene production (Fig. 4-3). Finally, transforming additional copies of both MK and PMK, which encodes phosphomevalonate kinase, resulted in a minimal additional increase in amorphadiene production (data not shown), suggesting that mevalonate kinase was the major bottleneck in the MBIS pathway.

Combination of modular vectors with increased mevalonate kinase expression

In order to expand upon the increased titers observed when supplementing the pathway with additional mevalonate kinase, we codon-optimized MK for use in *E. coli* (Supplementary Table 4-1) and subcloned the gene 30 of ADS in pADS, creating pAM94. This construct offered the advantage of expressing codon-optimized MK from a strong Trc promoter and a high-copy colE1 origin of replication, maximizing the benefits from MK over-expression. pAM94 was tested for amorphadiene production in the context of either the three-plasmid system with DH1 pAM25 pMBIS or in the modular vector system with DH1 pAM45 (Fig. 4-4). DH1 pAM25

pMBIS pAM94 increased amorphaadiene production approximately 2.3-fold over DH1 pAM25 pMBIS pADS. DH1 pAM45 pAM94 had the highest amorphaadiene production observed, with a peak specific productivity of 293mg/L/OD600 at 75 h, which represents a seven-fold increase over DH1 pMevT pMBIS pADS.

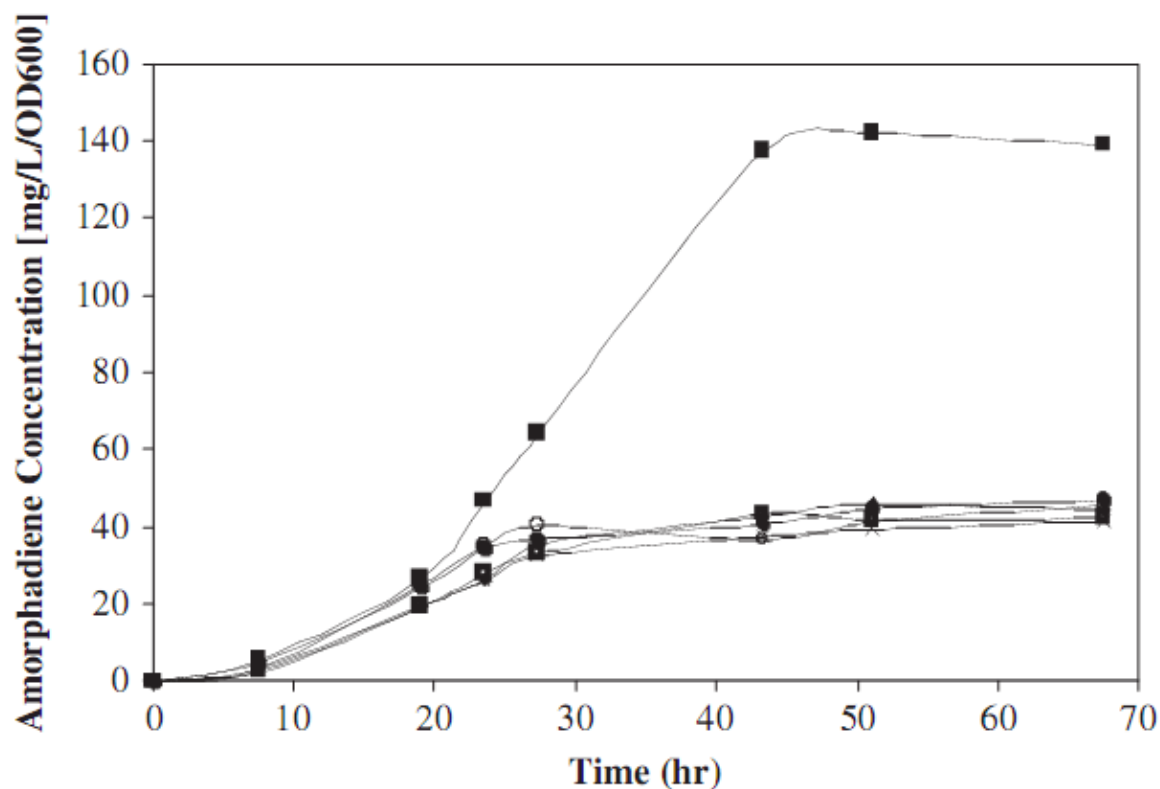


Fig. 4-3. Mevalonate kinase limits amorphaadiene production. Amorphadiene production from shake flask cultures of *E. coli* DH1 pMBIS pADS with an empty vector control (pAM29, $\blacklozenge$) or additional copies of MK ($\blacksquare$), PMK ($\bullet$), PMD ($\times$), idi ($\square$), or ispA ($\circ$). Each culture was supplemented with 20mM mevalonate. Samples were taken over a 68-h period for measurement of OD600 and amorphadiene titers.

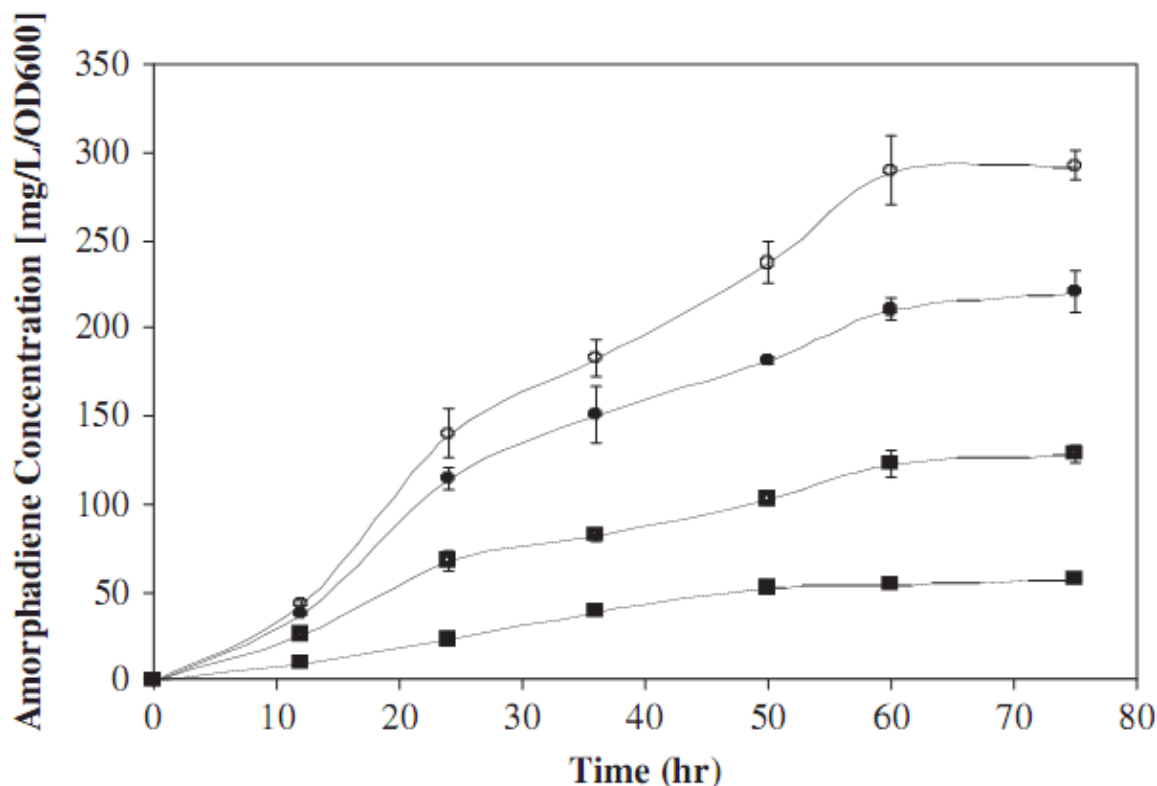


Fig. 4-4. Amorphadiene production increases when additional copies of mevalonate kinase are present. Amorphadiene production from DH1 harboring pAM25 pMBIS pADS (■), pAM25 pMBIS pAM94 (□), pAM45 pADS (●) or pAM45 pAM94 (○). Samples were taken over a 75-h period and assayed for OD600 and amorphadiene titers. Results are the average from three biological replicates with error bars indicating the average standard deviation from the mean.

Discussion and Future Direction

Engineering new or modified metabolic pathways into an organism is often a time-consuming and complicated process, especially when the optimal combination of expression elements for a given set of genes is not known. The wide varieties of plasmid copy number, promoter strengths and regulation, ribosome binding sites, and operon structure add to the complexity of successfully generating a high-flux, metabolic pathway. The vectors described in this work, which contain unique cloning sites utilizing compatible restriction enzymes, allowed for the optimization of the amorphadiene biosynthetic pathway. This approach minimizes both the time and effort needed to construct and test many alternative expression constructs. In addition, because multiple operons can be assembled into a single plasmid backbone, the total number of plasmids necessary to contain a complete pathway is reduced, which in turn lessens the metabolic burden on the host. The goal of our research was to construct plasmid vectors to both analyze pathway bottlenecks and optimize production of amorphadiene, a precursor to the anti-malarial drug artemisinin. Previous work to optimize this pathway involved modulation of 3-

hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) reductase activity to decrease levels of the toxic intermediate HMG-CoA³. However, codon optimization of genes and identification of additional limiting enzymes were previously not investigated. The first improvements in amorphaadiene production seen in this study were the result of increased expression from a codon-optimized MevT operon. Our cloning vector construct allowed for rapid screening of promoters, resulting in the replacement of the wild-type lac promoters previously used with two-fold stronger lacUV5 promoters upstream of both the MevT and MBIS operons. The use of a stronger promoter upstream of MevT resulted in a three-fold increase in production. In addition, the original MevT operon containing two *S. cerevisiae* genes (HMGS and tHMGR) was replaced with a codon-optimized version in the construct pAM45, resulting in an additional two-fold increase in production. The copy number of the MevT operon remained the same between our original strain and pAM45. However, this was not the case for the MBIS operon, whose copy number increased from 4–6 to 10–15 copies/cell. The increase in production seen from pAM45 suggested that there was a bottleneck in the MBIS pathway. Our experiments identified the first enzyme in the lower half of the pathway, mevalonate kinase, as limiting. By combining the MevT and MBIS operons on our single vector construct we were able to supplement an additional MK on a second plasmid to further improve amorphaadiene production (DH1 pAM45 pAM94). Previous studies have shown that the introduction of multi-copy plasmids in *E. coli* and expression of genes from these plasmids results in metabolic burden^{17, 18}. To maximize metabolic flux through the amorphaadiene pathway rather than into plasmid maintenance, we placed the MevT and MBIS operons and ADS on a single plasmid, pAM92. Interestingly, DH1 pAM92 did not produce more amorphaadiene than DH1 pAM45 pADS. We hypothesized that this decrease in production could be the result of insufficient ADS, because the copy number of this gene was reduced from ~30 to 10–15 copies/cell. To address this issue, we supplemented DH1 pAM92 with pADS, which increased the gene dosage of ADS to ~45 copies/cell. DH1 pAM92 pADS produced more amorphaadiene than pAM92, suggesting that ADS was limiting, but it still did not reach production levels observed with DH1 pAM45 pADS. Because DH1 pAM92 pADS and DH1 pAM45 pADS have the same mevalonate pathway and plasmid copy number, we suspected that the large size (16.2 kb) of pAM92 might contribute to plasmid instability. However, plasmid stability assays demonstrated that DH1 pAM92 maintained 90% of the plasmid over 96 h suggesting that the plasmid itself was relatively stable. Interestingly, the combination of pAM92 and pADS in the same host resulted in a 30% loss of plasmid over the same period. One hypothesis is that the presence of two strong promoter-gene cassettes (Ptrc-ADS) on two multi-copy plasmids increases the metabolic burden on the cell and results in physiological changes that lead to plasmid loss^{19, 20}. Through the use of these modular vectors, we obtained the highest amorphaadiene titers in shake flasks observed thus far (~300mg/L/OD600). The final strain, DH1 pAM45 pAM94, contained stronger lacUV5 promoters, codon-optimized MevT and ADS operons, and additional copies of MK and ADS. This vector system allowed for quick cloning and analysis of multiple components, resulting in a rapid increase in production over a short period of time. Due to the common methodologies in synthetic biology and pathway construction, we anticipate that this approach should be widely applicable to other metabolic engineering projects.

References

1. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., Keasling, J. D. Engineering the mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796-802 (2003).
2. Newman, J. D., Marshal, J. L., Chang, M. C., Nowroozi, F., Paradise, E. M., Pitera, D. J., Newman, K. L., Keasling, J. D. High-level production of amorpho-4,11-diene in a two phase partitioning bioreactor of metabolically engineered *E. coli*. *Biotechnol. Bioeng.* 95, 684-691 (2006).
3. Pitera, D. J., Paddon, C. J., Newman, J. D., Keasling, J. D., 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* 2, 193-207 (2007).
4. Barbirato, R., Grivet, J. P., Soucaille, P., Bories, A. 3-Hydroxypropionaldehyde, an inhibitory metabolite of glycerol fermentation to 1,3-propanediol by enterobacterial species. *Appl. Environ. Microbiol.* 62, 1448-1451 (1996).
5. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* 13, 247-216 (1995)
6. Pfleger, B. F., Pitera, D. J., Smolke, C. D., Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* 8, 1027-1032 (2006).
7. Zhu, M. M., Lawman, P. D., Cameron, D. C. Improving 1,3-propanediol production from glycerol in a metabolically engineered *E. coli* by reducing accumulation of sn-glycerol-3-phosphate. *Biotechnol. Prog.* 18, 694-699 (2002).
8. Yoshikuni, Y., Ferrin, T. E., Keasling, J. D. Designed divergent evolution of enzyme function. *Nature.* 440, 1078-1082 (2006).
9. Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R., Keasling, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature.* 440, 940-943 (2006).
10. Lutz, R., Bujard, H. Independent and tight regulation of transcriptional units in *E. coli* via the LacR/O, the TetR/O, and AraC/I1-I2 regulatory elements. *Nucl Acids Res.* 25, 1203-1210 (1997).
11. Stefano, J.E., Ackerson, J.W., Gralla, J.D. Alterations in two conserved regions of promoter sequence lead to altered rates of polymerase binding and levels of gene expression. *Nucl. Acids Res.* 8, 2709–2723 (1980).
12. Diaz Ricci, J. C., Hernandez, M. E. Plasmid effects on *Escherichia coli* metabolism. *Crit Rev Biotechnol.* 20, 79-108 (2000).
13. Paulsson, J., Ehrenberg, M. Trade-off between segregational stability and metabolic burden: a mathematical model of plasmid ColE1 replication control. *J Mol Biol.* 279, 73-88 (1998).
14. Rozkov, A., Avignone-Rossa, C. A., Ertl, P. F., Jones, P., O'Kennedy, R. D., Smith, J. J., Dale, J. W., Bushell, M. E. Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnol Bioeng.* 88, 909-915 (2004).
15. Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., Peterson, K. M. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene.* 166, 175-176 (1995).

16. Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M., Peterson, K. M. pBBR1MCS: A broad-host-range cloning vector. *Biotechniques*. 16, 800-801 (1994).
17. Flores, S., de Anda-Herrera, R., Gossset, G., F.G., B. Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnol. Bioeng.* 87, 485-494 (2004).
18. Jones, K. L., Keasling, J. D. Construction and characterization of F plasmid-based expression vectors. *Biotechnol. Bioeng.* 59, 659-665 (1998).
19. Bentley, W., Mirjalili, N., Andersen, D., Davis, R., Kompala, D. Plasmid-encoded protein: the principal factor in the metabolic burden associated with recombinant bacteria. *Biotechnol. Bioeng.* 35, 668–681 (2004).
20. Ow,D., Nissom,P. ,Philp,R. ,Oh,S. ,Yap,M., 2006. Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5a during batch fermentation. *Enzyme Microbial. Technol.*39, 391–398 (2006).
21. Hanahan, D. Studies of transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557-580 (1983).

Chapter 5: Study and optimization of the lower part of Heterologous mevalonate pathway in *Escherichia coli*; the cells' response

Introduction

Recent trends toward the production of complex chemicals and pharmaceuticals in engineered microbes require continual improvements in the design of non-native biosynthetic pathways. Using principles of metabolic engineering, multi-gene heterologous pathways have been incorporated into microorganisms for the production of many important classes of molecules^{1,2,3,4,5,6}. However, in the course of transferring enzymatic pathways from one organism to another, implemented regulations on the pathway are lost in the host organism, leading to imbalances in gene expression and protein activity^{7,8,9}. Over-expression of a gene may cause depletion of precursors that are necessary for cells growth¹⁰, while an imbalance in total activity of the enzymes in a pathway can limit carbon flux. This can result in a bottleneck in the biosynthetic pathway which reduces the production rate and can also lead to accumulation of intermediates that can inhibit pathway enzymes or are toxic to the cells.

This chapter describes how balancing carbon flux through heterologous mevalonate pathways is essential in the biosynthesis of isoprenoids. Detailed study of the heterologous mevalonate pathway through substrate feeding experiments suggested that *in vivo* accumulation of some of the mevalonate pathway intermediates is toxic and inhibits normal cell growth¹¹ (Figure 5-1). Therefore the largest obstacle in efficient microbial biosynthesis of isoprenoids is high level production of Farnesyl pyrophosphate (FPP) and quick conversion of this metabolite to amorphaadiene.

As described in the previous chapters, mevalonate pathway was transferred to *E. coli* to bypass the unstudied native regulation of FPP production and to increase the carbon flux towards production of isoprenoid precursors^{1,12-15}. Although this was beneficial in increasing the product yields, increased expression of the pathway can result in imbalances in genes expression and protein activity. Accompanied by lack of regulation, this can lead to accumulation of the pathway intermediates. We have confirmed that our engineered strains accumulate a significant amount of mevalonate. Furthermore, through initial optimization of expression levels, we have found that increasing the expression of enzymes that convert mevalonate to mevalonate phosphate and mevalonate diphosphate would increase the flux through the lower part of mevalonate pathway and result in higher-level amorphaadiene production¹⁶. In our attempt to further optimize the mevalonate pathway, we investigated the effect of ribosome binding site (RBS) strength on gene expression, production, and cell growth¹⁷.

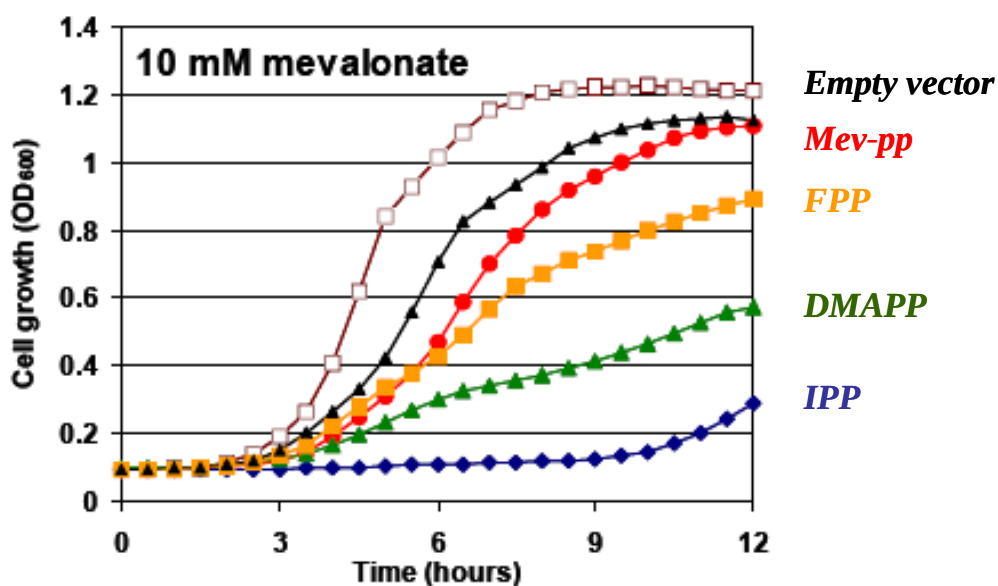


Fig. 5-1. Growth inhibition by high flux through the mevalonate bottom pathway. Growth curves of *E. coli* showing the inhibitory effect caused by accumulation of mevalonate pathway intermediates, IPP, DMAPP, and FPP.

For this purpose, different ribosome binding sites with various strengths were cloned in front of the genes of lower mevalonate pathway and the amorphaadiene synthase gene¹⁸. These ribosome binding sites were selected either based on their theoretical strengths or through predictions of a RBS calculator program that is designed to predict synthetic ribosome binding site sequences. Based on the desired strength and the translation rate is determined by multiple molecular interactions, including the hybridization of the 16S rRNA to the RBS sequence, the binding of tRNA^{fMET} to the start codon, the distance between the 16S rRNA binding site and the start codon (called spacing) and the presence of RNA secondary structures that occlude either the 16S rRNA binding site or the standby site¹⁹. Monitoring the growth and production level of the resulted constructs, significant differences in growth and production between some of the constructs was observed.

Through further analysis of mRNA levels and protein levels of all of the enzymes in mevalonate pathway, we could trace the protein levels to ribosome binding site strength. Measurement of pathway intermediates concentrations provides additional information on how the changes in protein levels translate to different production levels and growth rates. This helped us in putting different pieces of the puzzle together to have a better understanding of how different parts of this pathway work together and how despite all of the constraints enforced on an operon, the cells can control the expression levels to prevent further accumulation of a metabolite that is toxic to them.

These results demonstrate that balancing flux through the heterologous pathway and maintaining a steady growth condition is a key determinant in optimizing terpenoid production in microbial hosts. The results of this study are also of great value to

engineering microbes for production of any other molecule of therapeutic and industrial interest. Recent research in the field of metabolic engineering seeks to apply synthetic biology tools and well characterized genetic components to create novel biological functions. However in construction of large pathways from different characterized components, the combined effect is usually not predictable. It can lead to carbon imbalances that result in accumulation of intermediates that lower the efficiency of the system or are detrimental to the cell for often unforeseen reasons. Strains and plasmids used in this study are listed in table 4.

Table 4: Strains and plasmids used in this study.

H: High Activity RBS, M: Medium Activity RBS, L: Low activity RBS, O: Original RBS
All plasmids are pTrc99A derivatives with Ap^R

Strain/plasmid	Description	Reference
Plasmid		
pTrc.ADS-MBIS1	O-ADS, M-MK, M-PMK, M-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS2	O-ADS, H-MK, H-PMK, H-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS3	O-ADS, L-MK, L-PMK, L-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS4	O-ADS, H-MK, M-PMK, M-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS5	O-ADS, M-MK, H-PMK, M-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS6	O-ADS, M-MK, M-PMK, H-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS7	O-ADS, L-MK, M-PMK, M-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS8	O-ADS, M-MK, L-PMK, M-PMD, O-idi, O-ispA	This Study
pTrc.ADS-MBIS9	O-ADS, M-MK, M-PMK, L-PMD, O-idi, O-ispA	This Study
pTrc.5KADS-MBIS1	5000 units of activity RBS-ADS and MBIS1	This Study
pTrc.10KADS-MBIS1	10'000 units of activity RBS-ADS and MBIS1	This Study
pTrc.15KADS-MBIS1	15'000 units of activity RBS-ADS and MBIS1	This Study
pTrc.32KADS-MBIS1	32'000 units of activity RBS-ADS and MBIS1	This Study
pTrc.55KADS-MBIS1	55'000 units of activity RBS-ADS and MBIS1	This Study
pTrc.5KADS-MBIS2	5000 units of activity RBS-ADS and MBIS2	This Study
pTrc.10KADS-MBIS2	10'000 units of activity RBS-ADS and MBIS2	This Study
pTrc.15KADS-MBIS2	15'000 units of activity RBS-ADS and MBIS2	This Study
pTrc.32KADS-MBIS2	32'000 units of activity RBS-ADS and MBIS2	This Study
pTrc.55KADS-MBIS2	55'000 units of activity RBS-ADS and MBIS2	This Study
pTrc-MBIS2	pTrc99A derivative containing MBIS2 and no ADS gene	This Study
pMevT	pLac33 derivative containing MevT operon; Cm ^R	Martin et al (2003)
Strains		
DH1	F- supE44 hsdR17 recA1 gyrA96 relA1 endA1 thi-1 λ-	Hanahan (1983)
DH10B	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu) 7697 galU galK λ- rpsL nupG	Invitrogen

Materials and Method

E. coli strain DH10B was used for cloning and construction of MBIS constructs. *E. coli* strain DH1 (Invitrogen, Carlsbad, CA) was used for expression and production studies. Media components and chemicals were purchased from Sigma (St. Louis, MO) and

Fisher Scientific (Pittsburg, PA). All of the experiments were performed in Terrific Broth media (Sigma) with appropriate antibiotics for plasmid selection and glycerol as the carbon source. All strains were grown at 30 ° C.

For targeted proteomics experiments, mass spectrometric-grade Trypsin (Trypsin Gold) was obtained from Promega (Madison, WI) and prepared according to manufacture's instructions.

Operon and plasmid construction

The MBIS operon was previously cloned into pBBR1MCS, a low copy plasmid with an IPTG inducible *lac* promoter. To increase the expression of every gene on this operon, a higher copy number plasmid with a stronger promoter was used. All of the constructed MBIS operons were cloned into a pTrc99a derivative containing pBR322 origin, *lacI*^Q, and P_{TRC} promoter. Ribosome binding sites with different strengths were cloned in front of the first three genes of pMBIS (Mevalonate Kinase, *MK*, PhosphoMevalonate Kinase, *PMK*, and PhosphoMevalonate Decarboxylase, *PMD*) to construct the new MBIS operons. The substitute ribosome binding sites were chosen based on their theoretical strengths (Ref) so that we have a High activity, a Medium activity, and a Low activity RBS for each of the three genes mentioned above. The biobrick cloning strategy (figure 2) was used to put together different components of the pathway²⁰. The biobrick methodology for cloning presents an attractive platform for standardizing DNA construction. The biobrick strategy is based on the existence of restriction enzymes that recognize different restriction sites but create restriction fragments with complementary sticky ends. Therefore, after one rounds of digestion and ligation, a new site is created that is no longer recognized by either of the enzymes. This allows for reusing the enzymes and the procedure can be repeated to put different parts together.

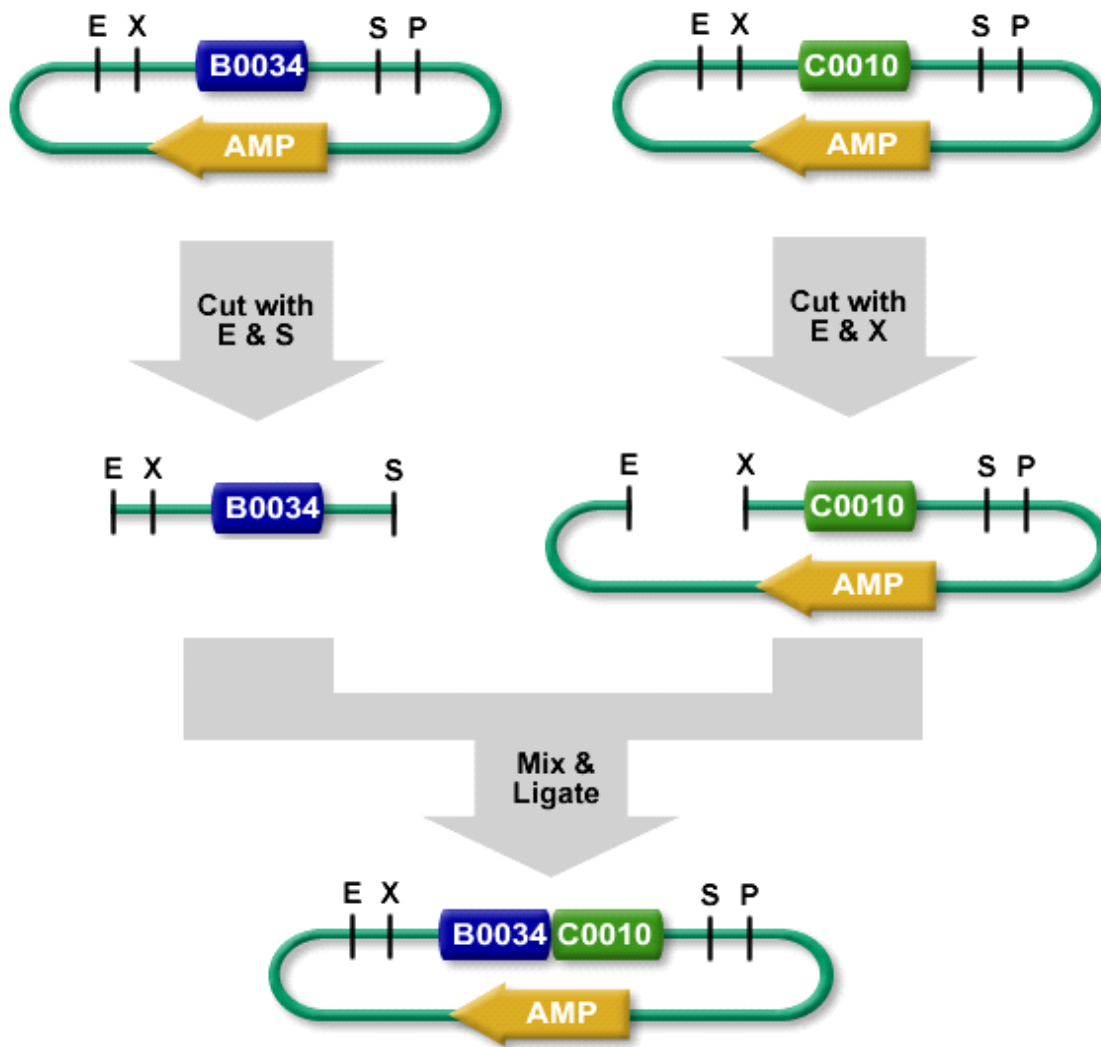


Fig. 5-2. Biobrick cloning method. Both parts have the same restriction sites on each end (E, X, S, and P). Restriction enzymes X and S create restriction fragments with complementary sticky ends. Taking advantage of this property of these two restriction enzymes

The first three genes of the lower mevalonate pathway, *MK*, *PMK*, and *PMD* were amplified by PCR using pMBIS as the template and gene specific primers that contain the three selected ribosome binding sites and restriction sites *EcoRI*, *BglII*, *BamHI* and *XhoI*. The last two genes, *idi* and *ispA*, were also amplified by PCR but only one RBS, Medium activity RBS, was used with these two genes (Figure 5-3).

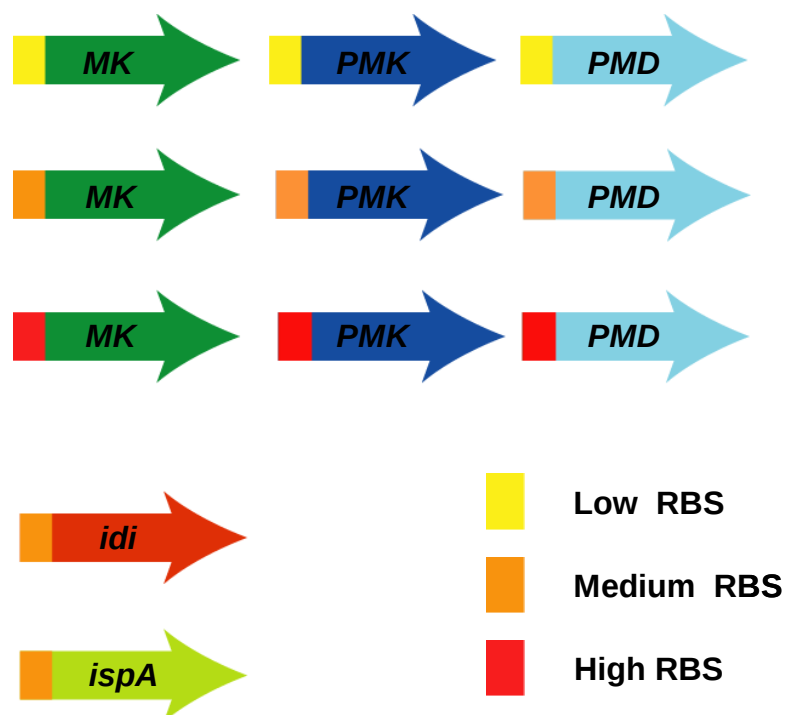


Fig. 5-3. Using PCR the five genes of the lower mevalonate pathway were amplified. By designing PCR primers that contain the selected RBS sequences, three RBSs with three different strengths were put in front of the first three genes of MBIS operon. Only one RBS of medium activity was used with the last two genes, *idi*, and *ispA*.

Every individual part was cloned into a biobrick backbone, pBca9145 (ColE1 origin of replication, Amp^R, and no promoter), to facilitate the construction of MBIS operons. We decided to combine the MBIS operon and amorphaadiene synthase gene with its original RBS into one operon, using pTrc99a as the backbone. Amorphaadiene synthase was digested with EcoRI/BamHI restriction enzymes, while the pBca9145-MBIS plasmids were digested with EcoRI/BglII restriction enzymes. The two digested fragments were ligated together to form the pBca9145-ADS.MBIS plasmids. Retaining all of the biobrock restriction sites, the pBca9145-ADS.MBIS plasmids were then digested with EcoRI and BamHI to be cloned into EcoRI-BamHI sites of pTrc99a expression vector to form pTrc-ADS.MBIS (figure 4).

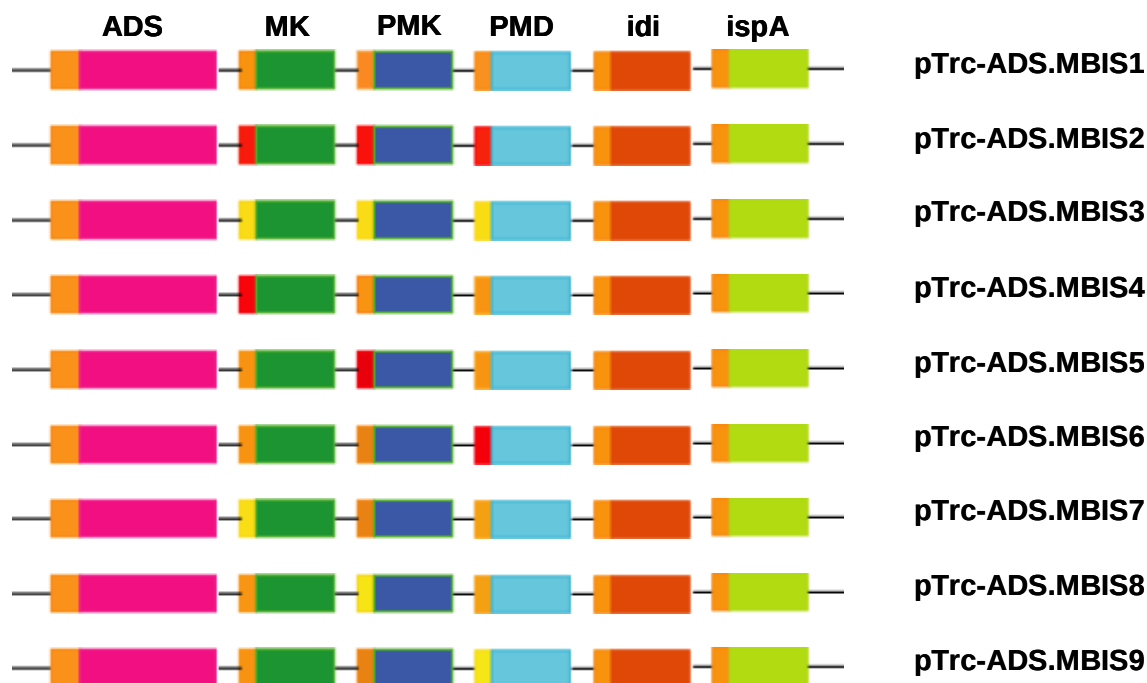


Fig. 5-4. The 9 constructed operons containing amorphadiene synthase (ADS) and MBIS operon cloned into pTrc99a backbone. These constructs were transformed into DH1 strain of *E. coli* along with pMevT operon and were tested for growth and amorphadiene production.

Out of 27 possible RBS-gene combinations to form pTrc-ADS.MBIS constructs, 9 constructs were chosen to be studied. These constructs are listed in table 5-2.

All of the resulted constructs were sequenced to ensure accuracy.

Growth Condition

DH1 strain of *E. coli* was transformed with the appropriate plasmids and plated on Luria Broth Agar plates containing appropriate antibiotics. Overnight cultures were started from fresh transformations into 5 ml of Luria Broth (Difco) with antibiotics. Overnight cultures were diluted to an OD₆₀₀ (optical density at a wavelength of 600 nm) of 0.05 in 40 ml of Terrific Broth (Difco) with 1% or 3% glycerol and overlaid with 20% (v/v) dodecane (Sigma-Aldrich). Experiments were initiated when cultures reached an OD₆₀₀ of 0.2–0.3 (~2–3 h after subculture) by the addition of 1 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG). Cultures were incubated for 48–72 h, and at each time point the OD₆₀₀ was read and samples from the dodecane phase were taken to determine amorphadiene production. Previous studies have shown that over 97% of the amorphadiene produced partitions into the organic dodecane overlay of cultures²².

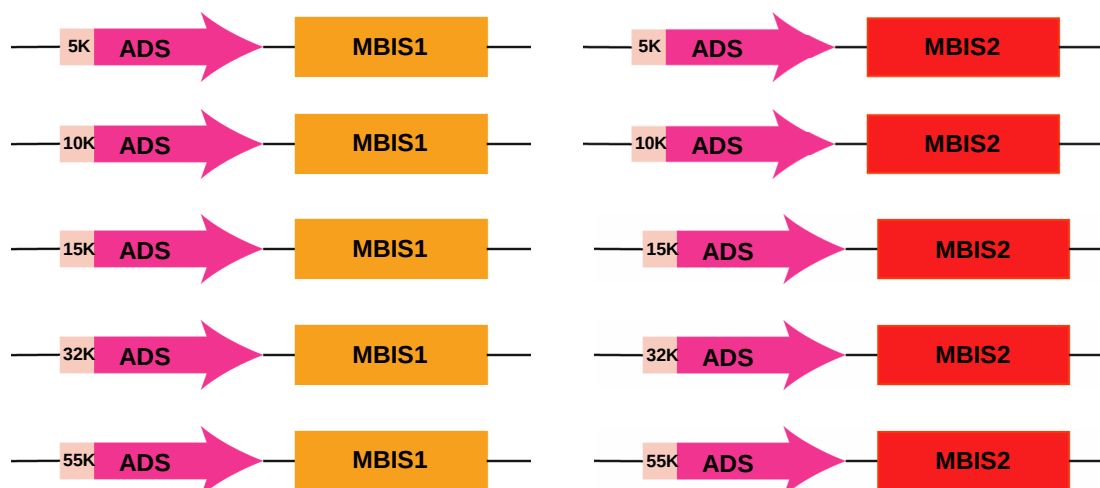


Fig. 5-5. The 10 constructs containing ADS with ribosome binding sites of different activity levels and two of the MBIS operons (MBIS1 and MBIS2). The two MBIS operons selected were the ones with all Medium activity and High activity RBSs in front of the first three genes of MBIS operon. These 10 yielded operons (pTrc-RBS.ADS-MBIS1 and pTrc-RBS.ADS-MBIS2) were transformed into DH1 strain of *E. coli* along with pMevT and were tested for growth and production.

Gas chromatograph-mass selective (GC-MS) analysis of amorphadiene

Samples (100 μ l) were removed from the dodecane layer of each culture, and centrifuged at 13,300g for 3 min at room temperature. After centrifugation, 10 μ l of the organic phase was added to 990 μ l of ethyl acetate (Sigma-Aldrich). To measure amorphadiene levels, 1 μ l of the sample was injected into an Agilent 6890 series gas chromatograph (GC) equipped with an Agilent 5973 mass selective (MS) detector and a cyclosyl-B (chiral) capillary column (30 m $\times$ 250 μ m $\times$ 0.25 μ m thickness, Agilent). The GC oven temperature program was as follows: 100 $^{\circ}$ C for 45 s, a ramp of 35 $^{\circ}$ C/min to 200 $^{\circ}$ C, a ramp of 100 $^{\circ}$ C/min to 250 $^{\circ}$ C, and a hold at 250 $^{\circ}$ C for 1 min. The MS was operated in selected ion monitoring (SIM) mode scanning for the molecular ions at 189 and 204 m/z. Purified amorphadiene was used to generate standard curves for obtaining production titers.

Proteomics Analysis

Proteomics Sample Preparation

Whole cell lysate proteomics samples were prepared as follows. The cells were pelleted via centrifugation at 8000 $\times$ g (4 $^{\circ}$ C) for 5 minutes, the supernatant was decanted, and cells were flash frozen in liquid nitrogen and stored at -80 $^{\circ}$ C. Protein extraction was done using a chloroform/methanol precipitation. Briefly, cells were allowed to thaw on ice and vortexed to homogenize the solution. 100 μ l of the concentrated cell pellet was placed into a fresh 1.7 ml tube and 400 μ l of methanol was added to the tube and vortexed. 100 μ l of chloroform was added and mixed, followed by the addition of 400 μ l of LC/MS grade water. Samples were vortexed and spun in a bench top centrifuge at 14,000 rpm for 1 minute. The upper layer was discarded carefully and 300 μ l of methanol was added and mixed. Samples were spun at 14,000 rpm for 2

minutes, allowing the protein to collect at the bottom of the tube. The supernatant was removed and protein pellets were allowed to dry in a vacuum concentrator for 30-60 minutes. 500 µl of 50mM ammonium bicarbonate in 10% acetonitrile was added and samples were placed in a sonic bath intermittently with continuous vortexing to resuspend the pellet. Protein concentration was determined with the Biorad DC Protein reagent (BioRad, Hercules, CA) according to manufacturer's instructions. Approximately 50 µg of protein was taken from each sample and diluted to a final concentration of 1 µg/µl in 50 mM ammonium bicarbonate. To reduce the disulfide bonds in the samples, 5 mM TCEP was added to the solution and allowed to sit at room temperature for 30 minutes. The proteins were alkylated by iodoacetic acid (200 mM IAA in 100 mM NaOH) in the dark for 30 min. Following alkylation, trypsin (1 µg/µl) was added to the sample to a final concentration of 1:50 (trypsin:sample) and incubated at 37°C overnight. The peptide samples were used without further purification. Standard proteins were dissolved in 50 mM ammonium bicarbonate and subjected to the same reduction, alkylation, and trypsinization conditions as the lysates.

LC-MS Analysis

Peptide samples were analyzed on a LC-MS/MS system consisting of an Eksigent TEMPO nanoLC-2D coupled to an Applied Biosystems (Foster City, CA) 4000Q-Trap mass spectrometer operating in MRM mode. The peptide samples were loaded onto a Pepmap100 µ-guard column (Dionex-LC Packings, Sunnyvale, CA) and were washed for 30 minutes (15 µl/min flow rate) prior to injection onto a Dionex Pepmap100 analytical column (75 µm I.d., 150 mm length, 100Å, 3 µm) at a flow rate of 300 nl/min. The analytical column was interfaced to the mass spectrometer with a Microionspray head with nebulizing gas and a 10 µm Picotip emitter (New Objective, Woburn, MA) operating in the positive mode (2250 – 2400 V). The LC conditions were as follows: following the wash period the sample was eluted with a gradient from 100 % buffer A (2% (v/v) acetonitrile, 0.1% (v/v) formic acid) to 65% buffer A, 35% buffer B (98% (v/v) acetonitrile, 0.1% (v/v) formic acid) over 120 min, followed by a ramp of 35 % to 80% (v/v) buffer B in 10 min. The solvent composition was held at 80% (v/v) B for 10 min followed by a short ramp back to 100% A and a re-equilibration period at 100% A for 15 min.

Initially, data were collected with Analyst™ 1.5 (Applied Biosystems) operating in Information Dependent Acquisition (IDA) mode and consisting of a MRM survey scan and two product ion (MS/MS) scans. The Q3 ions used as MRM transitions were selected by choosing high-abundant y-series fragments from the product ion scan above a molecular weight of 600 m/z. The total number of MRM transitions for all targeted proteins were limited to fewer than 4 for the purpose of maintaining a maximum dwell time of 100 ms per transition and a total cycle time of less than 4s. For a second round of analysis, an IDA scan with two MS/MS spectra per MRM scan was triggered for MRM transitions of greater than 1000 counts/s, to verify that the correct peptide signal was detected. Finally, the majority of the data was collected using an MRM only scan method that had been previously validated using the first two rounds of MS/MS selection. MS/MS spectra were collected for 2s over a mass range of 100-1600 m/z with Q1 resolution=LOW and rolling collision energy. MS/MS spectra were processed with the Paragon algorithm (ProteinPilot 2.0, Applied Biosystems) and searched against a database consisting of all the ORFs from *E. coli* K12 obtained from MicrobesOnline (www.microbesonline.com). Additionally, the database contained all exogenous proteins of interest, and common contaminants such as human keratins, trypsin, and commonly used standard proteins (bgal, bsa, cytochrome c, etc.). The data were searched with the following settings: protease digestion with trypsin, cysteine blocking with iodoacetic acid, confidence level was set to 95% (ProtScore =

1.3), and the Paragon algorithm was set to rapid so that only fully tryptic peptides with no post-translational modifications were considered.

Metabolomics Analysis

All chemical standards were obtained from Sigma-Aldrich, CA, USA. All chemicals used were of analytical and reagent grade, and all solvents used were of HPLC grade (Honeywell Burdick & Jackson, CA, USA).

The analysis of FPP by liquid chromatography and mass spectrometry (LC-MS)

FPP was purchased from Sigma-Aldrich, and was made up to 200 μ M, as the stock solution, in 50 % methanol and 50 % water. The separation of the FPP was conducted on a ZIC-HILIC column (150 mm length, 2.1 mm internal diameter, and 3.5 μ m particle size; from Merck SeQuant, and distributed via The Nest Group, Inc., MA., USA) using an Agilent Technologies 1200 Series HPLC system (Agilent Technologies, CA, USA). Injection volumes for the chemical standard and the metabolite were 2 and 10 μ L respectively. The temperature of the sample tray was maintained at 4 °C by an Agilent FC/ALS Thermostat. The column compartment was set to 32 °C. FPP was eluted isocratically with a mobile phase composition of 36 % of 50 mM ammonium acetate, in water, and 64 % of acetonitrile. A flow rate of 0.15 mL/min was used throughout.

The HPLC system was couple to an Agilent Technologies 6210 time-of-flight mass spectrometer (LC-TOF MS), by a 1/6 post-column split. Contact between both instrument set-ups was established by a LAN card in order to trigger the MS into operation upon the initiation of a run cycle from the MassHunter workstation (Agilent Technologies, CA, USA). Electrospray ionization (ESI) was conducted in the negative ion mode and a capillary voltage of - 3500 V was utilized. MS experiments were carried out in full scan mode, at 0.85 spectra/ second and a cycle time of 1.176 seconds, for the detection of $[M - H]^-$ ions. The instrument was tuned for a range of 50 – 1700 m/z. Prior to LC-TOF MS analysis, the TOF MS was calibrated via an ESI-L-low concentration tuning mix (Agilent Technologies, CA, USA). Internal reference mass calibration was utilized throughout the chromatographic run via an API TOF reference mass solution kit (Agilent Technologies, CA, USA). Data acquisition and processing were performed by the MassHunter software package. FPP from *E. coli* extracts were quantified via seven-point calibration curves ranging from 625 nM to 50 μ M. The R^2 coefficients for the calibration curves were ~0.9974.

The analysis of Mevalonate by liquid chromatography and mass spectrometry (LC-MS).

The chemical standard of mevalonate was prepared by treating mevalonolactone (Sigma-Aldrich, CA, USA) with 2 M potassium hydroxide (Martin et al 2003). Mevalonate was then dissolved in a solution of 50 % methanol and 50 % water. The separation of mevalonate was conducted on a Carbomix H-NP 10:8 % column (250 mm length, 4.6 mm internal diameter, and 5 μ m particle size; Sepax, CA, USA) using an Agilent Technologies 1100 Series HPLC system. A sample injection volume of 1 μ L was used throughout. The temperature of the sample tray was maintained at 4 °C by an Agilent FC/ALS Thermostat. The column compartment was set to 50 °C. Mevalonate was eluted isocratically with a mobile phase composition 0.1 % formic acid in water. A flow rate of 0.2 ml/min was used throughout.

The HPLC system was couple to an Agilent LC/MSD SL mass spectrometer (LC-MS), by a 1/6 post-column split. The LC-MS system was controlled by the Chemstation (Agilent Technologies, CA, USA) software package. Contact between both instrument set-ups was established by a LAN card in order to trigger the MS into operation upon the initiation of a run cycle from Chemstation. ESI-MS was conducted in the negative ion mode and a capillary voltage of - 3500 V was utilized. MS experiments were carried out in the selected ion monitoring mode, with a dwell time of 995 ms, for the detection of $[M - H]^-$ ions. The instrument was tuned for a range of 50 – 2000 m/z. The LC/MSD SL was calibrated externally by the Agilent ES tune mix (Agilent Technologies, CA, USA). Data acquisition and processing were performed by MassHunter. Mevalonate from *E. coli* extracts was quantified via seven-point calibration curves ranging from 312.5 nM to 100 μ M. The R^2 coefficients for the calibration curves were ~ 0.9972 .

Results

Initial steps of the lower mevalonate pathway limit carbon flux to amoradiene

To begin improving isopenoid production, from engineered *E. coli*, the limiting steps in the engineered pathway to amoradiene were determined. It was shown that over-expression of the first two genes of the lower mevalonate pathway, *mevalonate kinase* (MK) and *phosphomevalonate kinase* (PMK) increases the amoradiene production. There is also a considerable amount of mevalonate accumulating in the culture media, leading to the hypothesis that the optimal balance has not been obtained between the upper and lower mevalonate pathway and there is a bottleneck in the lower part limiting the flux through this portion of the pathway. Taking into consideration that the first three genes of our lower mevalonate pathway are taken from yeast *S. cerevisiae* while the last two genes are *E. coli*'s native genes and therefore should be functionally expressed in *E. coli*, we decided to focus our efforts on optimizing the expression of the first three heterologous enzymes (MK, PMK, and PMD) in *E. coli*. For this purpose, three ribosome binding sites with High, Medium, and Low activity levels were chosen and cloned in front of the MK, PMK, and PMD. It should be emphasized that the activity level of these ribosome binding sites are theoretical and the RBS with Low activity could in fact have very high activity levels in front of one gene and very low activity levels in front of another. Nine MBIS operons were constructed using different ribosome binding site combinations. The amoradiene synthase gene was cloned in front of these synthetic MBIS operons and the resulted ADS-MBIS operons were cloned into pTrc99a backbone. DH1 strain of *E. coli* was transformed with pMevT and pTrc-ADS.MBIS plasmids. Cultures were induced with IPTG and samples were taken at 24 hr and 48 hr time point to look at growth and amoradiene production level. We had shown that accumulation of pathway intermediates IPP, DMAPP, and FPP are cytotoxic to the cell and will cause growth inhibition. By comparing the growth between different strains we could check for any growth defect that could have been caused by accumulation of any of the toxic intermediates. Also if the protein levels were increased by changing the RBSs, we could expect to see an increase in amoradiene level due to the increased carbon flux through the bottom part of the pathway. The strains expressing these 9 MBIS synthetic operons demonstrate significant differences in growth and production. The strain expressing the MBIS operon with High activity ribosome binding sites for all three genes show growth inhibition at 24 hour that

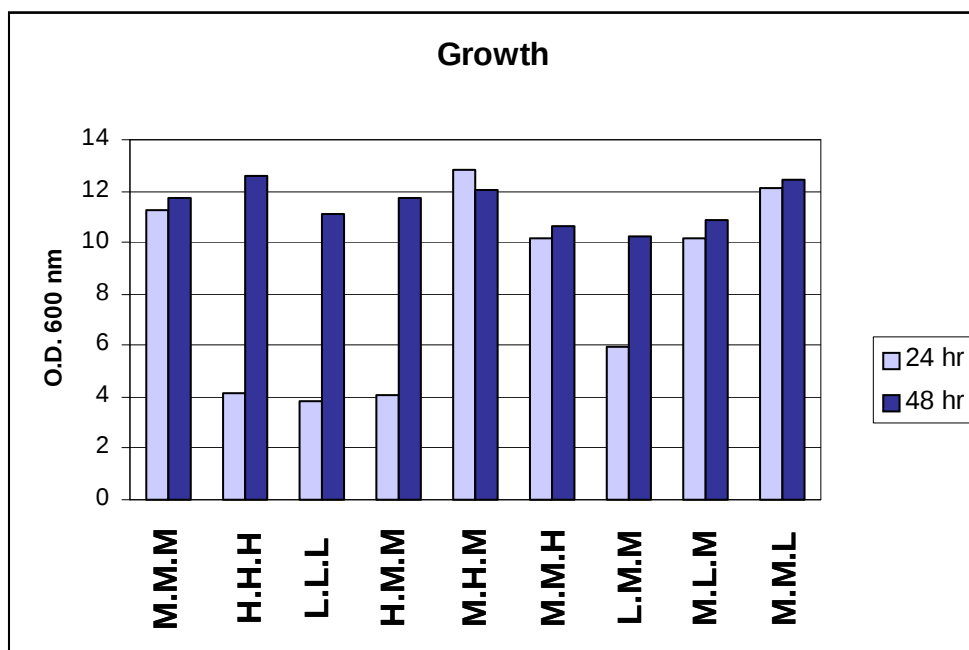


Fig. 5-6. Cell growth of *E. coli* DH1 strains expressing pMevT and pTrc-ADS.MBIS operons 24 hour and 48 hour after induction. Some of the strains show growth inhibition at 24 hour that is relieved at 48 hour.

is relieved at 48 hour (Figure 5-6). Also the amorphaadiene production level for this strain is not as high as the strains that do not show any growth inhibition. Comparing the growth and production levels for these 9 strains, we conclude that all of the strains with growth inhibition also exhibit decreased production levels. We chose the construct with all Medium activity RBSs (pTrc-ADS.MBIS1) as the control and all of the other constructs are compared to this construct. Out of the 9 constructs tested, only one resulted in a moderate improvement and increased the production over 50%. This increase in production could be contributed by the theoretically higher activity RBS in front of the second gene of the MBIS operon (PMK). Given that the PMK protein has been so low that is hardly detectable by LC-MS, even slight improvement in expression level could result in better amorphaadiene yields.

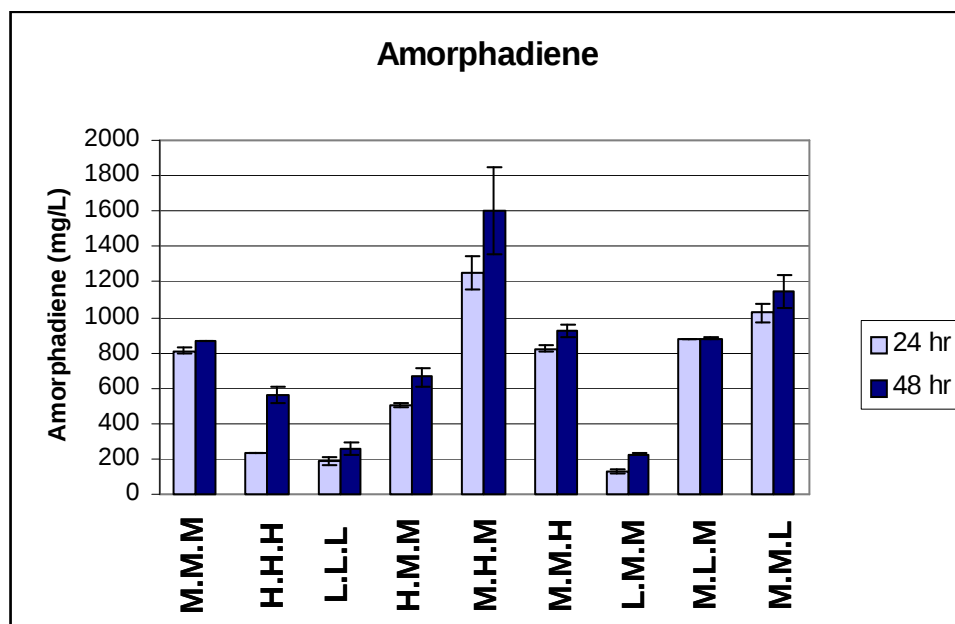


Fig. 5-7. Amorphadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-ADS.MBIS operons 24 hour and 48 hour after induction. The strains with growth inhibition at 24 hour also exhibit poor amorphadiene production.

From these observations we hypothesized that theoretical High activity ribosome binding sites in front of MK, PMK, and PMD could potentially be stronger ribosome binding sites that result in higher protein expression. The increase in enzyme level could in turn increase the carbon flux through the pathway and increase the FPP production. If there is not enough amorphadiene synthase enzyme to convert this extra FPP to amorphadiene, accumulation of FPP could cause growth defect and affect the production of amorphadiene.

Use of stronger RBS in front of ADS

We demonstrated that amorphadiene synthase is a limiting enzyme (chapter 4) in amorphadiene production. Attempts in engineering ADS for improved function to increase amorphadiene production proved futile. We suspect that the MBIS operon with High activity ribosome binding sites (pTrc-ADS.MBIS2) could be causing FPP accumulation. Therefore, we decided to increase enzyme concentration by use of strong synthetic ribosome binding sites that were specifically designed for ADS. A ribosome binding site calculator¹⁹ was used to design synthetic ribosome binding sites with desired strengths. Five ribosome binding site sequences with 5000, 10'000, 15'000, 32'000' and 55'000 units of activity were designed and cloned in front of ADS.

4944 units: **AGATCT** CCGCTCCTACCAAGAAATCACCGAGATAGAAGT ATG
 9273 units: **AGATCT** AACATAACAAATAACGATAAAAGGTAGGACA ATG
 14552 units: **AGATCT** GTCATTCTAACAATAACTACGAGGATACC ATG
 32543 units: **AGATCT** CCTACGCTCATACACACATTAAGGAGAAACGT ATG ...
 55027 units: **AGATCT** ACGTAATTAAAAGAACTAACAAGGAGAAAGTAAA ATG...

These RBS-ADS parts were then cloned into vectors carrying MBIS1 and MBIS2 operons. We chose these two MBIS operons since MBIS2 displays severe growth inhibition and is one of the constructs with low production levels. MBIS1 is chosen as the control and for comparison purpose. By expressing more of the ADS protein we hoped to increase amorphadiene production, specifically in the case of MBIS2 where higher substrate concentration (FPP) was suspected. DH1 strain of *E. coli* was transformed with pMevT and pTrc-RBS.ADS-MBIS plasmids. Cultures were induced with IPTG and samples were taken at 24 hr and 48 hr time point to look at growth and amorphadiene production level. The strains tested in this experiment are listed in table 5.

Table 5: Plasmids with different strength RBSs in front of ADS

Strain	Description	Reference
pTrc-5K.ADS-MBIS1	5000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-10K.ADS-MBIS1	10000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-15K.ADS-MBIS1	15000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-32K.ADS-MBIS1	32000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-55K.ADS-MBIS1	55000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-5K.ADS-MBIS2	5000 units of activity RBS in front of ADS with MBIS2	This Study
pTrc-10K.ADS-MBIS2	10000 units of activity RBS in front of ADS with MBIS2	This Study
pTrc-15K.ADS-MBIS2	15000 units of activity RBS in front of ADS with MBIS2	This Study
pTrc-32K.ADS-MBIS2	32000 units of activity RBS in front of ADS with MBIS2	This Study
pTrc-55K.ADS-MBIS2	55000 units of activity RBS in front of ADS with MBIS2	This Study

There is not a significant difference in growth between the strains expressing MBIS1 and MBIS2 operons (Figure 5-8). However there is growth variation between the strains expressing ADS with different strength

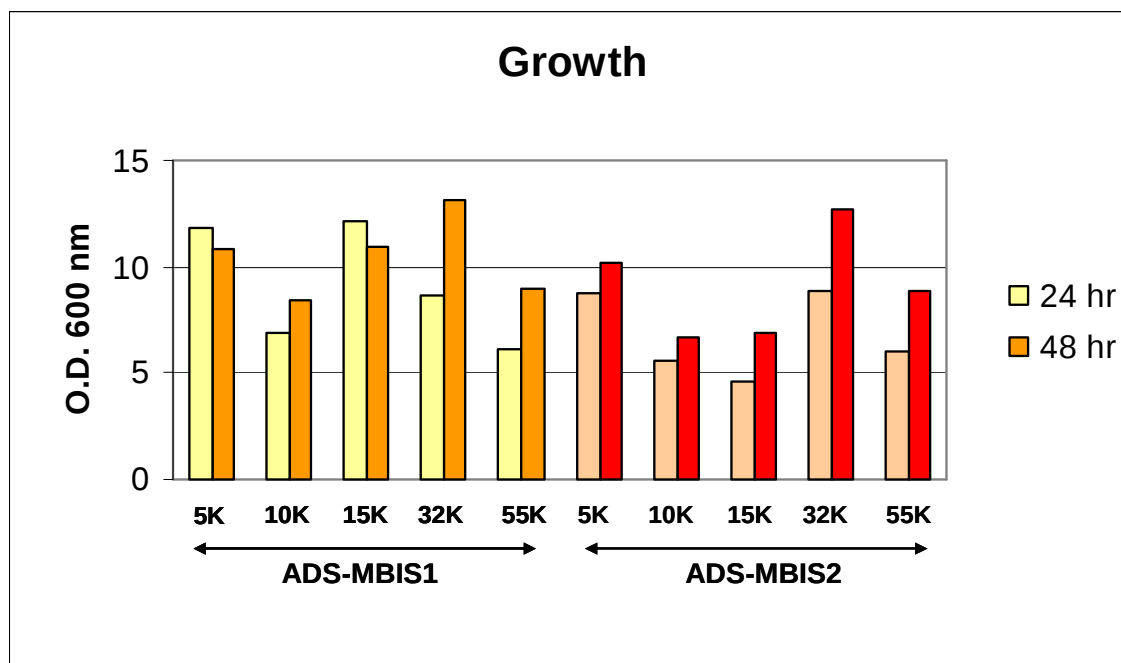


Fig. 5-8. Cell growth of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. Strains with different strength of RBS in front of ADS exhibit growth differences.

ribosome binding sites. Strains expressing ADS with ribosome binding sites of 10'000 and 50'000 activity units grow to lower ODs both at 24 hr and 48 hr time point. These ribosome binding sites could potentially be stronger than the original RBS in front of ADS and result in elevated enzyme concentration that pulls on FPP supply of the cells. As FPP is required for growth of *E. coli*, a decrease in FPP levels could have an impact on *E. coli*'s growth.

Amorphadiene production is also similar between strains expressing MBIS1 and MBIS2 operons (Figure 5-9). The RBS with 5000 activity units seems to be weaker than the original RBS and results in decreased amorphadiene levels (600 mg/L) compared to the original RBS (800 mg/L). While ribosome binding sites with 10'000 and 55'000 units of activity affect growth and result in a lower OD, there is moderate improvement in amorphadiene production compared to the original RBS (Figure 5-7 and Figure 5-9)).

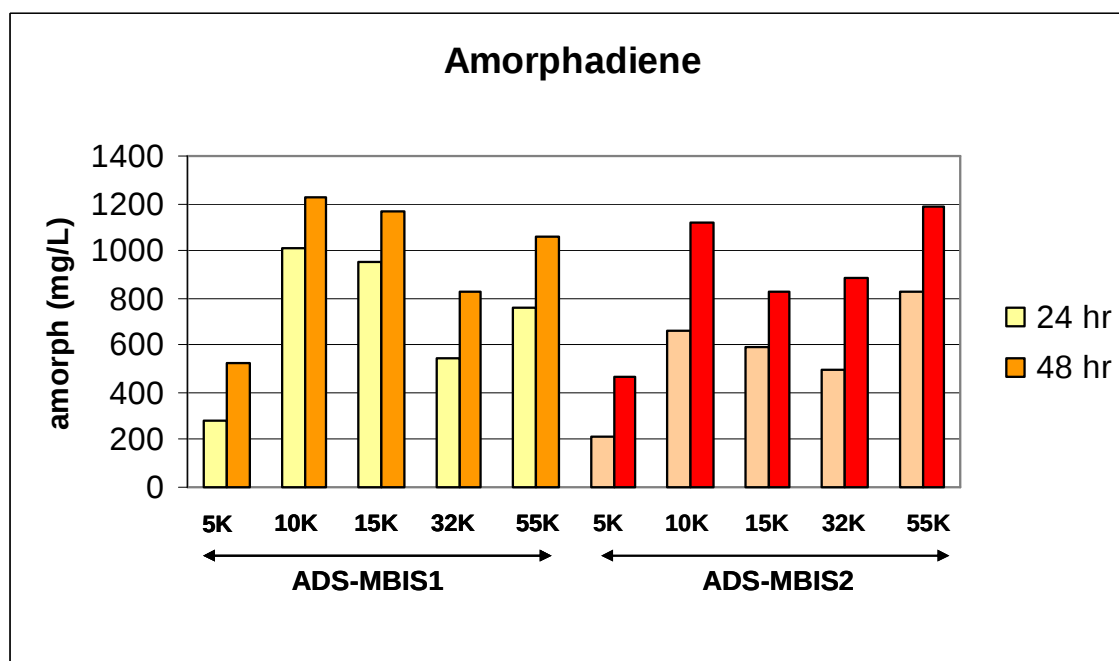


Fig. 5-9. Amorphadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. Strains with lower OD exhibit an increase in amorphadiene production compared to the strain with original RBS (800 mg/L, Figure 5-7)

There is no improvement in amorphadiene production in strains expressing MBIS2 operon even when stronger RBSs are used with amorphadiene synthase. One possibility was that there is a limitation of carbon source. The calculated maximum theoretical yield for production of amorphadiene when 1% glycerol has been used as the carbon source is approximately 3 g/L. By producing over 1g/L of amorphadiene, we are at 30% maximum theoretical yield. Taking into consideration that a significant amount of carbon is lost to CO₂ and biomass, it is not expected to have a yield much greater than 30%. To test this simple hypothesis, we increased the glycerol concentration from 1% to 3% and repeated the experiment. The strains that were studied in this experiment are listed in table 6.

Table 6: Plasmids tested with higher glycerol concentration

Strain	Description	Reference
pTrc-O.ADS-MBIS1	Original RBS in front of ADS with MBIS1	This Study
pTrc-5K.ADS-MBIS1	5000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-10K.ADS-MBIS1	10000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-O.K.ADS-MBIS2	Original RBS in front of ADS with MBIS1	This Study
pTrc-5K.ADS-MBIS2	5000 units of activity RBS in front of ADS with MBIS1	This Study
pTrc-10K.ADS-MBIS2	10000 units of activity RBS in front of ADS with MBIS2	This Study

DH1 strain of *E. coli* was transformed with pMevT and pTrc-RBS.ADS-MBIS plasmids. Cultures were induced with IPTG and samples were taken at 24 hr and 48 hr time point to look at growth and amorphaadiene production level.

The strains tested exhibit the same growth pattern in both 1% and 3% glycerol. DH1 strain expressing pTrc-O.ADS-MBIS2 shows growth inhibition at 24 hours that is relieved at 48 hour (Figure 5-6 and Figure 9). DH1 strain expressing pTrc-10K.ADS-MBIS1 and pTrc-10K.ADS-MBIS2 also exhibit slower growth rates and lower ODs compared to the control strain (pTrc-O.ADS-MBIS1).

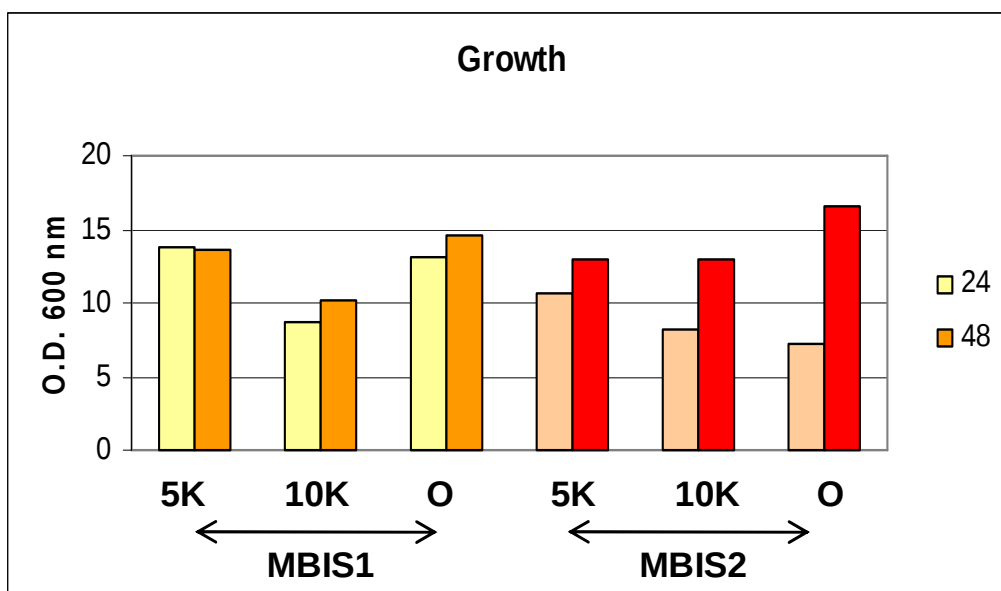


Fig. 5-10. Cell growth of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. Strains with different strength of RBS in front of ADS exhibit growth differences.

Analysis of the amorphaadiene concentration of the tested constructs however, yielded interesting results. At higher concentration of glycerol, all of the strains expressing MBIS1 operon have higher amorphaadiene yields than the strains expressing MBIS2 operon (Figure 5-11). Furthermore, *E. coli* strain expressing pTrc-O.ADS-MBIS1 produces a significant amount of amorphaadiene compared to any other strain we have ever tested. Although the amorphaadiene production for all of the strains studied here benefited from the added glycerol concentration, the improvement observed in case of pTrc-O.ADS-MBIS1 construct is not as significant for any of the other strains. By increasing carbon source, it was expected that the strains with elevated enzyme concentrations could synthesize more of the precursor metabolite FPP and convert more of this FPP to amorphaadiene. However our attempts in increasing enzyme concentrations could have lead to an imbalance in the total activity of the enzymes, resulting in accumulation of pathway intermediate and byproducts that inhibit pathway enzymes or are toxic to the cells.

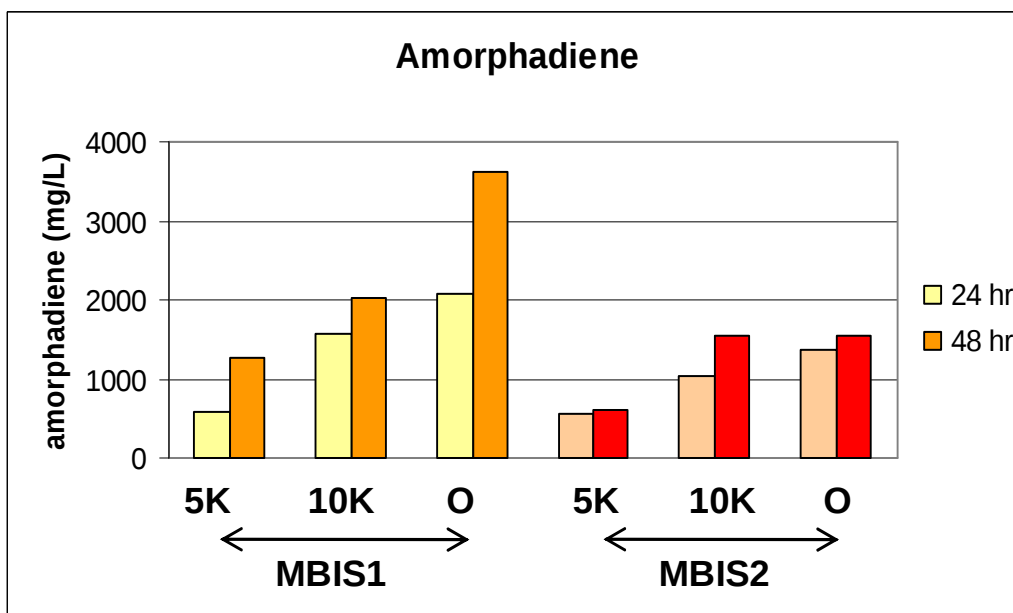


Fig. 5-11. Amorphadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction. While all the strains studied show moderate improvement in amorphadiene production when glycerol concentration is increased, the strain expressing pTrc-O.ADS-MBIS1 produces a significantly higher amorphadiene concentration.

In addition the amorphadiene concentration remained mainly unchanged between the two time points of 24 hour and 48 hour for strains expressing MBIS2 operon. This observation is more pronounced when specific production is considered (Figure 5-12). While the specific production increases from 24 hour to 48 hour for all strains expressing MBIS1 operon, it either remains the same or decreases for strains expressing MBIS2 operon. Comparing the two strains expressing pTrc-O.ADS-MBIS1 and pTrc-O.ADS-MBIS2, this claim can be most clearly observed (Figure 5-12). While the strain with pTrc-O.ADS-MBIS2 shows higher specific production at 24 hour than the strain with pTrc-O.ADS-MBIS1, there is no more amorphadiene being produced between these two time points for this strain. Based on these results, amorphadiene production appears to have been stopped after the 24-hour time point for strains expressing MBIS2 operon.

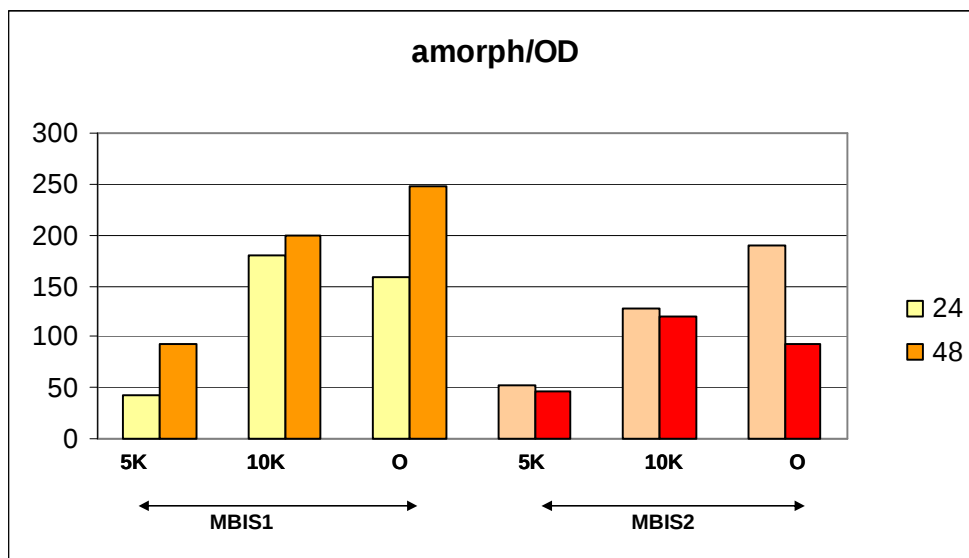


Fig. 5-12. Specific amorphadiene production of *E. coli* DH1 strains expressing pMevT and pTrc-RBS.ADS-MBIS operons 24 hour and 48 hour after induction.

Proteomics analysis: Measuring protein concentrations

To determine the actual strength of the selected ribosome binding sites, a proteomics approach was used to quantify protein levels for all of the enzymes in the pathway. By obtaining protein profiles for every enzyme in the mevalonate pathway, it could also be verified whether a decrease in enzyme concentration at 48 hour, in strains expressing MBIS2, could be impeding amorphadiene production after 24 hour time point.

E. coli DH1 harboring one of the following set of two plasmids (table 7) were incubated in TB + 3% glycerol, induced with 0.5mM IPTG at an optical density of 0.3. Two samples were taken at 24 hour and 48 hour and were prepared as explained in the Methods section.

Table 7: Plasmid sets used in this experiment:

Strain	Description	Reference
pMevT, pTrc-O.ADS-MBIS1	Original RBS in front of ADS with MBIS1	This Study
pMevT, pTrc-O.ADS-MBIS2	Original RBS in front of ADS with MBIS2	This Study
pMevT, pTrc-10.K.ADS-MBIS1	10000 units of activity RBS in front of ADS with MBIS1	This Study
pMevT, pTrc-10K.ADS-MBIS2	10000 units of activity RBS in front of ADS with MBIS2	This Study
pMevT, pTrc-5K.ADS-MBIS1	5000 units of activity RBS in front of ADS with MBIS1	This Study
pMevT, pTrc-5K.ADS-MBIS2	5000 units of activity RBS in front of ADS with MBIS2	This Study

Idi and ispA are the two enzymes of lower mevalonate pathway with the same ribosome binding site in both MBIS1 and MBIS2 operon. It is expected that when all of the expression conditions are similar for two enzymes, they both yield similar protein levels. Comparing the protein level for these two genes in strains harboring MBIS1 or MBIS2 operons, it appears that the MBIS2

operon is expressed at lower levels compared to MBIS1 at both 24 hr and 48 hr time points (Figure 5-13). MevT operon however, shows similar protein levels for all its constituents (AtoB, HmgR, and HmgS) between the strains expressing MBIS1 or MBIS2 operon at 24 hour time point. There is also a significant up-regulation of MevT operon observed at 48 hour time point in strain harboring pTrc-O.ADS-MBIS2 and not in the other strain. Another important observation is the increased MK protein level in the strain harboring pTrc-O.ADS-MBIS2 despite the general lower expression of this plasmid compared to pTrc-O.ADS-MBIS1. This confirms the preliminary assumption that the theoretical High activity ribosome binding sites selected for MK is in fact stronger and results in significant increase in protein expression. From proteomics analysis data, it is difficult to draw any conclusions in regard to the ribosome binding site strength for PMK and PMD. Such low PMK protein level makes any comparison and prediction unreliable.

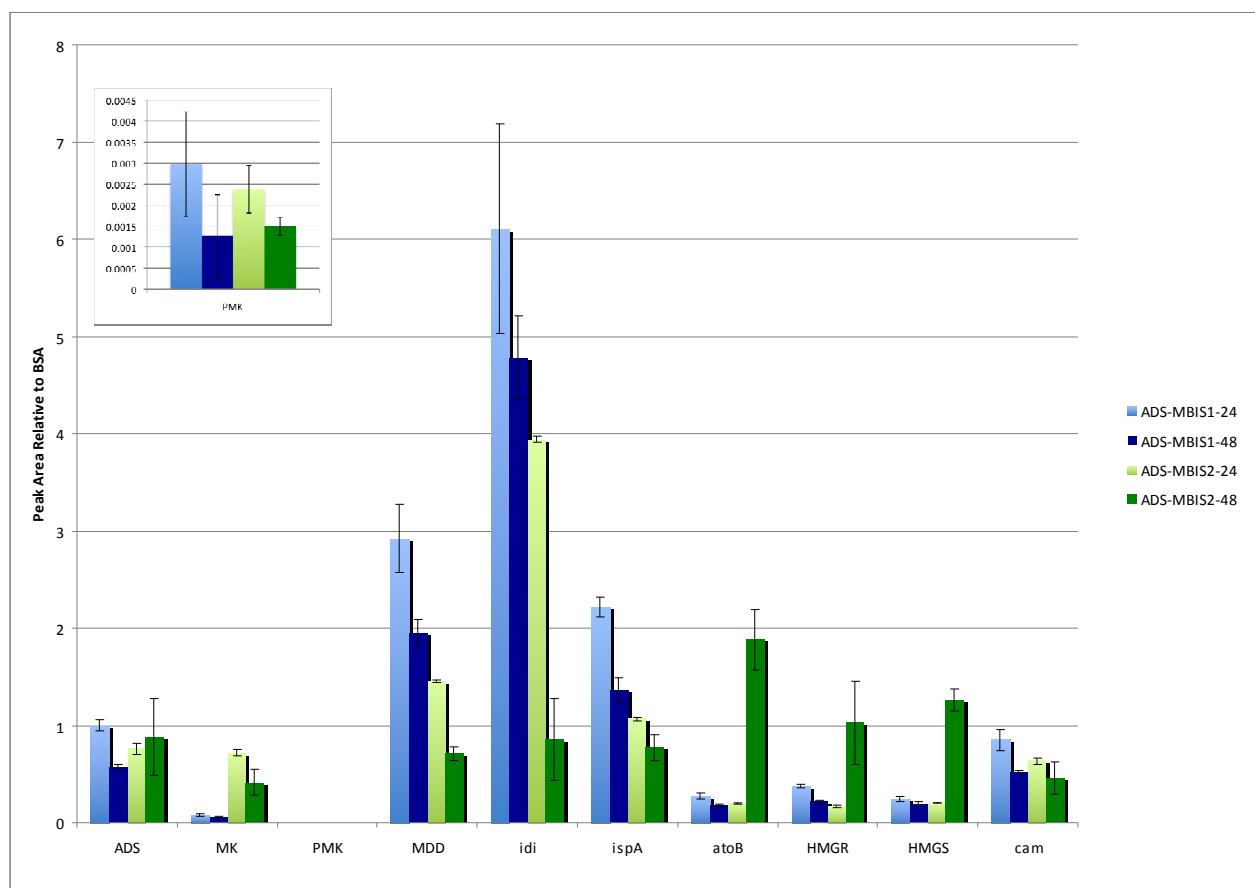


Fig. 5-13. Protein levels of all of the enzymes in the mevalonate pathway at 24 hr and 48 hr after induction. Blue bars represent DH1 strains harboring pMevT and pTrc-O.ADS-MBIS1 and green bars represent DH1 strains harboring pMevT and pTrc-O.ADS-MBIS2.

The general lower expression of pTrc-O.ADS-MBIS2 construct makes it difficult to make any statement on PMD ribosome binding site strength as well.

The proteomics analysis data for strains harboring pTrc-10K.ADS-MBIS1 and pTrc-10K.ADS-MBIS2 shows similar protein profile as the strain harboring pTrc-O.ADS-MBIS2. One objective

for studying these two strains was to compare the ribosome binding site strength of original RBS and the synthetic RBS with 10'000 units of activity. By comparing the peak area for ADS between strains harboring pTrc-O.ADS-MBIS1 and pTrc-10K.ADS-MBIS1, it can be observed that the RBS with 10'000 units of activity yields twice as much protein as the original RBS and hence is stronger.

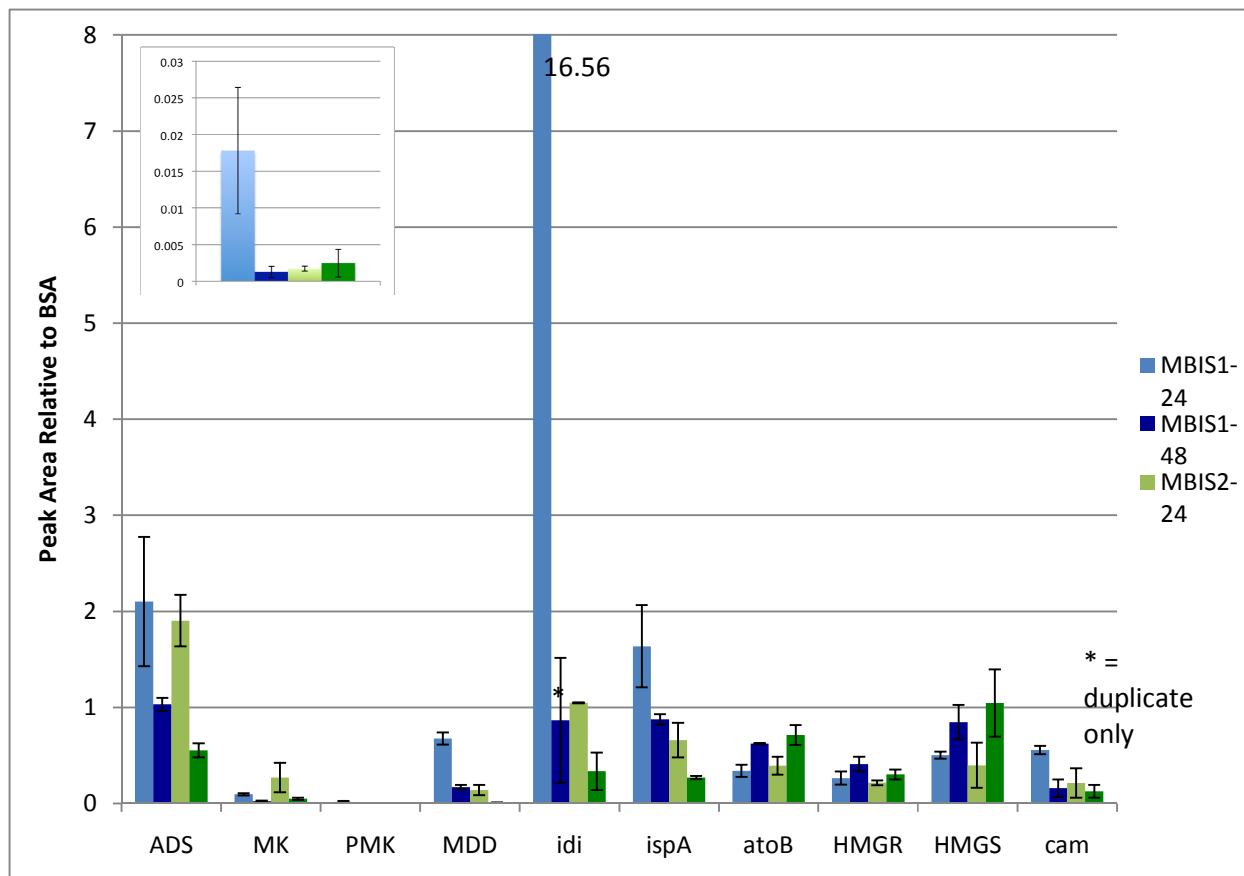


Fig. 5-14. Protein levels of all of the enzymes in the mevalonate pathway at 24 hr and 48 hr after induction. Blue bars represent DH1 strains harboring pMevT and pTrc-10K.ADS-MBIS1 and green bars represent DH1 strains harboring pMevT and pTrc-10K.ADS-MBIS2.

In case of both pTrc-10K.ADS-MBIS1 and pTrc-10K.ADS-MBIS2 the protein level for every gene in the pathway except for ADS is lower compared to constructs with original ribosome binding site in front of ADS (Figure 5-13 and Figure 5-14). The up-regulation of MevT operon at 48 hour is observed for both strains (Figure 5-14).

The growth inhibitions in cases of pTrc-O.ADS-MBIS2 and pTrc-10K.ADS-MBIS1 & 2 are potentially caused by different factors, accumulation of toxic pathway intermediates and over-consumption of cellular resources. However it can be suggested that when under stress, the cell responds by down-regulating the expression of MBIS operon. An up-regulation of MevT operon is also observed in all these cases.

Metabolomics analysis: Measuring metabolite concentrations

There is not sufficient kinetics information available on most of the enzymes in our engineered mevalonate pathway. For this reason, it is not feasible to determine the optimal protein level for each of the enzymes in the pathway. We expected to see improved amorphaadiene titers by choosing stronger ribosome binding sites that yield higher protein levels. However it appears that maintaining a balanced pathway where none of the intermediates is accumulated or depleted is the key determinant in optimizing isoprenoid biosynthesis.

By determining the concentration of intermediate metabolites, the bottlenecks in the pathway can be identified and it can be verified if accumulation of any of the intermediates is associated with toxicity and growth inhibition. These results could be employed to further optimize the isoprenoid biosynthetic pathway and provide us with a better understanding of how different parts of this pathway work together.

Growth and Intermediate metabolite concentration profiles: DH1: pMevT/ pTrc-MBIS2

As mentioned in this chapter, accumulation of FPP is toxic to the cell and causes growth inhibition. To determine the FPP concentration that is tolerable by *E. coli* without causing any growth inhibition, we transformed DH1 *E. coli* strain with pMevT and pTrc-MBIS2 plasmids. The amorphaadiene synthase gene was not expressed to allow for accumulation of FPP. We then took multiple samples at different time points during *E. coli*'s growth to monitor growth and analyze intermediate metabolite concentrations.

E. coli harboring pMevT and pTrc-MBIS2 grows at a normal rate until induction time. After induction, when the expression of pathway enzymes is initiated, the cells stop growing and exhibit a decreased optical density. It takes the cells over 10 hours to recover and restore growth (Figure 5-15).

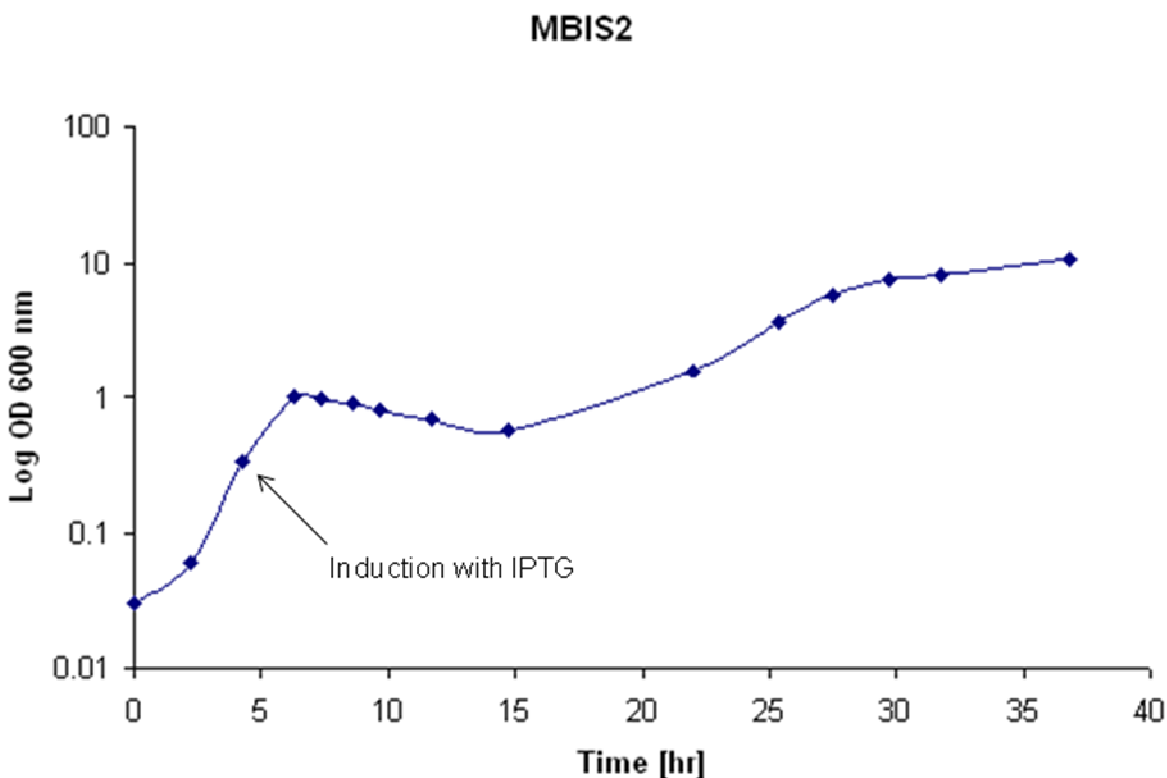


Fig. 5-15. Cell growth of *E. coli* DH1 strain expressing pMevT and pTrc-MBIS2 operons. While the cells exhibit normal growth until induction time, the cells cease to grow further for over 10 hours after induction. The growth inhibition is eventually alleviated and the cells grow to a normal final optical density.

Analyzing the samples taken at all of these time points to determine metabolite concentration, FPP and mevalonate were the only two metabolites detected (Figure 5-16). Given that the samples were pre-concentrated and considering the sensitivity of this approach, we can claim that not a significant amount of any of the other pathway intermediates is accumulated at any time point during this experiment.

By comparing the FPP concentration profile (Figure 5-16) and growth curve (Figure 5-15) of *E. coli* harboring pMevT and pTrc-MBIS2, it can be observed that the FPP production is initiated right after induction and peaks when the cells optical density begins to decline. The FPP concentration then begins to go down and by the time the cells have resumed growing again, FPP cannot be detected any more. After induction, during the time that some level of FPP accumulation occurs in the cell, there is no mevalonate build-up. As the FPP concentration diminishes and FPP becomes undetectable, mevalonate starts to accumulate. At the last time point the samples were taken no more FPP is detected by the mevalonate concentration is increasing and mevalonate has accumulate to very high levels (Figure 5-16).

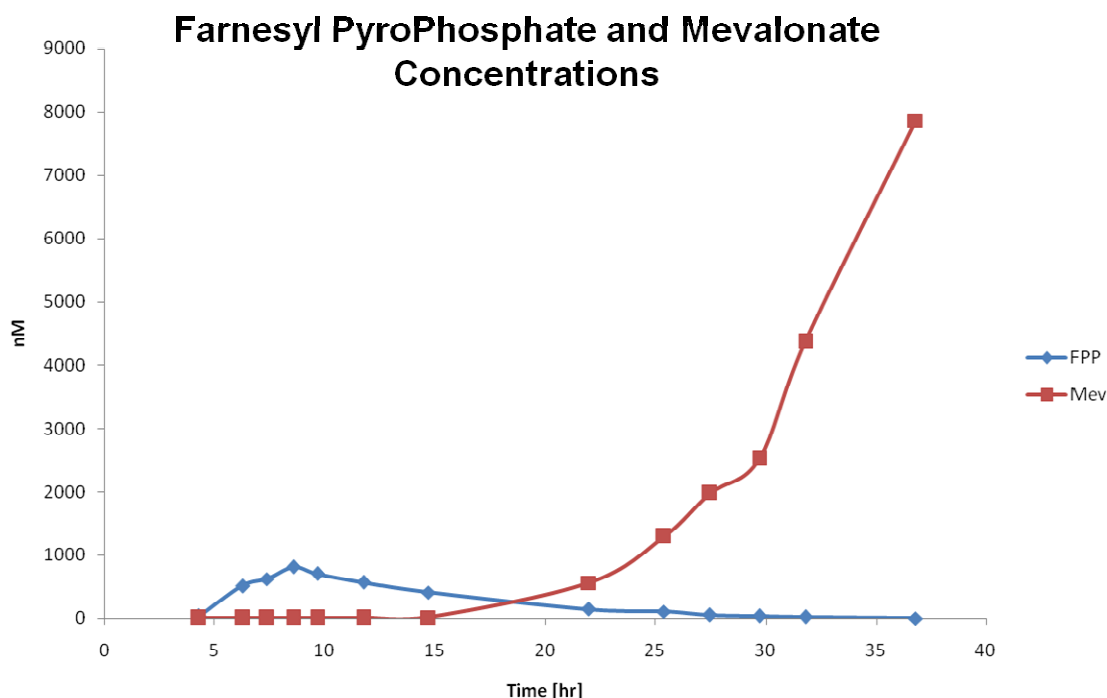


Fig. 5-16. Concentration of the intermediate metabolites in *E. coli* harboring pMevT and pTrc-MBIS2 plasmids. FPP and mevalonate are the only two metabolites detected. At earlier time points when FPP is accumulating, there is no build-up of mevalonate. As FPP levels start diminishing, mevalonate starts to accumulate and the mevalonate concentration increases to very high levels at the last time point taken.

Growth and intermediate metabolite concentration profiles:

DH1: pMevT/ pTrc-O.ADS-MBIS1 and DH1: pMevT/ pTrcO.ADS-MBIS2

It is shown here that *E. coli* cells can tolerate up to 1000 nM of FPP before the cell growth is affected and further production of FPP is prevented. In the strain studied above the amorphaadiene synthase gene was not expressed so that FPP can accumulate inside the cell. Since *E.coli* strain harboring pTrc-O.ADS-MBIS2 plasmid showed severe growth inhibition at 24 hour time point which was relieved at 48 hour, we decided to investigate the intermediate metabolite concentration profile of this strain and compare it with strain harboring pTrc-O.ADS-MBIS1.

E. coli DH1 strain was transformed with these two sets of plasmids:

pMevT/pTrc-O.ADS-MBIS1 and pMevT/pTrc-O.ADS-MBIS2

Multiple samples were taken at different time points during *E. coli*'s growth to monitor growth and analyze intermediate metabolite concentrations.

E. coli DH1 strain expressing the two plasmids pMevT and pTrc-O.ADS-MBIS1 exhibits normal growth rate and does not show any signs of growth inhibition (Figure 5-17). At the early time points *E. coli* DH1 strain expressing the two plasmids pMevT and pTrc-O.ADS-MBIS2 grows similarly to the strain harboring pTrc-O.ADS-MBIS1, however about 10 hours after induction the growth rate slows down considerably. The growth inhibition is later alleviated and the growth rate increases to normal levels at later time points. At 40 hour after inoculation both strains expressing pTrc-O.ADS-MBIS1 and pTrc-O.ADS-MBIS2 grow at a similar rate (Figure 5-17)

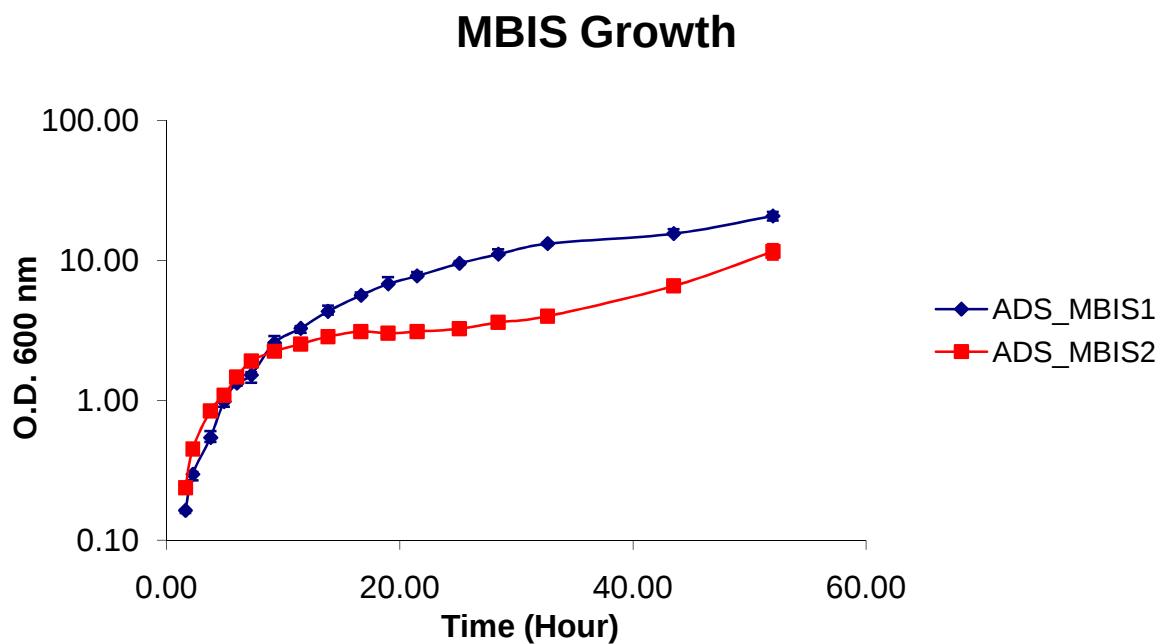
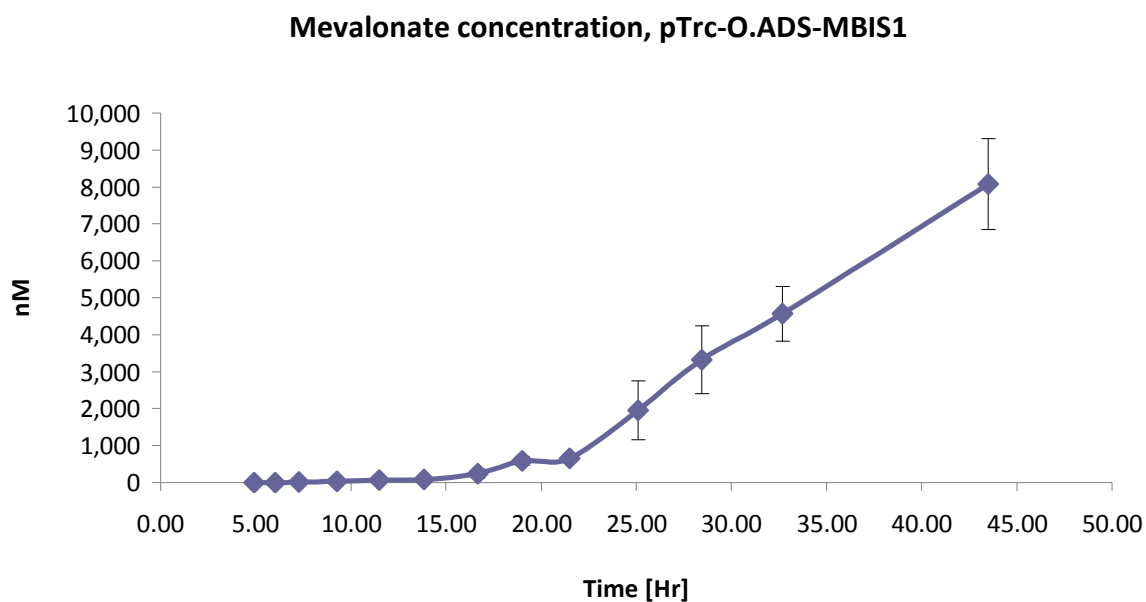


Fig. 5-17. Cell growth of *E. coli* DH1 strains expressing pMevT / pTrcO.ADS-MBIS1 and pMevT / PTrc-O.ADS-MBIS2 plasmids. Until 10 hour time point both strains grow at a normal rate, however the growth rate slows down for strain harboring pTrc-O.ADS-MBIS2. Later in the experiment the growth inhibition is finally alleviated and growth is restored.

A.



B.

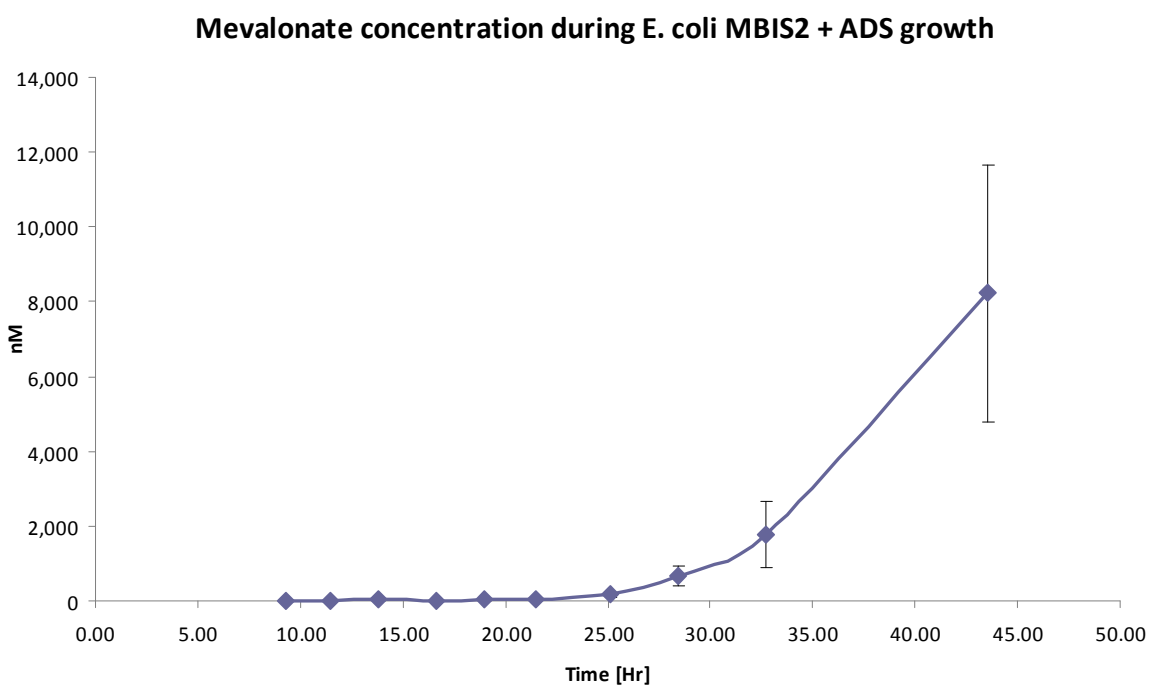
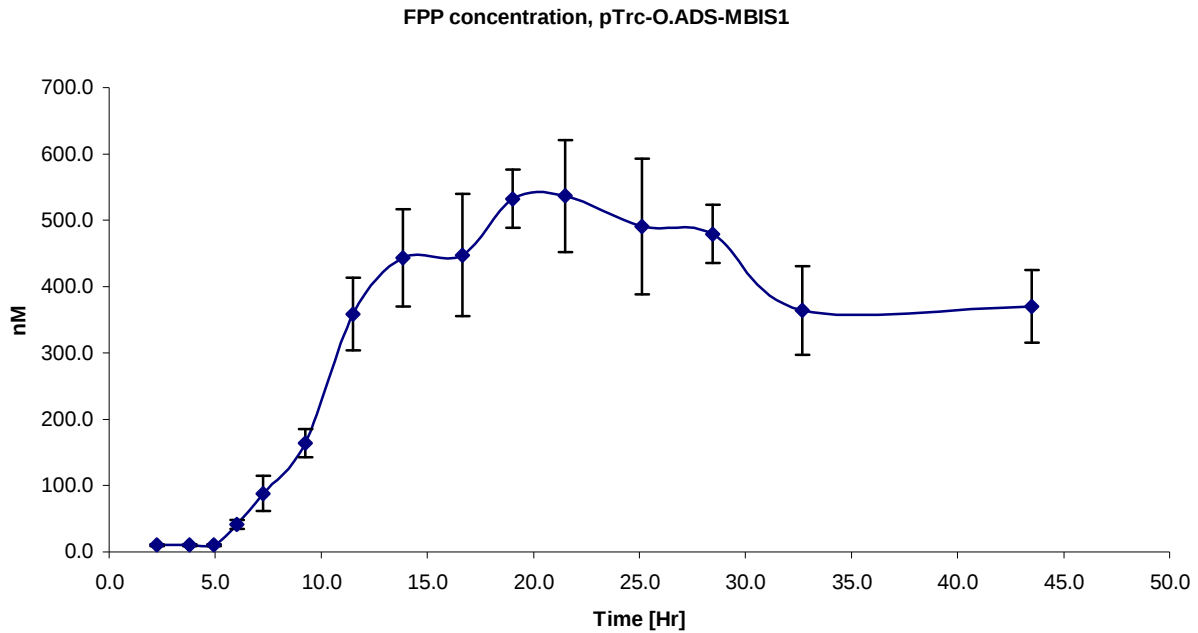


Fig. 5-18. Mevalonate concentration profile in DH1: pMevT / pTrc-O.ADS-MBIS1 (A) and DH1: pMevT / pTrc-O.ADS-MBIS2 (B) At final time point strain expressing MBIS2 operon produces higher mevalonate titers.

A.



B.

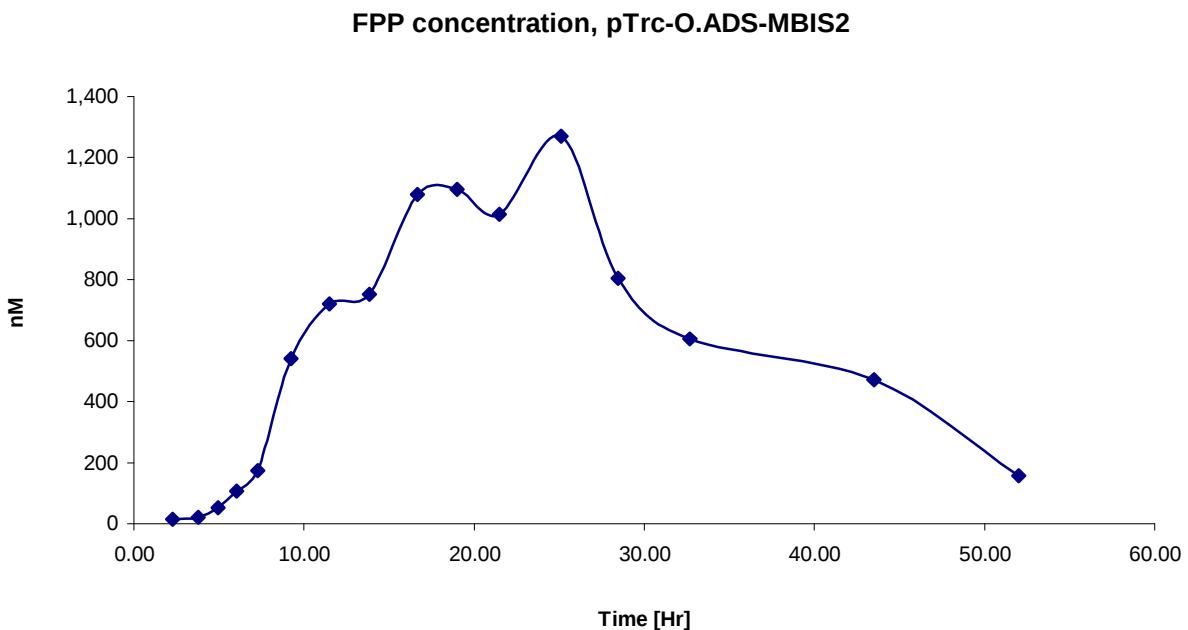


Fig. 5-19. FPP concentration profile in DH1: pMevT/ pTrc-O.ADS-MBIS1 (A) and DH1: pMevT / pTrc-O.ADS-MBIS2 (B). The strain expressing MBIS2 operon has a higher peak FPP Level compared to the strain expressing MBIS1 operon. While the strain expressing MBIS1 operon maintains a steady FPP concentration, the FPP concentration in strain expressing MBIS2 operon diminishes over time to reach an insignificant amount.

The mevalonate concentration profile for these two strains is shown in Figure 5-18. Both strains start to accumulate mevalonate in the late exponential phase. The internal mevalonate concentration is very similar between the two strains; however mevalonate is secreted out of the cell and accumulates in the culture at significant concentrations. In future experiments both intracellular and extracellular mevalonate concentrations should be determined to provide a basis for comparison of mevalonate production between the two strains.

The FPP concentration profile for these two strains is shown in Figure 5-19. DH1 strain expressing pMevT and pTrc-O.ADS-MBIS1 accumulates FPP to a maximum concentration of 500 nM (Figure 5-19.A), the FPP concentration then decreases and the cells maintain a steady FPP concentration of 400 nM. On the other hand, DH1 strain expressing pMevT and pTrc-O.ADS-MBIS2 accumulates FPP to a higher maximum concentration of 1200 nM. In this strain, FPP accumulates inside the cell during the period the growth rate has slowed down and cellular growth is inhibited. When the growth inhibition is alleviated and the normal growth rate is restored, the FPP concentration starts to decline and in the sample taken at final time point only a negligible amount of FPP is detected.

Discussion and Future Direction

Through optimization of the heterologous mevalonate pathway in *E. coli*, we have found that over-expression of the first two genes of lower mevalonate pathway leads to higher amorphaadiene titers. Also a significant amount of mevalonate appeared to be accumulating in the culture. These observations lead us to conclude that despite high titers of amorphaadiene produced by our engineered strains, the optimal balance of metabolites has not been obtained.

To address this flux imbalance and further optimize the lower mevalonate pathway, enzyme levels can be synthetically modified through transcriptional or translational control mechanisms. In the previous chapter various expression constructs were tested to both analyze pathway bottlenecks and optimize production of amorphaadiene. However, this approach suffers from several limitations. It is not feasible to predict transcriptional output as a function of promoter strength or plasmid copy number. Also the same control mechanism would apply to all of the genes in the pathway and it is not an option to increase the expression of one gene while decreasing the expression of another. In addition, all of the genes of lower mevalonate pathway are expressed from the same operon, under the same promoter and are located on the same transcript. It is most probable that the limitation is at translation level and not transcription. Therefore we took the approach of controlling expression level through the use of ribosome binding sites of different strength.

Different MBIS constructs that were studied in this work yielded different production levels and growth rates. The first MBIS construct that showed improved amorphaadiene titers without affecting growth was the construct that had a High activity ribosome binding site in front of the second gene of the lower mevalonate pathway (*PMK*) and Medium activity ribosome binding sites in front of all of the other genes (pTrc-ADS-MBIS5, Figure 5-7). By proteomics analysis of the enzymes of mevalonate pathway, we have shown that the protein level for *PMK* enzyme is so low that it is hardly detectable (Figure 5-13). Although we have not been able to determine the actual strength of the ribosome binding sites in front of *PMK* through proteomics analysis, the

High activity RBS appears not to form as strong RNA secondary structures as the Medium activity RBS and could potentially be slightly stronger than the Medium activity RBS. Considering the negligible PMK enzyme concentration, even slight improvement in protein level for this enzyme could potentially lead to higher flux through mevalonate pathway and improved amorphaadiene titers.

Through proteomics analysis data we have sufficient evidence to claim that the High activity RBS in front of MK is stronger than the Medium activity RBS. The High activity RBS results in 3 to 10 folds more of the MK protein compared to the Medium activity RBS (Figure 5-13 and Figure 5-14). As mentioned earlier, strains expressing MBIS2 operon consistently exhibit lower protein levels than strains expressing MBIS1 operon even for the genes with similar ribosome binding sites between the two operons. Taking into account the general down-regulation of the MBIS2 operon, the actual strength of the High activity RBS could be much greater than the Medium activity RBS. However, this increase in protein level for MK appears to inhibit growth and affect amorphaadiene production. The strains expressing the construct with High activity RBS in front of the first three genes of MBIS operon (pTrc-ADS-MBIS2) exhibit severe growth inhibition at 24 hour and lower production levels compared to the control strain expressing pTrc-ADS-MBIS1. It was assumed that the increased enzyme concentration directs more carbon flux through mevalonate pathway and leads to an increased FPP concentration which in turn can inhibit cellular growth. We expected to restore growth and increase amorphaadiene production even further by increasing ADS protein level.

A ribosome binding site calculator was used to select 5 ribosome binding sites of different strengths that were specifically designed by taking into account the sequence of ADS gene. At least two of these ribosome binding sites (10'000 and 55'000 units of activity) resulted in higher amorphaadiene titers when compared to the control strain DH1: pMevT/ pTrc-O.ADS-MBIS1 when cultured in TB + 1% glycerol (Figure 5-9). On the other hand, the strains expressing pTrc-10K.ADS-MBIS1 and pTrc-10K.ADS-MBIS2 both grow to much lower optical densities compared to the control strain (Figure 5-6 and Figure 5-8). The only difference between pTrc-O.ADS-MBIS1 and pTrc-10K.ADS-MBIS1 constructs is the RBS in front of ADS. Our initial assumption that the RBS with 10'000 activity units is stronger than the original RBS was later confirmed by proteomics analysis of these two strains. The higher amorphaadiene synthase protein level can deplete FPP supply of the cell and lead to growth inhibition. When the carbon source is increased to 3% glycerol, none of the strains with stronger ribosome binding sites and higher protein levels ends up to be the best producer. In fact, the strain with all Medium Activity ribosome binding sites, pTrc-O.ADS-MBIS1 (Figure 5-11), exhibits a significant improvement in amorphaadiene production (3.6 g/L). This demonstrates the significance of the balanced expression of the heterologous mevalonate pathway even further. While the strain with all High Activity ribosome binding sites produces more FPP, this FPP accumulation only results in growth inhibition and yields lower amorphaadiene titers.

From the metabolomics analysis, it was determined that *E. coli* cells can accumulate FPP up to 1000 nM before their growth is inhibited and the cells find a way to regulate or shut down FPP production (Figure 5-16). When the intermediate metabolite concentration profile is determined and the FPP concentration between the two strains expressing pTrc-O.ADS-MBIS1 and pTrc-O.ADS-MBIS2 is compared, the growth profile for these two strains can be explained.

The *E. coli* strain expressing pTrc-O.ADS-MBIS1 accumulates FPP to a peak concentration of approximately 500 nM and then stabilizes at a concentration of 350 nM at stationary phase (Figure 5-18). This concentration of 500 nM is below the FPP threshold concentration of 1000

nM and is well-tolerated by the cell. This strain grows at a normal growth rate and yields highest amorphaadiene production ever being reported in shake flask.

The *E. coli* strain expressing pTrc-O.ADS-MBIS2 accumulates FPP to a peak concentration of approximately 1200 nM, the FPP level then decreases to a negligible amount. This concentration of 1200 nM is above the FPP threshold concentration of 1000 nM and is not well-tolerated by the cell (Figure 5-19). This strain grows at a normal growth rate at early exponential phase during which the FPP concentration is well below 1000 nM. As FPP concentration increases, the growth rate significantly slows down. Eventually the growth is restored as the FPP level starts to go down. It appears that in an attempt to control FPP concentration at below the cell's threshold concentration, FPP production is shut down by the cell.

It appears that our attempts in increasing the flux through the lower mevalonate pathway have been successful: We showed that the MK protein level in MBIS2 construct has significantly increased despite general down-regulation of this operon and the strain expressing this operon produce and accumulate more FPP. We have also successfully increased the amorphaadiene synthase enzyme level. We hoped that by providing more of the ADS enzyme we could take advantage of this increased flux and convert the excess FPP to more amorphaadiene. However while both parts have yielded the desired outcome, the combined effect is not predictable. These heterologous enzymes work independently and may not have been designed to work together.

If the main obstacle is coordinating these enzymes and facilitate substrate channeling, the scaffolding approach²³ could be taken to co-localize these enzymes and prevent the diffusion of toxic intermediates.

We focused on changing the RBS strength of only the first three genes of the lower mevalonate pathway and ADS. However the optimal concentration of each of the enzymes of the lower mevalonate pathway is not known. By taking advantage of the RBS calculator¹⁹, ribosome binding sites with different strengths can be designed for every gene in the pathway. With the advent of higher throughput cloning and screening methods, different combinations of RBS strength can be tested for all the genes in the pathway to achieve optimal protein and metabolite balance.

We have standardized proteomics and metabolomics approaches for metabolic engineering of the isoprenoid pathway. We have also revealed some unknown aspects of the mevalonate pathway that can undoubtedly be helpful in further optimization of this pathway. We hope that the methods developed and knowledge gained during the course of these efforts will facilitate production of isoprenoid pharmaceuticals and other valuable compounds in *E. coli* in the future.

To sum up the major accomplishments of this work:

- We developed a mathematical model and all the experimental tools to determine the intracellular fluxes during exponential phase of *E. coli*'s growth. This model was expected to be used to further engineer our host for increased production of amorphaadiene. However it was later determined that the majority of amorphaadiene production occurs at stationary phase and no significant difference in production was observed among different strains producing different concentrations of amorphaadiene at exponential phase. The model in the other hand is only applicable to exponential growth phase and cannot be used at stationary phase. We also determined that we are most limited by our ability to express mevalonate pathway in *E. coli* and not by *E. coli*'s ability to provide the

substate and energy requirements of mevalonate pathway. Therefore cellular engineering of *E.coli* most likely will not result in a significant increase in amorphaadiene production.

- To overcome our limitations in optimal expression of mevalonate pathway, we first determined the bottlenecks of this pathway. We identified the two enzymes of lower mevalonate pathway, mevalonate kinase (MK) and phosphomevalonate kinase (PMK), as bottlenecks of mevalonate pathway. Amorphaadiene synthase was also shown to be limiting production of amorphaadiene from FPP.
- By using a High Activity RBS in front of MK, the flux through lower mevalonate pathway was increased and that resulted in higher intracellular concentration of FPP. However that resulted in growth inhibition and negatively affected amorphaadiene production. That is best explained by limitation of ADS is conversion of FPP to amorphaadiene.
- By using a RBS of moderate strength (Medium Activity) we achieved optimal growth and highest amorphaadiene production we have ever observed in shake flasks. This resulted in at least 3-fold increase in overall yield compared to our best engineered strain in chapter 4.
- We have developed a general approach for troubleshooting the production of secondary metabolites and expression of non-native pathways in *E.coli*. The tools developed through out this work are used and will be used to engineer biofuel-producing strains of *E.coli* at JBEI.

References

1. Fortman, J. L., Chhabra S., Mukhopadhyay A., Chou H., Lee T.S., Steen E., Keasling J.D. . Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol.* 26, 375–381 (2008)
2. Peralta-Yahya P.P., Keasling J.D. Advanced biofuel production in microbes. *Biotechnol. J.* (2010)
3. Steen E.J., Kang Y., Bokinsky G., Hu Z., Schirmer A., McClure A., Del Cardayre S.B., Keasling J.D. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature.* 463, 559-562 (2010)
4. Stephanopoulos, G. Challenges in engineering microbes for biofuels production. *Science* **315**, 801–804 (2007).
5. 2. Nakamura, C.E. & Whited, G.M. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **14**, 454–459 (2003).
6. 3. Khosla, C. & Keasling, J.D. Metabolic engineering for drug discovery and development. *Nat. Rev. Drug Discov.* **2**, 1019–1025 (2003).
7. Kizer, L., Pitera, D.J., Pfleger, B. & Keasling, J.D. Functional genomics for pathway optimization: application to isoprenoid production. *Appl. Environ. Microbiol* **74**, 3229–3241 (2008).
8. Zhu, M.M., Lawman, P.D. & Cameron, D.C. Improving 1,3-propanediol production from glycerol in a metabolically engineered *Escherichia coli* by reducing accumulation of sn-glycerol-3-phosphate. *Biotechnol. Prog.* **18**, 694–699 (2002).
9. Barbirato, F., Grivet, J.P., Soucaille, P. & Bories, A. 3-Hydroxypropionaldehyde, an inhibitory metabolite of glycerol fermentation to 1,3-propanediol by enterobacterial species. *Appl. Environ. Microbiol.* **62**, 1448–1451 (1996).

10. Harcum, S.W. & Bentley, W.E. Heat-shock and stringent responses have overlapping protease activity in *Escherichia coli*. Implications for heterologous protein yield. *Appl. Biochem. Biotechnol.* **80**, 23–37 (1999).
11. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., Keasling, J. D. Engineering the mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796-802 (2003).
12. Kim S.W., Keasling J.D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol. Bioeng.* **72**, 408-415 (2001)
13. Matthews, P.D. & Wurtzel, E.T. Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Applied Microbiology and Biotechnology* **53**, 396-400 (2000).
14. Harker, M. & Bramley, P.M. Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. *FEBS Lett* **448**, 115-119 (1999).
15. Kajiwarra, S., Fraser, P.D., Kondo, K. & Misawa, N. Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. *Biochemical Journal* **324**, 421-426 (1997).
16. Anthony J.R., Anthony L.C., Nowroozi F., Kwon G., Newman J.D., Keasling J.D. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab. Eng.* **11**, 13-19 (2009).
17. Pfleger B.F., Fawzi N.J., Keasling J.D. Optimization of DsRed production in *Escherichia coli*: effect of ribosome binding site sequestration on translation efficiency. *Biotechnol. Bioeng.* **92**, 553-558 (2005).
18. Hui A., Hayflick J., Dinkelspiel K., de Boer H.A. Mutagenesis of the three bases preceding the start codon of the beta-galactosidase mRNA and its effect on translation in *Escherichia coli*. *EMBO. J.* **3**, 623-629 (1984).
19. Salis H.M., Mirsky E.A., Voigt C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **10**, 946-950 (2009).
20. Shetty R.P., Endy D., Knight T.F. Jr. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **14**, 2-5 (2008).
21. Hanahan, D. Studies of transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580 (1983).
22. Newman, J. D., Marshal, J. L., Chang, M. C., Nowroozi, F., Paradise, E. M., Pitera, D. J., Newman, K. L., Keasling, J. D. High-level production of amorpha-4,11-diene in a two phase partitioning bioreactor of metabolically engineered *E. coli*. *Biotechnol. Bioeng.* **95**, 684-691 (2006).
23. Dueber J.E., Wu G.C., Malmirchegini G.R., Moon T.S., Petzold C.J., Ullal A.V., Prather K.L., Keasling J.D. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753-759 (2009).

Appendix 1

Mathematical Analysis of Flux Elucidation and results

Mathematical Analysis of Flux Elucidation

In the context of metabolic flux analysis (MFA) the description of a metabolic network requires the definition of the following sets, variables, and parameters.

Sets:

$I = \{i\}$ set of metabolites

$J = \{j\}$ set of reactions

$J^R = \{j\}$ set of reversible reactions

$I^F = \{i\}$ Metabolites present in growth medium

$I^E = \{i\}$ Metabolites that can cross cell boundaries

Parameters:

S_{ij} = stoichiometric matrix

Variables:

v_j = fluxes

b_i = exchange fluxes

The set I contains all of the metabolites present while set J enumerates all reactions composing the metabolic network. S_{ij} is the stoichiometric coefficient of metabolite i in reaction j . v_j quantifies the rate of reaction j , and b_i is the rate of transport (active or passive) of metabolites i across cellular boundaries. Reversible reactions are replaced by the difference of the corresponding pair of exchange reactions thus maintaining positivity of all reaction steps present in the model:

$$v_j = v_j^{\text{forward}} - v_j^{\text{backward}} \quad \forall j \in J^R \quad (0)$$

Using the principle of stoichiometric analysis along with the application of a pseudo-steady-state hypothesis to the intracellular metabolites [Vallino, 1993 #63], an overall flux balance can be written as follows:

$$\sum_j S_{ij} \cdot v_j = b_i \quad \forall i \in I \quad (1)$$

Metabolites that do not exchange across the cell's boundaries have no exchange flux. That is, $b_i = 0, \forall i \notin \text{exchange}$. Positive or negative values for b are allowed when a metabolite can enter or leave the cellular boundaries, respectively.

When ^{13}C substrate labeling is introduced, an additional layer of detail is needed to fully characterize the network. This information includes the substrate(s) labeling patterns and descriptions of the fate of the carbon atoms in each reaction. We express the labeling patterns using the concepts of isotopomer distribution vectors (IDVs) [Schmidt, 1997 #70; Wittmann,

2002 #36] and isotopomer mapping matrices (IMMs) (see subsection 2.2) that describe transitions of IDVs after each single-reaction step [Schmidt, 1997 #70]. The following additional sets, parameters and variables are required for the mathematical quantification.

Sets:

$K = \{k\}$ set of isotopomers

$N = \{n\}$ the number of carbon atoms

Parameters:

$IMM_{i' \rightarrow i, k' \rightarrow k}^j$ = isotopomer mapping matrix

Variables:

I_{ik} = isotopomer distribution vector

The set K enumerates all possible labeling patterns for a given metabolite. The fraction that a particular k isotopomer makes up of all the labeling patterns of a metabolite i , is given by variable I_{ik} . There are at most 2^n isotopomers for each metabolite containing n carbons, and the sum of all I_{ik} is unity for metabolite i . Parameter IMM links the specific isotopomers k' of reactant i' that contribute to the formation of product i in isotopomer form k through reaction j . The corresponding entry for such an indices combination is equal to one unless it refers to a symmetric molecule i' (e.g., succinate) for which the IMMs have a fractional entry to account for the fact that an isotopomer k' may map to more than one isotopomer k in the product molecule i . Given the above definitions the mass balance for every metabolite i in isotopomer form k can be written as follows:

$$I_{ik} \cdot b_i = \sum_{j|S_{ij}>0} \left(S_{ij} \cdot v_j \cdot \prod_{i' \in S_{ij}} \sum_{k'} IMM_{i' \rightarrow i, k' \rightarrow k}^j \cdot I_{i'k'} \right) \quad \forall i \in I, k \in K \quad (2)$$

$$+ \sum_{j|S_{ij}<0} S_{ij} \cdot v_j \cdot I_{ik}$$

Figure 3-5 pictorially illustrates the origin of each term present in the isotopomer balance equation. The balance equation as a mixing/splitting node where all generation terms through different reactions j of metabolite i in isotope form k are aggregated and then channeled through to all consuming reactions at a fixed isotopomer fraction. Note that the generation term gives rise to nonlinear terms. Unimolecular reactions yield products of metabolic fluxes times isotopomer fractions (bilinear terms) while bimolecular reactions contribute trilinear terms due to the presence of two separate isotopomer fractions in the product. Examples of these bi- and trilinear terms and the reactions that yield them are shown in Figure 3-5 on the left-hand side. Thus, the isotopomer balance in Eq. (2) can be used to calculate the I_{ik} of each intracellular metabolite in terms of the fluxes, the IMMs, and isotopomer distribution in the feedstock [Schmidt, 1997 #70].

From an experimental standpoint, techniques such as mass spectrometry can be used to obtain isotope labeling data. However, instead of providing isotopomer distribution I_{ik} of a compound directly, mass spectrometry provides data that can be written as mass distribution vectors

(MDV). Each member of the MDV contains a group of isotopomers that all have the same mass [Wittmann, 1999 #45]. That is, it contains information only about the total number of labeled carbons in a fragment generated by the mass spectrophotometer, but not their specific location. Linking the isotopomer distribution vectors and the mass distribution vectors requires the following additional definitions.

Sets:

$F = \{f\}$ set of fragments

$M = \{m\}$ set of mass fractions

$R = \{r\}$ set of replicants

$I^M = \{i\}$ set of measured metabolites

Parameters:

$IDVMDV_{f,k \rightarrow m}^i$ = isotopomer grouping matrix

Variables:

$MDV_{f,m,r}^i$ = mass distribution vector

The set F contains the fragments of a measured metabolite i in set I^M , the set M represents the mass fractions observed for a fragment f , and the set R represents each replicates of the measurement, as described above in the experimental section. Parameter $IDVMDV_{f,k \rightarrow m}^i$ links the specific isotopomers k in the I_{ik} of metabolite i that contribute to the mass distribution m of the corresponding MDV of the fragment f of metabolite i . That is, for a given fragment, all of the isotopomers that have the same number of labeled carbon atoms in them would be grouped together into the $MDV_{f,m,r}^i$.

Using the notation listed above, the problem of calculating the fluxes and labeling patterns for a given set of GC-MS data is formulated as a nonlinear optimization problem (FluxCalc).

$$\text{minimize } z = \left(\sum_{i,f} \sum_{m,r} \left(MDV_{f,m,r}^{i,exp} - MDV_{f,m}^{i,sim} \right)^2 \right)^{\frac{1}{2}} \quad (\text{FluxCalc}) \quad (3)$$

Subject to

$$\sum_j S_{ij} \cdot v_j = b_i \quad \forall i \in I \quad (1)$$

$$I_{ik} \cdot b_i = \sum_{j|S_{ij}>0} \left(S_{ij} \cdot v_j \cdot \prod_{i' \in S_{ij}} \sum_{k'} IMM_{i' \rightarrow i, k' \rightarrow k}^j \cdot I_{i'k'} \right) \quad \forall i \in I, k \in K \quad (2)$$

$$+ \sum_{j|S_{ij}<0} S_{ij} \cdot v_j \cdot I_{ik}$$

$$MDV_{f,m}^{i,sim} = \sum_k IDVMDV_{f,k \rightarrow m}^i \cdot I_{ik} \quad \forall m \in M, f \in F, i \in \text{measured} \quad (4)$$

$$\nu_j \geq 0, \quad \text{LB}_j \leq \nu_j \leq \text{UB}_j \quad \forall j \in J \quad (5)$$

$$0 \leq I_{ik} \leq 1 \quad \forall i \in I, k \in K, \sum_k I_{ik} = 1 \quad \forall i \in I \quad (6)$$

In (FluxCalc), the objective function z in Eq. (3) is the sum of the discrepancies between the experimental data (*exp*) and the simulated values (*sim*) for the MDVs of the measured amino acid fragments. Note that the simulated MDVs are the same for all replicates r . Eq. (1) enforces mass balance on metabolites and Eq. (2) does the same for individual isotopomers. The mapping of MDV onto I_{ik} is performed by Eq. (4). Eq. (5) constrains all the fluxes to be positive, and further restricts them between lower and upper bounds. Eq. (6) enforces that the I_{ik} remains between zero and one positive values and that, as mass fractions, their sum is unity for each metabolite i .

Given the nonconvex nature of (FluxCalc), only local optima solutions can be obtained. In order to enumerate multiple local optima, the additional formulation (FluxInit) was constructed to generate a set of random feasible flux distributions that are used as initial conditions for the solution of the NLP problem. (FluxInit) essentially minimally perturbs an original randomly chosen flux distribution so as to satisfy all metabolite balances. The resulting linear optimization problem is as follows.

$$\text{minimize } z = \sum_j e_j \quad \forall j \in J \quad (\text{FluxInit}) \quad (7)$$

Subject to

$$\sum_j S_{ij} \cdot \nu_j = b_i \quad \forall i \in I \quad (1)$$

$$\nu_j^{rand} - \nu_j \geq e_j \geq \nu_j - \nu_j^{rand} \quad \forall j \in J \quad (8)$$

$$\nu_j \geq 0, \quad \text{LB}_j \leq \nu_j \leq \text{UB}_j \quad \forall j \in J \quad (5)$$

Here, Eq. (7) minimizes the variable e_j that are the distances that are set in Eq. (8). Eq. (7) is a linear form of the absolute value between ν_j and the parameter ν_j^{rand} which is a uniform distribution of random fluxes. The fluxes are constrained as in (FluxCalc) by Eq. (6). The overall mass balance is applied as before by Eq. (1).

As indicated, the formulation (FluxCalc) yields ν_j , $MDV_{m,f}^{sim}$, and I_{ik} . The overall solution procedure is listed in a step-wise manner below.

Step 1: Populate model parameters parameters S_{ij} , $IMM_{i' \rightarrow i, k' \rightarrow k}^j$, $IDVMDV_{f, k \rightarrow m}^i$, $MDV_{m, f, r}^{i, exp}$, $I_{i \in feedstock, k}$ UB_j and LB_j based on the experimental system setup and metabolic network abstraction.

Step 2: Initialize v_j^{rand} using a random number generator to construct a uniform distribution such that $LB_j < v_j^{rand} < UB_j \quad \forall j$.

Step 3: Solve (FluxInit) to determine an initial feasible flux distribution $\rightarrow v_j$

Step 4: Initialize I_{ik} setting $I_{i,k=1} = 1, \forall i \notin \text{feedstock}$ $I_{i,k>1} = 0, \forall i \notin \text{feedstock}$.

Step 5: Solve the nonlinear optimization problem (FluxCalc) to generate the solution v_j , I_{ik} and $MDV_{m,f}^{sim}$.

(FluxCalc) is solved multiple times using CONOPT version 3 accessed within the GAMS modeling environment. The initial flux distributions are provided by solving (FluxInit) using CPLEX version 10, also accessed within GAMS.

Flux Distribution Results

Abbreviation ^b	Subsystem	Relaxed model ^a		Full model		Reduced model		Basic model	
		min ^c	max	min	max	min	max	min	max
ICL	Anaplerotic reactions	0	12.32	0	1.00	0	1.00	0	0
MALS	Anaplerotic reactions	0	12.32	0	1.00	0	1.00	0	0
ME1x	Anaplerotic reactions	0	75.90	0	4.12	0	3.83	0	0
PPA	Anaplerotic reactions	0.78	129.43	0.78	108.15	0.78	80.61	0.78	109.28
PPC	Anaplerotic reactions	0	130.07	1.09	6.77	1.11	5.46	1.42	5.58
PPCK	Anaplerotic reactions	0	128.65	0	5.36	0	4.04	0	4.10
ACONT	Citrate Cycle (TCA)	0.52	12.85	0.85	4.33	0.94	3.19	0.91	3.18
AKGD	Citrate Cycle (TCA)	0	12.32	0	2.70	0	2.67	0.06	2.65
CITL	Citrate Cycle (TCA)	0	128.65	0	128.53	0	128.65	0	0
CS	Citrate Cycle (TCA)	0.52	141.02	0.85	109.73	1.36	82.54	0.91	3.18
FRD2	Citrate Cycle (TCA)	0	68.73	0	58.3	0	0	0	0
FRD3	Citrate Cycle (TCA)	0	68.73	0	58.66	3.95	41.53	0	0
FUM	Citrate Cycle (TCA)	0.41	281.55	0.98	230.97	0.74	14.34	0.79	5.65
ICDH _y	Citrate Cycle (TCA)	0.52	12.85	1.12	3.23	0.52	3.19	0.91	3.18
MDH	Citrate Cycle (TCA)	-59.13	281.55	-1.43	230.97	-1.13	14.29	0.79	5.65
SUCD1i	Citrate Cycle (TCA)	0	77.27	2.62	60.47	5.09	43.81	0.38	2.65
SUCOAS	Citrate Cycle (TCA)	-12.26	18.41	-3.04	0.06	-2.60	11.61	-2.59	0
ENO	Glycolysis/Gluconeogenesis	-44.31	18.76	-40.26	18.38	16.47	18.38	6.82	18.76
FBA	Glycolysis/Gluconeogenesis	-0.27	9.63	0.17	9.63	7.82	9.63	7.89	9.63
FBP	Glycolysis/Gluconeogenesis	0	128.65	0	107.37	0	79.84	0	0
G1PP	Glycolysis/Gluconeogenesis	0	128.65	0	107.37	0	0	0	0
GAPD	Glycolysis/Gluconeogenesis	-32.30	18.97	-20.08	18.97	17.17	18.97	17.24	18.97
GLCP	Glycolysis/Gluconeogenesis	0	128.65	0	107.37	0	0	0	0
GLCS1	Glycolysis/Gluconeogenesis	0.10	128.75	0.10	107.47	0.10	0.10	0.10	0.10
GLGC	Glycolysis/Gluconeogenesis	0.10	128.75	0.10	107.47	0.10	0.10	0.10	0.10
HEX1	Glycolysis/Gluconeogenesis	0	136.15	0	115.96	0	10.00	0	10.00
PDH	Glycolysis/Gluconeogenesis	0	27.10	0	16.27	0	16.27	0	15.25
PFK	Glycolysis/Gluconeogenesis	0	138.28	0	116.23	7.82	88.34	7.89	9.63
PGI	Glycolysis/Gluconeogenesis	0	9.90	0	9.90	4.48	9.90	4.70	9.90
PGK	Glycolysis/Gluconeogenesis	-32.3	18.97	-20.08	18.97	17.17	18.97	17.24	18.97
PGM	Glycolysis/Gluconeogenesis	-44.31	18.76	-42.34	18.38	16.47	18.38	6.82	18.76
PPS	Glycolysis/Gluconeogenesis	0	128.65	0	108.43	0	79.78	0	0
PYK	Glycolysis/Gluconeogenesis	0	144.18	0	112.27	3.47	92.66	0	16.61
TPI	Glycolysis/Gluconeogenesis	-41.64	9.63	-29.43	9.63	7.82	9.63	7.89	9.63
ATPS4r	Oxidative phosphorylation	-24.18	108.45	-24.18	106.6	-24.15	60.29	-24.18	106.09
CYTBO3	Oxidative phosphorylation	27.85	77.15	31.12	77.15	31.10	43.72	31.05	77.15
FDH2	Oxidative phosphorylation	0	27.10	0	27.10	0	15.25	0	26.72
LDH_D2	Oxidative phosphorylation	0	68.73	0	54.47	0	0	0	0
NADH6	Oxidative phosphorylation	0	68.73	0	53.65	0	35.03	0	58.15
NADH7	Oxidative phosphorylation	0	68.73	0	58.75	0	0	0	0
NADH8	Oxidative phosphorylation	0	68.73	0	58.75	2.96	41.53	0	0
NADH9	Oxidative phosphorylation	0	68.73	0	58.32	0	0	0	0
NADH10	Oxidative phosphorylation	0	68.73	0	58.75	0	0	0	0
NADH12	Oxidative phosphorylation	0	68.73	0	54.81	0	35.03	0	57.55
POX	Oxidative phosphorylation	0	27.10	0	16.12	0	15.25	0	0
SUCD4	Oxidative phosphorylation	-0.12	77.15	2.26	60.35	4.96	43.72	0.26	59.28
THD2	Oxidative phosphorylation	0	257.30	0	214.75	0	159.67	0	217.01
THD5	Oxidative phosphorylation	0	272.60	0	229.59	0	163.93	0	231.14
TRDR	Oxidative phosphorylation	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
EDA	Pentose Phosphate Cycle	0	9.90	0	9.90	0	0	0	0
G6PDH _y	Pentose Phosphate Cycle	0	9.90	0	9.90	0	5.43	0	5.20
PGDH	Pentose Phosphate Cycle	0	9.90	0	5.45	0	5.43	0	5.20
PGDHY	Pentose Phosphate Cycle	0	9.90	0	9.90	0	0	0	0
PGL	Pentose Phosphate Cycle	0	9.90	0	9.90	0	5.43	0	5.20
RPE	Pentose Phosphate Cycle	-0.22	6.38	-0.22	4.51	-0.22	3.40	-0.22	3.25
RPI	Pentose Phosphate Cycle	-3.52	-0.22	-2.58	-0.22	-2.03	-0.22	-1.95	-0.22
TAL	Pentose Phosphate Cycle	-0.06	3.24	-0.06	2.3	-0.06	1.75	-0.06	1.68
TKT1	Pentose Phosphate Cycle	-0.06	3.24	-0.06	2.3	-0.06	1.75	-0.06	1.68
TKT2	Pentose Phosphate Cycle	-0.16	3.14	-0.16	1.66	-0.16	1.65	-0.16	1.57
ACT6	Transport, Extracellular	-12.32	0	-11.55	0	-11.51	-8.36	-11.52	0
GLCpts	Transport, Extracellular	0	10.00	0	10.00	0	10.00	0	10.00
GLCt2	Transport, Extracellular	0	10.00	0	10.00	0	10.00	0	10.00
AMDnt	Amorphadiene	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11

^aThe relaxed model only uses stoichiometric data and not the ¹³C labeling data and serves as the allowable bounds for the fluxes.

^bFlux abbreviations correspond to the names in Figure 3-8.

^cAll flux values have units of mM/hr and are relative to a glucose uptake rate of 10 mM/hr.

	<i>Number of fluxes</i>		
	Full	Reduced	Basic
Set by experimental conditions	3	3	3
Pruned by experimental conditions/reductions	33	47	57
Stoichiometrically fixed	92	102	101
Fully Resolved	0	0	0
Partially Resolved	126	109	85
Unresolved	14	7	22

Appendix 2

Protocols

The following protocols are included in this section:

- Fermentation protocol
- GC-MS of proteins for isotopomer analysis

Fermentation protocol

The 1-L Sartorius fermentor was assembled as described in the manual, with the following changes for two-phase culture in minimal medium:

- A wider diameter harvest line than that supplied with the fermentor was put through a fitting in the head plate. The height of this line was adjusted to the liquid level under the agitation and aeration conditions of our experiment. Constant volume was maintained in the fermentor by pumping the harvest line at a faster rate than the feed line, so the liquid level could not exceed the height of the harvest line. The feed and harvest lines were operated from the same pump drive. The feed was pumped using a Masterflex size 14 pump head and corresponding tubing, and the harvest line was pumped with a Masterflex size 16 pump head and corresponding tubing. With this configuration, at the same pump drive setting the harvest line allows a higher flow rate than the feed.
- Dodecane was added to the fermentor at 1/10 the feed flow rate. Black, solvent-compatible, Masterflex size 16 tubing was used for all lines coming into contact with dodecane (the dodecane supply to the fermentor and the harvest line). The dodecane was assumed to be sterile as supplied.
- All pumps were calibrated using water and set to achieve the appropriate flow rate. Volumetric flow rates were continuously verified throughout the experiment by collecting the harvest liquid in a 2-L graduated cylinder and recording the volume of the aqueous and dodecane phases in the cylinder at each sample time. The dodecane flow rate was verified by using a 250 ml graduated cylinder as a reservoir for the dodecane being pumped to the fermentor. The volume of dodecane in the cylinder was recorded at each sample time.
- The OD₆₀₀ was monitored over time by collecting small (~5 ml) samples using the sterile sampling port. At the time of induction and at all sampling times thereafter, amorphadiene samples were taken. Enough of the dodecane phase was removed from the sampling tube and transferred to a 0.6-ml microfuge tube so that 10 µl of dodecane could be sampled and transferred to a GC vial containing 700 µl of water. To this, 690 µl of ethyl acetate containing a caryophyllene internal standard was added. The resulting sample was vortexed, frozen at -80°C, and thawed later for amorphadiene analysis by GC-MS.
- The GasWorks off-gas analysis software and the Biocommand software for monitoring pH, dO, agitation and temperature were started at the time of inoculation and continued to record data throughout the fermentation. CO₂ evolution was heavily relied upon as an indicator of the condition of the culture.
- At the end of the fermentation, the spent culture was poured into a container and bleached. The fermentor vessel was only RINSED WITH WATER and WIPED WITH ETHANOL. Strong detergents and bleach can accumulate in vessels, so they should not be used for cleaning permanent parts of the fermentor.

GC-MS ISOTOPOMER ANALYSIS OF AMINO ACIDS

- Start with cell pellets from 5 ml of culture at OD 0.8-2, washed once with 0.9% NaCl and flash frozen at -80°C.
- Thaw the pellets and resuspend in 1 ml of sterile nanopure water.
- Transfer the resuspended pellets to 2 ml eppendorf tubes, and sonicate using the microtip for 3 min. with a 3 sec. on/1 sec. off cycle. Begin at power level one and turn up to power level three immediately.
- Precipitate the protein from the resulting lysate using trichloroacetic acid (TCA) (http://www.its.caltech.edu/~bjorker/TCA_ppt_protocol.pdf). Add 250 µl of TCA to the 1 ml of sample. Incubate at 4°C for 10 min. Centrifuge cold at 14,000 rpm for 5 min. Remove the supernatant, leaving protein pellet. Wash pellet with 200 µl cold acetone. Centrifuge cold at 14,000 rpm for 5 min. Repeat acetone wash. Dry pellet in 100°C oven for about 2 min.
- Resuspend the pellet in 900 µl of 6M HCl and transfer to a clear glass, screw-top GC vial. You will probably need to transfer a chunk of the pellet with a pipet tip as it will not fully resuspend. Wash the tube with another 900 µl of 6M HCl and transfer that to the glass vial, for a total volume of 1.8 ml HCl. Cap the vials and place them in the 100°C oven for 24 hours to hydrolyze all of the proteins into their amino acid monomers. **NOTE:** It is important to use a glass vial and not an Eppendorf tube. Conducting the acid hydrolysis and derivatization steps in plastic results in large contaminating peaks in the gas chromatogram and very low amino acid signals.
- Remove the lids and dry the samples completely under a stream of air or nitrogen. This will take approximately overnight.
- Partially dissolve the samples with 100 µl of tetrahydrofuran (THF). Add 100 µl of nTertbutyldimethylsilyl.. derivatization reagent. Incubate all samples in a water bath operated between 65 and 80°C for 1 hour. Vortex occasionally. Almost all of the brown solid material in the vial should dissolve. The samples will likely be brown.
- Transfer the material into a snap-lid GC vial containing a 200 µl insert.
- Prepare a THF blank.
- Analyze the samples using the GCMS, running a THF blank before and after the samples, and changing the injector wash solvents to THF. Use 1:5 or 1:10 split injections. Use the following GC-MS temperature program: hold at 150°C for 2 minutes; increase at 3°C per minute to 280°C; increase at 20°C per minute to 300°C and hold for 5 minutes.