Towards predictive metabolic engineering: kinetic modeling and experimental analysis of a heterologous mevalonate pathway in *E. coli*

Permalink
https://escholarship.org/uc/item/1ss913cv

Author
Weaver, Lane Justin

Publication Date
2013

Peer reviewed|Thesis/dissertation
Towards predictive metabolic engineering: kinetic modeling and experimental analysis of a heterologous mevalonate pathway in *E. coli*

by

Lane Justin Weaver

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

Joint Doctor of Philosophy

with the University of California, San Francisco

In

Bioengineering

in the

Graduate Division

of the

University of California at Berkeley

Committee in charge:

Professor Jay D. Keasling, Chair
Professor Tejal Desai
Professor Richard Mathies
Professor Matt Francis

Fall 2013
Abstract

Towards predictive metabolic engineering: kinetic modeling and experimental analysis of a heterologous mevalonate pathway in *E. coli*

By

Lane Justin Weaver

Joint Doctor of Philosophy in Bioengineering

University of California at Berkeley

Professor Jay D. Keasling, Chair

Owing to economic, political, and environmental concerns, the nature of finite natural resources will increasingly necessitate a transition to renewable resources over the next century. Biology, enabled by its use of enzymes, is a master craftsman and is thus well suited to utilize renewable resources, converting them into desired products. As such, the field of metabolic engineering has grown up over the past 20 or so years to harness biology’s catalytic and synthetic capabilities. In particular, the Keasling lab has successfully exploited the potential for the mevalonate pathway for many applications, most notably the production of Artemisinic acid, an anti-malarial compound.

However, despite many years of engineering efforts, there is still no definitive way to predict how expressing a certain enzyme at a certain level will effect pathway function, which can largely be attributed to the lack of availability of a model including a kinetic representation of pathway flux.

To aid in debugging efforts to increase yield, titer, and productivity of the pathway, an ordinary differential equation (ODE) model was built using constants culled from literature and enzyme concentrations derived from experiment. To account for uncertainty in model parameters, a global sensitivity analysis was performed. The model demonstrates that amorphadiene synthase (and in general terpene synthase) activity is limiting and needs to be addressed to increase pathway flux. Furthermore, the model predicts that in a local regime, the pathway is fairly insensitive to product inhibition, a hypothesis that had previously been posited to be limiting flux.

To experimentally test these specific predictions, three *E. coli* strains were constructed, the first acting as a base strain, the second encoding a homologous mevalonate kinase from *S. aureus* that is less susceptible to product inhibition, and a third strain in which the expression of amorphadiene synthase was increased. These strains were profiled by metabolomic and proteomic methods, which helped validate the predictions of the ODE model. Furthermore, they enabled the determination of an in vivo $k_{cat}$ value for amorphadiene synthase, which differed ~3-fold from its in vitro counterpart—
demonstrating that in vitro parameters are not always representative of in vivo conditions, further bolstering the case for ensemble-type kinetic modeling.

Having identified targets for engineering, high-throughput screens were developed in an attempt to leverage directed evolution for improvement of production. While a screen was unable to be developed for amorphadiene synthase, a platform for screening of another rate-limiting enzyme, Acinetobacter wax ester synthase, was demonstrated.

This type of analysis--starting with in silico pathway analysis, followed by in vivo characterization and finally targeted enzyme engineering--will become increasingly important as metabolic engineering is applied to the production of a variety of new targets. With this ability, metabolic engineers will more rapidly be able to move from concept to production of designer molecules in high titers, yields, and productivities.
# Table of contents

ABSTRACT 1

LIST OF FIGURES IV

ACKNOWLEDGMENTS VI

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION: 21ST CENTURY BIOLOGY AND SUSTAINABILITY 1

1.2 THE MEVALONATE PATHWAY 1

1.3 BACKGROUND: THE METABOLIC ENGINEERING CHALLENGE 2

1.3.1 STOICHIOMETRIC ANALYSIS AND PROCESS CONSIDERATIONS 3

1.3.1 STRATEGIES FOR METABOLIC ENGINEERING 5

YIELD 5

PRODUCTIVITY 8

TITER 12

1.4 DISSERTATION SCOPE 13

CHAPTER 2: KINETIC MODELING APPROACHES TO IDENTIFYING TARGETS FOR FLUX IMPROVEMENT IN THE MEVALONATE PATHWAY

2.1 INTRODUCTION 19

2.2 MATERIALS & METHODS 21

2.2.1 ODE MODELING 21

2.2.2 GLOBAL SENSITIVITY ANALYSIS MODELING 23

2.2.3 PROTEIN PURIFICATION & ACTIVITY VERIFICATION 23

3.2.4 IN VITRO PRODUCTION STUDIES 23

2.3 RESULTS 23

2.3.1 AN ODE MODEL OF THE BOTTOM PORTION OF THE MEVALONATE PATHWAY MAKES TESTABLE PREDICTIONS ABOUT PATHWAY FUNCTION 23

2.3.1.1 IMPACTS OF FPP FEEDBACK-LOOP 24

2.3.1.2 CAUSE OF DIFFERENCES IN AMORPHADIENE PRODUCTION 29

2.3.2 GLOBAL SENSITIVITY ANALYSIS CONFIRMS IMPORTANCE OF ADS LEVELS TO AMORPHADIENE PRODUCTION 32

2.3.3 IN VITRO PATHWAY RECONSTITUTION PROVIDES A SYSTEM FOR FUTURE TESTING OF MODEL 33

2.3.4 IN VITRO STUDIES POINT TO MevPP-BASED PRODUCT INHIBITION OF PMK 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>DISCUSSION</td>
<td>37</td>
</tr>
<tr>
<td>2.5</td>
<td>REFERENCES</td>
<td>38</td>
</tr>
</tbody>
</table>

**CHAPTER 3: IN VIVO CHARACTERIZATION OF HETEROLOGOUS MEVALONATE PATHWAY FUNCTION: A QUANTITATIVE –OMICS BASED APPROACH**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>INTRODUCTION</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>MATERIALS &amp; METHODS</td>
<td>42</td>
</tr>
<tr>
<td>3.2.1</td>
<td>CLONING/PLASMID CONSTRUCTION</td>
<td>42</td>
</tr>
<tr>
<td>3.2.2</td>
<td>GROWTH AND MEDIA</td>
<td>42</td>
</tr>
<tr>
<td>3.2.3</td>
<td>QCONCAT PROTEIN PURIFICATION</td>
<td>43</td>
</tr>
<tr>
<td>3.2.4</td>
<td>PROTEOMICS</td>
<td>43</td>
</tr>
<tr>
<td>3.2.5</td>
<td>METABOLITE ANALYSIS OF MEVALONATE PATHWAY INTERMEDIATES (LC-MS)</td>
<td>43</td>
</tr>
<tr>
<td>3.2.6</td>
<td>AMORPHADIENE PRODUCTION (GC-MS)</td>
<td>44</td>
</tr>
<tr>
<td>3.2.7</td>
<td>CONVERSIONS TO CELLULAR CONCENTRATIONS</td>
<td>44</td>
</tr>
<tr>
<td>3.3</td>
<td>RESULTS</td>
<td>45</td>
</tr>
<tr>
<td>3.3.1</td>
<td>CREATION OF A SYNTHETIC POLYPEPTIDE STANDARD FOR QUANTITATIVE PROTEOMICS AND MEASUREMENT OF ABSOLUTE CONCENTRATIONS OF PATHWAY ENZYMES</td>
<td>45</td>
</tr>
<tr>
<td>3.3.1.1</td>
<td>BUILD-UP OF FPP ENABLES DETERMINATION OF IN VIVO K&lt;sub&gt;cat&lt;/sub&gt; FOR AMORPHADIENE SYNTHASE</td>
<td>47</td>
</tr>
<tr>
<td>3.3.2</td>
<td>COMPARISON OF MODEL PREDICTIONS WITH EXPERIMENTAL DATA</td>
<td>48</td>
</tr>
<tr>
<td>3.3.2.1</td>
<td>RATES OF AMORPHADIENE PRODUCTION</td>
<td>49</td>
</tr>
<tr>
<td>3.3.2.2</td>
<td>EVIDENCE FOR A FEEDBACK LOOP</td>
<td>50</td>
</tr>
<tr>
<td>3.4</td>
<td>DISCUSSION</td>
<td>51</td>
</tr>
<tr>
<td>3.5</td>
<td>REFERENCES</td>
<td>51</td>
</tr>
</tbody>
</table>

**CHAPTER 4: TOWARDS A PLATFORM FOR DIRECTED EVOLUTION OF TARGETED ENZYMES**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>INTRODUCTION</td>
<td>52</td>
</tr>
<tr>
<td>4.2</td>
<td>MATERIALS &amp; METHODS</td>
<td>54</td>
</tr>
<tr>
<td>4.2.1</td>
<td>CLONING/PLASMID CONSTRUCTION</td>
<td>54</td>
</tr>
<tr>
<td>4.2.2</td>
<td>STRAIN CONSTRUCTION</td>
<td>54</td>
</tr>
<tr>
<td>4.2.3</td>
<td>GROWTH AND MEDIA</td>
<td>54</td>
</tr>
<tr>
<td>4.2.4</td>
<td>SURFACE DISPLAY</td>
<td>54</td>
</tr>
<tr>
<td>4.2.5</td>
<td>PYROPHOSPHATE-BASED ASSAY</td>
<td>55</td>
</tr>
<tr>
<td>4.2.6</td>
<td>COA-BASED ASSAY</td>
<td>55</td>
</tr>
<tr>
<td>4.2.7</td>
<td>CHIP EXPERIMENTS</td>
<td>55</td>
</tr>
<tr>
<td>4.2.8</td>
<td>PROTEIN PURIFICATION</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>RESULTS</td>
<td>56</td>
</tr>
</tbody>
</table>
4.3.1 Development of a Pyrophosphate Assay for Terpene Synthase Function 56
4.3.2 Miniaturized Assay for Pyrophosphate Detection on a Microfluidic Device 60
4.3.3 A Yeast Surface Display System for Droplet-Based Screening 61
4.3.4 Development of a Thiol-Reactive Fluorescence Assay for Esterification Reactions 63
4.3.5 Construction of Thioesterase-Deficient Screening Strain for Wax Ester Synthase Activity 65
4.3.6 Fluorescent Protein-Wax Ester Synthase Fusions for Normalized Activity Screening 66

4.4 Discussion 66

4.5 References 67

CHAPTER 5: CONCLUSION & FUTURE DIRECTIONS

5.1 Introduction 68

5.2 Modeling 69

5.2.1 Parameterization 69
5.2.2 Optimization 70

5.3 Directed Evolution 71

5.4 References 71

Appendix 73
List of Figures

Figure 1- Mevalonate pathway ........................................................................................................................................... 2
Figure 2- To achieve electron balance, reducing equivalents need to be combined with a terminal electron acceptor such as oxygen, or reintroduced into a pathway intermediate. .................................................. 3
Figure 3- An ideal cell would a) quickly grow up to the appropriate density and then become quiescent, decoupling growth and production phases. It would b) maintain high pathway flux for extended periods of time (minimizing the fraction of input carbon necessary for biomass formation) and remain viable even in the presence of high product concentration, transporting product outside of the cell (so that it is easy to harvest) and manifesting lower toxicity. Ideally, it would also be c) recyclable—further amortizing the biomass carbon via use over multiple fermentations .......................................................................................................................................................... 5
Figure 4- Microbial growth and product formation are dependent on the utilization of carbon, nitrogen, phosphate, and, depending on the pathways employed, oxygen. The utilization of these elements can be modeled with Monod kinetics, demonstrating that any one element can limit growth. Additionally, by examining the carbon mass balance (black arrows), a metabolic engineer can get gain insight into pathway performance .................................................................................................................................................. 6
Figure 5- Flux Balance Analysis. Pathways can be modeled with a system of ordinary differential equations. By assuming steady state (i.e. dX/dt=0), the system simplifies to a series of linear equations that is much easier to solve. However, since the system is underdetermined (i.e. more unknowns than equations), there is no unique solution for the unknown fluxes (in this example, only \( F_{\text{in}} \) is initially given). By adding an objective function to be maximized (here, simply maximizing the product efflux, \( F_{\text{out}} \)), a unique solution can be obtained .................................................................................................................................................. 7
Figure 6- Comparison of local and global sensitivity analyses. Left: (a) Enz 1 is initially limiting, as indicated by its control coefficient. (b) After increasing Enz. 1 expression, Enz. 3 becomes limiting. In this iterative fashion, a highly producing strain is reached. Right: Many pathways are generated by randomly sampling enzyme activities. The pathways are then simulated, and a statistical analysis then identifies the parameters that more often led to high product levels ........................................................................................................................................... 10
Figure 7- An idealized metabolic engineering workflow ........................................................................................................... 14
Figure 8- Linking inputs and outputs from different modeling domains .................................................................................................................. 20
Figure 9- Modeling workflow .......................................................................................................................................................... 24
Figure 10- Dependence of amorphadiene production and MevP synthesis on strength of feedback inhibition for base strain mbis3 .................................................................................................................................................. 25
Figure 11- Dependence of amorphadiene production and MevP synthesis on strength of feedback inhibition for strain 10kADS .................................................................................................................................................. 25
Figure 12- MK inhibition dependence on FPP for strain mbis3 .................................................................................................................. 26
Figure 13- Graphical representation of feedback loop dependence on ADS concentrations .................................................................................................................................................. 26
Figure 14- MevP levels by strain .......................................................................................................................................................... 27
Figure 15- Effect of MK product inhibition on MevP levels .................................................................................................................. 28
Figure 16- Effect of increasing MK levels in mbis3 .................................................................................................................................................. 29
Figure 17- Model predictions for amorphadiene production by strain .................................................................................................................. 30
Figure 18- Nonlinear response of amorphadiene production to ADS .................................................................................................................. 31
Figure 19- FPP concentrations in mbis3 as ADS is increased .................................................................................................................. 31
Figure 20- Amorphadiene titer for 1000 simulations of the pathway ................................................................................................. 32
Figure 21- Partial rank correlation coefficient for parameters of the bottom mevalonate pathway .................................................................................................................................................. 33
Figure 22- SDS-PAGE of mevalonate pathway proteins .................................................................................................................. 34
Figure 23- in vitro production of amorphadiene .................................................................................................................................................. 35
Figure 24- in vitro MK timecourse .................................................................................................................................................. 36
Figure 25- in vitro MK+PMK timecourse .................................................................................................................................................. 36
Figure 26- in vitro PMK timecourse .................................................................................................................................................. 37
Figure 27- Experimental system for examining pathway kinetics .................................................................................................................. 41
Figure 28- Strains used to test model predictions .................................................................................................................................................. 42
Figure 30- Qconcat peptide calibration curves .................................................................................................................................................. 46
Figure 31- Absolute concentrations of bottom mevalonate pathway proteins ........................................46
Figure 32-Intracellular FPP levels in engineered mbis strains ......................................................................47
Figure 33-Specific Amorphadiene production ..........................................................................................48
Figure 34-Comparison of in vivo and in vitro kcat values for ADS ..........................................................48
Figure 35- Computationally predicted and experimentally derived rates of amorphadiene production ..........................................................................................................................49

Figure 36-Intracellular MevP levels in engineered mbis strains .................................................................50
Figure 37-Inverse correlation between MevP and FPP for engineered mbis strains .....................................50
Figure 38-Luciferase-based PPI assay ........................................................................................................56
Figure 39-Luciferase-based standard curve for ADS .................................................................................57
Figure 40-Effect of additives on performance of lucifase-based assay in bulk emulsion ......................58
Figure 41-Fluorescence-based assay for PPI detection ..............................................................................58
Figure 42-Performance of fluorescence PPI assay at different ADS concentrations ............................59
Figure 43-Effect of pre-incubation time PPI readout ..............................................................................59
Figure 44-PiPer performance with ADS in buffer and cell lysate ...............................................................60
Figure 45-on chip PiPer reaction with purified ADS ......................................................................................61
Figure 46-chip vs. plate reader performance ................................................................................................61
Figure 47-Detection of surface displayed RFP ..........................................................................................62
Figure 48-Surface displayed RFP as detected by FITC-labeled anti-V5 antibody ..................................62
Figure 49-CoA-based detection assay .....................................................................................................64
Figure 50-CPM assay CoA standard curve in cell lysate ............................................................................64
Figure 51-Read-out of WES activity in cell lysate ......................................................................................65
Figure 52-background hydrolysis of palmitoyl-CoA in cell lysate ...............................................................65
Figure 53-Activities of purified WES-mCherry fusion ...............................................................................66
Figure 54-An idealized metabolic engineering workflow ........................................................................69
Figure 55-Representative mbis strain plasmid ..........................................................................................75
Figure 56-Representative surface display vector ......................................................................................75
Figure 57-Representative WES-mCherry vector ......................................................................................76
Figure 58-Qconcat 2.0 plasmid .................................................................................................................76
Acknowledgments

I have to start by thanking both of my PhD advisors, whose patience, encouragement, and confidence in my abilities has enabled any graduate school success I may have enjoyed.

First, Rich Mathies taught me how to approach a problem, break it down, analyze it, and communicate the subsequent results—skills that form the foundation of the PhD and will be employed in every endeavor I pursue in the future. Rich has an amazing depth of knowledge—when he has something to say, it is a real learning opportunity; I hope someday to gain such an expertise in a subject as to command the same amount of respect.

When I think about the lessons I’ve learned from Jay Keasling, a passage about Mervin Kelly, the long-time head of AT&T’s distinguished and prodigious Bell Labs, comes to mind,

“Mr. Kelly believed that freedom was crucial, especially in research. Some of his scientists had so much autonomy that he was mostly unaware of their progress until years after he authorized their work. When he set up the team of researchers to work on what became the transistor, for instance, more than two years passed before the invention occurred...In sum, he trusted people to create. And he trusted them to help one another create.”

Jay must have subscribed to this ethos of innovation, since the Joint Bioenergy Institute, where I have worked for the past four years, was largely built in its image. It has been an amazing place to do science and I have benefited greatly from the resources I have found there, particularly with respect to all the crazy projects Jay allowed me to pursue. I hope that I will have the opportunity to help inspire people in future leadership roles in the same way that Jay has been able to do in academic science.

My labmates have formed the bedrock of what has been both an enlightening and enjoyable past seven years---first Sam, Erik, and Richard, and later Steen, Jeff, Will, Luning, Tanveer, Becky, Zain, James, and many others. Also, my bioengineering classmates, especially my roommates Sam and Rick, have kept things fun outside the lab. And I’m particularly grateful for Numrin’s company; she has been there for me many times and has made the last few years of my PhD rich and enjoyable.

I’ve also got to thank my teachers and professors throughout the years, those at Hershey and Penn State. In particular, I’d like to thank Patricia Donaldson, my 12th grade AP English, who recently passed away. She quite unexpectedly awarded me the senior AP English distinction at a time when I thought my interpretation and communication skills were somehow inherently inferior to those of the class superstars. But she saw some undeveloped ability in me, and by highlighting it, gave me the confidence to push myself into new challenges (e.g. columnist for The Collegian newspaper); I’m sure this kernel of inspiration has spread its tendrils to all of my subsequent academic scholarships. And isn’t that the true mark of a teacher---identifying unrealized potential and empowering a student to achieve where his or her talents may have otherwise remained fallow? I never had the chance to thank Mrs. Donaldson, so I’ll instead dedicate this work in her memory.
1.1 Motivation: 21st century biology and sustainability

Mankind has harnessed microorganisms for production of chemicals for several millennia, but only with the advent of recombinant DNA technology in the past few decades has manipulation at the molecular level become possible. The ability to re-engineer biology to human specifications, termed synthetic biology, will play an important role in the 21st century, helping address issues of sustainability, access to natural resources, and economic solvency.

1.2 The mevalonate pathway

The mevalonate pathway converts acetyl-CoA, an intermediate of central carbon metabolism, into an activated 5-C molecule, isoprenyl pyrophosphate (IPP). IPP then serves as a basic building block that can be converted via condensation, cyclization, and oxidation unit operations into the terpenoid family of secondary metabolites, a group which consists of over 50,000 molecules and spans several categories of industrial relevance, including pharmaceutics, flavors, fragrances, solvents, biopolymers, and advanced biofuels.

Amorphadiene is a 15-C molecule that is formed through three successive condensations of IPP equivalents followed by a cyclization. As a precursor to the anti-malarial compound Artemisinin, it has served as a target for improvements in production.

The entire mevalonate pathway for the biosynthesis of amorphadiene, including metabolites, catalyzing enzymes, and inhibitory relationships, is shown in Figure 1. While in vitro evidence has demonstrated the inhibition of the mevalonate kinase reaction by FPP, it is currently unknown whether this has any in vivo relevance in heterologous (i.e. engineered) circumstances.
1.3 Background: The metabolic engineering challenge

Metabolic engineering is a subset of synthetic biology that focuses on the re-engineering of metabolism for the purpose of producing valuable chemicals. Generally, there are two approaches to the design of cellular metabolism; rational approaches that begin with a model (mathematical, heuristic, or otherwise) that provides a prediction for how the system will act as a function of defined control variables, and random approaches, which may recognize the control variables but do not posses a model for how they impact the relevant output. Correspondingly, the experimenter alters the variables in non-directed\(^2\) or potentially systematic (e.g. combinatorial/factorial\(^3\)) ways and selects or screens for the desired phenotype\(^4\). Given the complexity of even simple cells, this approach is both warranted and prudent, being particularly well-known in industry, where high-throughput techniques are routinely leveraged.

However, outside of a few examples\(^5\), the random approach is seldom followed by an analysis documenting the reasons for increases in production and subsequently using this knowledge for predictive model generation/refinement. Thus, while the technology pursuant to these studies increases at an ever-faster rate (e.g. -omics technologies, high-throughput screening), our understanding is somewhat lagging behind\(^6\).

The de-emphasis on mechanistic understanding extends to the initial choice of host organism or chassis for production. Once a target molecule is identified, an organism (e.g. Streptomyces, Aspergillus) that naturally produces the product can sometimes be found, followed by application of mutational techniques to increase the titer. While this may result in the fastest path to commercialization, it often trades expedition for comprehensive understanding. Partially for this reason (and partially for experimental convenience), many academic labs work in *E. coli* and *S. cerevisiae*, the two best characterized model organisms. Given the knowledge surrounding these organisms and the tools available for their study, they present excellent test-beds for model building. The ability to engineer any pathway of interest into these systems (with concomitant modification of the chasses where necessary)---or possibly in the future a minimal cell---presents the most acute test of our scientific knowledge, and, if successful, would put
metabolic engineering on par with other more mature engineering disciplines in terms of a systematic, predictable practice.

1.3.1 Stoichiometric analysis and process considerations

Perhaps the most important variable in a process is whether or not it takes place in an oxygen-containing environment. Aerobic processes are necessary when there is net demand for ATP in a pathway or when there is excess reducing equivalents. However, since aerobic processes add to operating costs, needing control systems for oxygen delivery and often cooling systems to dissipate heat generated via oxygen reduction, they are generally difficult to manage at very large scales. **Figure 2** shows viable processes adhering to redox constraints.

![Figure 2](image)

**Figure 2**-To achieve electron balance, reducing equivalents need to be combined with a terminal electron acceptor such as oxygen, or reintroduced into a pathway intermediate.

Besides decreasing the cost, the coupling between growth and product formation provides an additional benefit in favor of anaerobic processes. In these cases the product performs an analogous role to other native fermentative by-products (e.g. ethanol, lactate, succinate, formate) in that it precisely balances electron stoichiometry and prevents the accumulation of NADH, which would kinetically restrict glycolysis and biomass accumulation. When other fermentative pathways have been removed, this translates to an evolutionary benefit, since cells that do not have a highly functioning pathway will not grow (although if this leads to slower overall growth, contamination may pose a larger problem under these circumstances).

In instances where there are large profit margins (e.g. pharmaceuticals), the higher operating and feedstock costs (due to lowered yields) for an aerobic process may not translate to a material difference. However, for commodities with low margins, process considerations can be an important factor---and if an anaerobic process is possible within the constraints of biochemistry, it should be pursued.

How does one determine the potential for anaerobic process, and more generally the theoretical yield for a pathway? Examination of the balanced stoichiometric equation, with specific focus on cofactor usage (starting at the carbon source of interest), can determine the answers to these questions. The importance of using this simple analysis as a screening tool has been described previously⁷ (although the referenced study assumes NAD+}
regeneration via glycerol formation, despite the presence of the oxygen, the terminal electron acceptor).

As an illustration, the benefits and trade-offs of two different routes to produce isopentenyl pyrophosphate (IPP) are outlined below. The balanced equations for the mevalonate pathway, native to yeast, and the DXP pathway, native to bacteria, are listed in the box below.

Table 1-Stoichiometric comparison of mevalonate and DXP pathways for IPP synthesis

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Balanced equation</th>
<th>Redox?</th>
<th>Energy?</th>
<th>Yield (g IPP/g glucose)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mev</td>
<td>(1.5 \text{ glucose} \rightarrow 1\text{IPP} + 4\text{CO}_2 + 1\text{NAD}(P)\text{H})</td>
<td>Excess</td>
<td>Neutral</td>
<td>0.91</td>
</tr>
<tr>
<td>DXP</td>
<td>(1 \text{ glucose} + 2\text{ATP} + 1\text{NAD}(P)\text{H} \rightarrow 1\text{IPP} + \text{CO}_2)</td>
<td>Deficient</td>
<td>Net input</td>
<td>1.37</td>
</tr>
</tbody>
</table>

Examining electron cofactors, the DXP pathway does not generate excess reducing equivalent, while the mevalonate pathway generates 1 NADH (assuming that a NADH-dependent HMG-CoA reductase is used)---thus, an anaerobic process is inaccessible using the mevalonate route. If, however, a terminal reductase were available for the mevalonate pathway that utilized the final NADH, redox balance could be attained---opening up the intriguing possibility of anaerobic production (for a restricted number of products, e.g. farnasane).

However, the energy requirement imposes a further constraint on the process. The DXP pathway requires a net input of ATP, for instance, and thus must run aerobically, as substrate-level phosphorylation alone is not sufficient to form IPP. On the other hand, the mevalonate pathway is ATP-neutral, which does not preclude an anaerobic process (although maintenance requirements, which have not been accounted for here, would likely necessitate at least a microaerobic environment). In a well-aerated reactor, the energy advantage for the mevalonate pathway isn't likely to be substantial, as oxidation of glucose to \(\text{CO}_2\) provides ample ATP and most metabolic enzymes are saturated with respect to this cofactor\(^8\), thus kinetic considerations are likely to take precedence.

Thus an examination of cofactors suggests that aeration is necessary in both pathways, although with some engineering this oxygen requirement could be lessened for the mevalonate pathway.

With respect to yield, the balanced equations show that the DXP pathway is 33% more efficient than the mevalonate pathway. A more involved analysis would take into account the fraction of carbon necessary to form the additionally required cofactors in the DXP pathway, so even this theoretical yield is an overestimate. Still, much has been gleaned from this superficial stoichiometric analysis.

As an example of using balanced equations for pathway analysis, Gonzalez and coworkers observed that if they reversed the catabolic \(\beta\)-oxidation of fatty acids, alcohols could be
produced from glucose in a redox-balanced manner with net ATP generation\textsuperscript{9}, which led to very high yields for n-butanol (0.33 g n-butanol/g glucose). Liao and coworkers employed stoichiometric pathway analysis for a slightly different aim, using redox balance as a convenient selection method. With this idea, they were able to evolve a poorly performing trans-enoyl-CoA reductase into kinetically improved variant in a butanol producing pathway\textsuperscript{10}.

1.3.1 Strategies for Metabolic Engineering
There are three metrics that govern the success of a metabolic engineering project: (1) \textit{Yield}: the percentage of substrate converted to product, (2) \textit{Productivity}: the average rate of product formation, and (3) \textit{Titer}: the concentration of product. While only two are independent (given the fermentation time, average productivity and titer provide the same information), all three can independently impact the costs of a process, hence their importance. Figure 3 demonstrates characteristics of a so-called perfect cell that serve as end-points for engineering endeavors. These characteristics are explained in the following sections.

![Graphs and images explaining the characteristics of a perfect cell.]

Figure 3: An ideal cell would a) quickly grow up to the appropriate density and then become quiescent, decoupling growth and production phases. It would b) maintain high pathway flux for extended periods of time (minimizing the fraction of input carbon necessary for biomass formation) and remain viable even in the presence of high product concentration, transporting product outside of the cell (so that it is easy to harvest) and manifesting lower toxicity. Ideally, it would also be c) recyclable---further amortizing the biomass carbon via use over multiple fermentations.

Yield
The yield of a process is chiefly determined by the distribution of carbon at branch-points in cellular pathways. Conceptually, it is helpful to think of a mass balance for the destinations of carbon, as shown below. Substrate carbon can be converted into biomass, CO\textsubscript{2}, fermentation products, pathway intermediates (both intracellular and extracellular), or could remain as unused substrate in the media.

Generally speaking, yields for fuels and low-margin chemicals need to be very high (e.g. 90\%), as the carbon source constitutes a large percentage of the total cost of the operation.
Figure 4-Microbial growth and product formation are dependent on the utilization of carbon, nitrogen, phosphate, and, depending on the pathways employed, oxygen. The utilization of these elements can be modeled with Monod kinetics, demonstrating that any one element can limit growth. Additionally, by examining the carbon mass balance (black arrows), a metabolic engineer can gain insight into pathway performance.

**Identifying competing pathways: flux-based estimation of carbon distribution**

This distribution of carbon can be estimated using various flux methods, each carrying their own assumptions and strengths/weaknesses.

Flux balance analysis (FBA), for instance, can provide *a priori* estimations of carbon distribution based on maximization of an objective function (e.g. maximization of ATP or biomass) subject to mass-balance constraints (see Figure 5 for a basic example of how FBA works). While these objective functions apply well to native organisms which evolution has selected for based presumably on growth rates, they may not necessarily apply to engineered organisms.
Figure 5-Flux Balance Analysis. Pathways can be modeled with a system of ordinary differential equations. By assuming steady state (i.e. dX/dt=0), the system simplifies to a series of linear equations that is much easier to solve. However, since the system is underdetermined (i.e. more unknowns than equations), there is no unique solution for the unknown fluxes (in this example, only F_in is initially given). By adding an objective function to be maximized (here, simply maximizing the product efflux, F_out), a unique solution can be obtained.

$\begin{align*}
    dA/dt &= F_{in} - F_1 \\
    dB/dt &= F_{in} - F_2 - F_4 + F_2 \\
    dC/dt &= F_2 - F_3 + F_3 \\
    dP/dt &= F_3 - F_4 - F_{out} \\
\end{align*}$

1. set all dX/dt's = 0
2. solve for $F_{in}$ in terms of boundary fluxes
3. Maximize $F_{out}$

$F_{out} = F_{in} - F_8$

FBA methods have shown moderate success in identifying targets for increases in yield. For instance, a study attempting to increase sesquiterpene yield in S. cerevisiae predicted a 10-fold improvement in yield but was only able to realize <2-fold increase, which the authors attributed to the lack of inclusion of kinetics in their stoichiometric model. Other studies seeking to identify engineering targets for anaerobic production of products have generally shown good success in recapitulating previously known experimental solutions (e.g. knockout of all relevant native fermentative pathways), but have in a few cases have also identified unique solutions yet to be implemented.

**Elimination/attenuation of competing pathways**

With competing pathways identified (garnered either through computational analysis or via general biochemical knowledge), experimental methods can be applied to attenuate or eliminate these side-reactions. One way to accomplish this is to decrease the amount of carbon dedicated to biomass by imposing a cellular growth limitation based on a different element—for instance phosphate or nitrogen. The former is a major element in nucleic acids, while the latter is required for both nucleic and amino acid biosynthesis, so both are essential to cell growth. A phosphate limitation strategy, for instance, led to a two-fold increase in titer for production of amorpha-4,11-diene in a fed-batch fermentation study.
Growth may also be globally attenuated at the genetic level. By imposing a persister state where protein synthesis was arrested, Bokinsky et al. were able decrease biomass but maintain the same amount of production for a mevalonate overproducing strain. Unfortunately, the yield did not increase, as the extra flux appeared to be shunted to the TCA cycle, fully oxidizing the carbon to CO₂. In the future, precise regulation of the entry point into the TCA cycle (citrate synthase) may enable this approach to provide higher yields.

At a finer resolution, flux to competing pathways at branch points can be either eliminated (though chromosomal knock-out) or attenuated (through promoter replacement, addition of degradation tags, etc.) by targeting specific enzymes. This strategy is perhaps easiest to implement for products that can be synthesized anaerobically, as the competing pathways are the easily identified redox-balancing fermentation products. Many studies have gained significant improvements in yield by following this strategy17,18,19.

Sometimes, however, there are latent competing pathways that may initially be unknown20. A good strategy to uncover these pathways is to assay a cell lysate for its ability to catabolize spiked-in pathway intermediates21. Once catabolic/competing pathways are identified by this method, genetic interventions can then be employed as necessary.

Care should be taken with these approaches, however, as biomass-forming pathways cannot be completely abrogated. While even a very small amount of biomass could convert all substrate carbon to product, the time scale of the fermentation would make it entirely impractical. Thus in certain regimes yield must be weighed against productivity requirements, which technoeconomic analyses can illuminate22.

**Carbon amortization**

Ideally, the engineered cells would quickly grow up to the density appropriate for production, and then begin producing, such that the production time constituted a large fraction of the total fermentation time. This could be accomplished through judicious timing of transcriptional induction via small molecule inducers23, auto-induction via quorum sensing24, or light-based gene regulation25,26. Control could also be exerted at the protein level, through small molecule or light-based27 allosteric control.

This decoupling between growth and production would provide an ancillary evolutionary benefit, as the selective pressure against the engineered strain would be minimal, since it would not begin expressing pathway proteins or diverting carbon away from biomass until it reached the appropriate cell density.

Lastly, if the cells from a fermentation can be washed and recycled, biomass carbon can be further amortized over multiple batches28 (or continuously in the case of flow-based processes)29, further increasing the yield.

**Productivity**

Productivity is a proxy for the kinetics of the biosynthetic pathway of interest. Increasing
the activity of a pathway will likely increase the yield as well; as such, many of strategies outlined below will simultaneously impact the yield.

The total fermentation time is inversely proportional to the productivity, thus higher throughput and decreased capital costs are obtained with faster pathways. A technoeconomic analysis of ethanol production provided a rough time scale for industrial fermentations; the model suggested ~140 hour fermentation was about optimal\textsuperscript{30}.

**Regulation and pathway topology**

Native regulation can sometimes pose problems for production, especially when the initial enzymes in a pathway are allosterically regulated by the final product\textsuperscript{31,32}. A potential solution to this problem is to remove feedback inhibition through mutation or to find homologous enzymes that are feedback resistant\textsuperscript{33}.

Undesired regulation may also occur on the transcriptional or translational levels, for instance when a product decreases the precursor supply by repressing expression of genes upstream in the pathway, as occurs in tyrosine biosynthesis\textsuperscript{34}. These issues can be similarly dealt with by promoter replacement or transcript engineering.

On the other hand, regulation in a pathway can sometimes be advantageous\textsuperscript{35}. This is particularly true in split pathways where the distribution of flux needs to be carefully balanced between more than one branch to achieve stoichiometric yields in a final combination reaction. A good example was the production of fatty acid ethyl esters (FAEE), where the ethanol and acyl-CoA’s are combined in a final step to yield the FAEE. By imposing negative regulation on the ethanol biosynthesis genes that could be induced by fatty-acyl CoA’s, the authors created a dynamically regulated system that accumulated less ethanol, translating to lowered toxicity and a better balance of reactants for the final enzyme, the wax ester synthase\textsuperscript{36}.

In general, this auto-regulatory approach outlined above will be helpful for suppressing pathway intermediates that exhibit toxicity and for pathways that split but later converge for product formation. Intuition can guide the general regulatory topology necessary, but algorithms that sample many parameter sets to determine optimal regulatory components (e.g. promoter strengths, $K_D$’s, etc.) and control theory approaches\textsuperscript{37} will be needed to determine precise topologies necessary. To expand this strategy beyond fatty-acid derivatives, though, a more expansive toolbox of metabolite-responsive transcription factors will be needed to regulate a more diverse set of pathways\textsuperscript{38}.

**Identifying appropriate enzyme parameters**

In theory, if the kinetic parameters for all the enzymes in a pathway are known, a local sensitivity analysis can be performed in order to identify targets for engineering that will yield increases in flux. Metabolic Control Analysis (MCA)\textsuperscript{39} was a formalism developed in the 1970’s to address the topic of local sensitivity analysis in biochemical systems. One caveat is that this analysis is only valid in a small region of parameter space where linearization about a steady state solution is valid. Thus, the local sensitivities need to be
iteratively recalculated as components are manipulated to increase the flux (Figure 6, left panel).

To search a larger fraction of parameter space, global sensitivity analysis can be performed\(^40\) (Figure 6, right panel). The trade-off is that the information gained is statistical—so manipulating a parameter that correlates well to production, for example, is not guaranteed to translate to better kinetics/production, as it will depend on the context dictated by the other enzymes in that particular instance of the system.

![Figure 6-Comparison of local and global sensitivity analyses. Left: (a) Enz 1 is initially limiting, as indicated by its control coefficient. (b) After increasing Enz. 1 expression, Enz. 3 becomes limiting. In this iterative fashion, a highly producing strain is reached. Right: Many pathways are generated by randomly sampling enzyme activities. The pathways are then simulated, and a statistical analysis then identifies the parameters that more often led to high product levels.](image)

Given the relative strengths of the two approaches, a good strategy would be to start by performing a global sensitivity analysis in order to find the parameters that are statistically most important to changes in production. Carothers et al.\(^41\) used precisely this type of analysis to program RNA-regulated genetic devices whose output was gene expression. Then, a researcher could perform follow-up experiments and local sensitivity analyses in the regions of parameter space identified by the global analysis.

Kinetic analyses such as these are currently difficult to perform due to the lack of available parameters for pathways of interest. To address this limitation, in situ reactor systems are being developed to provide a rapid, experimental means of determining enzyme levels that
lead to high flux. A study utilizing a continuously stirred enzyme membrane reactor system to which cell-free extract was added enabled the identification of bottlenecks in production of dihydroxyacetone phosphate in E. coli, ultimately leading to three-fold higher production\(^ {42}\). Other studies have examined fatty acid biosynthesis in cell free systems, leading to identification of important factors, such as malonyl-CoA levels, to fatty acid production\(^ {43,44}\). In a further example, the biosynthetic pathway for propionyl-CoA formation was reconstituted in vitro, providing information about the rate-limiting step, which was then used to program in vivo expression to obtain a high intercellular concentration of the metabolic precursor\(^ {45}\).

Computational techniques are also emerging to cope with lack of available parameter data. Ensemble kinetic approaches\(^ {46}\), for instance, require only limited steady-state flux data to generate and parameterize predictive kinetic models.

**Finding enzymes with appropriate kinetic specifications**
Following topological, sensitivity, or in situ/vitro analyses that identify parameters leading to high pathway flux, actual enzymes need to be identified that can meet these requirements.

Ideally, design rules and algorithms will progress to the point where arbitrary enzymes can be generated de novo. While there are some exciting early examples of computational enzyme engineering utilizing the ROSETTA design platform\(^ {47,48}\), the field is still nascent. In particular, many of these designed enzymes exhibit very low turnover frequencies (k\(_{\text{cat}}\)), which may be due to the lack of inclusion of conformational dynamics in their design (however, a recent report suggests that these computationally designed enzymes are evolvable\(^ {49}\), so there may be an experimental recourse to increasing their kinetics). Until computational protein engineering becomes mature, though, genome mining and directed evolution will provide the best ways to find enzymes with the appropriate characteristics.

Several studies have leveraged genome databases, for instance, to find more optimal sets of enzymes for pathway function. Ma *et al.*\(^ {50}\) used five HMG-CoA reductases from disparate organisms to evaluate how cofactor preference impacted mevalonate production. Bond-Watts *et al.*\(^ {51}\) prospected to find a hydroxybutyryl-CoA dehydrogenase whose product better matched the stereochemical preference (hence increasing the kinetics) of the downstream enzyme.

Relatively fewer studies have utilized directed evolution, and most often for switching substrate specificity. One of the best examples of a directed evolution approach was the change of cofactor specificity of ketol-acid reductoisomerase from NADPH to NADH, enabling redox-balanced in anaerobic production of 2-methylpropan-1-ol\(^ {52}\). Another study demonstrated engineering of the rate parameters for increased production of the diterpene levopimaradiene\(^ {53}\). Through one round of saturation mutagenesis on targeted active site residue in levopimaradiene synthase, the authors were able to obtain approximately 10-fold increase in titer, presumably through alteration of the synthase’s k\(_{\text{cat}}\).
Finally, identification and selection of transporters is an important consideration for attaining high pathway flux, as high intercellular concentrations of the target molecule can lead to product inhibition. While overexpression of native pumps have demonstrated large increases in production, particularly for amino acid biosynthesis$^{54,55}$, it will be important going forward to identify a wider set of pumps specific to more classes of molecules.

**Titer**
Titer is an important parameter as it can impact purification costs. For instance, if the product is nonpolar and can surpass a critical concentration, phase separation will occur. Physical separation between the product and growth media is highly advantageous, and can be aided by employing a two-phase system where a carrier liquid, such as dodecane, is added at the beginning of the fermentation to enable *in situ* product removal. On the other hand, if the product is polar, a higher concentration of product will still translate to lowered purification costs, as downstream separations will be more efficient.

In general, titers of at least 1 g/L are needed before examining production at a larger scale.

Titer depends on both how many cells are present and on the normalized production of each cell:

\[
titer = cell\ density \times specific\ production
\]

Since specific production is largely embodied by the kinetics of a pathway, which has already been discussed, methods to increase the cell density by decreasing toxicity or increasing tolerance are discussed below.

**Pumps/transporters**
Many products of interest elicit solvent stress by affecting membrane function and unfolding proteins$^{56}$. Creating an extracellular sink for these products by introducing transmembrane efflux pumps is thus an emerging strategy for mitigating this stress. Dunlop et al.$^{57}$ used a competition assay to identify pumps that improved *E. coli* survival when challenged with seven different candidate biofuels. When expressed in a limonene producing strain, one of these pumps led to better growth and higher limonene titer.

**Adaptive Evolution**
Adaptive evolution, most commonly accomplished by performing serial transfers of an organism into media containing increasing concentrations of the target molecule of interest, can also be leveraged to increase resistance. Several studies have demonstrated this technique with respect to alcohols$^{58,59}$ leading to strains that can withstand higher product concentrations.

**Toxic intercellular intermediates**
Pathway intermediates can also exhibit toxicity and decrease cell density, particularly when they are heterologous and the host has not had the evolutionary opportunity to select against deleterious interactions with native cellular processes$^{60}$. The heterologous
mevalonate pathway, transferred from *S. cerevisiae* to *E. coli*, is a good case study of this effect. Martin *et al.*⁶¹ observed that accumulation of many of the diphosphate intermediates in the bottom portion of the pathway led to significant retardation of growth, demonstrating the importance of the balancing the activities of the enzymes in this part of the pathway. Pfleger *et al.*⁶² additionally observed this effect with the upstream portion of the pathway. Taking a combinatorial approach to transcript expression, they found that only when the ratio of HMG-CoA reductase:HMG-CoA synthase was high, were mevalonate titers and cell densities correspondingly high. This suggested that accumulation of the HMG-CoA intermediate was detrimental to growth, which was substantiated in a subsequent report⁶³, mechanistically elucidated using microarrays and targeted metabolite profiling⁶⁴, and addressed by employing protein scaffolding that facilitated substrate channeling of the HMG-CoA intermediate⁶⁵.

### 1.4 Dissertation scope

In light of the introductory sections of this dissertation, it is clear that while kinetic approaches to pathway analysis are still being developed, they will be necessary to convert metabolic engineering to a truly predictive discipline. In this dissertation, we move towards that goal by endeavoring to develop tools aiding the rational and systematic engineering of metabolic pathways by analyzing pathway kinetics, then using this information to identify targets for kinetic improvement. Specifically, in Chapter 2 we demonstrate how the experimental measurement of protein levels can be utilized in ODE models that can be used in predictive capacities to enable identification of enzymes that have large impacts on pathway flux. Then, in Chapter 3, using -omics techniques we experimentally test predictions made by these models and find that they are surprisingly congruent with Chapter 2’s predictions. Having identified targets for improvement, Chapter 4 then focuses on developing a platform that can be used to improve the activities of these target enzymes. The entire metabolic engineering process as envisioned by these Chapters can be seen schematically in Figure 7.
Figure 7- An idealized metabolic engineering workflow

1 Primak et al., “Characterization of a Feedback-Resistant Mevalonate Kinase From the Archaeon Methanosarcina Mazei.”
17 Shen et al., “Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in *Escherichia Coli*.”
18 Bond-Watts, Bellerose, and Chang, “Enzyme Mechanism as a Kinetic Control Element for Designing Synthetic Biofuel Pathways.”
20 Kim and Copley, “Inhibitory Cross-Talk Upon Introduction of a New Metabolic Pathway Into an Existing Metabolic Network.”
21 Bond-Watts, Bellerose, and Chang, “Enzyme Mechanism as a Kinetic Control Element for Designing Synthetic Biofuel Pathways.”
40 A. Saltelli et al., Global Sensitivity Analysis: The Primer (Wiley, Chichester, UK, 2008)
54 Park et al., “Rational Design of Escherichia Coli for L-Isoleucine Production.”
55 Park et al., “Metabolic Engineering of Escherichia Coli for the Production of L-Valine Based on Transcriptome Analysis and in Silico Gene Knockout Simulation.”


Chapter 2-Kinetic modeling approaches to identifying targets for flux improvement in the mevalonate pathway

2.1 INTRODUCTION

Despite the publication of many second-generation cloning methods deemed to be panaceas over the past five years, metabolic engineering efforts are still highly labor and time intensive. This is due both to the difficulty of automating next generation cloning techniques, but also due to relatively immature modeling frameworks capable of linking human-specifiable inputs (DNA sequences) all the way to human-desired outputs (product concentrations).

In terms of models the processes of the central dogma, two main tactics have been taken: (1) mechanistic, biophysical models and (2) a characterization-based approach, where the performance of library of genetic devices statistically measured and used and generate

2.2 MATERIALS & METHODS

2.2.1 ODE MODELING
2.2.2 GLOBAL SENSITIVITY ANALYSIS MODELING
2.2.3 PROTEIN PURIFICATION & ACTIVITY VERIFICATION
3.2.4 IN VITRO PRODUCTION STUDIES

2.3 RESULTS

2.3.1 AN ODE MODEL OF THE BOTTOM PORTION OF THE MEVALONATE PATHWAY MAKES TESTABLE PREDICTIONS ABOUT PATHWAY FUNCTION
2.3.1.1 IMPACTS OF FPP FEEDBACK-LOOP
2.3.1.2 CAUSE OF DIFFERENCES IN AMORPHADIENE PRODUCTION
2.3.2 GLOBAL SENSITIVITY ANALYSIS CONFIRMS IMPORTANCE OF ADS LEVELS TO AMORPHADIENE PRODUCTION
2.3.3 IN VITRO PATHWAY RECONSTITUTION PROVIDES A SYSTEM FOR FUTURE TESTING OF MODEL
2.3.4 IN VITRO STUDIES POINT TO MEVPP-BASED PRODUCT INHIBITION OF PMK

2.4 DISCUSSION

2.5 REFERENCES
confidence intervals on expression of new constructs\textsuperscript{71,72}. However, these models have largely only been applied to transcription and translation. To fully predict the behavior of a strain, we will need models for the conversion of starting materials or metabolites into products of interest. These models need not necessarily be whole-cell models, which would be rather unwieldy; they only need to be complicated enough to describe the flux of a desired pathway of interest with reasonable fidelity. Additionally, engineering strategies as outlined in the introduction can help minimize interactions between the pathway of interest and host cell metabolism, further lending credence to simple models.

The main issue plaguing the development of kinetic models to this point has been parameter uncertainty\textsuperscript{73}. In particular, an unresolved question is when and to what extent constants derived from \textit{in vitro} enzymology studies can be applied to \textit{in vivo} conditions, where the buffer components, pH, and molecular crowding\textsuperscript{74,75} may exert an effect on activity.

One approach that is gaining appeal is to allow parameters to fluctuate about some prescribed values, generating an ensemble of computational models that can then be compared against relevant experimental data\textsuperscript{76}.

Here, we take a three-pronged approach to try to obtain a mevalonate pathway model with reasonable predictive ability:

1. We cull steady-state, Michaelis-Menten parameters from the literature, and enzyme concentrations from experimental data, to obtain nominal starting values for a base computational model for mevalonate pathway flux

2. We perform a global sensitivity analysis to analyze dependence of model output (amorphadiene production) on input parameters

Figure 8-Linking inputs and outputs from different modeling domains
(3) Using purified enzyme, we reconstitute the pathway *in vitro* to validate up model assertions

We generate hypothesis about pathway function that are then examined experimentally in Chapter 3.

### 2.2 Materials & Methods

#### 2.2.1 ODE modeling

An ordinary differential equation model was built in Mathematica (Wolfram Research) using the kinetic parameters given in Table 2 and enzyme concentrations determined by quantitative proteomic analysis (Table 3). Initial metabolite concentrations were set to 100nM based on two factors (1) experiments performed on un-engineered DH1 cells were unable to detect any mevalonate pathway intermediates, suggesting they are under the limit of detection and are thus very low abundance species and (2) manual testing of the model demonstrated it to be insensitive to at least 10-fold changes of all intermediate species. The mevalonate concentration was set at (constant) 5mM, the average of experimental LC-MS data for the intracellular concentration over the time period of the model, which was 8 hours. Since *apparent* $K_m$’s were reported for all reactions utilizing ATP (i.e. $K_m$’s were not determined as a function of [ATP]), the ATP concentration was assumed to be constant and saturating for the entire simulation time.
Table 2-Kinetic parameters used in modeling studies

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m1$ (M)</th>
<th>$K_m2$ (M)</th>
<th>$K_i$ (M)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. cerevisiae</em> mevalonate kinase</td>
<td>38</td>
<td>1.31x10$^{-6}$</td>
<td>1.9x10$^{-6}$</td>
<td></td>
<td>Primak et al.$^{77}$</td>
</tr>
<tr>
<td><em>S. aureus</em> mevalonate kinase</td>
<td>6</td>
<td>41x10$^{-6}$</td>
<td>46x10$^{-6}$</td>
<td></td>
<td>Voynova et al.$^{78}$</td>
</tr>
<tr>
<td><em>S. cerevisiae</em> phosphomevalonate kinase</td>
<td>10</td>
<td>885x10$^{-6}$</td>
<td></td>
<td></td>
<td>Garcia et al.</td>
</tr>
<tr>
<td><em>S. cerevisiae</em> mevalonate diphosphate decarboxylase</td>
<td>4.9</td>
<td>123x10$^{-6}$</td>
<td></td>
<td></td>
<td>Krepkiy et al.$^{79}$</td>
</tr>
<tr>
<td><em>E. coli</em> Isopentenyl-diphosphate isomerase</td>
<td>0.33</td>
<td>8x10$^{-6}$</td>
<td></td>
<td></td>
<td>Hahn et al.$^{80}$</td>
</tr>
<tr>
<td><em>E. coli</em> farnesyl pyrophosphate synthase (reaction 1)</td>
<td>0.21</td>
<td>1.3x10$^{-6}$</td>
<td>29.3x10$^{-6}$</td>
<td></td>
<td>Ku et al.$^{81}$</td>
</tr>
<tr>
<td><em>E. coli</em> farnesyl pyrophosphate synthase (reaction 2)</td>
<td>0.47</td>
<td>10.3x10$^{-6}$</td>
<td>5.5x10$^{-6}$</td>
<td>74x10$^{-6}$</td>
<td>Ku et al.$^{82}$</td>
</tr>
<tr>
<td><em>A. annua</em> Amorphadiene Synthase</td>
<td>6.8x10$^{-3}$</td>
<td>3.3x10$^{-6}$</td>
<td></td>
<td></td>
<td>Picaud et al.$^{83}$</td>
</tr>
</tbody>
</table>

Table 3-Enzyme concentrations used in modeling studies

<table>
<thead>
<tr>
<th>Strain</th>
<th>Enzyme</th>
<th>mbs3 (μM)</th>
<th>saMK3 (μM)</th>
<th>10kADS (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>scMK</td>
<td>66.3</td>
<td></td>
<td>197.0</td>
<td></td>
</tr>
<tr>
<td>saMK</td>
<td></td>
<td>23.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scPMK</td>
<td>2.9</td>
<td>4.9</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>scPMD</td>
<td>2.2</td>
<td>2.7</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>ecIDI</td>
<td>18.4</td>
<td>15.3</td>
<td>21.6</td>
<td></td>
</tr>
<tr>
<td>ecISPA</td>
<td>35.0</td>
<td>36.8</td>
<td>24.2</td>
<td></td>
</tr>
<tr>
<td>aaADS</td>
<td>39.6</td>
<td>25.6</td>
<td>677.5</td>
<td></td>
</tr>
</tbody>
</table>
2.2.2 Global sensitivity analysis modeling
The method of Carothers et al. (2011)\textsuperscript{84} was used. Briefly, a custom perl script was written to generate random parameter sets sample from a uniform distribution ranging 10-fold above and 10-fold below each of the Michaelis-Menten and enzyme concentration parameters. A set of 10,000 of these models was simulated on a server cluster using Dizzy, a Java-based chemical kinetics simulation program\textsuperscript{85}, generating a 10,000 product outputs. The sensitivity to each parameter was analyzed with partial rank correlation coefficients using the R statistical programming language.

2.2.3 Protein purification & activity verification
\textit{S. cerevisiae} mevalonate kinase (MK), \textit{S. cerevisiae} phosphomevalonate kinase (PMK), and \textit{A. annua} amorphadiene synthase (ADS) were purified as previously described\textsuperscript{86}. \textit{S. cerevisiae} mevalonate diphosphate decarboxylase (PMD) and \textit{E. coli} isopentenyl-diphosphate isomerase (IDI) were purified using a phytic-acid dependent, autoprocessing enzyme-tag purification system.\textsuperscript{87} \textit{E. coli} farnesyl pyrophosphate synthase (ISPA) was purified with an N-terminal his tag.

Activity for PMD was measured by spectrophotometrically monitoring ATP consumption with a coupled assay\textsuperscript{88}. ISPA activity was confirmed by combining ISPA (0.2mg/mL final concentration) with 100μM DMAPP and 200μM IPP in 25mM Tris, 15mM MgCl\textsubscript{2} reaction buffer, reacting for 1 hour at 37°C, quenching with an equal volume of 100% methanol, and monitoring FPP appearance via LC-MS.

3.2.4 \textit{in vitro} production studies
The \textit{in vitro} production studies were conducted a buffer designed to mimic \textit{in vivo} conditions\textsuperscript{89}. Enzyme concentrations were chosen to match those of the base ODE model.

<table>
<thead>
<tr>
<th>Component</th>
<th>K+</th>
<th>PO\textsubscript{4}</th>
<th>Glu</th>
<th>Na+</th>
<th>Mg\textsubscript{2+}</th>
<th>Ca\textsubscript{2+}</th>
<th>Cl\textsuperscript{-}</th>
<th>Sulfate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration (mM)</td>
<td>200</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>[ATP]+5mM</td>
<td>.001</td>
<td>150</td>
<td>5</td>
</tr>
</tbody>
</table>

2.3 Results

2.3.1 An ODE model of the bottom portion of the mevalonate pathway makes testable predictions about pathway function
To test the validity of the model, predictions need to be made that can be compared to experiment. However, because it is currently a more difficult and a less reliable process to experimentally program in \textit{ad hoc} protein expression, we use the experimentally determined enzyme concentrations from Chapter 3 as the starting point for exploring the modeling behavior of this chapter. This is shown schematically in Figure 9.
2.3.1.1 Impacts of FPP Feedback-loop

While it is not the first committed step of the mevalonate pathway, the FPP-mediated inhibition of mevalonate kinase resembles the classic biochemical strategy of end-product inhibition\textsuperscript{90}. It is unknown, however, whether this feedback loop is strong enough to significantly affect flux to the final product, amorphadiene. This question was explored computationally; Figure 10 & Figure 11 shows the results, which indicate that there is ample precursor supply for the downstream enzymes of the pathway even if the feedback loop is functional. Specifically, at the nominal value for $K_i FPP, s. cerevisiae MK = 1.9 \mu M$, the feedback loop for base strain mbis3 is functional, while it is not functional for strain 10kADS, yet they have similar rates of amorphadiene production (the small discrepancy that they do show will be explained in the next section). Additionally supporting this assertion is that when the inhibition constant is titrated 1000-fold lower and higher within each strain, the rate of amorphadiene production again stays nearly constant, even through the switch-like transition that turns the feedback on or off (as measured by the rate of MevP synthesis, red markers).
The cause of the effect can be traced back to the FPP levels in each strain. Since the 10kADS strain has higher ADS levels and consequently lower FPP (its substrate) concentrations, its MK is uninhibited, while the converse holds for the mbis3 strain (Figure 12, Figure 13).
As just demonstrated, the presence of the feedback loop doesn’t have a strong effect on amorphadiene production, but it does have a sizable effect on MevP levels (Figure 14). Qualitatively the trends make sense, but quantitatively, the molar quantities of MevP predicted by the model would pose too much osmotic stress on the cell, likely resulting in lysis. This points to some missing feature of the pathway model. Given that MevP is the accumulating species, one possibility is that there is undocumented product inhibition by the MevP on mevalonate kinase.
Indeed, when a phenomenological power-law expression for MevP-mediated product inhibition of MK is introduced to the kinetic equations (Equation 1), MevP levels fall (Figure 15).

**Equation 1**

\[
\frac{d[MevP]}{dt} = \left(\frac{[MevP]}{0.005}\right)^p \frac{k_{cat} \times MK \times [Mev]}{K_m \times (1 + \frac{FPP}{K_{i,FPP\_MK}}) + [Mev]}
\]
The analysis on MevP is somewhat complicated by the fact that, because of experimental imprecision (likely some combination of transcriptional and translational coupling effects from the stronger ribosome binding site upstream of ADS), the 10kADS strain expresses more MK than the base mbis3 strain, and thus ADS levels aren’t the only variable quantity between the two strains. To test whether increasing MK expression by itself could lead to the observed MevP levels, the simulation for the base mbis3 strain was rerun, but with the increased levels of MK to match those found in the 10kADS strain, while leaving ADS (and by proxy FPP levels) unchanged. Figure 16 shows that while slightly more MevP is accumulated, the levels are still more than two orders of magnitude below the 10kADS case, thus supporting the notion that FPP, and not MK, is the controlling factor in MevP accumulation.
Figure 16: Effect of increasing MK levels in mbis3

Thus the model makes two predictions about the feedback loop:

1. FPP and MevP levels are inversely correlated
2. The feedback loop has no effect on the production of the final product, amorphadiene

2.3.1.2 Cause of differences in amorphadiene production

If the feedback loop doesn’t have an effect on amorphadiene production, what does? Figure 17 shows the predicted rate of amorphadiene production for the strains used in the experimental studies of Chapter 3.
As can be seen, amorphadiene production is dependent on ADS expression levels, but as Figure 18 shows, with diminishing returns as ADS increases. This suggests the precursor supply is no longer saturating ADS (Figure 19) and that flux control is shifting to other enzymes. Manual inspection of the model predicts that as ADS increases into this regime, more control is centered on Idi and IspA levels (data not shown).
Figure 18-Nonlinear response of amorphadiene production to ADS

Figure 19-FPP concentrations in mbis3 as ADS is increased
2.3.2 Global sensitivity analysis confirms importance of ADS levels to amorphadiene production

Based on the expression regime for base strain mbis3, the previous section suggested that amorphadiene production was most strongly dependent on ADS levels. However, one might ask if this result would still be the same if any of the parameters of the model were to change, due to, for example, measurement inaccuracies.

One way to answer this type of question is to allow the parameters of the model to vary randomly, generating an ensemble of models, whose outputs can be measured and statistically mapped back to the parameter sets that generated them; this is a global sensitivity analysis\(^9\).  

Shown in Figure 20 is the result of that analysis. When the data are mapped back to the generating sets of parameters through a partial rank correlation coefficient, the result is Figure 21. It shows that kcat7 and ADS have the highest coefficients (also equal, due to their cognate parameters’ mathematically equivalence), indicating that when amorphadiene titers are high, these parameters are most often high as well.

![Image of scatter plot representing global sensitivity analysis results]

*Figure 20*-Amorphadiene titer for 1000 simulations of the pathway
This finding again confirms in a more global sense (at least in a space that spans 10-fold one-sided changes in the parameters) that the control of the pathway is mostly on ADS.

### 2.3.3 In vitro pathway reconstitution provides a system for future testing of model

As a bridge to *in vivo* experiments, and as a model system more amenable to deconstruction and simplification, an *in vitro* model was constructed. Each of the enzymes in the pathway was purified, as shown in the SDS-PAGE gel in Figure 22 and as documented in Table 5, and each protein's activity was confirmed by a coupled assay or GC-MS (data not shown).
Table 5-Concentrations of purified mevalonate pathway proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Concentration (mg/mL)</th>
<th>Concentration (uM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. cerevisiae</em> MK</td>
<td>9.5</td>
<td>196</td>
</tr>
<tr>
<td><em>S. cerevisiae</em> PMK</td>
<td>4.9</td>
<td>97</td>
</tr>
<tr>
<td><em>S. cerevisiae</em> PMD</td>
<td>6.9</td>
<td>156</td>
</tr>
<tr>
<td><em>E. coli</em> IDI</td>
<td>6.8</td>
<td>332</td>
</tr>
<tr>
<td><em>E. coli</em> ISPA</td>
<td>6.3</td>
<td>196</td>
</tr>
<tr>
<td><em>A. annua</em> ADS</td>
<td>22</td>
<td>344</td>
</tr>
</tbody>
</table>

When all the enzymes are combined *in vitro*, the pathway is functional as demonstrated by the production of amorphadiene shown in Figure 23.
2.3.4 *In vitro* studies point to MevPP-based product inhibition of PMK

As previously mentioned, the *in vitro* system is useful for deconstructing the pathway and examining its constituent reactions either individually or in conjunction with other enzymes.

The high concentrations of phosphorylated mevalonate derivatives in the model led to the hypothesis that there might be undocumented product inhibitions for MK and PMK. Figure 24 shows a timecourse for the reaction of mevalonate kinase with mevalonate. Although the reaction goes to completion and therefore there doesn’t appear to be product inhibition, it may exist but not be manifest at concentrations in the low millimolar range, where this assay probed.

On the other hand, product inhibition is readily identifiable when MK is combined with PMK (Figure 25). To determine whether this was due to inhibition of MK or PMK, a timecourse was performed on the latter (Figure 26), which similarly leads to an incomplete reaction. The most likely explanation, then, is that PMK is strongly inhibited by its product, MevPP.
Figure 24 - *in vitro* MK timecourse

Figure 25 - *in vitro* MK+PMK timecourse
2.4 Discussion

The base model demonstrated that amorphadiene production was most sensitive to ADS expression/catalytic activity. However, the discrepancy between the Vmax and actual rate of amorphadiene production as ADS is increased is a reminder that not all control typically resides on one enzyme in a pathway. In particular, as the pathway moves from the concentration regime of base strain mbis3 to strain 10kADS, more control appears to shift to IDI and ISPA, as the rate of amorphadiene production diverges from $V_{\text{max, ADS}}$.

Also, while the negative feedback loop from FPP-mediated inhibition of mevalonate kinase does function to attenuate MevP production when sufficient FPP accumulates ($>>K_i$), the attenuation isn’t great enough to affect the precursor supply for the downstream enzymes, and thus doesn’t result in a difference in amorphadiene titer. On the other hand it does have a large effect on MevP levels, which accumulate to unrealistic concentrations.

As indicated by simulation, these MevP concentrations can be decreased with generic inhibition terms. In vitro studies support the existence of product inhibition for PMK, and cannot rule it out for MK at higher product concentrations. These results point to the importance of using the reversible form of the Michaelis-Menten equation for metabolic engineering simulations, as the high amount of flux diverted to and mismatch between catalytic efficiencies in heterologous pathways will often lead to accumulation of high amounts of pathway intermediates that will lead to enzymatic behavior not accounted for in initial-rate enzymological studies (where, by definition product concentration is 0). Thus, to accurately predict the dynamics of intermediates, more parameters will need to be measured (namely inhibition constants and constants from the reverse reaction).
However, there is still great utility in the model as it stands; for instance, it should be possible to walk through enzyme concentration space, iteratively increasing the expression of rate limiting enzyme(s) to increase the pathway flux.

2.5 References

79 Dmitriy Krepkiy and Henry M Miziorko, “Identification of Active Site Residues in Mevalonate Diphosphate Decarboxylase: Implications for a Family of Phosphotransferases,”


Chapter 3- *In vivo* characterization of heterologous mevalonate pathway function: a quantitative -omics based approach

3.1 INTRODUCTION

3.2 MATERIALS & METHODS
3.2.1 CLONING/PLASMID CONSTRUCTION
3.2.2 GROWTH AND MEDIA
3.2.3 QCONCAT PROTEIN PURIFICATION
3.2.4 PROTEOMICS
3.2.5 METABOLITE ANALYSIS OF MEVALONATE PATHWAY INTERMEDIATES (LC-MS)
3.2.6 AMORPHADIENE PRODUCTION (GC-MS)
3.2.7 CONVERSIONS TO CELLULAR CONCENTRATIONS

3.3 RESULTS
3.3.1 CREATION OF A SYNTHETIC POLYPEPTIDE STANDARD FOR QUANTITATIVE PROTEOMICS AND MEASUREMENT OF ABSOLUTE CONCENTRATIONS OF PATHWAY ENZYMES
3.3.2 BUILD-UP OF FPP ENABLES DETERMINATION OF *IN VIVO* $K_{cat}$ FOR AMORPHADIENE SYNTHASE
3.3.3 COMPARISON OF MODEL PREDICTIONS WITH EXPERIMENTAL DATA
3.3.3.1 RATES OF AMORPHADIENE PRODUCTION
3.3.3.2 EVIDENCE FOR A FEEDBACK LOOP

3.4 DISCUSSION

3.5 REFERENCES
3.1 Introduction

While considerable engineering effort has gone into improving production in the mevalonate pathway since its introduction in the Keasling lab over 10 years ago, the yield still hovers below 50% and is often unstable in replicate studies (likely due to toxicity imparted by intermediate accumulation). This lack of progress can be partly attributed to the unavailability of a model characterizing pathway flux as a function of experimentally relevant control variables. Specifically, there is currently no way to map a change in enzyme expression levels to changes in pathway output.

As such, this Chapter endeavors to more fully characterize pathway function by specifically interrogating pathway kinetics in a system largely independent from the rest of cellular physiology. This is accomplished by feeding exogenous mevalonate to strains encoding different versions of the bottom part of the mevalonate pathway (Figure 27). Protein levels are measured and are unchanged through the course of the analysis, and since the only metabolite linking pathway function to the rest of cellular metabolism is ATP, whose levels are saturating throughout the experiment, pathway function can be decoupled and studied independently from cellular metabolism, which typically complicates analysis, as when glucose is used as the carbon source. Thus, for the purposes of this Chapter, the cell has largely been rendered a bag of enzymes.

![Figure 27-Experimental system for examining pathway kinetics](image)

Three specific hypothesis predicated by the ODE model are tested here

(1) FPP-mediated feedback inhibition of mevalonate kinase is insufficient to significantly affect product titers, thus there should be little difference between amorphadiene production in the mbis3 and sa3 strains

(2) An increase in ADS expression levels should lead to an increase in amorphadiene production concomitant with model prediction
(3) *In vitro* kinetic parameters accurately match their *in vivo* counterparts

![Diagram](image)

Figure 28 – Strains used to test model predictions

To accomplish this, absolute concentrations of all relevant enzymes and metabolites must be measured. While commercial sources are available for the latter that can be used to generate standard curves for LC-MS studies, there is no commercial equivalent for the former. With this in mind, a synthetic polypeptide concatenating the best performing peptides for all pathway proteins was purified from a synthesized gene and used to construct a standard curve.

### 3.2 Materials & Methods

#### 3.2.1 Cloning/plasmid construction

The pMBIS3 plasmid was constructed by F. Nowroozi and was used as the backbone for the Circular Polymerase Extension Cloning (CPEC)-based insertion of a *M. mazei* mevalonate kinase, which was amplified by PCR with 30bp vector overlaps. The pMBIS3_10Kads plasmid was constructed with round-the-horn mutagenesis with a forward primer containing a 10k-strength ribosome binding site (as calculated by the RBScalculator\(^{92}\)) and a reverse primer binding just upstream of the RBS. Prior to PCR, the primers were phosphorylated with Polynucleotide Kinase (PNK), which provided phosphates for the blunt-end, circularization ligation reaction performed on the PCR amplicon after a gel purification clean-up.

The mevalonate pathway qconcat gene was synthesized (Genscript) with flanking 5’ Ndel and 3’ Xhol sites. The gene was cloned into a T7 expression plasmid, pET29b, via the aforementioned sites.

#### 3.2.2 Growth and media

Plasmids were freshly transformed into DH1 cells for each experiment. Colonies were picked the next day and used to inoculate overnight cultures in LB. The following day,
200μL overnight culture was diluted in 50 mL LB supplemented with 50mg/L carbenicillin and 0.5% (w/v) glucose and induced with 1mM IPTG when OD reached 0.6. Two hours later, 2mL 1M mevalonate was added (40mM final concentration) and 20% dodecane was overlayed to capture the volatile sesquiterpene. Time points were sampled every two hours as follows: 0.25mL culture was used to measure OD, and 1.5mL of culture was harvested through centrifugation (8000g | 8 min | 4°C) for various analyses. For extracellular metabolites, 250μL of the supernatant was added to 250μL ice cold methanol, filtered through a 3 kDa membrane, and stored at -20°C until LC-MS analysis. The remaining supernatant was discarded and the resultant pellet was resuspended in 300μL ice-cold methanol, and stored at -20°C until work-up on LC-MS and MS for intracellular metabolites and proteins, respectively.

3.2.3 Qconcat protein purification
The Qconcat protein was purified under denaturing conditions using Nickel immobilized metal affinity chromatography (Qiagen). It was stored in 50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0 buffer, and buffer exchanged via 3kDa spin filter (Amicon) into 100mM ammonium bicarbonate prior to trypsin digestion.

3.2.4 Proteomics
240μL of water and 150μL of chloroform was added to the 300μL methanol pellet, vortexed, and centrifuged (14000g | 10min | 4°C) to create 3 phases --540 mL of aqueous (methanol/water) on top, a protein interface, and 150μL chloroform on the bottom. The top layer was harvested for metabolite analysis, while the remaining protein and chloroform mixture was resuspended in 300μL methanol, vortexed, and centrifuged (14000g | 2min | RT). The supernatant was discarded and the pellet was dried in the speed-vac for 45 min.

The dried protein pellet was resuspended in 10μL methanol and 90μL of 100mM ammonium bicarbonate. This solution was vortexed and the concentration measured by a colorimetric method similar to the lowry assay (DC Protein Assay, BioRad), using BSA to construct a standard curve. Protein samples were normalized to 1mg/mL in 100mM ammonium bicarbonate and 10% (v/v) methanol. The sample was then reduced with 5mM TCEP (final concentration) for 30min at room temperature, and cysteines were subsequently blocked with 4mM (final concentration) iodoacetic acid. The samples were digested with for 16 hours at 37°C with 1/50 the concentration of sample protein, and frozen until analysis.

Proteomic Selected Reaction Monitoring (SRM) was performed on a LC-MS/MS system as previously described⁹³.

3.2.5 Metabolite analysis of mevalonate pathway intermediates (LC-MS)
Chemical standards were made up to 200 μM, as the stock solution, in methanol-water (50:50, v/v). The separation of the mevalonate pathway intermediates was conducted on a ZIC-pHILIC column (150 mm length, 4.6-mm internal diameter, and 5-μm 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). The sample
injection volumes was 3 µL. The temperature of the sample tray was maintained at 4°C using an Agilent FC/ALS Thermostat. The column compartment was set to 40°C. The mobile phase was composed of A) 10 mM ammonium carbonate and 0.5% ammonium hydroxide in acetonitrile-water (2:8, v/v) and B) 10 mM ammonium carbonate and 0.5% ammonium hydroxide in acetonitrile-water (8:2, v/v). Mevalonate pathway intermediates were eluted isocratically with a mobile phase composition of 33% mobile phase A and 67% of mobile phase B. A flow rate of 0.45 mL/min was used throughout.

The HPLC system was coupled 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 using 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 0.86 spectra/second 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). Data acquisition and processing were performed by the MassHunter software package. Mevalonate pathway intermediates from E. coli extracts were quantified via eight-point calibration curves ranging from 625 nM to 200 µM. The R² coefficients for the calibration curves were ≥0.99.

### 3.2.6 Amorphadiene production (GC-MS)

Amorphadiene production was measured by GC-MS by diluting 10 µL of the dodecane overlay into 990 µL ethyl acetate spiked with a caryophyllene internal standard.

The GC-MS method was employed as previously described.

Amorphadiene concentrations were quantified by comparing caryophyllene-normalized sample peak areas to those of an authentic (amorphadiene) standard curve.

### 3.2.7 Conversions to cellular concentrations

To convert concentrations measured in culture volume to intracellular concentrations, the method of Volkmer and Heinemann was used to generate the following formulae:

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Correction factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular metabolites</td>
<td>$C_{intracellular} [M] = \frac{C_{LC-MS}[\mu M]}{OD_{600} \times 3.6 \times 10^5}$</td>
</tr>
<tr>
<td>Proteins</td>
<td>$C_{intracellular} [M] = \frac{m_{peptide, MS}[pmol] \times 0.324}{OD_{600} \times m_{totalprotein}}$</td>
</tr>
</tbody>
</table>
3.3 Results

3.3.1 Creation of a synthetic polypeptide standard for quantitative proteomics and measurement of absolute concentrations of pathway enzymes

To quantitate the proteins of the mevalonate pathway, a 31kDa Qconcat polypeptide constituted of the peptides listed in Table 7, with trypsin sites inserted between each, was gene synthesized, expressed, and purified.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mevalonate pathway Qconcat</td>
<td>atoB, scHMGS, saHMGS, scHMGR, scMK, mmMK, saMK, scPMK, scPMD, ecIDI, ecNUDB, ecISPA, agGPPS, msLS, agBIS, AaADS,</td>
</tr>
</tbody>
</table>

Error! Reference source not found. Figure 29 is an SDS-PAGE gel of the protein, which was purified under denaturing conditions, and by inspection is more than 95% pure.

![Figure 29-Qconcat protein](image)

The peptides from the Qconcat were verified on a triple-quadrupole mass spectrometer, selecting for the appropriate mass/charge ratios. A linear dilution series was performed from which the calibration curves in Figure 30 were derived. As can be seen, almost all fragments produce excellent fits to linear curves (18 out of 19 have R^2 ≥ .97).
Using the equations derived from these fits, the absolute concentrations of the proteins in the bottom portion of the mevalonate pathway were determined, as shown in Figure 31.
These values were used for the modeling studies described in Chapter 2 and the determination of *in vivo* parameters described in the next section.

### 3.3.1.1 Build-up of FPP enables determination of *in vivo* $k_{\text{cat}}$ for amorphadiene synthase

Intracellular metabolite data was collected by LC-MS, and using the factors in Table 6, corrected to intracellular concentrations, as shown in Figure 32. Also plotted is the $K_m$ for Amorphadiene synthase, which is 3 μM. The intracellular concentrations range from 30-fold (10kADS strain) to ~1000-fold (saMK3 strain) above $K_m$ for all time points, thus ADS is reliably saturated for the time-span of the experiment.

![Intracellular FPP levels in engineered mbis strains](image)

**Figure 32-Intracellular FPP levels in engineered mbis strains**

Having determined that the velocity is in an asymptotic regime from the FPP data, Equation 2 applies.

**Equation 2**

$$v = V_{\text{max,ADS}} = k_{\text{cat,ADS}} \times [\text{ADS}]$$

Thus, by using amorphadiene production rates derived from the specific amorphadiene production (M/cell) shown in Figure 33 along with the ADS concentrations determined by proteomics, an *in vivo* $k_{\text{cat,ADS}}$ can be determined, which is compared with the *in vitro* value in Figure 34.
3.3.2 Comparison of model predictions with experimental data
Several predictions were made in Chapter 2 about mevalonate pathway function under different genetic perturbations:
• A feedback loop was predicted to be manifest from the action of FPP in both mbis3 and SA3 strains, although their Ki’s weren’t sufficient to distinguish between the two in terms of metabolite accumulations. On the other hand, the 10kADS strain, by decreasing FPP concentration beyond a critical point, should accumulate more MevP. Thus 10kADS strain should accumulate more MevP as a result of lower FPP levels, while the mbis and SA3 strains should accumulate less MevP as the result of higher FPP levels.

• In terms of amorphadiene production, 10kADS should produce at a greater rate, but with diminishing returns, as FPP synthase and IPP isomerase gain more control as ADS is increased (i.e. a 10-fold increase in ADS expression should only lead to a 3-5 fold increase in rate of production).

3.3.2.1 Rates of amorphadiene production
Figure 35 shows the rates of amorphadiene production for experimental data as compared to predictions from the model using (1) the in vitro $k_{cat}$ for ADS (orange) and (2) the in vivo $k_{cat}$ (dark grey) as determined in the previous section. The in vivo $k_{cat}$ values, all within experimental error, yield less aberrant predictions than their in vitro counterparts.

![Predicted vs. experimental rates of Amorphadiene production](image_url)

**Figure 35-** Computationally predicted and experimentally derived rates of amorphadiene production
3.3.2.2 Evidence for a feedback loop

As expected, the mbis3 and saMK3 strains have much lower concentrations of MevP due to the action of feedback inhibition (Figure 36). Additionally supporting the hypothesis of feedback inhibition is the inverse correlation between MevP and FPP observed for the strains (Figure 37).

**Figure 36-Intracellular MevP levels in engineered mbis strains**

**Figure 37-Inverse correlation between MevP and FPP for engineered mbis strains**
3.4 Discussion

The experimentally observed rates of amorphadiene production are in good agreement with those predicted by the model. The accuracy is perhaps due to the fact that in the current expression regime, ADS activity is limiting amorphadiene production, and thus only one measurement (the concentration of ADS) needs to be made to provide accurate predictions. Had the pathway been in a regime where MK or PMK had been limiting, the model may not have been as accurate, due to undocumented/un-parameterized product inhibition, as explained in the previous chapter. As the model predicts that the next limiting factors will be ISPA and IDI, it would be interesting to see if this is experimentally borne out.

Also clear from this work is the importance of performing experiments with standards so absolute concentrations can be determined. For instance, up to this point, metabolic engineers had only used integrals of peak areas to make relative comparisons of protein expression between strains. This can lead to spurious conclusions mainly due to unpredictable differences in ionization efficiencies between peptides. For example, the discrepancy between relative and absolute measures is most notable when for phosphomevalonate kinase, which was previously believed to be very under-expressed, but when accurately measured, is in fact in more than ample supply. Thus, in the future all proteomic studies should be conducted with Qconcat proteins for quantitation. This additionally helps correct for differences in detector sensitivities, so data from multiple instruments in different labs can be compared. And finally, this in turn provides more data for model building, which will hopefully result in a virtuous cycle of experimentation and computation.

3.5 References

Chapter 4-Towards a platform for directed evolution of targeted enzymes

4.1 INTRODUCTION

4.2 MATERIALS & METHODS

4.2.1 CLONING/PLASMID CONSTRUCTION
4.2.2 STRAIN CONSTRUCTION
4.2.3 GROWTH AND MEDIA
4.2.4 SURFACE DISPLAY
4.2.5 PYROPHOSPHATE-BASED ASSAY
4.2.6 COA-BASED ASSAY
4.2.7 CHIP EXPERIMENTS
4.2.8 PROTEIN PURIFICATION

4.3 RESULTS

4.3.1 DEVELOPMENT OF A PYROPHOSPHATE ASSAY FOR TERPENE SYNTHASE FUNCTION
4.3.2 MINIATURIZED ASSAY FOR PYROPHOSPHATE DETECTION ON A MICROFLUIDIC DEVICE
4.3.3 A YEAST SURFACE DISPLAY SYSTEM FOR DROPLET-BASED SCREENING
4.3.4 DEVELOPMENT OF A THIOL-REACTIVE FLUORESCENCE ASSAY FOR ESTERIFICATION REACTIONS
4.3.5 CONSTRUCTION OF THIOESTERASE-DEFICIENT SCREENING STRAIN FOR WAX ESTER SYNTHASE ACTIVITY
4.3.6 FLUORESCENT PROTEIN-WAX ESTER SYNTHASE FUSIONS FOR NORMALIZED ACTIVITY SCREENING

4.4 DISCUSSION

4.5 REFERENCES

4.1 Introduction

With the identification of amorphadiene synthase as the most sensitive node in the mevalonate pathway, the natural question becomes how to improve its activity. As the maximal rate of product formation in the steady-state for an enzymatic reaction is proportional to both enzyme concentration and turnover-number ($k_{cat}$), two strategies are available.

$$V_{max} = [E] \cdot k_{cat}$$

Equation 3
In the first strategy, activity can be increased by increasing enzyme expression level. This can typically be accomplished through some combination of transcriptional (e.g. stronger promoter, increased mRNA stability) and/or translational (e.g. alteration of ribosome binding site) engineering. However, based on the amount of protein already overexpressed in the pathway and the physio-chemical constraints of the cell (both cell volume, additional energy needed for protein synthesis, and protein solubility), this strategy is likely of limited value. Particularly, preliminary experiments placing bisabolene synthase (with similar kinetic parameters to amorphadiene synthase) on a T7 promoter failed to show an improvement in production (personal communication Taek Soon Lee). This could be due to a solubility limit or perhaps to a limit of heterologous protein able to be expressed in E. coli. Both hypotheses suggest a strategy of improving catalytic efficiency, rather than enzyme expression level, as the most tenable solution to improving flux. This option seems even more sensible when downstream industrial applications are considered, in which the pathway would be genomically integrated rather than plasmid-borne, due to issues of genetic stability. While plasmids offer copy numbers of ≥ 300/cell, genomic integration limits copy number to only a few/cell. Thus, many arguments point to engineering catalytic efficiency rather than just increasing expression level as the most fruitful long-term strategy for increasing pathway flux.

For this reason, a directed evolution approach to increasing $k_{cat}$ was pursued here. In this way, flux through the pathway could be increased at equivalent or perhaps lower enzyme expression levels, mitigating many of the concerns present with overexpression strategies.

To evolve an enzyme, an appropriate screen or selection is necessary. However, since amorphadiene (and any sesquiterpene in general) is a hydrocarbon consisting of no functional groups, it is difficult to develop a chemical or biological sensor to directly detect its production. However, researchers have used a coupled assay based the release of pyrophosphate (PPI) to measure steady state kinetics of bisabolene synthase, and this assay should be amenable to screening in general.

To increase the likelihood of finding catalytically improved variants from a mutagenic library, the throughput of the screen needs to be high. Droplet-based microfluidics offer fluorescence-based sorting in the kilohertz range, enabling the screening of approximately 10 million variants in a day; thus, our screens were prototyped in the bulk phase with an eye to moving eventually to the microscale.

We determined that measuring activity in cell lysate interfered with the PPI assay by an undetermined mechanism, so instead a yeast surface display system was pursued that would obviate the need for lysis. While we were able to develop a functional on-chip assay with purified enzyme, we were unable to surface display amorphadiene synthase or any other enzyme of interest.

In search of alternative enzymes limiting flux through other pathways of interest, we came across Wax Ester Synthase/Acyl Coenzyme A:Diacylglycerol Acyltransferase (WES), which catalyzes the terminal step of the fatty acid pathway producing biodiesel. Experimental evidence demonstrates significant build-up of the substrates for the enzymes, supporting the need for improving catalytic efficiency.
To develop a screen for WES activity, we took a similar tact to the pyrophosphate assay, detecting the co-product, Coenzyme A, (CoA), which provides a convenient fluorescence-based read-out. Furthermore, we determined this assay was compatible with cell lysate. We then constructed a background thioesterase knock-out strain to attenuate competing reactions and purified the enzyme to further prototype the screen. With this progress, we are well-positioned to carry out the directed evolution of WES.

4.2 Materials & Methods

4.2.1 Cloning/Plasmid construction
For creation of the surface display series, the appropriate genes (Appendix) were amplified by PCR using forward primers containing the KpnI restriction site and reverse primers containing the XhoI restriction site. These PCR amplicons were, along with the pYD1 backbone, digested with KpnI and XhoI, gel-purified, ligated, and transformed into tg1 cells.

The wax-ester synthase (WES) expression vector was created by PCR amplying the AtfA gene with an NdeI restriction site followed by a 6X-his tag incorporated into the non-annealing (5’) end of the forward primer and a BamHI restriction site incorporated into the 5’ end of the reverse primer. The amplicon was cloned into a pET30a expression vector.

The mCherry-WES variants were creating via circular polymerase extension cloning (CPEC)\textsuperscript{100}, with 30bp of backbone-overlapping sequence incorporated into the forward and reverse primers amplifying mCherry.

4.2.2 Strain construction
The triple-thioesterase knockout strain was created by moving the TesA and TesB kanamycin-gene disruptions from the appropriate Keio collection strain via P1 phage transduction, flipping out the antibiotic cassette with the pcp20 plasmid between each transduction. The FadM locus was subsequently disrupted by the Datsenko-Wanner method.

4.2.3 Growth and Media
E. coli Tg1 and T7express cells were used for general cloning and protein expression, respectively. S. Cerevisiae EBY100 cells were used for all surface display experiments.

LB media was used to culture E. coli, while SC-Trp-Ura supplemented with either 2% glucose or 2% galactose was used in all yeast experiments.

4.2.4 Surface display
EBY100 colonies containing the appropriate pYD1 vector were inoculated in SC-Trp-Ura+2% glucose media overnight at 30°C. The OD\textsubscript{600} was measured, and when it was between 2 and 5, the cells were centrifuged at 3000g for 5min at room temperature. The cells were then resuspended in SC-Ura-Trp+2% galactose media to an OD\textsubscript{600} of 0.5. The cells were then incubated at 20°C with shaking for up to 72 hours, with time points taken every 12 hours.
Staining was performed by centrifuging 1mL of culture from each time point at 3000g for 5 min at 4°C, resuspending in 1X PBS, re-centrifuging as before, resuspending in 250 μL 1X PBS, 1mg/mL BSA, and 1 μg FITC-labeled antibody, and incubating for 30 min on ice with occasional mixing. To wash, cells were centrifuged as before and resuspended in 1X PBS (2X), and finally resuspended in 40 μL PBS. 5 μL of this cell suspension was spotted on an agar pad and observed microscopically.

4.2.5 Pyrophosphate-based assay
Luciferase assays were conducted using the PPILight (Lonza) according to the manufacturer’s instructions. Fluorescent pyrophosphate detection was performed using PiPer (Invitrogen), again following manufacturer's instructions.

4.2.6 CoA-based assay
WES activity assays were carried out either in 125mM sodium phosphate buffer (pH 8) or cell lysate, depending on the experiment. The reaction were incubated for the time corresponding to the plots in this Chapter, at which point the 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM, Sigma-Aldrich) was added.

Table 8-CoA assay mix

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer or lysate</td>
<td>96</td>
</tr>
<tr>
<td>20mM palmitoyl-CoA</td>
<td>1</td>
</tr>
<tr>
<td>300mM butanol</td>
<td>1</td>
</tr>
<tr>
<td>2mg/mL WES</td>
<td>1</td>
</tr>
<tr>
<td>6mM CPM</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
</tr>
</tbody>
</table>

4.2.7 Chip experiments
Chip experiments were performed in the Abate lab at UCSF using custom-built PDMS chips. Fluorescence measurements were collected by microscope objective.

4.2.8 Protein purification
The his-tagged WES, N-terminal mCherry-fused WES, and C-terminal mCherry-fused WES proteins were purified via Nickel immobilized metal affinity chromatography from Qiagen\textsuperscript{101}. The proteins were stored in 50 mM NaH\textsubscript{2}PO\textsubscript{4}, 300 mM NaCl, pH 8.0. Protein concentrations were determined via absorbance measurements at 280nm, using the theoretical extinction coefficients\textsuperscript{102}.
4.3 Results

4.3.1 Development of a pyrophosphate assay for terpene synthase function
The first step to developing an assay that could be used for in vivo directed evolution efforts was to develop an assay that could read out activity in vitro from purified enzyme. Towards this goal, a luciferase assay was first attempted, shown schematically in Figure 38.

![Figure 38-Luciferase-based PPI assay](image)

Figure 39 shows the results of performing a standard curve with amorphadiene synthase as the independent variable. Blue bars indicate a 6-hour pre-incubation with amorphadiene synthase and its substrate, FPP, prior to the addition of the luciferase coupled-assay reagents for detection. Red bars indicate a 3-hour pre-incubation. The results demonstrate that the luciferase assay has excellent dynamic range and was suitable for pursuing further.
Another concern with the eventual chip-based methodology was the compatibility of the FPP substrate and enzyme with the oil that surrounds the aqueous droplets. To test whether substrate or enzyme would be sequestered/unfolded along the oil-aqueous interface, reactions were performed in bulk emulsions with various additives to try to attenuate non-specific interfacial adsorption. The emulsions were then broken and combined with the coupled enzyme detection reagents. The results are shown in Figure 40. The addition of both BSA and PEG brought the performance of the assay in line with the positive control (i.e. assay with no emulsion, brown bars), providing the largest fold change in luminescence between the 0nM ADS and 200nM ADS samples.
While the assay performed well in bulk, we unfortunately observed no signal when we attempted the reaction on-chip, likely due to insufficient sensitivity from chemiluminescence.

For this reason, an alternative, fluorescence-based assay was developed, as outlined in Figure 41.

**Figure 40**-Effect of additives on performance of lucifase-based assay in bulk emulsion

**Figure 41**-Fluorescence-based assay for PPI detection
The dynamic range and sensitivity of this assay was not as good as the luminescence based assay (Figure 42). However, we did expect small changes in activity from directed evolution, so it was deemed sufficient to proceed with.

![Figure 42-Performance of fluorescence PPI assay at different ADS concentrations](image)

The amount of time the enzyme was allowed to react with the substrate prior to the addition of coupled-detection reagents, or pre-incubation time, was also investigated, with the hypothesis that longer times would lead to more PPI accumulation and thus better assay performance. In fact, longer pre-incubation times negatively impacted the sensitivity of the readout.

![Figure 43-Effect of pre-incubation time PPI readout](image)
As a next step, we investigated whether the assay could be performed in cell lysate. The results from Figure 44 demonstrated that some unspecified component of the lysate was interfering with the assay.

![Purified ADS std curves](image)

This was then the motivation for the surface display section of experiments, attempting to mitigate the effects of cell lysate on the assay readout.

### 4.3.2 Miniaturized assay for pyrophosphate detection on a microfluidic device

Prior to the cell lysate experiment, we tested whether the results from the bulk experiments could be reproduced on-chip.

Figure 45 shows three superimposed traces for the 0nM ADS sample (red) and three traces for a 10nM ADS sample (green), each consisting of 10,000 data points. The data is normalized and re-plotted in Figure 46, showing that the same fold-change is observed on-chip as on a plate reader.
4.3.3 A yeast surface display system for droplet-based screening
Since prior experiments had demonstrated the incompatibility of the fluorescence assay with cell lysate, an alternative approach was pursued whereby the enzymes of interest would be surface displayed as a fusion to a mating peptide of *S. cerevisiae*\(^{103} \). To test our ability to display heterologous proteins, we first tested an RFP fused to a v5 epitope tag. This is shown graphically in Figure 47.
When carried out, EBY100 cells were able to display a high amount of the RFP, which could additionally be detected by staining with a FITC-labeled anti-v5 antibody, as shown in Figure 48.
However, when the same experiment was attempted with plasmids containing other candidate genes for directed evolution efforts, only one (NudB) was able to be displayed. The results are outlined in the following table.

Table 9—Summary of surface display results

<table>
<thead>
<tr>
<th>Protein</th>
<th>Size (kDa)</th>
<th>Glycosylation/disulfides?</th>
<th>Successfully displayed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>NudB</td>
<td>25</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>RFP</td>
<td>17</td>
<td>no</td>
<td>Yes</td>
</tr>
<tr>
<td>Trichodiene synthase</td>
<td>44</td>
<td>n/a</td>
<td>No</td>
</tr>
<tr>
<td>Wax Ester Synthase</td>
<td>52</td>
<td>no</td>
<td>No</td>
</tr>
<tr>
<td>Amorphadiene Synthase</td>
<td>64</td>
<td>no</td>
<td>No</td>
</tr>
<tr>
<td>Bisabolene synthase</td>
<td>93</td>
<td>no</td>
<td>No</td>
</tr>
</tbody>
</table>

4.3.4 Development of a thiol-reactive fluorescence assay for esterification reactions

In spite of promising on-chip results, the inability to conduct a pyrophosphate assay in cell lysate and the inability to surface display any enzyme of interest precluded droplet-based microfluidics approach for the terpene synthase family of enzymes.

However, the lab had a CoA-releasing enzyme in need of improvement, and a thiol-based assay had previously been described[^104], so we began developing the assay in Figure 49.
Adding free CoA into *E. coli* cell lysate (Figure 50) shows that the free-thiol background wasn’t sufficient to mask readout by CPM in an concentration regime relevant to screening efforts (approx. 10-1000 μM).

Spiking purified wax ester synthase at a concentration relevant for screening (≥1 μM) along with its substrates (palmitoyl-CoA and butanol) additionally supported the ability to screen for activity in cell lysate.
4.3.5 Construction of thioesterase-deficient screening strain for wax ester synthase activity

While the lysate didn’t provide as much background for the CoA-CPM assay as it did for the pyrophosphate-based assay, it nonetheless does contribute to it in a time dependent manner, as shown in Figure 52.

The most likely enzymes contributing to this background are native E. coli thioesterases: TesA, TesB, and FadM. To attenuate this background, we knocked the first two using P1 phage transduction and the last using the lambda-red recombinase system\textsuperscript{105}.
4.3.6 Fluorescent protein-wax ester synthase fusions for normalized activity screening

In order to correct for the variability of protein expression on a single cell level, which will be required in a droplet screening format, enzyme activity needs to be normalized to the enzyme's concentration (i.e. the kcat needs to be measured, not the Vmax). To accomplish this, two different fluorescent protein fusions were made to WES: a C-terminal mCherry fusion and an N-terminal mCherry fusion. The proteins were purified and the activity of both variants was measured and compared to the un-fused enzyme, to ensure that native activity hadn’t been abrogated. As shown in Figure 53, both variants retain activity, and in fact the C-terminal fusion appears to increase activity.

![Purified N- and C-terminal WES fusion activities](image)

**Figure 53-Activities of purified WES-mCherry fusion**

4.4 Discussion

In this Chapter we attempted to close the loop on the modeling-target identification-target improvement process of metabolic engineering outlined in previous Chapters.

While we were able to develop screening assays for ADS, they unfortunately had critical shortcomings. For instance, the luciferase assay showed much promise, but we weren’t able to detect its signal on-chip. Considering the difference in integration times between the plate reader (~1s) vs the chip (~5ms) reveals why might be. While the integration time on chip could be increased, it would unfortunately come at the expense of throughput. Similarly, the fluorescence assay suffered from problems as well—in addition to its poor
dynamic range and high background, it was incompatible with cell lysate. Since ADS could not be surface displayed, there was no recourse to continue efforts against this target. However, having developed the conceptual approach and technology necessary for microscale screening, we pursued another problematic enzyme, WES. Here, we made more progress, ultimately showing that enzyme-dependent activity could be measured in cell lysate, and that a fluorescent protein tag could serve as a normalization signal for comparison of library members. With these results, we are now primed to perform mutagenesis and screening for this important enzyme.

4.5 References

Chapter 5-Conclusion & future directions

5.1 INTRODUCTION

5.2 MODELING

5.2.1 PARAMETERIZATION

5.2.2 OPTIMIZATION

5.3 DIRECTED EVOLUTION

5.4 REFERENCES

5.1 Introduction

This body of work has demonstrated progress towards predicting engineered pathway behavior in E. coli and using this knowledge to perform targeted engineering. Specifically, we able to identify the most important factor to amorphadiene production, Amorphadiene synthase, and we were additionally able to rule out the FPP-metabolite feedback loop as mechanism for decreasing flux. In doing so, we’ve helped focus future engineering efforts toward projects that will be more impactful.

The ODE system of equations developed here will additionally serve as a base model for pathway behavior that can be added to, amended, and manipulated as more information is discovered.

In terms of the targeted engineering, we were unsuccessful in our efforts to develop a screen for amorphadiene synthase, but this led us to target another significant enzyme in a different pathway, for which we did make considerable progress.

These efforts have highlighted the main features necessary for performing a targeted metabolic engineering workflow, as outlined in the introductory Chapter. To close the loop on Figure 54, however, further work is needed. The following sections delineate the steps necessary to make this more of a systematic process in the future.
5.2 Modeling

5.2.1 Parameterization
George E.P. Box, the eminent statistician, was quoted as saying, “all models are wrong, but some are useful.”\textsuperscript{106} That is, there is always some level of detail or abstraction at which a model will fail to represent reality or experimental data, but if it doesn’t matter in the context of the experiment, it is irrelevant. In this respect, the models constructed in this dissertation were useful, in that they helped identify the most critical node in amorphadiene production---but they clearly aren’t complete, since many of the intermediate time-courses do not exactly match those from experiments.

There are several reasons why this might be occurring, whose causes and potential solutions are outlined in the table below.
Table 10- Causes & potential remedies for model-experiment disagreement

<table>
<thead>
<tr>
<th>Cause for model-experiment disagreement</th>
<th>Potential remedy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme processes modeled as irreversible reactions without product inhibition</td>
<td>Measure both forward and reverse reactions and use to parameterize the reversible michaelis-menten equation</td>
</tr>
<tr>
<td>Apparent constants used for multi-substrate reactions (in particular ATP)</td>
<td>Perform kinetic experiments to determine constants for both substrates (e.g. $K_{m,\text{mevalonate}}$ and $K_{m,\text{ATP}}$ rather than just $K_{m,\text{app,mevalonate}}$)</td>
</tr>
<tr>
<td>Insoluble (inactive) protein not accounted for in enzyme concentrations</td>
<td>Determine the fraction of soluble protein by SDS-PAGE</td>
</tr>
<tr>
<td>In vivo constants differ from in vitro constants</td>
<td>Perform parameter fitting to fit model parameters to experimental data</td>
</tr>
</tbody>
</table>

If accurate dynamic and steady-state data are necessary, which might be the case if a model incorporating growth were to be constructed—since it was demonstrated that intermediate concentrations can have a large impact on growth—then the reversible Michaelis-Menten equation should be employed. While its analytical form can be derived, it unfortunately introduces many extra parameters that currently are unavailable in the literature and so will have to be determined experimentally. During growth, ATP concentrations are typically saturating (approximately 10 mM), and so apparent $K_m$'s for the binding of mevalonate, mevalonate-phosphate, and mevalonate diphosphate (which were measured under conditions of saturating ATP) will likely capture the relevant kinetics of the pathway. If, however, it is desired to characterize pathway behavior as a function of more diverse physiological states (e.g. stationary or persister-like states) in which ATP levels may significantly fluctuate, $K_m$'s will need to be determined for both substrates.

A final way to increase the accuracy of the model would be to catalogue all the inhibitory interactions between highly accumulating species and pathway enzymes. While the correct way to accomplish this will be to experimentally determine the type (e.g. non-competitive) and strength of the interaction, the experimental design will quickly become unwieldy. Another option is to use phenomenologically constructed inhibition terms, as was done in Chapter 2, and then perform parameter fitting. While this approach recently met with success for glycolysis, it is still unclear whether it will apply to a diverse range of experimental conditions/expression regimes.

5.2.2 Optimization

Optimizations, whereby some objective function is maximized subject to constraints, is a technique widely used in linear programming, which in the context of metabolic engineering falls under the auspices of flux modeling.

However, optimization should also be possible with kinetic equations, in which product concentration is maximized, while, for instance, the total amount of enzyme to yield a
desired productivity is minimized. Computational techniques like monte-carlo could aid in this type of analysis.

Another complimentary, experimental approach would be to perform metabolic control analysis with purified enzymes in a Continuous Stirred Tank Reaction (CSTR)-like system that is coupled to a mass spectrometer, as was done by Bujara et al. With control coefficients determined in real-time, one can experimentally climb through local regions of expression space by sequentially injecting the enzyme that has the greatest control coefficient. Being flow based, this type of system has the added benefit of being able to keep multiple substrates and cofactors (e.g. glucose, mevalonate, ATP) in a physiologically relevant steady-state. This is beneficial since "batch"-like experiments need to have cofactors in vast excess in order to remain at saturating levels (especially if only apparent K_m's are measured, as previously alluded to), but this may lead to inhibitions (e.g. substrate, competitive) that yield aberrant dynamical behavior.

5.3 Directed evolution

Unfortunately, a robust assay amenable to miniaturization was unable to be developed for sesquiterpene production. Cytochrome p450s are one of the few enzymes that can bind to sesquiterpenes (often performing tailoring reactions such as oxidations), so perhaps some type of biosensor (e.g. hybrid transcription factor) could be developed based on it.

More promising was the work to develop a screen for wax-ester synthase activity. Having developed a lysate-compatible assay where activity can be normalized to enzyme concentration through the use of the fluorescent tag fusion, all that’s left to do is to (1) construct an error-prone PCR library of the enzyme, (2) screen on a microfluidic device, (3) take the best-selected variants and validate using an in vivo production platform.

5.4 References


## Appendix

### Table 11-Strains and plasmids used in this work

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH1</td>
<td>Wildtype E. coli: endA1 recA1 gyrA96 thi-1 glnV44 rpoA1 hsdR17[rK- mK+] λ-</td>
<td>-</td>
</tr>
<tr>
<td>Tg1</td>
<td>F' [traD36 proAB' lacI' lacZΔM15]supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5, (rKmK)</td>
<td>-</td>
</tr>
<tr>
<td>T7express</td>
<td>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet)2 [dcm] R(zgb-210::Tn10--Tet) endA1 Δ(mrcC-mrr)114::IS10</td>
<td>NEB</td>
</tr>
<tr>
<td>NEBexpress</td>
<td>fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet)2 [dcm] R(zgb-210::Tn10--Tet) endA1 Δ(mrcC-mrr)114::IS10</td>
<td>NEB</td>
</tr>
<tr>
<td>T7expressTEko</td>
<td>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet)2 [dcm] R(zgb-210::Tn10--Tet) endA1 Δ(mrcC-mrr)114::IS10 ΔTesA ΔTesB</td>
<td>This work</td>
</tr>
<tr>
<td>EBY100</td>
<td>MATa ura 3-52 trp 1 leu2Δ1 his3Δ200 pep4:HIS3 prb1A1.6R can1 GAL</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pMBIS3</td>
<td>Trc promoter-ADS-s.c.MK-PMK-PMID-IDI-ISPA ColE1 Amp</td>
<td>Built by F. Nowroozi</td>
</tr>
<tr>
<td>pMBIS3mmMK</td>
<td>Trc promoter-ADS-m.m.MK-PMK-PMID-IDI-ISPA ColE1 Amp</td>
<td>This work</td>
</tr>
<tr>
<td>pMBIS3_10Kads</td>
<td>Trc promoter-10krbsADS-s.c.MK-PMK-PMID-IDI-ISPA ColE1 Amp</td>
<td>This work</td>
</tr>
<tr>
<td>pIDI_CPD</td>
<td>T5 promoter Idi-CPD-His Kan ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pPMD_CPD</td>
<td>T5 promoter PMD-CPD-His Kan ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pISPA_his</td>
<td>T5 Ispa-His Kan ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pWES_his</td>
<td>T7 promoter-His-WES Kan ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pWES_nMC_his</td>
<td>T7 promoter-His-mCherry-WES Kan ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>Vector Name</td>
<td>Description</td>
<td>Source</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>pWES_cMC_his</td>
<td>T7 promoter WES-mCherry-his Kan ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-MCS-V5-His Trp Amp ColE1</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pYD1_ADS</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-ADS-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_Bis</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-Bis-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_FS</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-FS-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_IspA</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-IspA-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_NudB</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-NudB-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_RFP</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-RFP-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_Trichodiene</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-TS-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pYD1_WES</td>
<td>GAL promoter-Aga2sigPep-Aga2-SerGly-WES-V5-His Trp Amp ColE1</td>
<td>This work</td>
</tr>
<tr>
<td>pMevQconcat</td>
<td>T7 promoter-Stag-MevQconcat-His Kan ColE1</td>
<td>This work</td>
</tr>
</tbody>
</table>
Figure 55-Representative mbis strain plasmid

Figure 56-Representative surface display vector
Figure 57 - Representative WES-mCherry vector

Figure 58 - Qconcat 2.0 plasmid
Mathematica code

ClearAll["Global\"*"]

(*parameters for strains*)
pMBIS3 = {mk -> .000066, pmk -> .0000029, pmd -> .0000024, idi -> .0000184,
ispA -> .000035, ads -> .000039, kcat1 -> 38, klffp -> 0.0000019,
km1mev -> .000130};
pSA3 = {mk -> .000023, pmk -> .0000049, pmd -> .0000027, idi -> .000015,
ispA -> .000037, ads -> .000026, kcat1 -> 6, klffp -> 0.0000046,
km1mev -> .000041};
p10kADS = {mk -> .000197, pmk -> .000014, pmd -> .0000023, idi -> .0000216,
ispA -> .0000242, ads -> .0000678, kcat1 -> 38, klffp -> 0.0000019,
km1mev -> .000130};

(*state variables*)
variables = {MevP, MevPP, IPP, MAP, GPP, FPP, Amo}; p = 0;

(*turnover frequencies*)
kcat2 = 10; kcat3 = 4.9; kcat4a = 0.33; kcat4b = \ .33; kcat5 = 0.21; kcat6 = .47; kcat7 = 0.006*3;

(*Michaelis constants*)
km2 = .0001; km3 = .0001; km4 = .000008; km5ipp = \ .0000013; km5dmappp = .0000293; km6ipp = .0000013; km6gpp = .0000055; klgpp = \ .000071; km7 = .000003;

(*initial metabolite concentrations*)
Mev = .005; initialConditions = {MevP[0] == 0.000001, MevPP[0] == 0.00001, IPP[0] == 0.000001,
MAP[0] == 0.000001, GPP[0] == 0.000001, FPP[0] == 0.000001,
Amo[0] == 0.07};

(*Simulation time*) ts = 2000000;(*time in plots*) tp = 30000;

(*Pathway ODE's*) pathway = {
MevP[t] == (MevP[t]/.005)^(-p)*kcat1*mk*
Mev/(km1mev*(1 + (FPP[t]/klffp)) + Mev) - (MevPP[t]/.005)^(-p)*kcat2*
pmk*MevP[t]/(km2 + MevP[t]),
MevPP[t] == (MevPP[t]/.005)^(-p)*kcat2*pmk*MevP[t]/(km2 + MevP[t]) -
kcat3*pmd*MevPP[t]/(km3 + MevPP[t]),
IPP'[t] ==
kcat3*pmd*MevPP[t]/(km3 + MevPP[t]) - kcat4a*idi*IPP[t]/(km4 + IPP[t]) - kcat5*ispA*IPP[t]*
MAP[t]/(km5ipp*MAP[t] + km5dmapp*IPP[t] + MAP[t]*IPP[t]) +
kcat4b*idi*MAP[t]/(km4 + MAP[t]) -
kcat6*ispA*GPP[t]*
IPP[t]/(klgpp*km6ipp + km6ipp*GPP[t] + km6gpp*IPP[t] + GPP[t]*IPP[t]),
MAP'[t] ==
kcat4a*idi*IPP[t]/(km4 + IPP[t]) - kcat4b*idi*MAP[t]/(km4 + MAP[t]) - kcat5*ispA*IPP[t]*
MAP[t]/(km5ipp*MAP[t] + km5dmapp*IPP[t] + MAP[t]*IPP[t]),
GPP'[t] ==
kcat5*ispA*IPP[t]*
MAP[t]/(km5ipp*MAP[t] + km5dmapp*IPP[t] + MAP[t]*IPP[t]) -
kcat6*ispA*GPP[t]*
IPP[t]/(klgpp*km6ipp + km6ipp*GPP[t] + km6gpp*IPP[t] + GPP[t]*IPP[t]),
FPP'[t] ==
kcat6*ispA*GPP[t]*
IPP[t]/(klgpp*km6ipp + km6ipp*GPP[t] + km6gpp*IPP[t] + GPP[t]*IPP[t]) -
kcat7*ads*FPP[t]/(km7 + FPP[t]),
Amo'[t] == kcat7*ads*FPP[t]/(km7 + FPP[t]),
initialConditions);

(*solve pathway system of equations*)
sol = NDSolve[pathway /. pMBIS3, variables, {t, 0, ts}, AccuracyGoal -> 30];
(*Compare rates to determine flux control*)
"= D[Amo]/dt" First[
D[Amo[t] /. sol, t] /. t -> 10000]
"= Vmax" kcat7*ads /. pMBIS3

(*Plot metabolite[t] solutions*)
Plot1 = Plot[Evaluate[{Mev}], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["Mev", Bold]}];
Plot2 = Plot[Evaluate[{MevP[t]} /. sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["MevP", Bold]}];
Plot3 = Plot[Evaluate[{MevPP[t]} /. sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["MevPP", Bold]}];
Plot4 = Plot[Evaluate[{IPP[t]} /. sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["IPP", Bold]}];
Plot5 = Plot[Evaluate[{MAP[t]} /. sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["MAP", Bold]}];
Plot6 = Plot[Evaluate[{GPP[t]} /. sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["GPP", Bold]}];
Plot7 =

78
Plot[Evaluate[{FPP[t]} /. sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["FPP", Bold]}];
Plot8 = Plot[Evaluate[{(Amo[t]) /. sol}], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["Amo", Bold]}]; Plot9 =
sol], {t, 0, tp},
AxesLabel -> {Style["t", Bold], Style["S intermediates", Bold]}];
GraphicsGrid[{{Plot1, Plot2, Plot3}, {Plot4, Plot5, Plot6}, {Plot7, Plot8, 
Plot9}}, ImageSize -> {900, 500}]

(*Export data*)
Export["data.csv",
Table[Flatten[{t, Amo[t], MevP[t], FPP[t]} /. sol], {t, 0, 28800, 1000}]]