Metabolic Engineering for the Production of Functionalized Terpenoids in Heterologous Hosts

by

Steven McBride Edgar

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2017

© Massachusetts Institute of Technology 2017. All rights reserved.

Author .... Signature redacted ......
Department of Chemical Engineering
December 6, 2016

Certified by....... Signature redacted ....
Gregory Stephanopoulos
Willard Henry Dow Professor of Chemical Engineering and Biotechnology
Thesis Supervisor

Accepted by ..... Signature redacted ....
Daniel Blankschtein
Chairman, Department Committee on Graduate Theses

Metabolic Engineering for the Production of Functionalized Terpenoids in Heterologous Hosts

by

Steven McBride Edgar

Submitted to the Department of Chemical Engineering
on December 6, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Terpenoids are a large and diverse group of secondary metabolites which natively play essential biochemical roles, and have been adopted for a wide range of uses by humans. These molecules are of interest scientifically due to their often complex structures, as well as commercially, with market value exceeding $50 billion annually. Many of these compounds are derived from natural sources, such as plants, which are prone to long life cycles, low productivity, as well as climate and market variability. As such, production in heterologous microbial systems through metabolic engineering stands as an attractive alternative to truly natural production and allows for increased control over biosynthetic pathways enabling improved yields, titer, and productivity.

The anti-cancer molecule Taxol (Paclitaxel) stands as one of the most medically and economically important terpenoids. However, despite decades of extensive study, its biosynthesis remains poorly understood. In this work we report on investigations into the reconstruction of the early Taxol bioysnthetic pathway. Early successes toward pathway reconstitution within *E. coli* and *S. cerevisiae* led to the identification of an early bottleneck, identified to arise from unpredictable behavior of the first oxygenation enzyme, taxadiene-5α-hydroxylase (CYP725A4), which produces a range of undesired products. We present chemical and biological evidence of an unreported epoxidase activity of taxadiene-5α-hydroxylase that puts into question the previously proposed radical-rebound mechanism. We demonstrate that the poor selectivity of taxadiene-5α-hydroxylase arises from the non-selective degradation of an epoxide intermediate produced via a selective oxidation step, rather than from promiscuous oxidation, as previously proposed. We support these conclusions by demonstrating variable enzyme behavior in differing hosts and conditions, similarity of products and product ratios generated from chemical epoxidation and by taxadiene-5α-hydroxylase, and differing enzymatic activity on alternative taxadiene isomers.

We next systematically investigate three methodologies, terpene cyclase engineering, P450 engineering, and hydrolase-enzyme screening to overcome this early pathway selectivity bottleneck. We demonstrate that engineering of taxadiene synthase, upstream of the promiscuous oxidation step, acts as a practical methodology for se-

lectivity improvement. Through mutagenesis we achieve a 2.4-fold improvement in yield and selectivity for an alternative cyclization product, taxa-4(20)-11(12)-diene; and for the Taxol precursor taxadien-5α-ol, when co-expressed with CYP725A4. This work lays the foundation for the elucidation, engineering, and improved production of Taxol and early Taxol precursors.

In addition to bottlenecks in the downstream production of terpenoids such as Taxol, upstream flux derived from the microbial methylerythritol phosphate (MEP) pathway also imposes limitations on culture productivity and titer. Thus, efforts were also focused on developing alternative biosynthetic routes to terpenoids. This led to the investigation of a novel route, which relies on phosphorylation of the bulk chemical feedstock isopentenol, to generate isopentenyl disphosphate (IPP), the building block of all terpenoids. Following proofs-of-concept experiments demonstrating pathway functionality, we turned our attention to methodologies to improve flux through this pathway. As few high-throughput methodologies for the quantification of terpenoid production have been previously described, this led to the development of a genetically-encoded terpenoid-responsive biosensor. Furthermore, we develop theoretical framework and proof-of concepts for the application of evolution-based approaches to the optimization of the pathway.

Thesis Supervisor: Gregory Stephanopoulos
Title: Willard Henry Dow Professor of Chemical Engineering and Biotechnology

Acknowledgments

The path to a PhD is a long one, one more challenging than I had imagined upon first starting and I owe a great debt to those who have helped me along the way. As such, I first want to thank my thesis adviser, Professor Gregory Stephanopoulos. His optimism and enthusiasm is both unmatched and contagious, and his passion for science has renewed my own on many occasion. He has encouraged me to think big picture, to not only ask questions, but to ask what the value of the question is, so that my work may have the greatest impact. Additionally, he has fostered an environment in which I have been allowed to grow as a scientist. I have been free to pursue my ideas and develop as an independent scientist, capable not only of executing experiments and projects, but also of conceptualizing them and developing resources needed to complete them. I would also like to thank those who have served on my thesis committee - Professors Kris Prather, Catherine Drennan, Jing-ke Weng for their support and guidance. Each member of my committee brought unique knowledge and insights and has offered advice at many a critical juncture in my work.

The Stephanopoulos lab is large with a broad field of study and high turnover. As such, I have had the benefit of meeting many good people and great scientists with diverse skill sets. I had the pleasure of working with Vikram Yadav, who helped me to get started in the Lab. I sat adjacent and worked closely with Kangjain Qiao, whose ability to think big picture inspired me to do the same and who was the source of endless (and endlessly useful) advice. I learned from Kang Zhou that most procedures can be shortened if you are careful, and though I never reached his 1-minute transformation protocol, it inspired me to be more cognizant of the time I spent and how to use it so that I could do more science in the same hours. I also want to thank Peter Enyeart for insights into science, how to do it well, and finding your place in it.

Though my projects often did not overlap with other lab members, I enjoyed countless research discussions with Niju Narayanan, Ben Woolston, Jason King, Hao-ran Zhang, and Devin Currie. Many older students and post-docs including Tom

Wasylenko, Andrew Silverman, Brian Pereira, and Mark Kiebler provided me with lively discussions on graduate life and beyond (even on late Friday nights in lab) and reminding me at every opportunity that the failure of one experiment was only a minor setback because I'm "only an Nth year". Furthermore, Rosangela dos Santos has treated the members of our lab more like family than coworkers during my stay. Her spectacular work allowed the lab to run smoothly and for me, and others, to retain our focus on research.

It has been a pleasure to watch many of the younger students: Zhe Zhang, David Emerson, Wentao Dong, and Nian Liu mature as scientists and I look forward to seeing the many great things they do in their careers. I also had the pleasure of working with a younger graduate student, Alkiviadis Chatzivasileiou, with whom I have watched from entry into the lab through to his development to a fully-capable graduate student in his own right. Alkis has been a wonderful student to mentor, and then collaborate with, and has been instrumental to the work performed in the latter half of the thesis, and will continue this work himself. I can leave the lab in good spirits knowing the work I devoted several years of my life to lives on in his capable hands.

I owe a great deal to my classmates and cohort, who have provided a great deal of support outside of lab (and relief from lab). From lunch breaks, a Vegas trip, and exploring the city we called home these past years I am thankful to have met this group of people including Eric Zhu, Mark Weidman, Sean Hunt, Lionel Lam, Xiao Su, Rosanna Lim, Thomas Williams, Karthik Narsimhan, and Helen Luo.

My daughter Olivia, although a late addition to my graduate school career, taught me many lessons, not the least of which was the value of using one's time wisely. It is through her I have been allowed to grow not only as a scientist, but as a person in these past years. Beyond all else, I want to thank my parents, Gary and Debbie Edgar. They have provided their unconditional support that enabled me not only pursue my passion for science, but also to achieve goals I never thought possible. I am continually inspired by my mother's energy and my father's insight, and they have provided me not only with a model which to emulate, but the tools to succeed. For

that I thank them, for they are the foundation upon which I am built.

I am truly grateful for all of the wonderful people I have met during my PhD and for all the kindness and support these people have supplied me.

Contents

1 Introduction **21**

1.1 Terpenoids 22

1.2 The Industrial Relevance of Terpenoids 23

1.3 Thesis Objectives and Overview 24

1.4 Bibliography 25

2 Reconstructing Early Paclitaxel Biosynthesis **27**

2.1 Introduction 27

2.2 The Use of Heterologous Hosts for Natural Product Biosynthesis . . . 29

2.3 The Taxol Biosynthetic Pathway 31

2.4 Building a Generalizable System 35

2.5 Pathway Elucidation 41

2.6 Conclusions and Proposals for Future Work 45

2.7 Materials and Methods 45

2.8 Bibliography 47

3 Understanding Early Paclitaxel Biosynthesis **55**

3.1 Introduction 55

3.2 Cytochrome P450 Enzymes 56

3.2.1 Previous Studies on Taxadiene-5α-Hydroxylase 59

3.3 Reevaluating the Mechanism 64

3.3.1 Effect of Further Biotransformation 64

3.3.2 Effect of Enzyme Fusions 65

3.4 Effect of Enzyme Environment on Selectivity 67
3.5 Chemical Synthesis 70
3.5.1 Characterization of Epoxide Products 72
3.6 Alternative Taxadiene Isomer Utilization Supports Epoxidation Route 78
3.7 Evaluation of Alternative Oxidants 80
3.8 Plant Cells 83
3.9 Further Remarks and Comparison to Literature 84
3.10 Conclusions 87
3.11 Materials and Methods 89
3.12 Bibliography 94

4 Engineering Early Paclitaxel Biosynthesis 101
4.1 Introduction 101
4.2 Engineering Selectivity Through Addition of Plant Enzymes 103
4.3 Taxadiene-5α-Hydroxylase Engineering For Improved Selectivity . . . 105
4.4 Taxadiene Synthase Engineering for Increased Isomer Production . . 107
4.4.1 The Mechanism of Taxadiene Synthase 112
4.4.2 Investigation of Stereospecific Base 112
4.4.3 Saturation Mutagenesis of Select Residues 116
4.5 Conclusions 127
4.6 Materials and Methods 129
4.7 Bibliography 133

5 Developing An Alternative Route to Isoprenoids 137
5.1 Introduction 137
5.2 Evaluating Alternative Pathways 138
5.2.1 The Isopentenol Utilization Pathway 140
5.2.2 Pathway Feasibility and Economic Analysis 142
5.3 Construction of an Isopentenol-based pathway 146
5.3.1 Selection of Candidates 147
5.3.2 Detection of IP 148

5.3.3 In-Vitro Screening of Candidates 149
5.3.4 *In-Vivo* Pathway Validation 153
5.4 The Need for High-Throughput Screening Methodologies 154
5.5 Evolution-Based Pathway Optimization 155
5.5.1 Applying Terpene Degradation in Pseudomonads 156
5.5.2 Transformation, Expression, and Evolution in *P. citronellolis* . 161
5.5.3 Isoprenoid Auxotrophic E. coli 162
5.6 Conclusions and Proposals for Future Work 164
5.7 Materials and Methods 164
5.8 Bibliography 166

6 Development of a Terpene-Responsive Biosensor 171
6.1 Introduction 171
6.2 AtuR and Genetic Basis of Regulation 172
6.3 Initial Attempts at Promoter Design 173
6.3.1 Implementation of Two-Plasmid System 175
6.3.2 Re-Investigating the Binding Motif 176
6.3.3 T7 Promoter 183
6.3.4 Determining Ligand Responsiveness 185
6.3.5 Consolidation of Promoter 187
6.4 Conclusions and Future work 187
6.5 Materials and Methods 189
6.6 Bibliography 191

7 Conclusions 195
7.1 Thesis Summary 195
7.1.1 Taxol Pathway Elucidation and Bottleneck Alleviation 195
7.1.2 Development of a Biocatalytic Route to Terpenoids 196
7.2 Future Directions 197
7.2.1 Taxol Pathway Elucidation and Optimization 197

7.2.2 Optimization of the Isopentenol Utilization Pathway and Application of Evolutionary-Based Approaches 198
7.2.3 Optimization of an Isoprenoid-Responsive Biosensor 199
7.3 Concluding Remarks 200
7.4 Bibliography 201

A P450 Sequences from Transcriptome Analysis 203

B NMR Spectra 243

List of Figures

2-1 Chemical Structure of 10-Deacetylbaccatin III and Taxol 29

2-2 Characterized Steps in Early Paclitaxel Biosynthesis 32

2-3 Pathway Elucidation Cycle 35

2-4 Multiple Library Methodology for Pathway Elucidation 38

2-5 Single Library Methodology for Pathway Elucidation 38

2-6 Cladogram of Aligned Contigs 40

2-7 CYP Protein Phylogenetic Tree 42

2-8 CO Difference Spectra for Purified Taxadiene-5α-hydroxylase 43

2-9 Pathway Elucidated from Proof-of-Concept Screening 43

2-10 *In-Vitro* Acetylation of Taxadiene-5α-ol 44

2-11 Mass Spectra from Yeast Expressing Three Functional Pathway Steps 44

3-1 Reported Selectivity of Taxadiene-5α-Hydroxylase 57

3-2 Consensus Radical-rebound based mechanism for Cytochrome P450st 60

3-3 Previous Support for a Radical-Mediated Reaction 61

3-4 Racical-Rebound Mediated Taxadiene Hydroxylation 62

3-5 Molecular Docking Poses of taxadien-5α-ol within taxadiene-5α-hydroxylase 63

3-6 Depiction of All Catalytic Routes to Final Products by Taxadiene-5α-Hydroxylase 64

3-7 Observed Selectivity of Strains With and Without Acetylase Expression 66

3-8 Desired Product Production of Strains With and Without Acetylase Expression 66

3-9 Selectivity observed from taxadiene-5α-hydroxylase 67

3-10 Selectivity of Taxadiene-5α-Hydroxylase in Varying Hosts 68
3-11 Selectivity of Taxadiene-5α-Hydroxylase in Varying Solvents 70
3-12 Chemical Analysis of Taxadiene Epoxidation. 71
3-13 GCMS Trace of Chemical Epoxide Synthesis 71
3-14 Reflux of OCT 73
3-15 Structure of Compound 11 76
3-16 Proposed Reaction Cascade to Compound 11 77
3-17 Acid Treatment of Compound **11** 79
3-18 Treatment of *E. coli* Crude Lysate with Compound 11 80
3-19 Feeding of Compound 5 to Lab Strains. 81
3-20 Feeding of Taxadiene Isomers 82
3-21 Chemical Epoxidation of isotaxadiene 82
3-22 Plant Cell Feeding Time Course. 85
3-23 OCT Formation in Plant Cells 85
3-24 Possible Epoxide Breakdown Pathways. 86
3-25 Barton et al. Chemical Epoxidation Experiments 88

4-1 Relative Selectivity of Alanine-Scanning Mutants 103
4-2 Proposed Mechanism for Conversion of Epoxide to Taxadien-5α-ol . . 104
4-3 Possible Mechanisms for Di-Hydroxylated Product Formation 106
4-4 Sequence Alignment of Taxadine-5a-Hydroxylase to downstream Taxol P450s. 108
4-5 Homology Model of Taxadien-5a-Hydoxylase. 109
4-6 Mass Spectra and Retention time of Observed Diol. 110
4-7 Observed Selectivity from T380S Mutant 111
4-8 Proposed Pathway for Formation of Compound 12 111
4-9 Mechanism of Taxadiene Synthase 112
4-10 Select Active Site Residue Locations 113
4-11 Molecular Docking Results for Taxadiene 114
4-12 Absolute Selectivity of Alanine-Scanning Mutants 115

4-13 Relative Selectivity of Alanine-Scanning Mutants 115
4-14 Mass Spectra Obtained from Compound 10 116
4-15 Proposed Expanded Taxadiene Synthase Catalytic Cascade 117
4-16 Mutagenesis Coverage Observed From CLIVA Methodology 118
4-17 Total Titer Observed from Taxadiene Synthase Mutants. 119
4-18 Selectivity and Titer Observed From All Terpene Cyclase Mutants. . 120
4-19 Distribution Observed for Saturation Mutagenesis In Different Protein Regions. 121
4-20 Observed Selectivity for Individual Mutations at Residue S587. 122
4-21 Observed Selectivity for Individual Mutations at Residue Q609. . . . 123
4-22 Observed Selectivity for Individual Mutations at Residue Y684. . . . 124
4-23 Observed Selectivity for Individual Mutations at Residue Y688. . . . 125
4-24 Observed Selectivity for Individual Mutations at Residue M830. . . . 126
4-25 Cyclization Selectivity Observed From Select Mutants 128
4-26 Oxidation Selectivity Observed From Select Mutants 128
4-27 Relative Selectivity of Alanine-Scanning Mutants 129

5-1 Depiction of Pathways to IPP and DMAPP 139
5-2 Proposed Isopentenol Utilization Pathway 140
5-3 Proposed Industrial Application of The Biocatalytic Process 141
5-4 Cofactor Requirements for MEV and MVA Pathways 145
5-5 Methodology for Enzymatic Detection of IP 149
5-6 Enzymatic Transformation of MVAP to IP 150
5-7 Initial Screen of Isopentenol Kinase Candidates 150
5-8 Initial Screen of Isopentenol Kinase Candidates 151
5-9 Chemical Structure of Isopentenol and Choline 151
5-10 Representation of Choline Kinase Active Site 152
5-11 Relationship of Enzymatic Activity to Productivity 153
5-12 Initial Screen of Isopentenol Kinase Candidates 154
5-13 Methodologies for Selection and Library Generation 156

5-14 Pathways Evaluated for Optimization of IU Pathway. 158
5-15 Full Pathway for Geraniol Production and Utilization from Isopentenol 159
5-16 Flow Cytometry on Transformed *P. Citronellois* 162

6-1 Terpenes Acting as Sole-Carbon Sources in Pseudomads 172
6-2 Organization of Atu Gene Cluster in *Pseudomnoas citronellolis* 173
6-3 Genetic Basis for Repression in *Pseudomonas* 173
6-4 Testing of Initial Synthetic Biosensor Construct 174
6-5 Results from First Synthetic Promoter 175
6-6 Demonstration of AtuR Expression 176
6-7 Response obtained from pSEN020 Biosensor Construct 177
6-8 Orientation of Binding Motifs in Tested Constructs 177
6-9 Typical TetR Repressor Binding Motifs 178
6-10 Literature Results obtained from AtuR Titration 179
6-11 Alignment of AtuR AtuA intergenic Regions for Motif Analysis. . . . 180
6-12 Motif Sequence Comparison 181
6-13 Design of Extended Motif Constructs 182
6-14 Repressor Performance with Extended Binding Motif 182
6-15 Construction and Testing of an T7-AtuR Biosensor System 184
6-16 EMSA of AtuR Binding to Short Oligomer 186
6-17 EMSA Displaying Responsiveness to Acyclic Terpenes 186
6-18 Sensor Responsiveness to Acyclic Terpenes. 187
6-19 Use of Constitutive Promoters for Repressor Expression. 188
6-20 Response of Promoter to Geraniol. 188

B-1 1H NMR of Compound 11 244
B-2 13C NMR of Compound 11 245
B-3 gCOSY NMR of Compound 11 246
B-4 HMBC NMR of Compound 11 247
B-5 HSQC NMR of Compound 11 248
B-6 1H NMR of Verticilla-3,7,12-triene 249

B-7 13C NMR of Verticilla-3,7,12-triene 250
B-8 DEPT NMR of Verticilla-3,7,12-triene 251
B-9 HHCOSY NMR of Verticilla-3,7,12-triene 252
B-10 HMBC NMR of Verticilla-3,7,12-triene 253
B-11 HSQC NMR of Verticilla-3,7,12-triene 254
B-12 1H NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol 255
B-13 13C NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol 256
B-14 HHCOSY NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol 257
B-15 HMBC NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol 258
B-16 HSQC NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol 259

List of Tables

3.1 Yields Obtained from Chemical Epoxidation of Taxadiene 70

3.2 Chemical Shifts for Compounds **11** 75

3.3 Strains Used in This Chapter 95

5.1 Feedstock Requirements Needed to Generate One Mole of IPP 144

5.2 Maximum cost ratio ($ per kg isopentenol / $ per kg glucose) to allowed for industrial feasibility at theoretical yield. This is calculated by assuming a cost of 1 unit cost per kg of glucose. 144

Chapter 1

Introduction

While traditional synthetic chemistry has long served as a method for the production of chemicals for human use, the use of biotechnology is a comparatively young field. Often biology offers not only a "greener" approach, but a more economically favorable one. Production of secondary metabolites with multiple chiral centers or complex chemistry can often be achieved more efficiently through biological production than traditional chemistry.

Although the transition from first transplanting genes from one organism to another, pioneered by Cohen & Boyer in 1973 [1], to the founding of Genentech in 1976 and the first production of a human protein (somatostatin) in *E. coli* the following year was extremely rapid, the field of metabolic engineering remained limited by available molecular biology tools, focusing on single genes or operons. Early work culminated in the publication of two seminal papers and creating a true systematic practice in 1991 [2, 3]. Currently, metabolic engineering has come to stand as the pinnacle of modern biotechnology. It aims not only to express genes and produce proteins within an organism, but to fundamentally rewire the metabolism and/or physiology of that organism for the production of specific compounds. This differs vastly from the production of recombinant insulin or other protein products, wherein only a single peptide must be expressed.

1.1 Terpenoids

One family of molecules that metabolic engineering has achieved great success with are terpenoids (also referred to as isoprenoids). Terpenoids are one of the most diverse and important classes of secondary metabolites, and arise from functionalization of molecules referred to as terpenes. These molecules are immensely important in biochemical processes, forming the basis of archeal cell membranes and steroids such as cholesterol and testosterone. These compounds are also produced by a wide variety of plants, as well as some animals, and are found in fruits, vegetables, flowers, and plant resins. Their roles in plants are diverse, including defense against infection and feeding, communication, in photosynthesis as pigments, as hormones, and as attractants and repellents. Furthermore, it is often these molecules which are responsible for the characteristic fragrance or taste of a given plant or plant product. Currently, more than 55,000 terpenoids have been isolated, with the number growing yearly [4].

Biosynthetically, terpenes are derived from five-carbon isoprene units, beginning from the polymerization of dimethylallyl diphosphate (DMAPP) and isopentyl phosphate (IPP). After initial polymerization, isoprene units are added in an incremental manner, allowing classification of terpenoids based on the number of isoprene units they contain, the primary groupings being monoterpenes (C10), sesqiterpenes (C15), diterpenes (C20), sesterterpenes (C25), and triterpenes (C30). Each of these families is large in their own right, giving rise to thousands of molecules. In addition to this broad diversity, terpenes also typically undergo similar biosynthetic processes including chain elongation, cyclization, and functionalization. Furthermore, similar enzymes catalyze these biotransformations; thus, by studying the biosynthesis of a single terpenoid, many learnings may be applied to terpenes in general. In this thesis, we will discuss both the late-stage biosynthetis of a single compound, Taxol, as well as work upstream that acts to enable high-level production of all terpenoids.

1.2 The Industrial Relevance of Terpenoids

Although not immediately obvious, terpenoids have a vast range of industrial applications in addition to their native biological ones and are a primary target of a new wave of synthetic biology companies which have garnered over $4.1 billion in funding since 2009 [5, 6]. While natural rubber, a long-chain isoprenoid polymer, is by far the single most important terpenoid commercially produced with annual sales ranging between $26 and $40 billion [7, 8], many other terpenoids are also of tremendous economic importance. Terpenoid production for the flavor and fragrance industry exceeds $1 billion per year, Taxol production exceeds $4 billion per year, and artemisinin sales exceed $1 billion per year [9]. In addition to these molecules, nutraceuticals such as the terpenoid class known as ginkgolides (dervied from the diterpene levopimaradiene), are present within a number of consumer products as well.

Due to their economic importance, terpenoids also stand as one of the primary targets of the modern commercial synthetic biology. Commercial synthetic biology terpenoid products range from the sesquiterpene next-generation biofuel, farnesene, produced by Amyris; the sesquiterpene anti-malarial drug artemisinin produced by Sanofi; complex fragrance mixtures produced by Gingko Bioworks utilizing Amyris' isoprenoid production chassis; as well as the flavor and fragrance molecules valencene and nootkatone produced by Evolva, the latter also undergoing FDA approval as an insecticide to combat mosquito-borne diseases such as the Zika virus.

However, the molecules produced through synthetic biology are only a small fraction of the industrially relevant isoprenoids. In fact, the vast majority of commercialized molecules ranging from natural rubber, *Ginkgo biloba* extract, lycopene and carotenoids, natural artemenisin, Taxol, and numerous flavors and fragrance molecules are still obtained from their native production system, often plants. Furthermore, a wide variety of potential products lie untapped due to low natural compound availability. As such, research on the production of isoprenoids is important both medically and commercially, and is far from complete.

1.3 Thesis Objectives and Overview

The main objective of this thesis is to apply metabolic engineering for the production of terpenoids. Initially, we aim to reconstruct and study the Taxol biosynthetic pathway for production of Taxol precursors within a heterologous host. We then focus our attention on the upstream pathway and methodologies to improve the availability of IPP and DMAPP for the production of any terpenoid. Finally, we aim to facilitate pathway optimization and enzyme screening through the development of high-throughput screening methodologies including a terpenoid-responsive biosensor.

Chapter 2 opens to provide an overview of the importance, biosynthesis, and current production methodologies for the anti-cancer drug Taxol. We then review methods commonly used for pathway reconstruction in general, and those used specifically for Taxol and discuss their shortcomings. We then discuss our early attempts at pathway reconstruction beginning with *in-vitro* assays and moving forward to a generalizable yeast system based on a rapid assembly method and available transcriptome data. Following method development, we identified a previously-unidentified pathway botteneck preventing pathway elucidation due to low pathway flux.

Chapter 3 focuses on investigating the source of this bottleneck. We identify that it arises from poor selectivity of the first hydroxylation step, catalyzed by the cytochrome P450, taxadiene-5α-hydroxylase. To debottleneck this system, we discuss feasible mechanisms for this enzyme, and investigate these through both biological and chemical methods. The results obtained allow us to propose through multiple lines of evidence that taxadiene-5α-hydroxylase proceeds via an epoxide intermediate, rather than a previously-proposed radical-mediated route.

In Chapter 4 we aim to leverage the above mechanistic insights to improve the selectivity of the first hydroxylation step in paclitaxel biosynthesis. This is performed by investigating three routes to improvement including epoxide hydrolysis engineering, P450 engineering, and terpene cyclase engineering. We then demonstrate the success of one of these methods, and illustrate how altering the selectivity of the terpene cyclase, upstream of the promiscuous oxidation step, can improve whole-pathway

selectivity.

Chapter 5 shifts focus from the final product Taxol and its early biosynthesis to further upstream. Here we discuss limitations of upstream pathways leading to IPP and DMAPP, the universal precursors to all isoprenoids, and methodologies in which they may be overcome. We propose an alternative pathway and perform proof of concept experiments to validate this pathway. We then discuss the shortcomings of available screening methodologies and propose and design tools for the implementation of an evolution-based screening methodology.

Chapter 6 focuses on development of a terpenoid-responsive biosensor. We first demonstrate multiple approaches to heterologous biosensor construction and demonstrate a functional sensor. We next improve the dynamic range of this sensor by first studying and then applying alternative binding motifs. An alternative, inducible, promoter for reporter expression and a constitutive promoter for repressor expression was next applied in order to finalize sensor construction. We conclude by validating ligand responsiveness and demonstrate final biosensor functionality.

In Chapter 7 we address the overall implications of the work and future directions for study. This focuses specifically on the application of methodologies developed in Chapters 5 and 6 for high-throughput optimization of the newly developed biocatalytic isoprenoid pathway. Additionally, remaining bottlenecks for Taxol pathway elucidation are addressed and approaches to overcome them are discussed.

1.4 Bibliography

[1] S. N. Cohen, A. C. Chang, H. W. Boyer, and R. B. Helling, "Construction of biologically functional bacterial plasmids in vitro," *Proceedings of the National Academy of Sciences*, vol. 70, no. 11, pp. 3240–3244, 1973.

[2] G. Stephanopoulos and J. Vallino, "Network rigidity and metabolic engineering in metabolite overproduction," *Scuebce*, vol. 259, p. 9646, 1991.

[3] J. E. Bailey, A. A. for the Advancement of Science, *et al.*, "Toward a science of metabolic engineering," American Association for the Advancement of Science, 1991.

[4] P. K. Ajikumar, K. Tyo, S. Carlsen, O. Mucha, T. H. Phon, and G. Stephanopoulos, "Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms," *Molecular pharmaceutics*, vol. 5, no. 2, pp. 167–190, 2008.

[5] C. Harrison, "Reviewing the synbio startup scene in 2016 - synbiobeta," Aug 2016. Accessed: 2016-09-30.

[6] A. Kunjapur, "An introduction to start-ups in synthetic biology." http://blogs.plos.org/synbio/2015/09/08/an-introduction-to-start-ups-in-synthetic-biology/, Sept 2016. Accessed: 2016-09-30.

[7] C. Bennett, "Natural rubber bounces back for farmers." http://www.agweb.com/article/natural-rubber-bounces-back-for-farmers-naa-chris-bennett//, Dec 2015. Accessed: 2016-09-30.

[8] A. Takada, "Natural rubber faces global surplus for fourth straight year," May 2015. Accessed: 2016-09-30.

[9] S. Wu, M. Schalk, A. Clark, R. B. Miles, R. Coates, and J. Chappell, "Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants," *Nature biotechnology*, vol. 24, no. 11, pp. 1441–1447, 2006.

Chapter 2

Reconstructing Early Paclitaxel Biosynthesis

2.1 Introduction

The plant natural product Taxol (Paclitaxel) has attracted the interest of biologists and chemists for decades due to its potent anti-cancer properties as well as its complex, highly functionalized structure, shown in Figure 2-1 [1, 2, 3, 4, 5]. This molecule, originally derived from the Pacific Yew tree (*Taxus brevifolia*), was found to be a potent anti-cancer therapeutic agent in the 1960s. Currently, Taxol is used primarily in the treatment of metastatic carcinoma of the ovary, metastatic breast cancer, and non-small cell lung cancer [6]. Its method of action (MoA) involves binding to the surface of a microtubule, promoting polymerization, thereby inhibiting cell proliferation by stabilizing the spindle during mitosis, resulting in apoptosis [7]. Its mode of action has also resulted in recent studies for its use against a range of other diseases which require microtubule stabilization such as Psoriasis and Alzheimers yielding positive preliminary results [8, 9]. In addition to Taxol itself, structural analogs may have significant value. This has been illustrated with the semi-synthetic analog Taxotere, and it is imaginable even more potent derivatives of Taxol may arise in the near future. By 2003, over 400 natural taxoids (derived from the universal precursor Taxadiene) had been characterized, some with potential to be more effective

than Taxol, but with recovery from plant tissue too low to be commercialized [10, 11, 12].

Furthermore, Taxol production currently exceeds $1 billion per year [6], while semi-synthetic derivative Docetaxel (Taxotere) had sales of over $3 billion in 2009 [13]. Abraxane, a humuan albumin bound form of paclitaxel, was recently approved as a monotheraphy in metastatic breast cancer and has projected sales exceeing $3 billion/year. Taxoprexin consists of docosahexaenoic acid bound paclitaxel, which is activated by cleavage of the fatty acid. Another formulation, Paclica, contains poly-(l-glutamic acid) conjugated paclitaxel and is in Phase II trials in nonsmall lung cancer, as well as others. In addition to these, other formulations exist including Paccal and ANG1005 [14].

Currently, production of Taxol occurs via a number of methods. Harvesting and extraction of the final product, Taxol, from the bark of Yew trees was the initial mode of production. However, this method required up to three trees to obtain approximately 10 kg of bark, with a final Taxol yield of 1 gram, leading to rapid depletion of this slow-growing natural resource [15]. Later, a less destructive method of production was developed by harvesting of *taxus* needles to obtain the compound 10-deacetlylbaccatin III (Figure 2-1), and performing chemical semi-synthesis to reach Taxol [16]. Despite advances in other areas, extraction from trees remains in use industrially. Bristol Myers was reported in 2000 to maintain a farm of over a billion yew trees while a European company (Indena) recently patented a semisynthetic route from 10-deacetylbaccatin III isolated from cultivated trees the company maintained in 2007 [17].

More recently, research focus has shifted toward the use of plant cell culture. Although plant cell culture has a variety of drawbacks (discussed in more detail below), it has become a viable option due to the disfavorable economics of harvesting from mature trees. However, even with significant advances, expensive fermentation techniques and media components, production titers from plant cell fermentations remain below 150 mg/L after extended culture times [6, 18]. Thus, engineering of native plant metabolism or the heterologous production of Taxol in microbial hosts

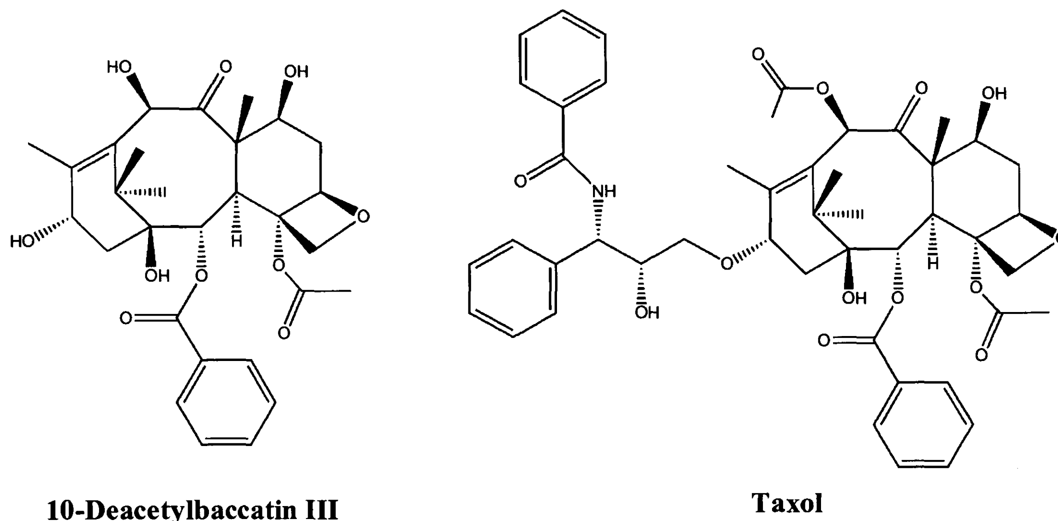

Figure 2-1: **Chemical Structure of 10-Deacetylbaccatin III and Taxol.**

could substantially improve Taxol production [19].

2.2 The Use of Heterologous Hosts for Natural Product Biosynthesis

The choice between plant cell culture and heterologous microbial expression depends on a number of factors. The advantages and disadvantages of these systems will be discussed here in reference to production of Taxol for simplicity, although many of the statements are generalizable. Plant cell culture has the primary advantage of native, endogenous production of the desired molecule with no genetic engineering of the plant cells and with no knowledge of the production pathway required. However, even with the use of inducers such as methyl jasmonate, terpenoids only make up between 2% and 3% of dry cell weight [20], with Taxol making up approximately 6% of these taxoids [21], resulting in yields which fail to exceed 137/5 mg/L following 42 days of fermentation [22]. Other difficulties arising in plant cell culture include a tendency to form aggregates in suspension, resulting in heterogeneity of environment such as oxygen and nutrient availability, reducing reproducability [23]. Additionally, variability and cell line instability are two of the largest obstacles to plant cell culture,

plant cells from different lines (or even within the same line) exhibit extremely high levels of variability and cell lines quickly degrade [20].

In contrast to plant cell culture, production of Taxol in microbial hosts requires an understanding of pathway architecture and the specific genes involved, which poses a non-trivial challenge. Additionally, functional expression of the pathway enzymes has at times proven difficult [24], and a large amount of troubleshooting must be performed to establish a minimally functional pathway. However, once established, this method of production has several advantages. First, unlike plant cells, host metabolism is well understood and genetic manipulation may be made with relative ease in order to improve yields. The metabolism of *Taxus* cells is much less well characterized, and although minor successes have been achieved in genetic modifications, this is an area still in its infancy and changes to host metabolism are likely to remain a challenge. However, when using *E. coli*, large number of modifications to the host metabolism and the heterologous pathway can be combined and evaluated rapidly [20]. Additionally, in *E. coli*, few natural product synthesis pathways exist natively, and none containing P450's. This eliminates 'cross talk' and prevents issues with native host metabolism and reduces complexity of engineering as well as separation. Furthermore, many microbial platforms are extensively studied, allowing one to leverage a significant body of literature when performing engineering, and well-established methodologies exist in order to divert carbon flux into secondary metabolic pathways. As an example of the power of heterologous systems, high level production of amorpha-4,11-diene (a precursor to Artemisinin) has been achieved at 25 g/L; if 0.2% of this yield was achieved for a late-stage Taxol intermediate, this would be superior to current plant based systems [25].

Another primary advantage of heterologous systems over those for plant systems is the ease and simplicity of running the resulting bioprocess. Among the most important aspects are that in *Taxus* cell culture production of taxoids is growth coupled, only being exhibited during exponential phase [26]. In heterologous hosts, growth may be uncoupled, allowing for a growth phase to build up cell mass prior to induction and a production phase. Additionally microbes such as *E. coli* are much more

resistant to sheer stress, allowing for better agitation, mixing, and aeration, and have a much shorter doubling time, shortening the potential length of a run [20].

This also raises the possibility of utilizing microbial synthesis as a platform for pathway discovery and modification. In the construction of a biosynthetic pathway in a heterologous host, only desired enzymes are added. This allows for the steps to be added iteratively, and for rapid redesign of the pathway as the function of additional enzymes are characterized. This enables the creation and screening of new, potentially improved drugs. Many studies have been performed on Taxol and its analogs and a great deal is known about the clinical relevance of its functional groups, allowing for potential improvement of the drug. As an example, a recent study has isolated a natural analog with anti-tumor properties, as well as improved bioavailability, however this is substituted at both carbons 13 and 14, and no known enzymes exist to perform this functionality [27]. These compounds are typically present at much lower quantities in plant cells, raising issues about titer and economic feasibility. In contrast, if a second-generation taxoid drug is proven effective, biochemical characterization of pathway enzymes rapidly enables production when using a heterologous system. Semi-synthesis of similar compounds with similarly improved properties has also been investigated [28].

2.3 The Taxol Biosynthetic Pathway

An understanding of the Taxol biosynthetic pathway is essential to production in heterologous hosts. Extensive work by Rodney Croteau's lab has led to the identification of several genes in the pathway. However, many steps remain uncharacterized, and the order of known functionalizations remains a mystery. A diagram of the early biosynthetic pathway is given in Figure 2-2, which shows all previously characterized enzymes and the substrates used to test these enzymes, although from this diagram it is clear that knowledge of the pathway remains imprecise.

OPP

GeranylgeranylPP (GGPP)

Taxadiene Synthase
(Source)

Taxadiene
5α-Hydroxylase
(Rontein 2007)

11
12
3
5

5(12)-Oxa-3(11)-cyclotaxane (OCT)

Taxa-4(20),11(12)-diene
Incoperated into Taxol in *Taxus* cell culture

Taxa-4(20),11(12)-dien-7b-ol
(Chau 2004)

Taxa-4(20),11(12)-dien-14b-ol
(Jennewein 2003)

Taxadiene
5α-Hydroxylase
(Jennewein 2004)

Taxadiene
14b-Hydroxylase
(Jennewein 2003)

Taxa-4(20),11(12)-dien-5a-14b-diol
(Off Pathway Product)

Taxa-4(20),11(12)-dien-5a-ol
Incoperated into Taxol in *Taxus* Cell culture

Taxa-4(20),11(12)-dien-14b-ol
(Jennewein 2003)

Mutiple Acetyl-Transferases
(Chau 2004-b)

Taxadiene
13β-Hydroxylase
Jennewein 2001

Taxadiene
14b-Hydroxylase
(Preferred by a factor of 6 over 5a-OH
(Jennewin 2003)

Taxa-4(20),11(12)-dien-5a-14b-diol-5-acetate

Taxa-4(20),11(12)-dien-5a-yl-acetate
Incoperated into Taxol in *Taxus* Cell culture

Taxa-4(20),11(12)-dien-7b-ol
(Chau 2004)

Taxadiene
10β-Hydroxylase
Jennewein 2001
(Two genes exist,
one can only use this route)

Followed by

Mutiple
Acetyl-Transferases
(Chau 2004-b)

Taxa-4(20),11(12)-dien-5a-13a-diol
Incoperated into Taxol in *Taxus* Cell culture

Taxa-4(20),11(12)-dien-5a-13a-14b-triol
(Jennewein 2003)

Taxa-4(20),11(12)-dien-5a-10b-13a-triol
(Jennewein 2001)
Acetylation Followed by 10b Hydroxylation not Tested

Taxadiene
13β-Hydroxylase
Jennewein 2001
(Jennewein 2001)
6% of tybe rate of 5a-OH

Taxadiene
10β-Hydroxylase
Jennewein 2001
(Two genes exist, only
one has neem stidoed beyond
proving functionality
on 5-Ac)

Taxa-4(20),11(12)-dien-5a-13a-diol-5-acetate

Taxa-4(20),11(12)-dien-5a-10b-13a-triol-5-acetate
(Jennewein 2001)
Acetylation Followed by 10b Hydroxylation not Tested

Taxa-4(20),11(12)-dien-5a-10b-diol-5-acetate
Incoperated into Taxol in *Taxus* Cell culture

Taxa-4(20),11(12)-dien-5a,10b,13a-triol-5-acetate
(Jennewein 2001)

Taxa-4(20),11(12)-dien-5a,7b-diol-5-acetate
(Chau 2004)

Figure 2-2: **Characterized Steps in Early Paclitaxel Biosynthesis**

The Taxol biosynthetic pathway is proposed to proceed via 19 enzymatic steps from geranylgeranyl diphosphate (GGPP) [29]. This complex pathway includes 8 oxygenation steps catalyzed by cytochrome P450s, three acylations, and a range of additional transformations including side chain modification and attachment during late-stage biosynthesis. Biosynthesis begins with the cyclization of the isoprenoid precursor geranylgeranyl pyrophosphate (GGPP) to form taxa-4(5),11(12)-diene (taxadiene) by taxadiene synthase [30]. Taxadiene is then hydroxylated at the 5α position by the cytochrome P450 taxadien-5α-hydroxylase (CYP725A4) [31, 32]. This product, taxa-4(20),11(12)-dien-5α-ol, has been shown to undergo further hydroxylation at the 13α, 10β, and 14α positions, with hydroxylation at the 13α or 10β positions capable of biotransformation into Taxol in the native host [33, 34, 35].

Previous work in our lab has led to the production of taxadiene at titers of 1 g/L via over-expression of rate-limiting MEP pathway steps, GGPP synthase, and taxadiene synthase [1]. This work represents the highest production titer of any pathway intermediate, and also the furthest progression down the pathway produced without exogenous feeding late-pathway intermediates. In another recent attempt to reconstruct the pathway in a heterologous system, a multi-step pathway consisting of 8 pathway steps (GGPP synthase, taxadiene synthase, an acetyl-transferase, and enzymes to perform hydroxylations at the 5α ,10β , 2α, and 13α positons) was recently expressed in a *Saccharomyces cerevisiae* expression system, however only low levels of the taxadiene (1 mg/L) and taxadien-5α-ol (25 ug/L) were obtained [36].

To understand the difficulties in reconstructing the early biosynthetic pathway, it is important to note that the Taxol biosynthetic pathway is proposed to be a branching, nonspecific pathway. This has important implications as many enzymes may act on a given substrate, but when performed in an improper order, lead to dead-end metabolites. For example, it is known that acetylation at 5α most likely takes place early in the pathway, but at which point is unknown [37]. Further work later revealed a large number of acetyl-transferases, each with differing activities on a range of substrates, but with many having overlapping activity including the ability to aceylate the 5α position with or without functional groups at other locations [38]. Hydrox-

ylation at the 10β step has also been demonstrated. This enzyme has been shown to have a strong preference for the 5α-acetylated form over the hydroxylated form, suggesting that this step follows acetylation [39]. Alternatively, branches exist even very early on in the pathway, as a possible branch following initial hydroxylation leads to hydroxylation at the 13α position. In contrast to the 10β hydroxylase, this step has been shown to have a strong preference for the hydroxylated form of 5α over the acetylated form. An important note is that hydroxylation at the 13α position appears to prevent hydroxylation at the 10β position, although work with an acetylated form has not been attempted [40].

Unfortunately, due to the low yield of these pathway steps and the uncertainty of the steps following them, pathway elucidation has not proceeded past this step using these substrates. However, it has been demonstrated that taxadiene, taxadien-5α-ol, taxadien-5α-yl-acetate, taxadien-5α,10β-diol, taxadien-5α,13α-diol, and taxadien-5α,10β-diol-5α-acetate are all incorporated into Taxol when fed to plant cell culture [40, 41]. Therefore, even if these metabolites are not on the primary pathway utilized by plants, a pathway can at least be constructed that includes these first steps. Work to understand the subsequent pathway steps has involved examining the oxidation frequency of taxoids found in cell cultures. The patterns of oxidation frequency indicate the remaining pathway steps proceed in the order of hydroxylation at the 5α, 10β, 2α, 9α, 13α, and 7β positions, followed by epoxidation at the C4-C5 position and hydroxylation at the 1α position [42]. This work is also reinforced by the phylogenetic relationship between the pathway enzymes [43]. Interestingly, an enzyme leading only to off-pathway products was recently characterized as well. This enzyme, performed hydroxylations at the 14β position. Taxol has no hydroxylation at this position and thus it is unlikely that this step lies on the pathway [44]. However, knowledge of this enzyme and its regioselectivity may enable future work on engineering enzyme specificity.

The branching, promiscuous, nature of the Taxol biosyntetic pathway has important implications for pathway elucidation. Moving past the 10β-hydroxylating enzyme little is known about the pathway. This has historically been due primarily

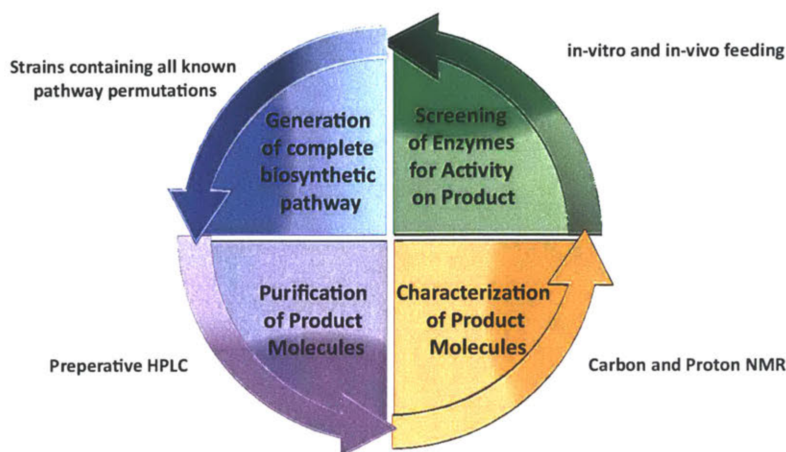

Figure 2-3: **Pathway Elucidation Cycle**

to the availability of early-stage pathway precursors, which previous researchers have primarily attempted to synthesize retrosynthetically via the defunctionalization of advanced precursors [45]. This has the disadvantage that the created products may not actually be precursors in the pathway, but can be used due the promiscuity in substrate range of many enzymes. This highlights the importance of building a more reliable system for Taxol pathway elucidation.

2.4 Building a Generalizable System

We next aimed to construct a generalizable system for the rapid pathway elucidation and screening of a library of cytochrome P450s and transacetylases. This approach was designed with several parameters in mind. First, synthetic, rather than retrosynthetic, compound generation is utilized to ensure any compounds tested can be feasibly generated via a known pathway. This is in contrast to previous work, which relies on de-functionalizing late stage intermediates to yield products which may or may not lie on the pathway. The general approach taken is illustrated in Figure 2-3; beginning with a high-producing background strain, sequential addition of genes allows for production of previously unreachable metabolites and may allow for rapid pathway elucidation.

With this approach in hand, we next sought to design a system allowing for

rapid library construction and cloning, modular library assembly as new information becomes available, and rapid screening for functionality and expression. To ensure rapid library construction and modularity, we began by selecting *S. cerevisiae* yeast strain BY4741 as our expression host. This was done for multiple reasons; first, yeast is superior to other hosts, such as *E. coli* for the expression of complex and membrane-bound proteins in a functional state. Next, this yeast strain possesses four auxotrophic markers (His, Leu, Met, Ura) [46]. Thus, we are capable of making up to four independent chromosomal integrations, allowing for multiple pathway steps to be integrated and screened in tandem if required. We then selected four previously characterized integration sites [47] and constructed backbone plasmids within which genes may be expressed:

- YorwΔ22 - URA
- YPRCΔ15 - LEU
- ADH6 - HIS
- PDC6 - MET

Importantly, in the design of these plasmids, promoter and terminator sequences remain identical, thus only a single set of primers is required to clone a P450 PCR fragment into any backbone. We also selected CLIVA [48] as our assembly method of choice due to reduced cost compared to Gibson assembly, with cost originating in primer generation, greatly increasing modularity and decreasing library generation costs. A primary advantage of this methodology is that PCR mixtures can be obtained for a given candidate and these can be combined with any generic backbone, thus only a single set of intial PCR's is required for the elucidation of any pathway consisting of 4 or fewer steps, and other yeast strains can be used beyond this.

It is important to note that the route to Taxol, due to the functional plasticity of enzymes, may not be a linear pathway but rather a network of interconnected routes with several common nodes [49]. This presents the opportunity to select one route and optimize it to eliminate side products and improve yield. However, it also complicates

pathway elucidation. Thus, we initially proposed two separate methodologies for pathway elucidation.

The first methodology involves sequential pathway construction, and is depicted in Figure 2-4. Initially, a library is fed the primary substrate (taxadiene) and screened for activity. Positive hits are isolated, and metabolites are characterized. New libraries are then generated which consist of the P450 library expressed in a strain that also contains a positive hit from the previous round, with one library required for each positive hit for the previous round. In this regime, for each metabolitic route, a unique library is constructed, with multiple libraries required for branching pathway. However, this has the advantage that once a branch leading to a desired metabolite is generated, the pathway is known to be functional, expressed, and operating, and thus only needs to be optimized rather than built. This methodology is agnostic to complex interactions such as multiple actions by a single enzyme, or competition between enzymes for a single substrate, but is unable to provide information on individual enzyme activity. Additionally, this has the advantage that only a single metabolite (taxadiene) must be fed to each library, and thus large scale generation of only this metabolite is required.

The second methodology requires the generation of only a single library, and screening of this library against compounds as they are discovered, this is represented in Figure 2-5. This has the advantage that enzyme properties are characterized in isolation and only a single round of cloning must be performed. However, each metabolite must be prepared not only in a quantity required for characterization, but also for subsequent studies.

We next set out to identify library P450s and transacetylases for pathway elucidation. Fortunately, several transcriptomic data sets are publicly available and thus we relied on these to begin drafting our library. We first selected sequences bearing similarity to known P450s and acetyltransferases. However, many sequences were short contigs rather than full genes, and thus we required additional information to obtain full gene sequences for our libraries. We began by pooling the following data sets:

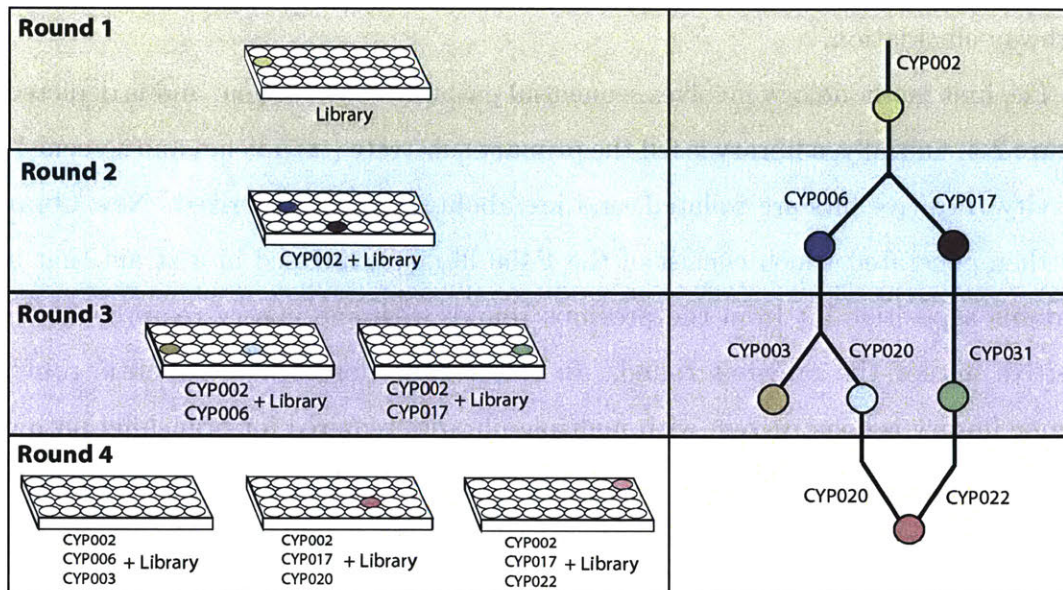

Figure 2-4: **Multiple Library Methodology for Pathway Elucidation.** Enzymatic activity is depicted via black lines, while metabolites are shown as filled circles.

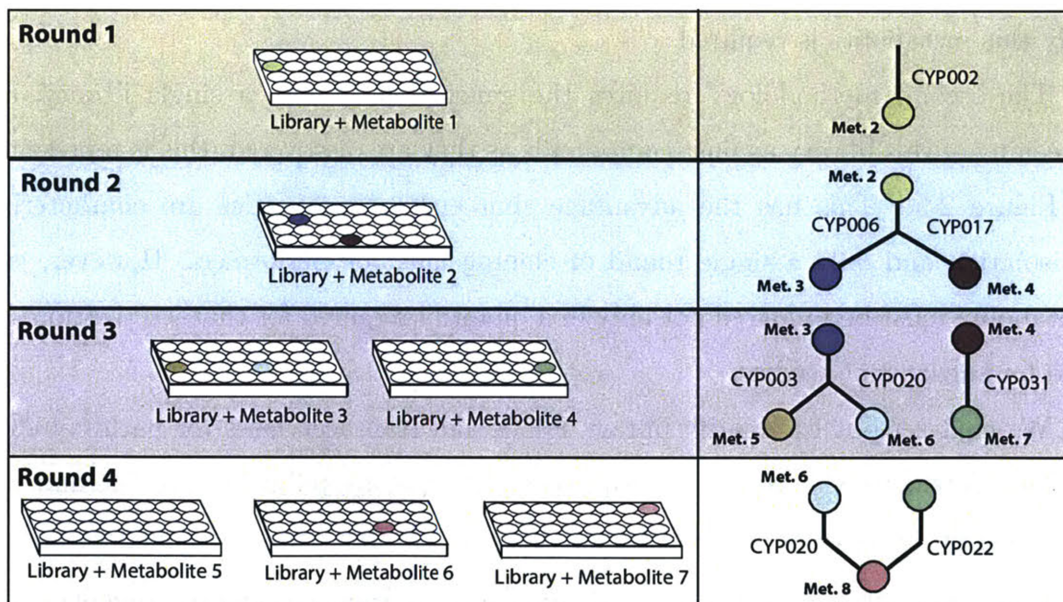

Figure 2-5: **Single Library Methodology for Pathway Elucidation.** Enzymatic activity is depicted via black lines, while metabolites are shown as filled circles.

- Patents Issued by Rodney Croteau [50]
- Previous Publications Characterizing P450s [15, 16, 39]
- Differential Tissue Transcriptomics [51]
- Transcriptome Derived from Meristamitc Stem Cells [52]
- Transcriptomic Data from Differential Induction [53]
- An EST-Based Methodology from Differential Induction [54]

Flagged P450s and acetyltransferases were subjected to reciprocal BLAST and placed into groupings. These groupings were confirmed by a phylogenetic assessment of all sequences, which confirmed correct grouping of proposed P450s (Figure 2-6). These groupings were then aligned and assembled into full-length P450 sequences. This yielded the full library of sequences. Interestingly, when examining P450 sequences, a phylogenetic cluster consisting of all 7 known taxoid-acting P450s was observed (as shown in in Figure 2-7). We thus hypothesize that this cluster contains high-probability targets for further study. CYP09 in particular has not been screened enzymatically and is the only non-characterized gene that exists within a cluster containing the 2α ,5α ,7β, and 14α genes, and is thus of high interest. Full sequences for all library genes are available in Appendix A.

Figure 2-6: **Cladogram of Aligned Contigs.** All contigs used for recroprical BLAST, colored according to grouping.

2.5 Pathway Elucidation

In order to speed pathway elucidation, we first aimed to confirm functional expression of P450s in our system prior to screening. To do this, we relied on the fact that P450s have unique carbon monoxide difference spectra [55]. Thus, we are able to quantify functional P450 expression. As shown in Figure 2-8, taxadiene-5α-hydroxylase was confirmed to be functionally expressed within our system. This is a useful tool as it allows quantification of P450 function without prior knowledge of substrates or products.

We next sought to perform proof-of-concept experiments on our system by applying it to known pathway steps and a small enzymatic library. This library included two known acetyl transferases, taxadiene-5α-hydroxylase, two taxadiene-10β-hydroxylases, and a taxadien-13α-hydroxylase [56]. Following initial characterization, taxadiene-5α-hydroxylase was found to be the sole enzyme acting on taxadiene (depicted in Figure 2-9, 'Round 1'). In the next iteration, this gene was included in the background strain and the remaining library genes were added. As shown in Figure 2-9, 'Round 2', only a single acetyltransferase was found to have activity upon this substrate in our experiments. This experiment revealed production of a new metabolite with GCMS retention time and mass spectra matching that of a mono-acetylated compound, confirming taxadien-5α-ol could be acetylated in our system to generate taxa-4(20),11-dien-5alpha-yl acetate, as shown in Figure 2-10. Thus making use of both of our proposed methodologies.

Following this step, the process was repeated (Figure 2-9, 'Round 3'), yielding a compound matching predicted retention time and mass spectra of a di-hydroxylated mono-acetylated compound (Figure 2-11). However, following complete reconstruction of the pathway in *E. coli*, and then in a yeast co-culture system [57], this compound was not available at sufficient quantity, even in scaled-up reactors, to allow for characterization.

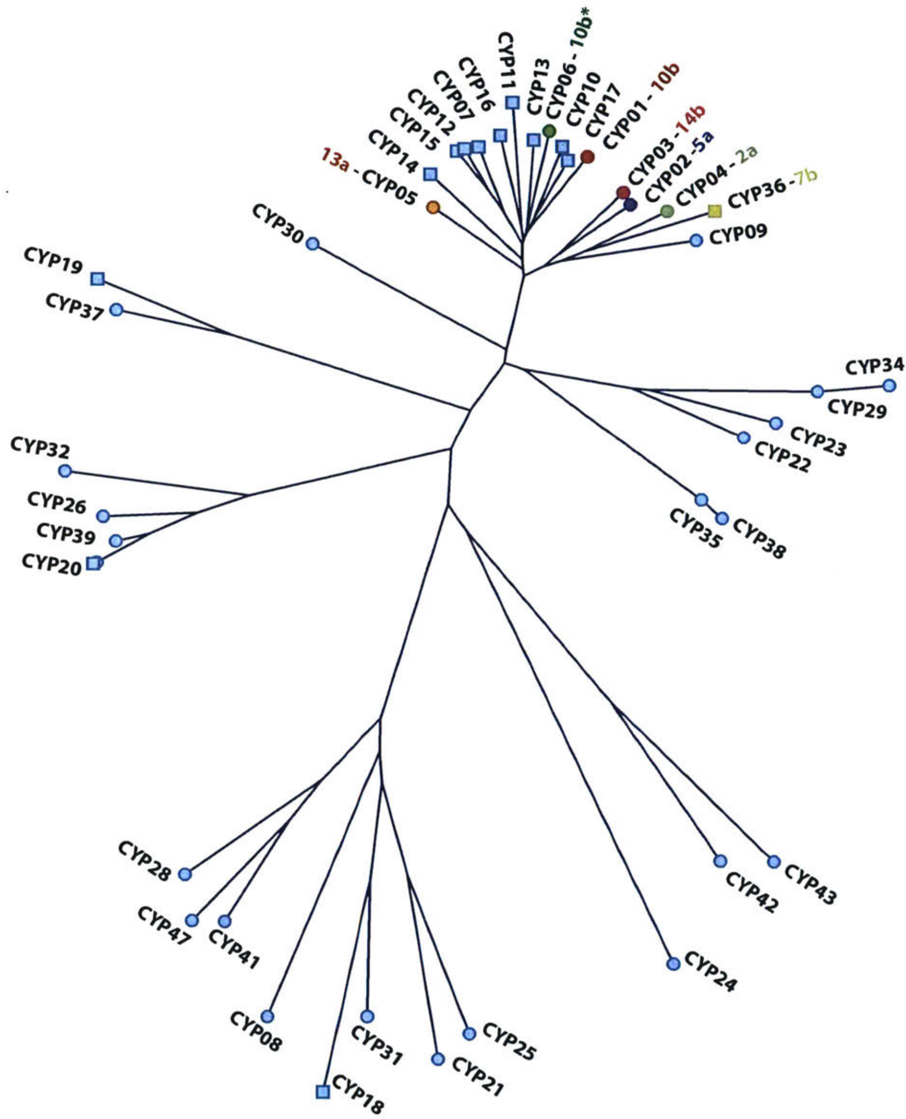

Figure 2-7: **CYP Protein Phylogenetic Tree.** Non-blue circles represent CYPs with characterized functionality, and locations of hydroxylation are noted. Of note is the cluster of 17 sequences in which all 7 known taxoid-acting CYPs exist, indicating the other CYPs in this cluster may also encode for taxoid hydroxylases. CYPs marked with circles are one which are found in previous patent literature, while those with squares have not.

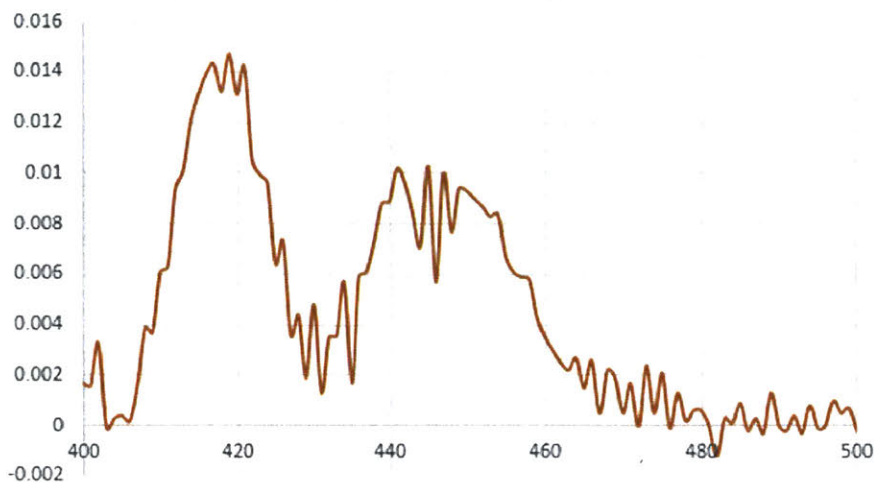

Figure 2-8: **CO Difference Spectra for Purified Taxadiene-5α-hydroxylase.** TSpectra given is from fully purified protein, during which partial denaturation occured (peak at 420nm), similar peaks are observed *in-vivo*.

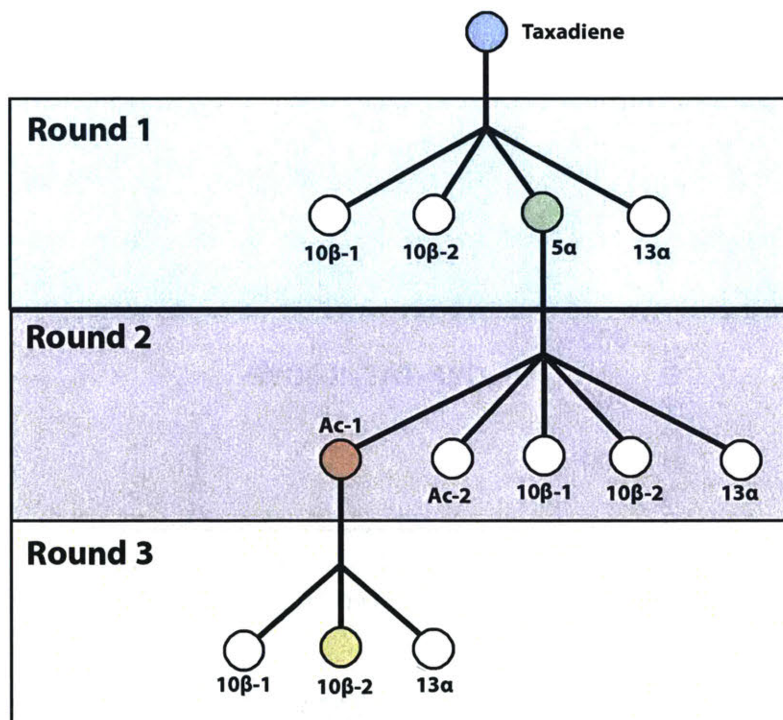

Figure 2-9: **Pathway Elucidated from Proof-of-Concept Screening.** Filled circles indicate metabolites observed following library screening, unfilled circles indicate a lack of activity.

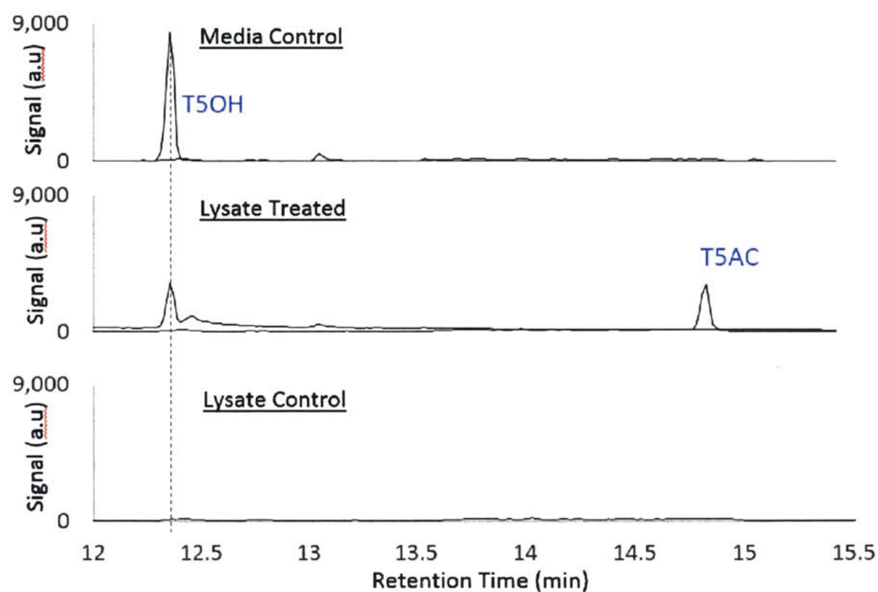

Figure 2-10: ***In-Vitro* Acetylation of Taxadiene-5α-ol.** Peaks labeled T5OH and T5AC represent taxadiene-5α-ol and its acetylated version, respectively.

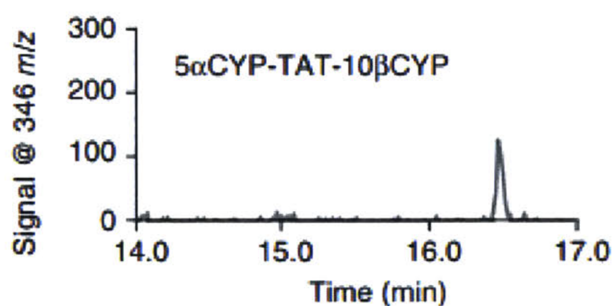

Figure 2-11: **Mass Spectra from Yeast Expressing Three Functional Pathway Steps.** Peak at 346 m/z indicates a di-hydroxylate mono-acetylated compound. Figure Reproduced from [57]

2.6 Conclusions and Proposals for Future Work

Using the methodologies outlined in this chapter, we were able to reconstruct the 5α-hydroxylase, 5α-acetyltransferase, and 10β-hydroxylase steps in a heterologous system for the first time. However, despite the success of reconstituting 5 pathway steps in *E. coli*, and the generation of a generalizable screening system, it rapidly became infeasible to continue pathway elucidation. This was due to a loss of yield during pathway elucidation.

Returning to *in-vitro* data we obtained on our pathway, it was revealed that this loss in yield arises primarily from the first hydroxylation step catalyzed by taxadiene-5α-hydroxylase. Large amounts of the first cyclization product (taxadiene) were biotransformed with only 5% to 15% selectivity in many reconstituted systems. This indicated selectivity of the oxygenation step by taxadiene-5α-hydroxylase stood as a key obstacle to heterologous Taxol production. In order to continue pathway elucidation, improvements in early pathway flux are essential. Thus, for the following chapter we began an investigation into this pathway step and the source of this poor selectivity with the goal of improving it.

It is vital to note that once such a bottleneck is alleviated, pathway elucidation is now a straightforward task using the tools and methodologies developed within this chapter. Additionally, for simple pathway elucidation, recent work by Baran et al. [58] details methodologies to prepare gram-scale quantities of taxadiene, which may be much less labor intensive than multiple bioreactor runs. Additionally, as discussed later in Chapter 4, generation of an isomeric form of taxadiene through chemical or biological synthesis can be used to overcome this bottleneck and thus may be utilized as well.

2.7 Materials and Methods

Candidate Enzyme Selection. The CLC Genomics software suite was used for all reciprocal blast, alignment, and phylogenetic analysis. Reciprocal blast results were

considered positive when a percent identity of 90% or above, with 40 base pairs or more of overlap were present, and each contig was the top hit for its partner. This was performed iteratively using each of the 4 datasets. Following this, groups were aligned and all alignments were manually confirmed.

Strain Construction. For generation of cDNA, RNA was extracted from Taxus baccata suspension cell cultures following methyl jasmonate induction and 48 hours of incubation. Plant cell cultures were homogenized manually with mortar and pestle prior to RNA extraction. A portion of plant cell culture was also subjected to LCMS analysis to confirm Taxol production, if Taxol production was not observed, extracted cDNA was discarded. 5' and 3' race were performed according to previously established protocols [59] when necessary. CLIVA was utilized in order to assemble vectors. *S. cerevisiae* was transformed using the standard lithium acetate transformation protocol with 10 mg/mL boiled sheared salmon sperm DNA.

***In-Vitro* Enzyme Characterization.** Crude lysates for enzymatic assays were performed by diluting cells to an equal optical density followed by lysis with BPER II. NADPH cycling31 was used to ensure adequate NADPH supply by supplying NADP, Glucose-6-Phosphate, and Glucose-6-Phosphate Dehydrogenase when employing P450s. Reaction was initiated upon addition of taxadiene, with reactions performed without additional supplied NADPH showing no turnover. Products were extracted using ethyl acetate and analyzed via GCMS. Small scale fermentations.

For E. coli cultures, a single colony of E. coli was inoculated into LB media (10g/L tryptone, 5 g/L yeast extract, 10g/L NaCl, pH 7) and shaken overnight (37°C,250 r.p.m.). 20ul of overnight culture were inoculated into 1 mL of small-scale culture media (5 g/L yeast extract, 10 g/L tryptone, 15g/L glycerol, 10g/L NaCl, 100 mM HEPES, pH 7.6), unless otherwise stated, in a 14 mL pyrex glass tube. Cultures were induced upon inoculation with 0.1 mM IPTG and shaken (22°/250) for 5 days prior to analysis of product distribution. For feeding, taxadiene was added as a solution in DMSO (1% final culture volume, final concentration of 100mg/L) at the time of inoculation.

Quantification of Isoprenoids. At the conclusion of fermentations, 200µL of fer-

mentation broth was sampled and added to 200µL ethyl acetate containing internal standard and 100µL of 0.1mm glass beads. This solution was then vortexed at 4°C for 30 minutes and centrifuged at 18,000g for 10 minutes. The top, organic, layer was extracted and 1µL was subjected to analysis via GCMS. The GCMS consisted of a Varian Saturn 3800 GC containing an HP-5ms (Agilent Technologies, USA) column, and Varian 2000 MS, with helium at a flow rate of 1mL/min as the carrier gas. The injector and transfer lines were maintained at 250°C. For oxygenated taxanes, the column oven was maintained at 100C for 1 minute and increased at a rate of 15°C/min to 175°C, then at 4°C/min to 220°C, then at 50°C/min to 290°C, and held at this temperature for 1 minute. For analysis of taxadiene isomers the column oven was maintained at 100°C for 1 minute and increased at a rate of 15°C/min to 165°C, then at 2°C/min to 170°C, held at this temperature for 7 minutes, increased at 50°C/min to 290°C, and held at this temperature for 1 minute. Caryophyllene (90mg/L) was used as an internal standard. Analysis of dodecane and DINP was performed by diluting 100-fold into internal standard.

2.8 Bibliography

[1] P. K. Ajikumar, W.-H. Xiao, K. E. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, and G. Stephanopoulos, "Isoprenoid pathway optimization for taxol precursor overproduction in escherichia coli," *Science*, vol. 330, no. 6000, pp. 70–74, 2010.

[2] K. Nicolaou, Z. Yang, J. J. Liu, H. Ueno, P. Nantermet, R. Guy, C. Claiborne, J. Renaud, E. Couladouros, K. Paulvannan, *et al.*, "Total synthesis of taxol," *Nature*, vol. 367, no. 6464, pp. 630–634, 1994.

[3] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos, "Distributing a metabolic pathway among a microbial consortium enhances production of natural products," *Nature biotechnology*, vol. 33, no. 4, pp. 377–383, 2015.

[4] A. Mendoza, Y. Ishihara, and P. S. Baran, "Scalable enantioselective total synthesis of taxanes," *Nature chemistry*, vol. 4, no. 1, pp. 21–25, 2012.

[5] S. Soliman and Y. Tang, "Natural and engineered production of taxadiene with taxadiene synthase," *Biotechnology and bioengineering*, vol. 112, no. 2, pp. 229–235, 2015.

[6] S. Malik, R. M. Cusidó, M. H. Mirjalili, E. Moyano, J. Palazón, and M. Bonfill, "Production of the anticancer drug taxol in taxus baccata suspension cultures: a review," *Process Biochemistry*, vol. 46, no. 1, pp. 23–34, 2011.

[7] T. K. Yeung, C. Germond, X. Chen, and Z. Wang, "The mode of action of taxol: apoptosis at low concentration and necrosis at high concentration," *Biochemical and biophysical research communications*, vol. 263, no. 2, pp. 398–404, 1999.

[8] A. Ehrlich, S. Booher, Y. Becerra, D. L. Borris, W. D. Figg, M. L. Turner, and A. Blauvelt, "Micellar paclitaxel improves severe psoriasis in a prospective phase ii pilot study," *Journal of the American Academy of Dermatology*, vol. 50, no. 4, pp. 533–540, 2004.

[9] B. Zhang, A. Maiti, S. Shively, F. Lakhani, G. McDonald-Jones, J. Bruce, E. B. Lee, S. X. Xie, S. Joyce, C. Li, *et al.*, "Microtubule-binding drugs offset tau sequestration by stabilizing microtubules and reversing fast axonal transport deficits in a tauopathy model," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 1, pp. 227–231, 2005.

[10] O. Expósito, M. Bonfill, E. Moyano, M. Onrubia, M. Mirjalili, R. Cusido, and J. Palazon, "Biotechnological production of taxol and related taxoids: current state and prospects," *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, vol. 9, no. 1, pp. 109–121, 2009.

[11] H. Itokawa and K.-H. Lee, *Taxus: the genus Taxus.* CRC Press, 2003.

[12] D. G. Kingston, "The shape of things to come: structural and synthetic studies of taxol and related compounds," *Phytochemistry*, vol. 68, no. 14, pp. 1844–1854, 2007.

[13] A. L. Demain and P. Vaishnav, "Natural products for cancer chemotherapy," *Microbial biotechnology*, vol. 4, no. 6, pp. 687–699, 2011.

[14] C. Khanna, M. Rosenberg, and D. Vail, "A review of paclitaxel and novel formulations including those suitable for use in dogs," *Journal of Veterinary Internal Medicine*, vol. 29, no. 4, pp. 1006–1012, 2015.

[15] K. C. Nicolaou, W.-M. Dai, and R. K. Guy, "Chemistry and biology of taxol," *Angewandte Chemie International Edition in English*, vol. 33, no. 1, pp. 15–44, 1994.

[16] A. Boa, P. Jenkins, and N. Lawrence, "Recent progress in the synthesis of taxanes," *Contemporary Organic Synthesis*, vol. 1, no. 1, pp. 47–75, 1994.

[17] S. Chandra, "Endophytic fungi: novel sources of anticancer lead molecules," *Applied microbiology and biotechnology*, vol. 95, no. 1, pp. 47–59, 2012.

[18] F. Nosten and N. J. White, "Artemisinin-based combination treatment of falciparum malaria," *The American journal of tropical medicine and hygiene*, vol. 77, no. 6 Suppl, pp. 181–192, 2007.

[19] V. Hale, J. D. Keasling, N. Renninger, and T. T. Diagana, "Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs," *The American journal of tropical medicine and hygiene*, vol. 77, no. 6 Suppl, pp. 198–202, 2007.

[20] S. C. Roberts, "Production and engineering of terpenoids in plant cell culture," *Nature chemical biology*, vol. 3, no. 7, pp. 387–395, 2007.

[21] R. Ketchum and R. Croteau, "The taxus metabolome and the elucidation of the taxol® biosynthetic pathway in cell suspension cultures," in *Plant Metabolomics*, pp. 291–309, Springer, 2006.

[22] H.-K. Choi, S.-I. Kim, J.-S. Son, S.-S. Hong, H.-S. Lee, and H.-J. Lee, "Enhancement of paclitaxel production by temperature shift in suspension culture of taxus chinensis," *Enzyme and microbial technology*, vol. 27, no. 8, pp. 593–598, 2000.

[23] R. Ketchum and D. Gibson, "Paclitaxel production in suspension cell cultures of taxus," *Plant cell, tissue and organ culture*, vol. 46, no. 1, pp. 9–16, 1996.

[24] M. C. Chang, R. A. Eachus, W. Trieu, D.-K. Ro, and J. D. Keasling, "Engineering escherichia coli for production of functionalized terpenoids using plant p450s," *Nature chemical biology*, vol. 3, no. 5, pp. 274–277, 2007.

[25] H. Tsuruta, C. J. Paddon, D. Eng, J. R. Lenihan, T. Horning, L. C. Anthony, R. Regentin, J. D. Keasling, N. S. Renninger, and J. D. Newman, "High-level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in escherichia coli," *PLoS One*, vol. 4, no. 2, p. e4489, 2009.

[26] V. Srinivasan, L. Pestchanker, S. Moser, T. J. Hirasuna, R. A. Taticek, and M. L. Shuler, "Taxol production in bioreactors: kinetics of biomass accumulation, nutrient uptake, and taxol production by cell suspensions of taxus baccata," *Biotechnology and bioengineering*, vol. 47, no. 6, pp. 666–676, 1995.

[27] M. I. Nicoletti, T. Colombo, C. Rossi, C. Monardo, S. Stura, M. Zucchetti, A. Riva, P. Morazzoni, M. B. Donati, E. Bombardelli, *et al.*, "Idn5109, a taxane with oral bioavailability and potent antitumor activity," *Cancer research*, vol. 60, no. 4, pp. 842–846, 2000.

[28] C. Ferlini, M. Distefano, F. Pignatelli, S. Lin, A. Riva, E. Bombardelli, S. Mancuso, I. Ojima, and G. Scambia, "Antitumour activity of novel taxanes that act at the same time as cytotoxic agents and p-glycoprotein inhibitors," *British journal of cancer*, vol. 83, no. 12, p. 1762, 2000.

[29] M. Hezari and R. Croteau, "Taxol biosynthesis: an update," *Planta medica*, vol. 63, no. 04, pp. 291–295, 1997.

[30] M. R. Wildung and R. Croteau, "A cdna clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis," *Journal of Biological Chemistry*, vol. 271, no. 16, pp. 9201–9204, 1996.

[31] J. Hefner, S. M. Rubenstein, R. E. Ketchum, D. M. Gibson, R. M. Williams, and R. Croteau, "Cytochrome p450-catalyzed hydroxylation of taxa-4 (5), 11 (12)-diene to taxa-4 (20), 11 (12)-dien-5a-ol: the first oxygenation step in taxol biosynthesis," *Chemistry & biology*, vol. 3, no. 6, pp. 479–489, 1996.

[32] S. Jennewein, R. M. Long, R. M. Williams, and R. Croteau, "Cytochrome p450 taxadiene 5α-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis," *Chemistry & biology*, vol. 11, no. 3, pp. 379–387, 2004.

[33] S. Jennewein, C. D. Rithner, R. M. Williams, and R. B. Croteau, "Taxol biosynthesis: taxane 13α-hydroxylase is a cytochrome p450-dependent monooxygenase," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13595–13600, 2001.

[34] A. Schoendorf, C. D. Rithner, R. M. Williams, and R. B. Croteau, "Molecular cloning of a cytochrome p450 taxane 10β-hydroxylase cdna from taxus and functional expression in yeast," *Proceedings of the National Academy of Sciences*, vol. 98, no. 4, pp. 1501–1506, 2001.

[35] S. Jennewein, C. D. Rithner, R. M. Williams, and R. Croteau, "Taxoid metabolism: taxoid 14β-hydroxylase is a cytochrome p450-dependent monooxygenase," *Archives of biochemistry and biophysics*, vol. 413, no. 2, pp. 262–270, 2003.

[36] C. Ferlini, M. Distefano, F. Pignatelli, S. Lin, A. Riva, E. Bombardelli, S. Mancuso, I. Ojima, and G. Scambia, "Antitumour activity of novel taxanes that act at the same time as cytotoxic agents and p-glycoprotein inhibitors," *British journal of cancer*, vol. 83, no. 12, p. 1762, 2000.

[37] K. Walker, A. Schoendorf, and R. Croteau, "Molecular cloning of a taxa-4 (20), 11 (12)-dien-5α-ol-o-acetyl transferase cdna from taxus and functional expression in escherichia coli," *Archives of biochemistry and biophysics*, vol. 374, no. 2, pp. 371–380, 2000.

[38] M. Chau, K. Walker, R. Long, and R. Croteau, "Regioselectivity of taxoid-o-acetyltransferases: heterologous expression and characterization of a new taxadien-5α-ol-o-acetyltransferase," *Archives of biochemistry and biophysics*, vol. 430, no. 2, pp. 237–246, 2004.

[39] A. Schoendorf, C. D. Rithner, R. M. Williams, and R. B. Croteau, "Molecular cloning of a cytochrome p450 taxane 10β-hydroxylase cdna from taxus and functional expression in yeast," *Proceedings of the National Academy of Sciences*, vol. 98, no. 4, pp. 1501–1506, 2001.

[40] S. Jennewein, C. D. Rithner, R. M. Williams, and R. B. Croteau, "Taxol biosynthesis: taxane 13α-hydroxylase is a cytochrome p450-dependent monooxygenase," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13595–13600, 2001.

[41] R. E. Ketchum, T. Horiguchi, D. Qiu, R. M. Williams, and R. B. Croteau, "Administering cultured taxus cells with early precursors reveals bifurcations in the taxoid biosynthetic pathway," *Phytochemistry*, vol. 68, no. 3, pp. 335–341, 2007.

[42] K. Vongpaseuth and S. C. Roberts, "Advancements in the understanding of paclitaxel metabolism in tissue culture," *Current pharmaceutical biotechnology*, vol. 8, no. 4, pp. 219–236, 2007.

[43] K. Walker and R. Croteau, "Taxol biosynthetic genes," *Phytochemistry*, vol. 58, no. 1, pp. 1–7, 2001.

[44] S. Jennewein, C. D. Rithner, R. M. Williams, and R. Croteau, "Taxoid metabolism: taxoid 14β-hydroxylase is a cytochrome p450-dependent monooxy-

genase," *Archives of biochemistry and biophysics*, vol. 413, no. 2, pp. 262–270, 2003.

[45] E. Baloglu and D. G. Kingston, "The taxane diterpenoids," *Journal of natural products*, vol. 62, no. 10, pp. 1448–1472, 1999.

[46] C. B. Brachmann, A. Davies, G. J. Cost, E. Caputo, J. Li, P. Hieter, and J. Boeke, "Designer deletion strains derived from saccharomyces cerevisiae s288c: a useful set of strains and plasmids for pcr-mediated gene disruption and other applications," *YEAST-CHICHESTER-*, vol. 14, pp. 115–132, 1998.

[47] D. Bai Flagfeldt, V. Siewers, L. Huang, and J. Nielsen, "Characterization of chromosomal integration sites for heterologous gene expression in saccharomyces cerevisiae," *Yeast*, vol. 26, no. 10, pp. 545–551, 2009.

[48] R. Zou, K. Zhou, G. Stephanopoulos, and H. P. Too, "Combinatorial engineering of 1-deoxy-d-xylulose 5-phosphate pathway using cross-lapping in vitro assembly (cliva) method," *PloS one*, vol. 8, no. 11, p. e79557, 2013.

[49] M. Chau and R. Croteau, "Molecular cloning and characterization of a cytochrome p450 taxoid 2α-hydroxylase involved in taxol biosynthesis," *Archives of biochemistry and biophysics*, vol. 427, no. 1, pp. 48–57, 2004.

[50] R. B. Croteau, A. Schoendorf, and S. Jennewein, "Cytochrome p450 oxygenases and their uses," Sept. 7 2004. US Patent 6,787,343.

[51] G. Sun, Y. Yang, F. Xie, J.-F. Wen, J. Wu, I. W. Wilson, Q. Tang, H. Liu, and D. Qiu, "Deep sequencing reveals transcriptome re-programming of taxus$\times$ media cells to the elicitation with methyl jasmonate," *PloS one*, vol. 8, no. 4, p. e62865, 2013.

[52] E.-K. Lee, Y.-W. Jin, J. H. Park, Y. M. Yoo, S. M. Hong, R. Amir, Z. Yan, E. Kwon, A. Elfick, S. Tomlinson, *et al.*, "Cultured cambial meristematic cells as a source of plant natural products," *Nature biotechnology*, vol. 28, no. 11, pp. 1213–1217, 2010.

[53] G. Ge, P. Xiao, Y. Zhang, L. Yang, *et al.*, "The first insight into the tissue specific taxus transcriptome via illumina second generation sequencing," *PLoS one*, vol. 6, no. 6, p. e21220, 2011.

[54] S. K. Lenka, N. Boutaoui, B. Paulose, K. Vongpaseuth, J. Normanly, S. C. Roberts, and E. L. Walker, "Identification and expression analysis of methyl jasmonate responsive ests in paclitaxel producing taxus cuspidata suspension culture cells," *BMC genomics*, vol. 13, no. 1, p. 1, 2012.

[55] J. B. Schenkman and I. Jansson, "Spectral analyses of cytochromes p450," *Cytochrome P450 protocols*, pp. 25–34, 1998.

[56] M. Girhard, K. Machida, M. Itoh, R. D. Schmid, A. Arisawa, and V. B. Urlacher, "Regioselective biooxidation of (+)-valencene by recombinant e. coli expressing cyp109b1 from bacillus subtilis in a two-liquid-phase system," *Microbial cell factories*, vol. 8, no. 1, p. 1, 2009.

[57] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos, "Distributing a metabolic pathway among a microbial consortium enhances production of natural products," *Nature biotechnology*, vol. 33, no. 4, pp. 377–383, 2015.

[58] N. C. Wilde, M. Isomura, A. Mendoza, and P. S. Baran, "Two-phase synthesis of (-)-taxuyunnanine d," *Journal of the American Chemical Society*, vol. 136, no. 13, pp. 4909–4912, 2014.

[59] E. Scotto-Lavino, G. Du, and M. A. Frohman, "5âĂš end cdna amplification using classic race," *Nature protocols*, vol. 1, no. 6, pp. 2555–2562, 2006.

Chapter 3

Understanding Early Paclitaxel Biosynthesis

3.1 Introduction

Portions of this chapter are adapted from [1].

As discussed in the previous chapter, taxadiene-5α-hydroxylase currently stands as the primary known bottleneck to high-level Taxol production. However, prior to investigating this pathway bottleneck in depth, it is informative to examine the evidence in support of this enzyme as catalyzing the next step in Taxol biosynthesis to ensure this is not a misguided approach. Analysis of pathway products and intermediates reveals that no known oxygenated taxoids posses the 4(5) double bond present in taxadiene, whereas nearly all taxoids possess either an alcohol group, or a functionalization likely derived from an alcohol group, at the C5 position. This, combined with relative abundance of oxygenated taxanes, appears to indicate that migration of this alkene group and hydroxylation occur either as the first step, or as an early step in the Taxol biosynthetic pathway [2, 3, 4]. Further studies were performed to validate this hypothesis in which radiolabeled taxadiene was exogenously suppled to cultured *Taxus* cells. The only hydroxylated metabolite observed was in fact taxadien-5α-ol, confirming the hypothesis and demonstrating that the first metabolite on the Taxol biosynthetic pathway leads to the production of this metabolite [5].

Following this result, Jennewin et al. employed a homology-based cloning strategy to obtain a P450-enriched cDNA library. Resultant genes were then screened for activity on taxadiene and a single clone yielding biotransformation to the desired product was obtained when expressed in *Saccharomyces cerevisiae* strain WAT11 containing a heterologously expressed cytochrome P450 reductase (CPR) domain [6, 7]. Although the step appeared to be fully characterized, confirmation of the experiment by Rontein et al. failed to yield the same product, and in fact yielded a different product, cyclic ether 5(12)-oxa-3(11)-cyclotaxane (OCT, **7**) [8]. This result is especially concerning as the strain was identical across experiments, with the exception being that the assay was performed from yeast microsomes rather than whole cells when OCT was produced. Furthermore, additional host strains, assay conditions, and research groups obtained still more variable selectivity results. These results are summarized in Figure 3-1. Due to these conflicting reports on the selectivity of taxadiene-5α-hydroxylase, along with evidence supporting this step as a key bottleneck in Taxol biosynthesis, a greater understanding of its mechanism and mode of operation is needed in order to allow for engineering and improvement of early Taxol biosynthesis.

3.2 Cytochrome P450 Enzymes

Cytochromes P450s are a large enzyme class of heme-containing metalloenzymes, similar to the chloroperoxidase (CPO) and nitric oxide synthase (NOS) enzyme families. These enzymes were first discovered with the demonstration of an enzymatic molecular oxygen transfer by an iron-containing enzyme in 1955 by Hayaishi et al. These enzymes were the named for their definitive carbon monoxide difference spectra, yielding a distinctive peak at 450nm [10]. In the ensuing decades, considerable effort has been devoted to understanding this enzyme class due both to the wide variety of catalysis they perform in addition to its ability to perform highly selective catalysis, even in the presence of additional functional groups; a task which would be difficult to perform using traditional chemical synthesis. In fact, porphorin rings

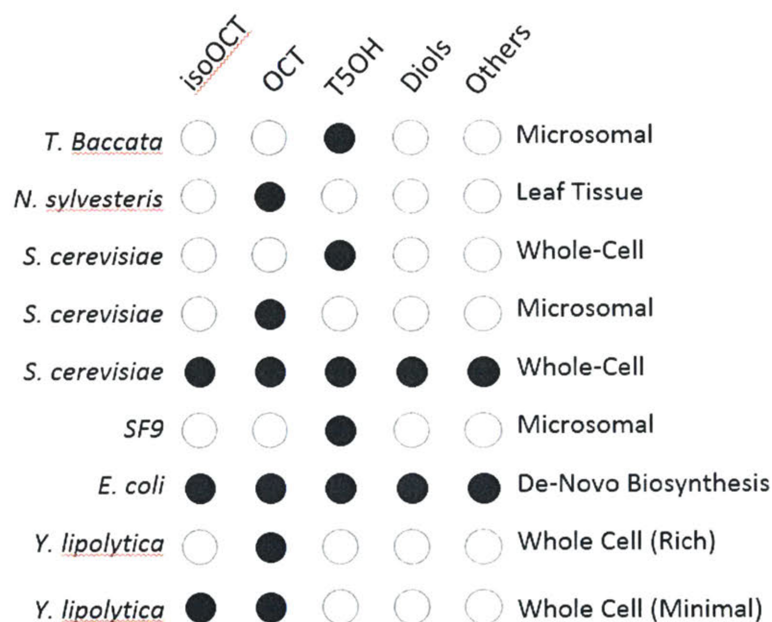

Figure 3-1: **Reported Selectivity of Taxadiene-5α-Hydroxylase.** Summary of literature and experimental results on Taxadiene-5α-Hydroxylase selectivity in various systems. Dark circles indicate products observed when exogenous taxadiene is supplied. Of note is that an identical *S. cerevisiae* strain (WAT11), when assayed using microsomes [8] or whole cells [9] produced differing selectivity profiles, indicating assay conditions may influence selectivity.

such as Fe(TPP)Cl, which resemble the heme moiety of P450s, have recently been adopted for oxidation reactions due to their catalytic abilities [11, 12].

Structurally, P450 active sites contain a heme-iron prosthetic group, covalently linked to the protein through conserved residues centered around a cysteine thiolate ligand. These enzymes are incredibly versatile, performing reactions including hydrocarbon hydroxylation, heteroatom oxidation and dealkylation, olefin and acetylene oxidation, oxidation of aromatic rings, dehydrogenation, and carbon-carbon bond cleavage between oxygenated carbons, alpha to oxygenated carbons, and alpha to carbons bearing a nitrogen atoms [13]. Distinctively, these incorporate molecular oxygen itself, rather than oxygen from water, into their substrates. The consensus catalytic cycle of P450's is outlined below and depicted in the companion figure, Figure 3-2 [14].

1. The P450 exists in a resting state in which a water molecule is coordinated to the heme moiety. This is proposed to prevent electron transfer to the heme by maintaining a low-spin state of the iron.

2. Substrate diffuses in and binds within the active site, inducing a conformational change and displacing a water molecule. This displacement induces a spin change in the heme iron from low spin to high spin [15].

3. The CPR reductase partner donates a single electron, derived from NADPH.

4. Molecular oxygen then binds to the ferrous heme to produce a dioxygen complex ($Fe-O_2$).

5. A second electron transfer occurs from the cytochrome P450 reductase (CPR), followed rapidly by protonation which yields the Fe(III)-hydroperoxy complex, referred to as **Cpd0**.

6. Protonation and heterolytic cleavage of the O-O bond, releasing one water molecule and forming the primary reactive iron-oxo species, compound 1 (**Cpd1**). This intermediate, although highly reactive, was isolated in 2010 [16].

7. Reaction occurs, such as hydroxylation (as depicted in Figure 3-2) returning the heme center to its original state, following diffusion of substrate out. The return of a water molecule then completes the cycle.

It is also of value to note that P450's are a large and diverse enzyme class. These enzymes, and their reductase partners, come in many varieties and can exist as a single enzyme, or paired enzymes. Microsomal P450s, such as those taxadiene-5α-hydroxylase belongs to, obtain electrons from NADPH which is transferred by a cytochrome p450 reductase (CPR). Other classes exist such as mitochondrial p450s which employ adrenodoxin and adrenodoxin reductase for transfer of electrons from NADPH; bacterial which employ ferredoxin and ferrodoxin reductase; cytochrome b5 reduction based systems; FMN-based systems in which a FMN-domain containing reductases exists as a fusion protein with the CYP, as well as P450 only systems in which no CPR equivalent is required. In *Taxus*, the CYP and CPR are independent enzymes. This necessitates the expression of both enzymes in potential microbial hosts for optimal activity, and leads to lower reaction rates as compared to bacterial fusion enzymes due to the need for enzyme association. Additionally, this leaves open the possibility for optimizing the balance between the expression of these two enzymes.

Engineering work on cytochrome P450s has primarily focused on the *Bacillus megaterium*-derived P450, BM3 [17, 18, 19]. This is largely because it and its reductase partner exist as a single, soluble, protein and it exhibits some of the highest turnover of any P450. Other common P450s studied include P450cam, P450-119, and P450-eryF. However, little in the way of truly systematic P450 engineering exists outside of these enzymes and we mechanistic considerations into taxadiene-5α-hydroxylase can only rely on consensus P450 mechanistic information.

3.2.1 Previous Studies on Taxadiene-5α-Hydroxylase

The biotransformation of taxadiene to taxadien-5α-ol is unique as, in addition to the hydroxylation typical for P450s, it also involves an allylic rearrangement. Previous

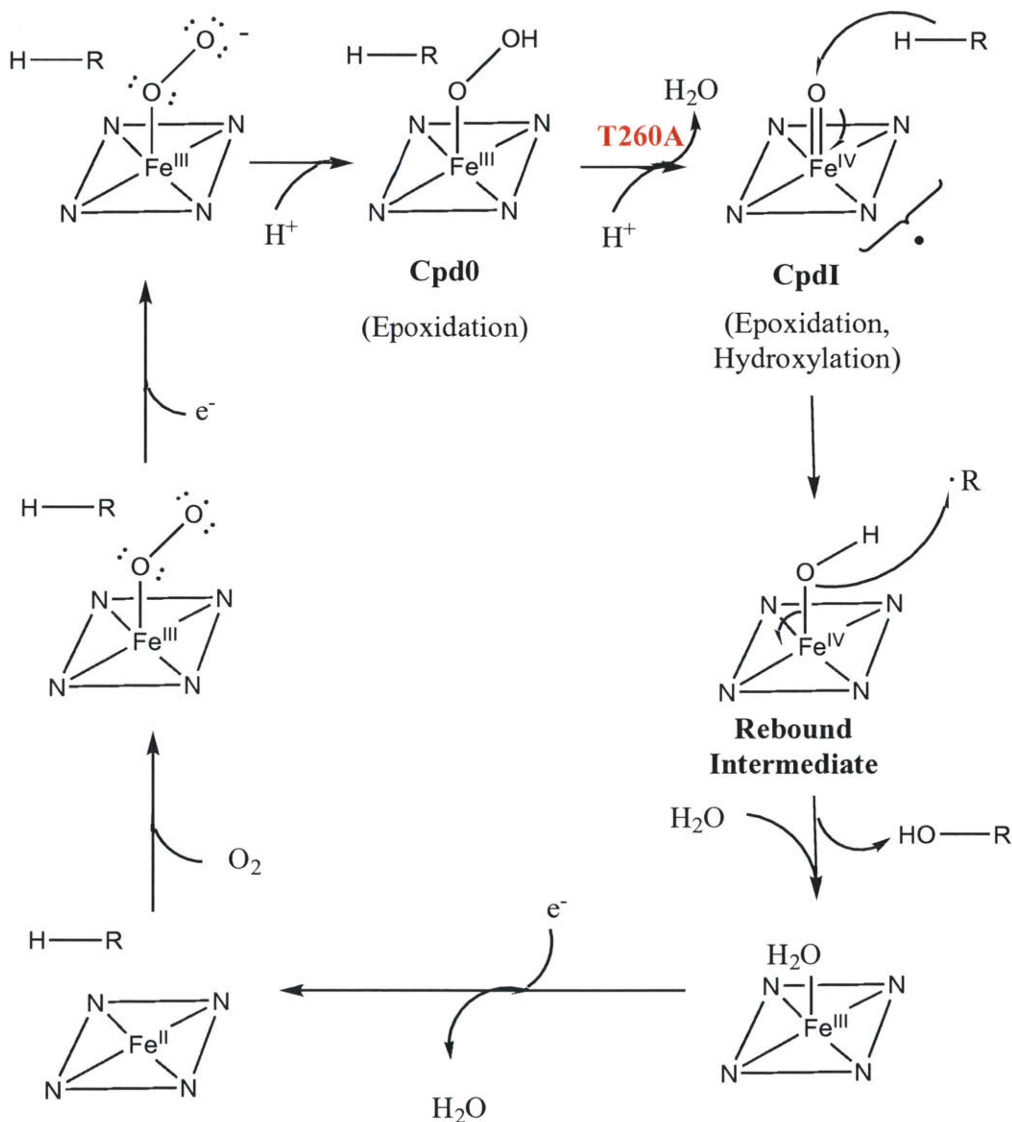

Figure 3-2: **Consensus Radical-rebound based mechanism for Cytochrome P450s** Activation of substrate is shown along the right-hand side beginning with hydrogen radical abstraction, forming a radical intermediate, followed by rebound. The heme-cofactor is then regenerated for the next catalytic cycle. Importantly, Cpd0 and CpdI are shown along the top. As indicated, a highly-conserved threonine is proposed to be responsible for hydrogen donation and the formation of Cpd0 from CpdI. Disruption of this residue through mutagenesis (here notated as T260A) has been proposed to increase the relative proportion of Cpd0 to CpdI, yielding information about the active species performing a given hydroxylation.

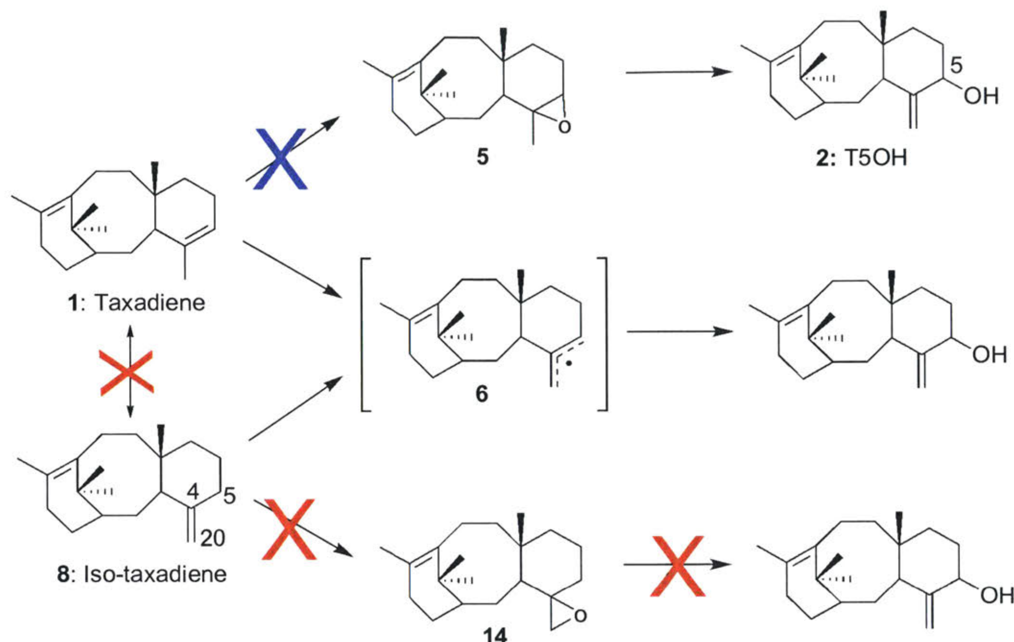

Figure 3-3: **Previous Support for a Radical-Mediated Reaction** All feasible biotransformation pathways from cyclized precursors to hydroxylated products are shown. Pathways which are ruled out via direct experimental evidence are identified via a red 'X', and pathways which are ruled out via induction are identified via a blue 'X' [6]

research has aimed to understand the mechanism of this reaction [6]. Two primary possibilities were evaluated: hydroxylation and allylic rearrangement via a radical-mediated mechanism, and epoxidation. This early research concluded that the enzyme operated via a radical-mediated mechanism (which would proceed as depicted in Figure 3-4), using three primary lines of evidence:

1. An isomer of taxadiene (**8**, Figure 3-3) was capable of producing the desired product, a process which is not feasible via epoxidation.

2. No isomeration of taxadiene isomers was observed.

3. No taxadiene-epoxide (**5**) was detected via HPLC or GCMS.

Importantly, while this evidence does lend support to a radical-mediated mechanism, it is largely indirect evidence. Furthermore, shortcomings can be seen as only

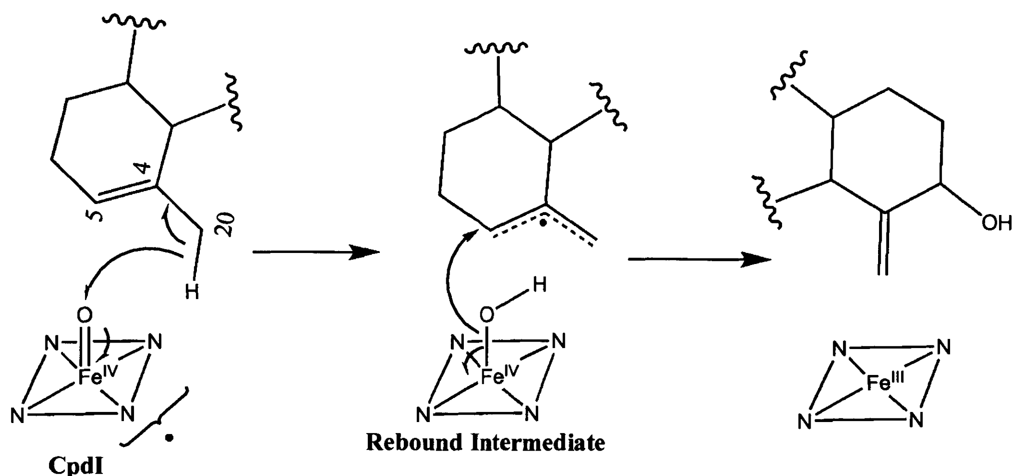

Figure 3-4: **Racical-Rebound Mediated Taxadiene Hydroxylation** As previously proposed [6], taxadiene may undergo a radical-mediated transformation. Under this regime, the H20 proton is abstracted, yielding an allylic radical which can then undergo rearrangement.

the isomer of taxadiene (**8**) has been demonstrated not to proceed via epoxidation. The implication that taxadiene itself is not epoxidized is based almost solely on its radical, rather than on affirming the radical state.

This enzymatic promiscuity also has support beyond direct mechanistic evidence as secondary metabolic pathways typically have a degree of in-built promiscuity. These pathways make a variety of products ranging from a few compounds to hundreds or more, rather than a single one as might be expected from enzymes such as those participating in central carbon metabolism. It has been previously proposed that enzyme promiscuity is an adaptive trait in plants, improving fitness by generating a varity of products, thus increasing the likelihood that any single product may confer an advantage, such as an attractant for pollinators, or as deterrent to feeding [20, 21, 22]. In this view, the promiscuity of taxadiene-5α-hydroxylase is to be expected, and explains what we interpret as 'poor' selectivity as being in-built functionality instead.

This concept motivated our early approaches [23, 24]. Substrate and product range is typically determined by substrate pose and orientation, which allow the iron-oxo reactant to access various portions of the substrate. Thus, we initially sought

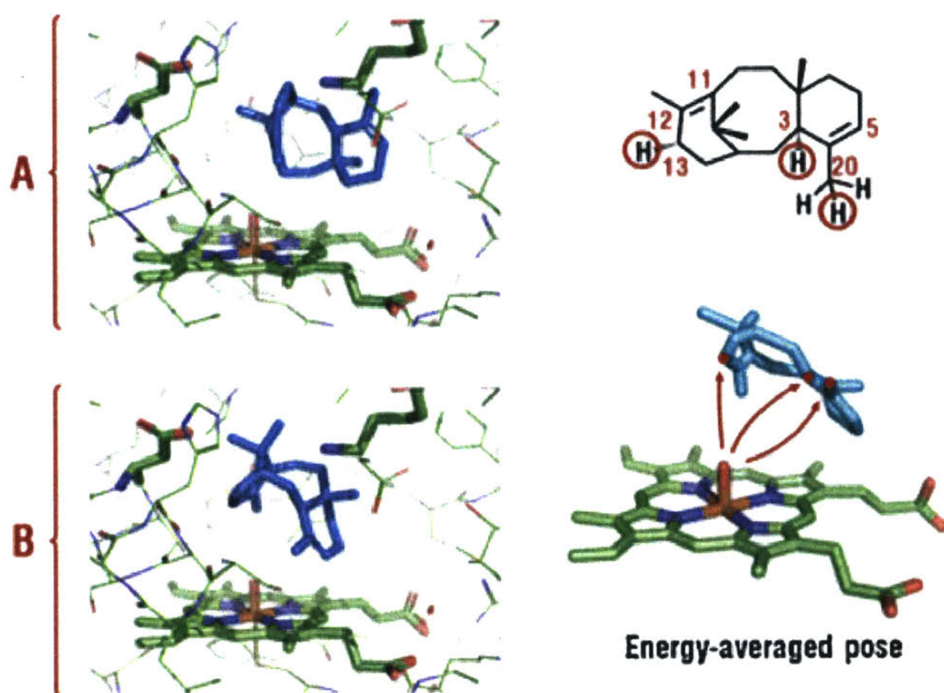

Figure 3-5: **Molecular Docking Poses of taxadien-5α-ol witin taxadiene-5α-hydroxylase** The energy-averaged structure of the two likeliest orientations of taxadiene in the active site (poses A and B) suggests that as many as 3 proton abstractions might be occurring. Figure reproduced from [24].

to alter this pose through site-directed mutagenesis. To achieve this goal, a computational approach was taken in which active site residues were mutated *in-silico*, localized energy minimization was performed, and substrate docking energetics were determined. Mutations which were predicted to alter substrate pose dramatically (typically between two primary predicted orientations, Figure 3-5) were passed onto the next round and screened experimentally. Of the 53 mutants screened, no single mutant displayed statistically significant differences in selectivity.

The failure of this approach or the mutants generated to yield changes in selectivity (positive or negative) is informative in itself as it is rare for such a large number of active site mutants to have no detectable change in product profile. This information, combined with the array of selectivity observed by differing research groups, indicated we had insufficient information about the enzyme itself to enable engineering. Thus,

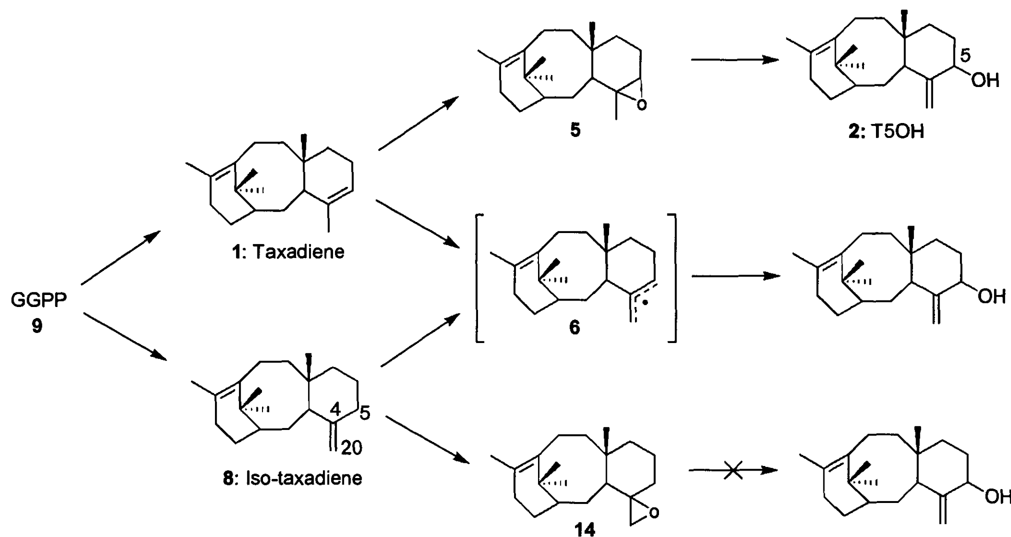

Figure 3-6: **Depiction of All Catalytic Routes to Final Products by Taxadiene-5α-Hydroxylase.**

we sought to gain a deeper understanding into the mechanism of the enzyme.

3.3 Reevaluating the Mechanism

We thus began by reevaluating the mechanism in its entirety, as shown in Figure 3-6. This figure illustrates several mechanistic routes are possible when taking both taxadiene isomers into account. The first of these possibilities is that the two isomers proceed via similar catalytic routes, beginning with a hydrogen radical abstraction and proceeding through a radical intermediate, as previously proposed [6]. Also feasible is that the two isomers undergo separate biotransformations and one (or both) proceeds via an epoxidation route. Furthermore, as it is not intuitive where by-products such as OCT derive from, they may branch from any given point, or even result from the action of native host metabolism on the final product taxadien-5α-ol.

3.3.1 Effect of Further Biotransformation

We thus set out to determine if byproducts may arise from functionalization or degradation of the final product, taxadien-5α-ol [24]. This situation would not be appar-

ent *in planta* due to the existence enzymes catalyzing the biotransformation of this compound to downstream products prior to off-pathway reactions. The proposed degradation could be catalyzed by host metabolism in the heterologous system, result from thermal degradation, or products may even exist in equilibrium *in-vivo*. We thus aimed to test this hypothesis by demonstrating that the observed products are not in equilibrium *in-vivo* and are incapable of isomerization to taxadien-5α-ol. This was performed by examining the effect further enzymatic functionalization has on observed selectivity. If by-products are derived from taxadien-5α-ol, expression of a downstream step would deplete the pool of this compound, and a selectivity shift in which all by-products are lowered would be observed. In contrast, if taxadien-5α-ol is a terminal product, only this product would decrease and a concomitant increase in the downstream pathway product would be observed, without changes in byproduct formation.

To perform this experiment, we selected the taxadien-5α-acetyltransferase, discussed in Chapter 2. This enzyme was cloned into an *E. coli* strain expressing the entire biosynthetic pathway to generate TaxE14. Fermentation of this strain led to near-complete conversion of taxadien-5α-ol to its acetylated form, without a similar decrease in the production of other products (Figure 3-7). Furthermore, no dramatic change was observed in the total taxadien-5α-ol and taxadien-5α-yl-acetate selectively following expression of the acetyltransferase (Figure 3-8). This effectively discounted both the proposal that these compounds may be in equilibrium, and that utilizing a rapid-acting downstream enzyme may correct selectivity.

3.3.2 Effect of Enzyme Fusions

In the native host, the cytochrome P450 and its reductase partner exist as two separate membrane-bound proteins; however, in our heterologous expression system, taxadiene-5α-hydroxylase was expressed as a fusion protein with its reductase. This protein contained a 5 amino acid flexible synthetic linker sequence due to earlier proposals that a fused enzyme would possess high catalytic rates and would reduce protein expression burden due to reduced membrane space of *E. coli* as compared

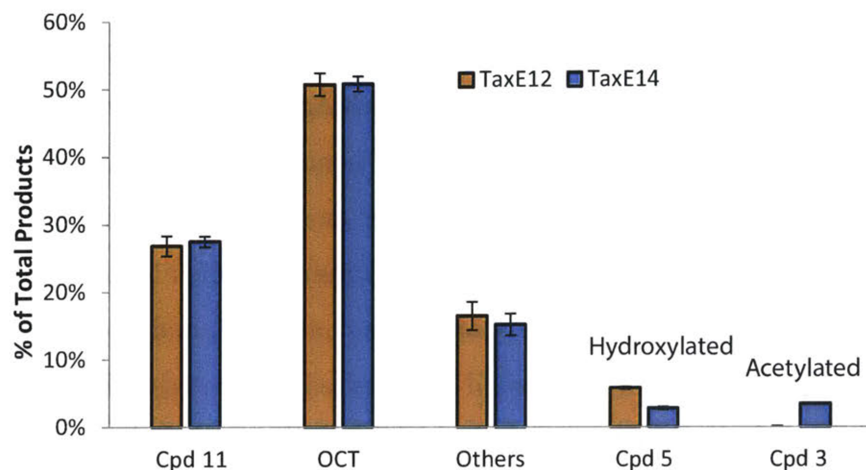

Figure 3-7: **Observed Selectivity of Strains With and Without Acetylase Expression.** Strain with acetylase expression (TaxE14) and without (Tax12) show no difference in level of by-products produced.

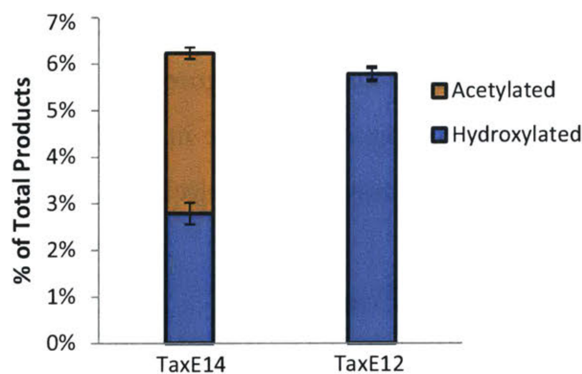

Figure 3-8: **Desired Product Production of Strains With and Without Acetylase Expression.** Total selectivity (percentage of total oxygenated products) for taxadien-5α-ol and taxadien-5α-yl-acetate obtained in fermentations. It can be seen that only minor changes are observed.

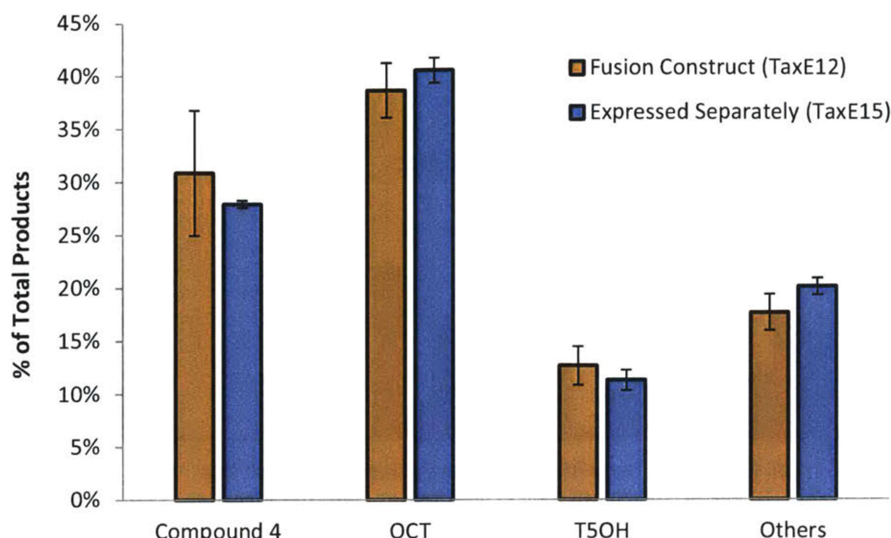

Figure 3-9: **Selectivity observed from taxadiene-5α-hydroxylase fused (TaxE12) to its reductase (CPR) and expressed as a separate enzyme (TaxE15).** The two forms of the enzyme show similar selectivity indicating that the choice of a fused or unfused enzyme system does not alter reaction selectivity.

to the native host [25]. To investigate if this enzyme fusion resulted in enzymatic promiscuity we expressed the enzymes independently and compared product profiles. As shown in in Figure 3-9, this did not result in significant selectivity changes, eliminating this as a possible source.

3.4 Effect of Enzyme Environment on Selectivity

To further understand factors influencing the selectivity of taxadiene-5α-hydroxylase, we began by comparing conflicting reports on the product profile obtained from expression of the enzyme (Figure 3-1). These reports suggested assay conditions play a key role in observed selectivity, and thus we set out to probe the effect systematically, with expression host selected as an initial point of investigation. Taxadiene-5α-hydroxylase and its native redox partner (CPR) were co-expressed in three common lab strains: *E. coli* (TaxE9), *S. cerevisiae* (TaxS1), and *Yarrowia lipolytica* (TaxY1), and cultured in their corresponding rich and minimal media. Taxadiene was supple-

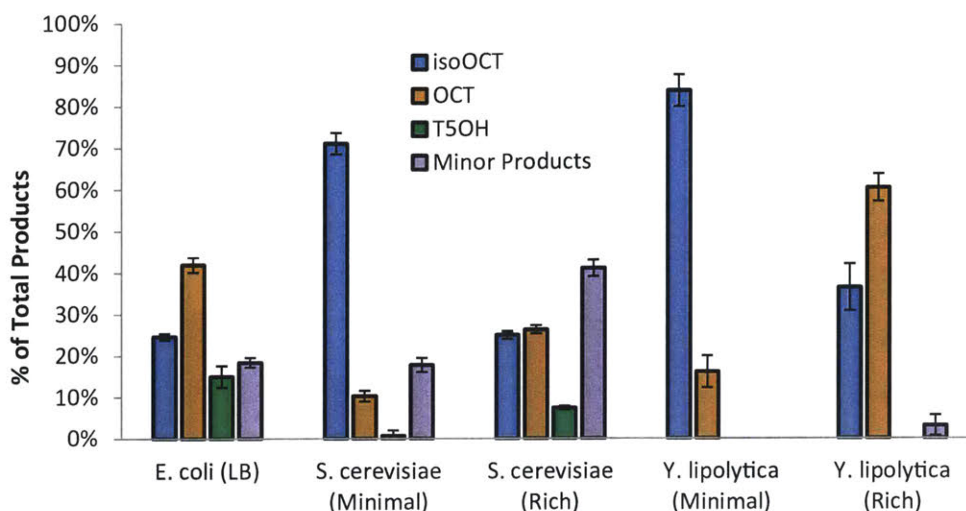

Figure 3-10: **Selectivity of Taxadiene-5α-Hydroxylase in Varying Hosts** Selectivity for oxygenated products obtained from feeding of taxadiene to whole-cells expressing taxadiene-5α-hydroxylase in different systems, normalized to total production. Strains depicted are: *E. coli* (TaxE9), *S. cerevisiae*(TaxS1), and *Y. lipolytica* TaxY1; rich media denotes YPD and minimal media denotes YNB. Three primary products are shown; minor products, due to their low abundance, are grouped. These minor compounds include only observed compounds with parent mass of 288 m/z, indicating mono-hydroxylation.

mented exogenously and product distribution was determined via GCMS. As depicted in Figure 3-10, multiple products were observed, including three major and four minor mono-hydroxylated compounds (parent ion 288 m/z), as well as low levels of compounds with mass spectra indicating dihydroxylation (parent ion 304 m/z, not depicted). As shown, different expression hosts and media conditions both had a dramatic impact on selectivity of taxadiene-5α-hydroxylase. There is little precedent for heterologous enzymes to exhibit differential selectivity of this nature; thus we hypothesized that changes in selectivity were not attributable to changes in the enzyme itself, but rather to the presence of a reactive intermediate such as an epoxide. We propose that the transformation of taxadiene to taxadien-5α-ol would require multiple steps, between which host metabolism or intracellular conditions may intervene to alter product distribution, giving rise to the observed differences in enzyme selectivity.

To further probe the existence of a multi-step pathway, we examined the effect

that extractive fermentation, or the addition of an organic solvent to the fermentation medium, would have on product distribution. In multi-step pathways or situations in which enzymes catalyze sequential reactions, addition of an organic solvent has been shown to affect product profile, presumably by altering the concentrations of reactants, products, and pathway intermediates [26, 27]. We thus cultured *E. coli* cells containing the entire biosynthetic pathway (TaxE12) in the presence of dodecane, diisononyl phthalate (DINP), or a solid C18 chromatography resin. Each extractive method displayed a unique product distribution (Figure 3-11). Although one may feasibly propose dodecane and DINP influence selectivity via direct interaction with the taxadiene-5α-hydroxylase, the ability of the solid C18 resin to alter selectivity argues against this, since the particles (40-60μm) are too large to enter the cell or directly interact with the enzyme itself. This result, and the one above, indicates the presence of intermediates which exist outside of the enzyme itself. This is again irreconcilable with radical rebound, in which the radical form is a short-lived intermediate present only within the active site of the enzyme [9].

We thus hypothesized that hydroxylation of taxadiene by taxadiene-5α-hydroxylase occurs via an epoxide intermediate (Figure 3-6, **5**), as is common in many P450s. In this scenario the pathway from taxadiene consists of two distinct steps: epoxidation catalyzed by taxadiene-5α-hydroxylase followed by degradation to the observed products. Under this scheme, the variation in selectivity between hosts can be explained by adventitious enzyme activity on epoxide **5** or intracellular host conditions, and variations observed in extractive fermentation can be attributed to properties of the solvent in which the epoxide is sequestered.

Enzyme selectivity observed from a purified-enzyme system further supports this result. Following purification of the fusion enzyme and enzymatic assays, the selectivity observed was shown to be nearly identical to that obtained from whole-cell *E. coli* assays. Furthermore, recent results from another group independently validate this result [28].

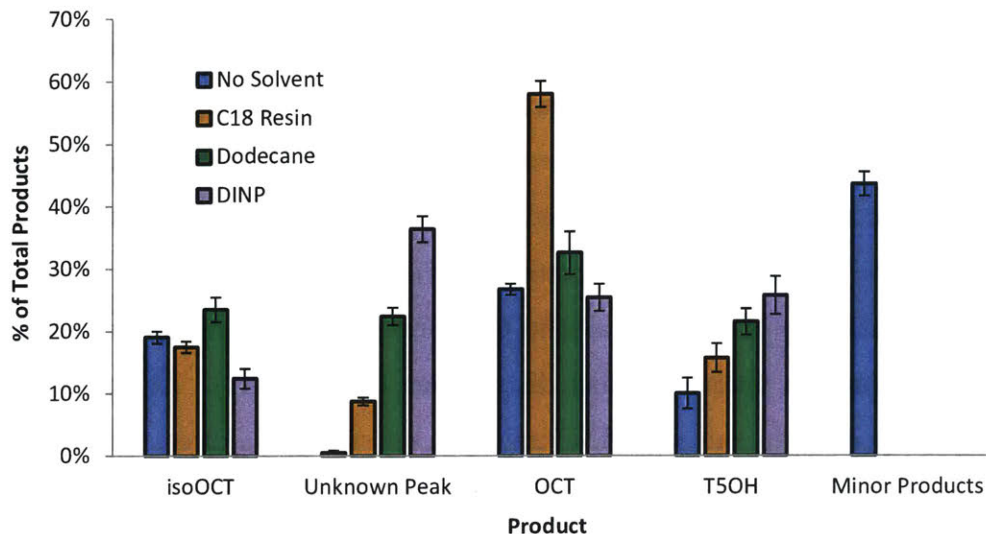

Figure 3-11: **Selectivity of Taxadiene-5α-Hydroxylase in Varying Solvents** Selectivity for oxygenated products obtained from co-culturing taxadiene-5α-hydroxylase expressing cells (TaxE1) in various solvents. Only the threee primary products are shown; minor products, due to their low abundance, are grouped. These minor compounds include only observed compounds with parent mass of 288 m/z, indicating mono-hydroxylation.

Table 3.1: Yields Obtained from Chemical Epoxidation of Taxadiene

	mCPBA	DMDO
IsoOCT (11)	9.4%	23.5%
Cpd 13	5.7%	13.6%
OCT (7)	13.2%	31.4%
taxadien-5α-ol (2)	6.2%	14.9%

3.5 Chemical Synthesis

We next set out to confirm that the product distribution observed in *E. coli* can result from degradation of a putative epoxide intermediate. Thus, we sought to chemically synthesize taxen-C4(C5)-monoepoxide **5** for feeding to *E. coli* lacking the taxadiene biosynthetic pathway and taxadiene-5α-hydroxylase (TaxE0). This strategy would remove any influence of taxadiene-5α-hydroxylase on the oxidized taxadiene pool and test whether the epoxide is capable of degradation to the observed products in-vivo.

Conformational analysis of taxadiene depicts a steric preference for epoxidation on the internal face of the C4-C5 olefin (Figure 3-12). We therefore anticipated that

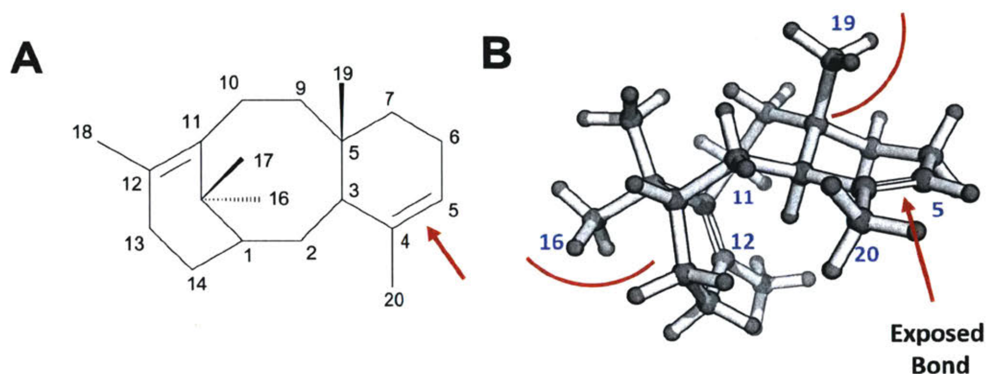

Figure 3-12: **Chemical Analysis of Taxadiene Epoxidation.** Taxadiene with predicted reaction site of chemical epoxidation indicated by an arrow. (B) Three-dimensional representation of taxadiene displaying steric hindrance arising from the C16 methyl group and concave face of the molecule preventing nucleophilic attacks at the C11-C12 double bond, and from the C19 methyl group preventing attack at 5β position.

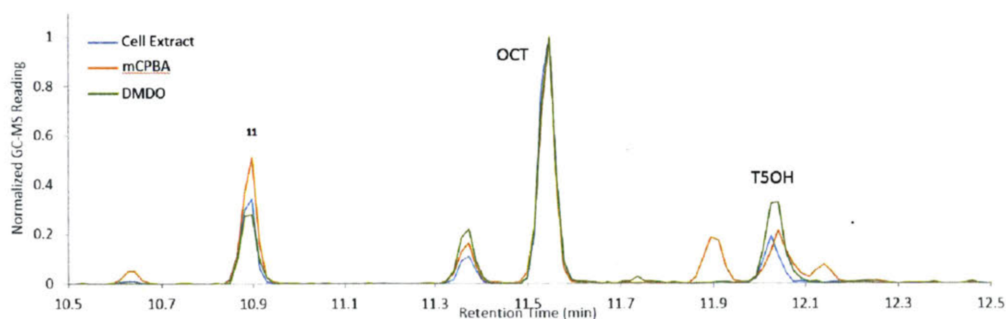

Figure 3-13: **GCMS Trace of Chemical Epoxide Synthesis** GCMS traces of reaction products from DMDO (Green), mCPBA (Orange), compared to extracts from *E. coli* (TaxE12) grown in the presence of dodecane (Blue); normalized to maximum observed TIC. *E. coli* in the presence of solvent was chosen for comparison due to its proposed ability to sequester unstable intermediates, allowing approximation of the true selectivity of the enzyme.

mono-epoxidation of taxadiene with a mild, non-selective epoxidation reagent such as dimethyldioxirane (DMDO) or meta-chloroperoxybenzoic acid (mCPBA) would yield only the mono-epoxide product 1 in a regio- and stereo-selective manner. However, following the reaction, GCMS analysis of the crude products revealed multiple peaks rather than a single epoxide. The retention times, mass spectra, and total ion count (TIC) ratios of these peaks matched those previously observed using *E. coli* TaxE12 in the presence of solvent (Figure 3-13). Reaction optimization, including the use of an acetone-free DMDO preparation performed at at -78°C, failed to alter selectivity or produce only the desired epoxide. We believe this most likely indicates epoxide degradation upon warming, impeding our ability to isolate and directly characterize the epoxide. Furthermore, a combination of poor stability and structural complexity of taxadiene and the observed reaction products impedes the absolute characterization of the epoxide **5** via direct methods such as in-situ low-temperature NMR during the reaction.

The observation that these biologically derived isomerization products are also products of chemical epoxidation of taxadiene implicates **5** as a common intermediate for both the chemical and biological reactions. This further suggests that the biosynthetic products result from degradation of the epoxide and not from multiple reaction pathways of a promiscuous enzyme. Furthermore, these experiments highlight the instability of the cup-shaped epoxide molecule, which one might surmise promotes nucleophilic attack of the C12-C13 olefin into the newly formed epoxide, thus preventing its characterization.

3.5.1 Characterization of Epoxide Products

We next set out to characterize the four major reaction products observed in both the biological and in the chemical reactions above. Two of the observed products match GCMS retention times and mass spectra to previously reported taxadiene derivatives, taxadien-5α-ol and OCT (retention time = 12.3 min and 11.8 min, respectively) [8, 29, 25]. The two remaining compounds could be isolated individually from the DMDO and mCPBA reactions, as well as biological extracts, via HPLC Compounds 11 and

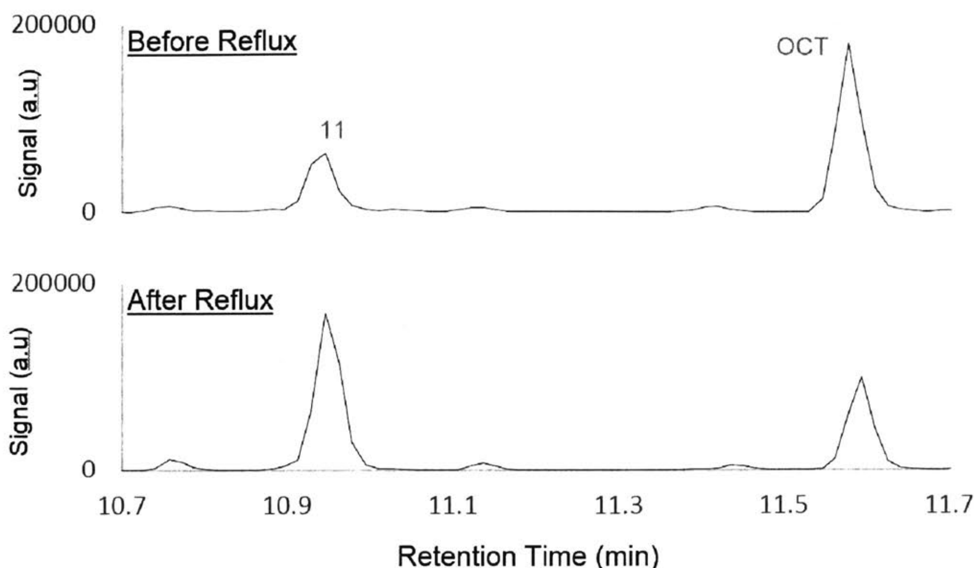

Figure 3-14: **Reflux of OCT.** Following refux in acetonitrile, OCT is shown to be transformed to compound **11**.

13; Figure 3-13, confirming that they are, indeed, stable reaction byproducts and not simply artifacts of thermal degradation during GCMS analysis.

We aimed to purify **11** from the crude epoxidation reaction material; however, fractionation via HPLC yielded two isolates containing **11**. One of these isolates contained **11** in pure form, while another fraction contained it as a mixture with OCT. As the quantity of **11** observed in the pure fraction was less than 10% of the total **11** observed in the mixture with OCT, we hypothesized that **11** could also be generated from degradation of OCT. To support this, we warmed the HPLC-isolated (semi-pure) OCT to reflux in acetonitrile (t = 2 hours). Following reflux, the concentration of **11** was observed to have increased while the quantity of OCT had decreased, indicating that **11** could be generated, at least in part, through OCT decomposition (Figure 3-22). Next we isolated compound **11** and solved its structure with NMR spectroscopy (Table 3.2, Figure 3-15). Similar to OCT, compound **11** exhibits a polycyclic carbon framework and a cup-forming tetrahydrofuran ring (Figure 3-15). This structure allows one to propose a feasible mechanism by which compound **11** may be derived from OCT or directly from taxadiene, as illustrated in Figure 3-16.

The formation of the OCT and **11** in the biological reaction is difficult to reconcile with the radical rebound mechanism. Previous reports predicted that a covalently bound heme-taxane carbocation intermediate rearranges to form OCT yet this mechanism is without precedent [8]. In contrast, epoxidations catalyzed by P450s are common; while epoxidation followed by concomitant ether ring expansions and carbon-carbon bond formation, such as those proposed here for OCT and **11**, are well represented in terpenoid biosynthesis [30]. Additionally, comparison of **11** to other known taxadiene oxidation products allows the proposal of a common mechanism by which all known products may be obtained from epoxide **5** (Figure 3-16).

Interestingly, the other unknown compound **13** was observed only under extractive fermentation regimes and from chemical transformations. Although instability precluded its isolation and characterization, we aimed to investigate the possibility that **13** either acts as a precursor to other observed products, or is epoxide **5**, by studying its decomposition. Following HPLC purification and incubation of **13** in the presence of acid, two previously undetected compounds, rather than any of the previously observed products, were detected leading us to believe that **13** was not epoxide **5** (Figure 3-17). To rule out differences arising from biological systems, we then assayed its transformation by crude lysates of *E. coli.* Crude lysates were able to rapidly convert **11** to a compound with a parent ion of 290 m/z (Figure 3-18, dissimilar to all other observed products, but again sufficient quantity could not be obtained for characterization. Feeding **11** to whole cells of various species showed that the appearance of this compound varied between host organisms (Figure 3-19). Taken together, these results indicate that extractive fermentation leads to the accumulation of compound **11** due to its effective removal from the cell, thus preventing its degradation by host metabolism. Furthermore, these results illustrate that this compound is not likely a precursor to other observed products.

Table 3.2: Chemical Shifts for Compounds **11**

Position	d1H ppm	COSY	HSQC	gHMBC	d13C ppm
1	1.56	2a, 14b	46.8	2, 3, 15	46.8
2a	1.35 dd (3.2, 14.6)	1, 2b	29.5	1, 3, 4	29.5
2b	1.72	1, 2a, 4	29.5	1, 3, 14	29.5
3	-	-	53.2	-	53.2
4	2.34	20	43.1	2, 3, 5, 7, 8, 20	43.1
5	4.03, t (5.5)	2b, 4	77.1	3, 6, 12	77.1
6a	1.71	6b, 7a, 7b	28.7	4, 7, 8	28.7
6b	1.83	6a, 7a	28.7	4, 5, 7, 9	28.7
7a	1.27	6a, 6b, 7b	34.0	5, 6, 8, 9	34.0
7b	1.55	6a, 7a	34.0	6, 8, 9, 19	34.0
8	-	-	47.7	-	47.7
9a	1.29	9b, 10a	44.7	3, 7, 11	44.7
9b	1.78	9a, 10b	44.7	19	44.7
10a	2.03, ddd (8.5, 12.2, 14.7)	9a, 10b	38.4	9, 11	38.4
10b	2.13, dd (10.2, 14.7)	9b, 10a	38.4	3, 8, 9, 11, 12	38.4
11	-	-	97.6	-	97.6
12	-	-	51.0	-	51.0
13a	1.43, ddd (4.7, 12.3, 14.6)	13b, 14a, 14b	34.1	1, 11, 14, 18	34.1
13b	2.30	13a	34.1	11, 12, 14, 15	34.1
14a	1.26	13a, 14b	27.2	2	27.2
14b	1.76	13a, 14a	27.2	2	27.2
15	-	-	43.9	-	43.9
16	0.97, s	17	20.5	1, 12, 15, 17	20.5
17	0.89, s	16	27.2	1, 12, 15, 16	27.2
18	1.17, s		17.1	11, 12, 13, 15	17.1
19	0.95, s		27.8	3, 7, 8, 9	27.8
20	0.87, d (7.4)	4	9.8	3, 4, 5	9.8

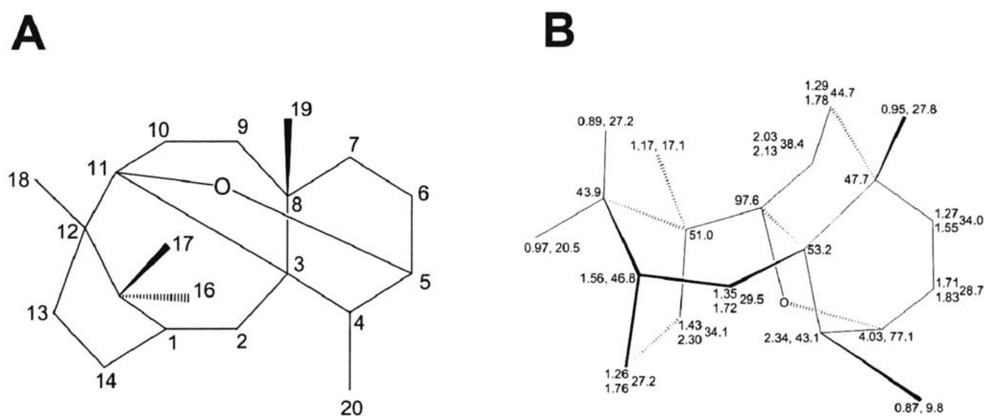

Figure 3-15: **Structure of Compound 11** A) Stick representation of the newly elucidated isoOCT structure. As can be seen, the center ring has expanded from an 8-membered to a 9-membered ring and a C-C bridge between C3 and C11 is present, as in OCT cite17. B) Three-dimensional representation with observed carbon and hydrogen chemical shifts are shown.

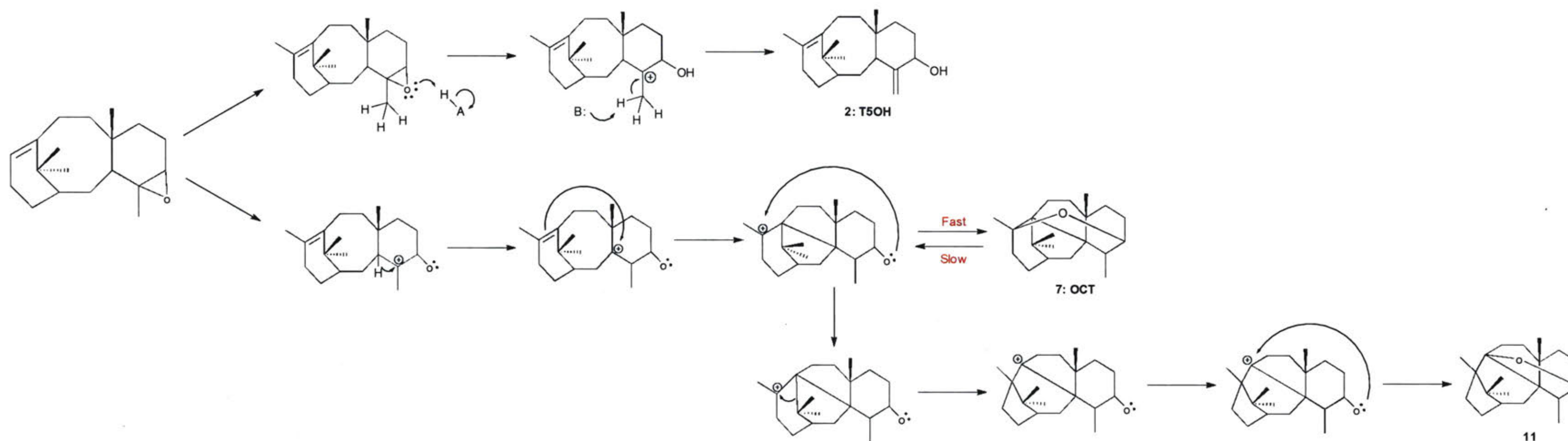

Figure 3-16: **Proposed Reaction Cascade to Compound 11** Proposed cascade to known products from epoxide intermediate **5**, showing feasibility of formation of OCT, taxadien-5α-ol, and **11**. We depict a route from which **11** can be generated both directly from the epoxide intermediate, as well as from OCT, in line with existing data.

3.6 Alternative Taxadiene Isomer Utilization Supports Epoxidation Route

An alternative taxadiene isomer, taxa-4(20),11(12)-diene (isotaxadiene, **8**), is capable of transformation into taxadien-5α-ol by taxadiene-5α-hydroxylase, but cannot feasibly do so via the proposed epoxidation mechanism. To reconcile our assertion that taxadiene proceeds via epoxidation with the fact that isotaxadiene cannot reasonably yield taxadien-5α-ol via this route, we proposed that the two isomers undergo distinct oxidation processes catalyzed by taxadiene-5α-hydroxylase. In this regime, taxadiene undergoes epoxidation while isotaxadiene undergoes hydroxylation via the aforementioned radical rebound mechanism (Figure 3-6). Such substrate-controlled preference for hydroxylation versus epoxidation has been proposed in other systems [31, 32].

To test this hypothesis, we performed substrate feeding of both taxadiene and its isomer to TaxE9 (Figure 3-20). Isotaxadiene was detected as a cyclization by-product of GGPP via GCMS, with mass spectra, retention time, and product ratio to taxadiene matching those previously reported [33], purified via preparative-scale HPLC, and used in feeding experiments. As before, when feeding taxadiene, multiple products were observed. However; when feeding was performed with isotaxadiene, only the taxadien-5α-ol product was observed. As epoxidation of isotaxadiene cannot feasibly yield taxadien-5α-ol (Figure 3-6), it can be reasoned that this isomer of taxadiene undergoes radical rebound hydroxylation. The proposal that epoxidation of isotaxadiene does not yield taxadien-5α-ol was also confirmed via chemical epoxidation, as shown in Figure 3-21.

Significantly, through examination of the reaction scheme for radical rebound, it can be seen that if the two isomers share the same common radical intermediate **6**, they should theoretically yield identical product profiles regardless of expression system. However, as taxadiene oxidation does not result in the same product distribution as isotaxadiene, we can deduce that these substrates undergo different reaction mechanisms, with distinct reaction intermediates. These assertions - of differing mechanisms and radical rebound for iso-taxadiene - when taken together, support the

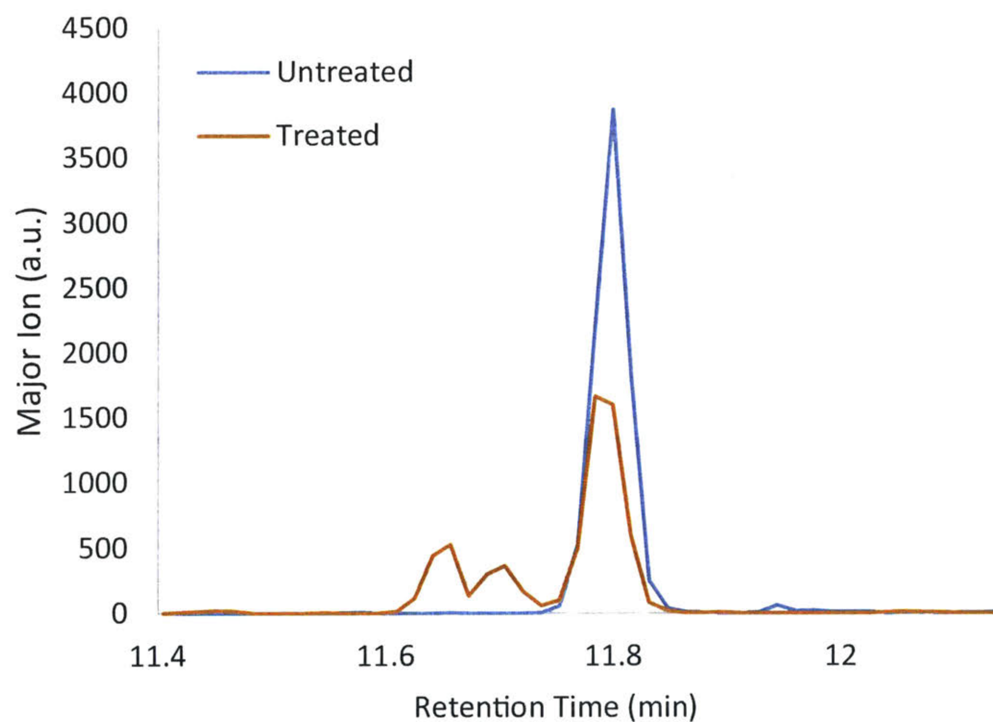

Figure 3-17: **Acid Treatment of Compound 11.** GCMS chromatogram of products generated via acid hydrolysis of Cpd 11, demonstrating two new peaks, which do not co-elute with any previously observed compounds.

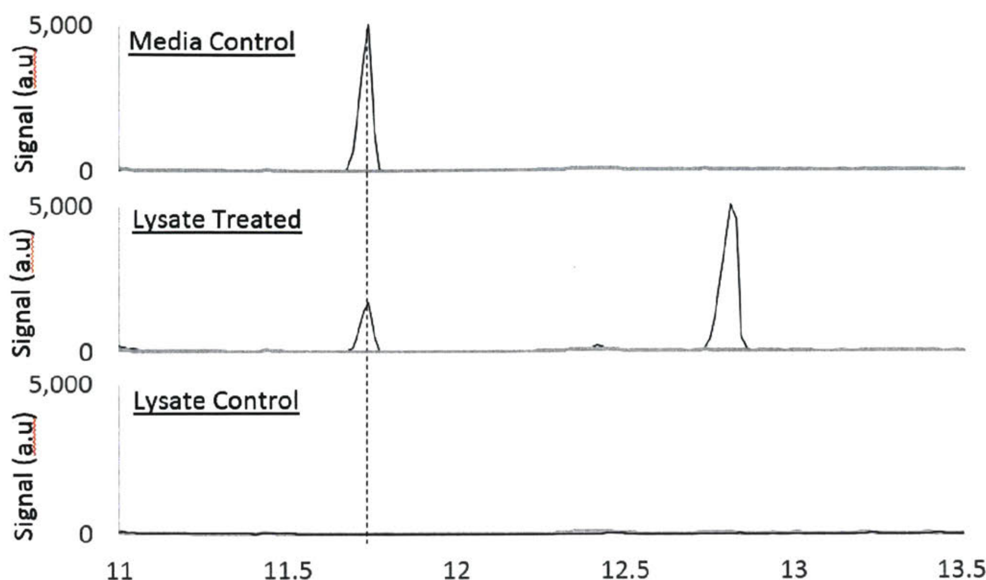

Figure 3-18: **Treatment of *E. coli* Crude Lysate with Compound 11** GCMS chromatogram of product generated via treatment of *E. coli* crude lysate with compound **11**, demonstrating two new peaks, which do not co-elute with any previously observed compounds.

hypothesis that taxadiene undergoes epoxidation by taxadiene-5α-hydroxylase.

3.7 Evaluation of Alternative Oxidants

We next sought to further understand the catalytic cycle of taxadiene-5α-hydroxylase and elucidate why the two isomers undergo different oxidation processes. P450s have been proposed to utilize multiple reactive species to perform oxidative reactions, with the two supported by the most evidence being the so-called Cpd0 (Fe^{+3}-OOH), and CpdI (vide supra) (Figure 3-2) [34]. Cpd0 has been reported to favor epoxidation and its use would support an epoxidation-based mechanism. Additionally, it has been reported that when P450s utilize the same oxidizing specie (CpdI) to catalyze both hydroxylation and epoxidation, the reaction path is decided primarily by the ionization energy of the substrate rather than enzyme properties [31, 32, 35]. Such substrate-driven selectivity would further bolster the epoxidation-based mechanism as it aids in explaining the different paths the alternative taxadiene isomers take en

6A

Wild Type	5αCYP-Expressing
E. coli (1)	E. Coli (2)
S. cerevisiae (3)	S. Cerevisiae (4)
Y. Lipolytica (5)	Y. Lipolytica (6)
Pichia pastoris (7)	
Ralstonia eutropha (8)	
Shewanella oneidensis (9)	

6B

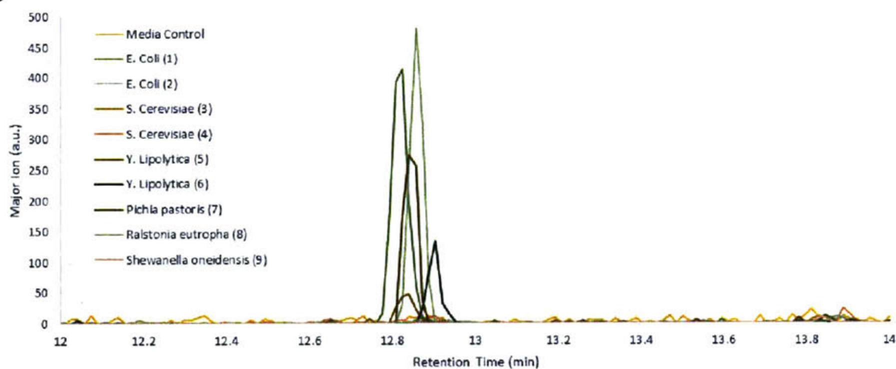

Figure 3-19: **Feeding of Compound 5 to Lab Strains.** (A) Lab strains to which **compound 13** was fed, both wild-type and taxadiene-5α-hydroxylase expressing cells are performed for *E. coli*, *S. cerevisiae*, and *Y. lipolytica*. (B) GCMS Chromatograph (parent ion 290 m/z shown), displaying production of a novel compound produced from transformation of compound **13** in select organisms.

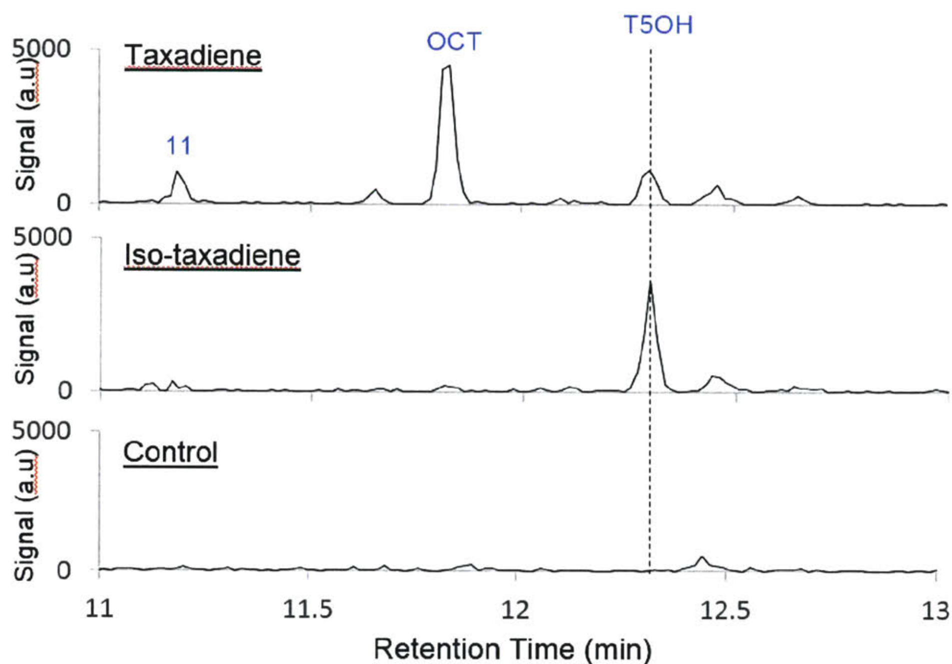

Figure 3-20: **Feeding of Taxadiene Isomers.** Formation of oxygenated products catalyzed by engineered *E. coli* expressing taxadiene-5α-hydroxylase when supplemented with taxadiene, isotaxadiene, or no substrate, as confirmed by GCMS analysis

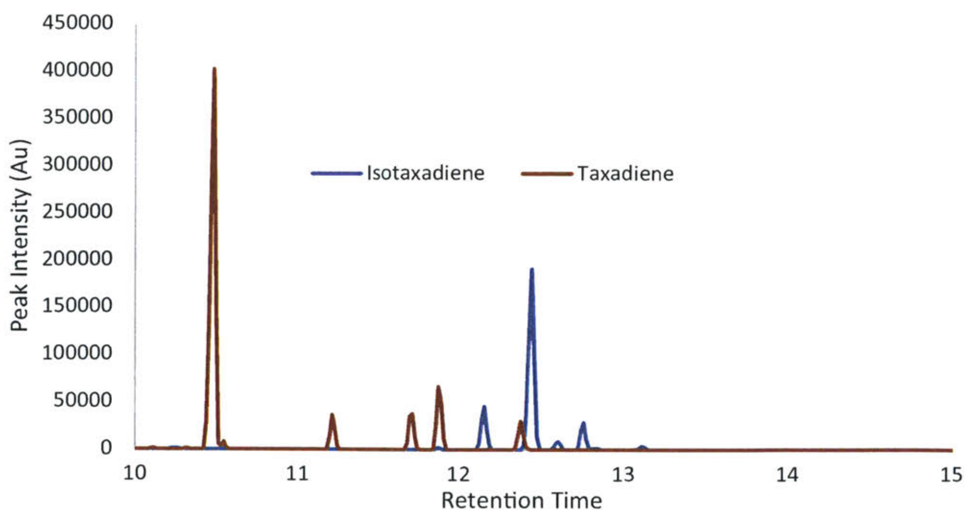

Figure 3-21: **Chemical Epoxidation of isotaxadiene.** GCMS chromatogram of the chemical transformation of both Taxadiene and isotaxadiene, showing that isotaxadiene yields a unique set of products following epoxidation and no taxadien-5α-ol is observed

route to final products.

A single highly conserved acid-alcohol pair has been demonstrated to be critical for the transition from Cpd0 to CpdI (Figure 3-2) [32]. If turnover rates for different substrates decrease proportionally upon mutagenesis of the conserved alcohol (typically a threonine, here Thr260) to a nonpolar amino acid, it is likely they utilize the same oxidizing species. We thus mutagenized taxadiene-5α-hydroxylase and compared the catalytic activity between a T260A mutant and its wild-type parent on the taxadiene and its isomer. We observed that turnover in crude-lysates was reduced to less than 5% of the wild type enzyme for both isomers, thus implicating a common oxidant and supporting the substrate control hypothesis. This result discounts the role of Cpd0 in oxidation by taxadiene-5α-hydroxylase and adds further weight to the proposal that enzyme preference for epoxidation versus hydroxylation is primarily substrate-driven. Following the publication of these results, additional studies were performed by Barton et al. which further bolster this claim of a substrate-driven epoxidation, and discussed in greater detail on Page 89 [36].

3.8 Plant Cells

We next aimed to determine if OCT and other observed compounds are products of heterologous expression or can also be produced in the native host, *Taxus cuspidata.* For screening, methyl jasmonate induced *T. cuspidata* cell cultures were supplemented with exogenous taxadiene in order to boost generation of taxadiene-5α-hydroxylase products. These cells were then screened for the appearance of the primary oxidation byproduct, OCT. A clear peak matching the retention time and mass spectra of OCT was present in all cultures, with the concentration in fed cells increasing with increased supplementation (Figure 3-23). This result demonstrates that products other than taxadien-5α-ol are in fact produced natively. The low quantity of OCT (0.5 mg/L) in the native host implies a more favorable product distribution for taxadien-5α-ol by taxadiene-5α-hydroxylase, or further metabolism of OCT either through degradation or further oxidation to final products. For example, OCT may act as a precursor

to compounds containing a C-3(11) bridge, of which there are 18 currently known [8, 37]. Additionally, as OCT is produced both natively and heterologously and the selectivity of taxadiene oxidation is substrate controlled (see above), it is difficult to reconcile that product profiles are influenced by reductase partner; however, the possible influence of various redox partners on substrate and product affinities and the concomitant changes to the observed product distribution cannot yet be predicted or discounted.

Furthermore, improved selectivity of the epoxide degradation step may be a function of intracellular conditions, or the presence of additional enzymes. Feasibly, two enzymatic routes are possible that may generate taxadien-5α-ol from epoxide **5**. The first consists of a two-step reaction catalyzed by an enzyme similar to lemonine epoxide hydrolase [38], which cleaves the epoxide to vicinal diols. A second enzyme would then perform a dehydration reaction to form taxadien-5α-ol; such activity has precedent in terpenoid biosynthesis [39]. The alternative route, consisting of a single enzymatic step, would cleave the epoxide directly to the desired product (Figure 3-24). As such an enzyme would enable the improved biosynthesis of Taxol precursors in heterologous systems, their discovery warrants further study.

3.9 Further Remarks and Comparison to Literature

Another key piece of supporting data arises from the original work performed on the enzymatic mechanism [6]. Crucially, the authors observed an inverse kinetic isotope effect (KIE) when deuterating the C20 position. If abstraction at this location is essential, as would be in radical rebound, a KIE should be observed. However, if this does not take place in the rate-limiting step, likely within the enzyme, then no KIE should be observed. Thus, this data, which was origionally discounted, stands to further support an epoxidation-based mechanism. Further support would be gained by obtaining KIE data for the taxadiene isomer. This isomer should theoretically possess a KIE as we propose it undergoes radical rebound, but obtaining this compound for testing is impractical within our laboratory.

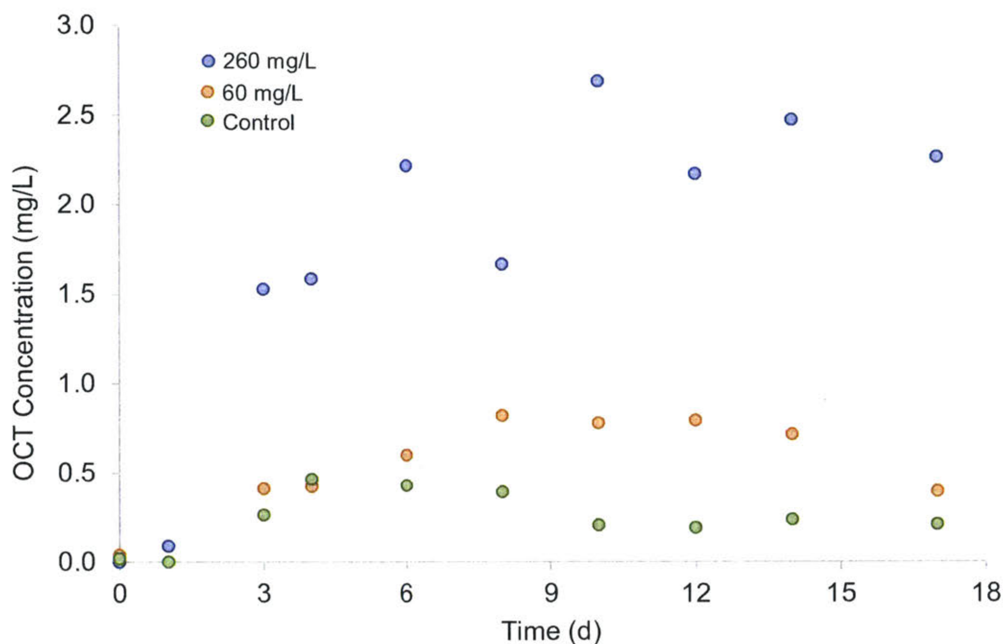

Figure 3-22: **Plant Cell Feeding Time Course.** Time course study of OCT formation by *T. cuspidata* cultures supplemented with exogenous taxadiene, displaying the formation of OCT occurs in native plant cells, and can be increased when exogenous taxadiene is fed.

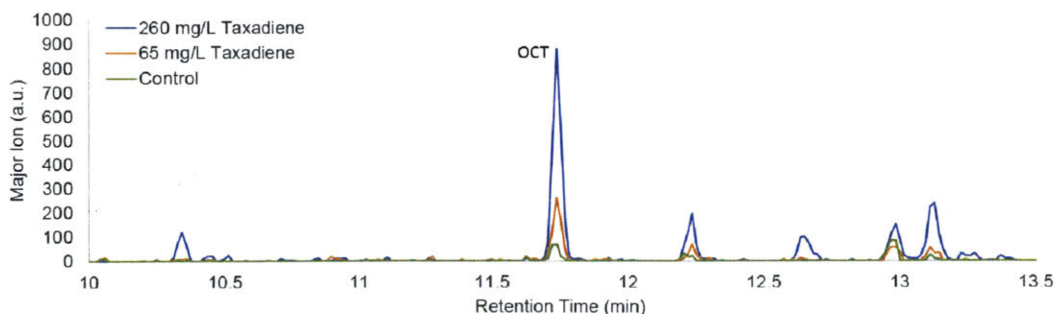

Figure 3-23: **OCT Formation in Plant Cells.** GCMS chromatogram of taxanes displaying OCT, shown with primary ion (191 m/z), obtained from plant cell cultures supplemented with exogenous taxadiene, displaying OCT formation

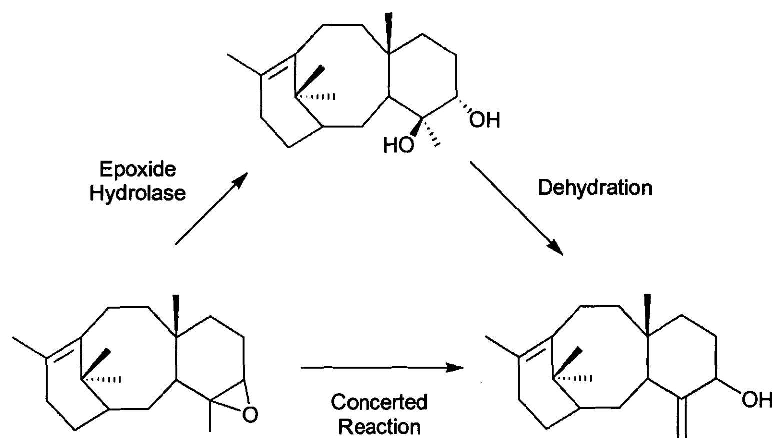

Figure 3-24: **Proposed routes from the putative epoxide intermediate 1 to the desired product, T5OH.** Two routes are feasible. Upper route would utilize enzymes which have present within terpenoid biosynthesis including an epoxide hydrolase and a dehydratase, both of which have characterized enzymes known for other terpenoids.

Furthermore, following publication of this data, Barton et al. supplied additional chemical evidence in support of an epoxide intermediate. The key findings are shown in Figure 3-25 [36]. Their work confirmed epoxide instability, particularly in the presence of acids, with degradation proceeding to OCT. This is not only reported, but is also demonstrated as reactions with $TMPAlEt_2$ were required to be performed using the epoxide in crude form. Furthermore, this reaction was observed to occur under direct observation via NMR. Treatment of the epoxide with a different acid (pTSA) yielded taxadien-5α-ol as the primary product, demonstrating conclusively the alcohol can be produced via acid-catalyzed rearrangement. They also demonstrated porphyrin-mediated rearrangement of the epoxide to OCT. This porphyrin species closely mimics the heme moiety of the P450 in question, indicating the P450 itself may catalyze this rearrangement. Additionally, specifically to address the instability of the taxadiene-epoxide, a cyclogeraniol derived epoxide does not rearrange under the same conditions fostering rearrangement of the taxadiene epoxide. This highlights that the instability of taxadiene is due to its overall structure, possibly due to the large amount of ring strain, or the nucleophilic 12(13) alkene group. Furthermore, it was demonstrated the porphyrin-ring mediated oxidation is capable of not only

epoxidizing, but also of catalyzing the rearrangement to OCT.

3.10 Conclusions

This chapter provides compelling evidence for the existence of an epoxide intermediate in the mechanism of taxadiene-5α-hydroxylase which poses a significant, and previously unknown, obstacle to high-level Taxol biosynthesis. Though epoxide instability precluded its direct isolation and characterization, we were able to utilize indirect evidence to establish its existence. These include: (1) by varying hosts, media conditions, and solvents, we demonstrated a change in the oxidized taxadiene pools; (2) identical products and similar product ratios were generated through chemical epoxidation and enzymatically by taxadiene-5α-hydroxylase; (3) product structures and analysis of possible mechanistic pathways showing a clear route from the proposed epoxide intermediate; and (4) different product profiles were obtained upon feeding of two taxadiene isomers to whole cell taxadiene-5α-hydroxylase catalysts, indicating substrate control on the reaction outcome that is not reconcilable with the previous assertions of taxadiene oxidation by taxadiene-5α-hydroxylase.

Other lines of evidence including support for CpdI as the primary oxidant, further supporting the substrate-driven selectivity preference, inability of further functionalization to shift product profiles, and the detection of multiple products in the native plant host provide further support for the proposed mechanism of the enzyme. Although instability of the putative epoxide precluding characterization via NMR seems at odds with the above findings of intermediate extraction from whole cells by C18 resin; consideration of the different time-scales for extraction, which derive from diffusional timescales, as compared to timescales required for purification and NMR characterization reconciles these contrasting observations.

Furthermore, the array of data in support of an epoxidation-based mechanism is more compelling than that previously used to support one based on radical-rebound. Three primary lines of evidence were previously used to support the radical-rebound mechanism (1) No free epoxide was detected; (2) taxadiene-5α-hydroxylase is capa-

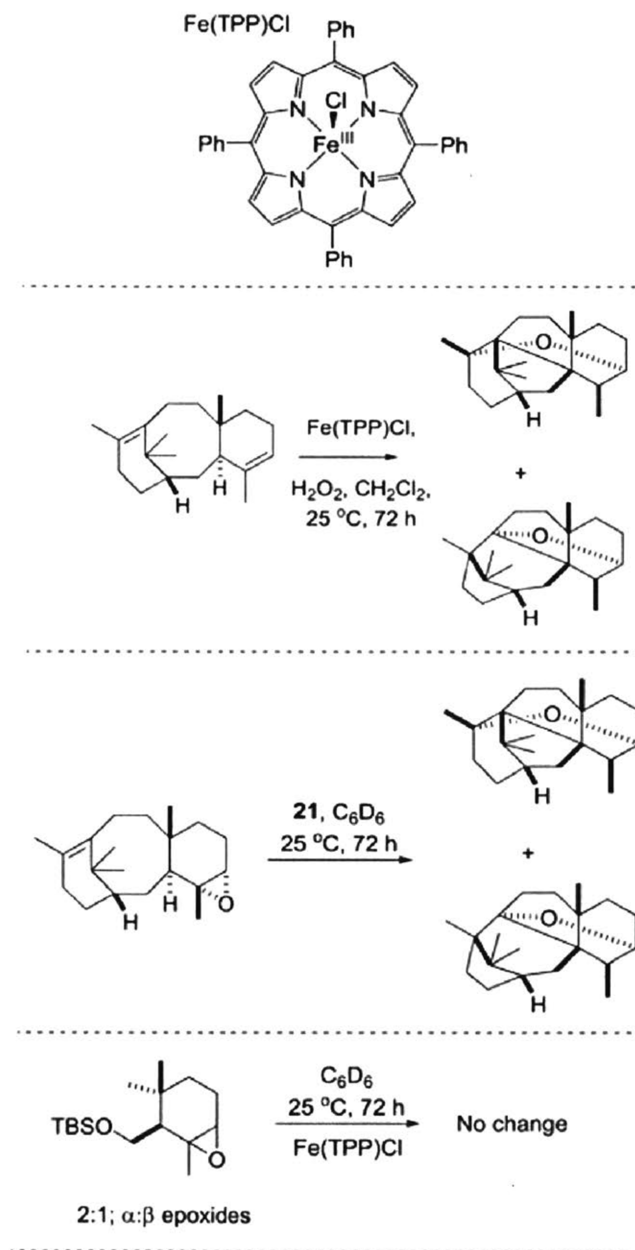

Figure 3-25: **Barton et al. Chemical Epoxidation Experiments.** Data From [36]

ble of utilizing alternative taxadiene isomers; and (3) no isomerization of taxadiene isomers was observed. We are able to address each of these points as we show the thermochemical instability of the epoxide precludes its easy detection as it rapidly decomposes, alternative isomers are not required to undergo the same oxidation process, and that no isomerization between taxadiene isomers is required. Furthermore, the previously proposed, radical-based, allylic rearrangement is without precedent in previous literature, whereas epoxidation of terpenoids by P450s is well documented [30]. Through proposal of the epoxide intermediate we were able to rectify previous, conflicting, reports on the selectivity of taxadiene-5α-hydroxylase.

In addition to offering improved understanding of taxadiene-5α-hydroxylase, these insights enable future work in improving the biosynthetic yield of Taxol and its analogues. This can be seen as by assuming a radical-based mechanism only a single route to improvement is available (P450-mutagenesis); however, with a substrate-driven epoxidation route, one is able to intervene to improve selectivity at multiple steps within the pathway. These locations include prior to the enzyme, by supplying different substrates, as demonstrated by our feeding of alternative isomers; and the addition of an enzyme capable of metabolizing the epoxide to the desired product.

3.11 Materials and Methods

Strain construction. *E. coli* strains TaxE0, TaxE1, TaxE9, TaxE12 and the *S. cerevisiae* strain TaxS1 were previously constructed by Kang Zhou (National University of Singapore) and Chin Giaw Lim (Manus Bio, Cambridge Massachusetts, USA) in our laboratory [40]. TaxE12 contains taxadiene-5α-hydroxylase and *T. cuspidata* CPR domain as a fusion construct, as in previous literature [25]. TaxE14 was generated by modification of the plasmid from TaxE12 (pSC101 origin of replication) to include an operon containing the acetyltransferase, and cloning into TaxE11. The non-fused P450 expression strain, Tax15, was generated by replacing the linker domain of Tax12 fusion construct with a ribosomal binding site and a membrane targeting domain as was performed to taxadiene-5α-hydroxylase; no difference in selectivity was observed

between TaxE14 and TaxE15. *Y. lipolytica* stain TaxY1 was constructed by integrating the expression cassette harboring taxadiene-5α-hydroxylase and CPR under the control of TEF intron promoter. Strains used in this study are listed in Table 3.3.

Small Scale Fermentations For cultures of *S. cerevisiae* or *Y. lipolytica*, a single colony of the desired strain was inoculated into 1mL of YNB (Yeast nitrogen base with casamino acids and ammonium sulfate, supplemented with 20g/L glucose) or YPD (10g/L yeast extract, 20 g/L peptone, 20g/L glucose) and cultured overnight (30°/250 r.p.m). 20µl of the overnight culture was used to inoculate 1 mL of the desired culture media (YPD or YNB) in a 14 mL Pyrex glass tube for culturing. For feeding experiments, taxadiene was fed exogenously upon inoculation (0.1 mg in 10µl DMSO). Cultures were shaken (22°C/250 r.p.m.) for 5 days prior to analysis of product distribution.

For *E. coli* cultures, a single colony of *E. coli* was inoculated into LB medium (10g/L tryptone, 5 g/L yeast extract, 10g/L NaCl, pH=7) and shaken overnight (37°,250 r.p.m.). 20µl of overnight culture were inoculated into 1 mL of small-scale culture media (5 g/L yeast extract, 10 g/L tryptone, 15g/L glycerol, 10g/L NaCl, 100 mM HEPES, pH 7.6), unless otherwise stated, in a 14 mL pyrex glass tube. Cultures were induced upon inoculation with 0.1 mM IPTG and shaken (22°/250) for 5 days prior to analysis of product distribution.

For feeding of taxadiene and iso-taxadiene, these compounds were added as a solution in DMSO (1% final culture volume, final concentration of 100mg/L) at the time of inoculation. In the extractive-fermentative regime, *E. coli* strain TaxE12 was grown in the presence of 1000µl of DINP, 1000µl of dodecane, or 50 mg of C18 chromatography resin. All the plasmids constructed in this study were validated by means of sequencing.

Preparation of Taxadiene, iso-Taxadiene, Compound 11, and Compound 13. A 13-Liter Bioflow bioreactor (New Brunswick) was used for semi-preparative scale production of desired compounds. Seed cultures, generated by inoculating a single colony into 100mL LB and shaking overnight (37°C,250 r.p.m.), were inoculated into 10 liters of fermentation media (5 g/L yeast extract, 20 g/L glu-

cose, 13.3 g/L KH2PO4, 4 g/L (NH4)2HPO4, 1.7 g/L citric acid, 0.0084 g/L EDTA, 0.0025 g/L CoCl2, 0.015 g/L MnCl2, 0.0015 g/L CuCl2, 0.003 g/L H3BO3, 0.0025 g/L Na2MoO4, 0.008 g/L Zn(CH3COO)2), 0.06 g/L Fe(III) citrate, 0.0045 g/L thiamine, 1.3 g/L MgSO4, pH 7.0). Temperature was maintained at 22°C and pH was maintained at 7.0 via addition of NaOH.

For taxadiene and iso-taxadiene production, cultures were performed under anaerobic conditions to reduce cell density, aiding in product recovery. For oxygenated taxanes such as compound 11, oxygen was supplied as sterile-filtered air at 1 L/minute with agitation adjusted to maintain DO at 10%. Induction with 0.1 mM IPTG was performed when OD reached 0.8. Taxanes were obtained from culture broth in a one-step extraction using a C18 chromatography resin and 100% ethyl acetate for elution (100 mL). The effluent was then dried to completion, suspended in acetonitrile and used for HPLC purification.

Preparation of **11** for NMR characterization was performed by epoxidizing biologically-produced taxadiene (HPLC purified), and refluxing in acetonitrile for 4 hours to increase the quantity of Compound **11** present. Semi-preparative HPLC purification was performed using a Supelco DiscoveryC18 (25cm x 10mm x 5µm) HPLC column. **11** was then purified under isocratic conditions (60% Acetonitrile in water at 6 mL/min) with **11** possessing an elution time of 42 minutes. Taxadiene and Isotaxadiene were purified under isocratic conditions (90% Acetonitrile in water, 6 mL/min) and had elution times of 19.1 and 20.2 minutes, respectively.

Chemical Reaction Experiments. In a typical mCBPA reactionion [41], taxadiene (1 mg, 0.0036 mmol) was weighed into a 14 mL pyrex glass tube containing CH2Cl2 (1mL), NaHCO3 (0.5 mg, 0.05 mmol). A magnetic stir bar was added and placed in a salt water/ice bath on a magnetic stir plate. The reaction was initiated with the addition of mCPBA (20µl of a 25mmolar stock, calculated assuming a nominal purity of 70%) and was allowed to proceed for 2 hours. The crude product mixture was washed twice with saturated sodium carbonate and once with NaS_2O_3. The organic layer was then removed, dried to completion, and products subjected to analysis. Direct analysis of the organic layer via GCMS yielded identical product

ratios, indicating wash steps did not alter product distribution.

A typical DMDO reaction was carried out by adding 0.34 mg taxadiene (0.00125 mmol) to 250 µl EA and 250 µl Water (buffered with 0.66mg sodium carbonate) with 10 µl acetate. Oxone in water (4.8 mg in 1 mL) and a solution of acetone and sodium carbonate in water (18 µl and 0.66 mg, respectively, in 1mL) were added dropwise as separate solutions over 2 hours. Sodium carbonate was chosen as the final buffering reagent as the reactions exhibited maximum yield when operated at high pH. Additionally, scaling up or transition to an acetone-free DMDO preparative method[1] did not alter the product distribution observed via GCMS or HPLC. HPLC-analysis of chemical reaction experiments was performed using a Waters Spherisorb ODS2 (25cm x 4.6mm x 3um) column using a gradient from 50% to 70% acetonitrile in water. For acid-hydrolysis of compound 5, 0.1 mg of HPLC-purified 5 was incubated in the presence of 5% HCl at 37°C for 1 hour. Following incubation the products were analyzed via GCMS.

Crude Lysate Enzymatic Assays. Crude lysates for enzymatic assays were performed by diluting cells to an equal optical density followed by lysis with BPER II. NADPH cycling [42] was used to ensure adequate NADPH supply by supplying NADP, Glucose-6-Phosphate, and Glucose-6-Phosphate Dehydrogenase when employing P450s. Reaction was initiated upon addition of taxadiene, with reactions performed without additional supplied NADPH showing no turnover. Products were extracted using ethyl acetate and analyzed via GCMS.

OCT Detection in Plant Cell Cultures. *Taxus Cuspidata* cell suspension cultures were grown in B5 sucrose media. Two days after passaging, cells were induced with 40µl/L methyl jasmonate and supplemented with the appropriate quantity of taxadiene. Cells were harvested 7 days post-induction and extracted, and isoprenoids were quantified.

Quantification of Isoprenoids. At the conclusion of fermentations, 200µL of fermentation broth was sampled and added to 200µL ethyl acetate containing internal standard and 100µL of 0.1mm glass beads. This solution was then vortexed at 4°C for 30 minutes and centrifuged at 18,000g for 10 minutes. The top, organic, layer was

extracted and 1µL was subjected to analysis via GCMS. The GCMS consisted of a Varian Saturn 3800 GC containing an HP-5ms (Agilent Technologies, USA) column, and Varian 2000 MS, with helium at a flow rate of 1mL/min as the carrier gas. The injector and transfer lines were maintained at 250°C. For oxygenated taxanes, the column oven was maintained at 100C for 1 minute and increased at a rate of 15°C/min to 175°C, then at 4°C/min to 220°C, then at 50°C/min to 290°C, and held at this temperature for 1 minute. For analysis of taxadiene isomers the column oven was maintained at 100°C for 1 minute and increased at a rate of 15°C/min to 165°C, then at 2°C/min to 170°C, held at this temperature for 7 minutes, increased at 50°C/min to 290°C, and held at this temperature for 1 minute. Caryophyllene (90mg/L) was used as an internal standard. Analysis of dodecane and DINP was performed by diluting 100-fold into internal standard.

NMR Characterization of 4. All NMR spectra were collected using a 500 MHz Varian INOVA spectrometer with deuterated benzene as the solvent and internal reference. All spectra were detected using a z-gradient 5mm inverse broadband probe. The 1H 1-D NMR spectrum was collected using a 8.6µs RF pulse with a sweep width of 4024.1 Hz centered at 3.955 ppm. The time-domain signal was collected for 3 seconds, corresponding to 24,146 points (12,073 complex). No apodization function was applied. The transformed data was loaded into an array with 128k complex points. Phase balancing and baseline correction were applied.

The 1H-1H 2-D gCOSY NMR spectrum was collected using 8.037 µs 1H RF pulses. The sweep width was 4029 Hz, and 1024 points were collected over 0.254 seconds. The spectral window was centered at 3.955 ppm. A relaxation delay of 1 second preceded each scan. 256 t1 time increments were used to collect 256 FIDS, and linear prediction was used to extend the t1 time domain data out to 512, thereby doubling the effective t1 evolution time. Two scans were averaged for each FID. The total experiment time was 11 minutes and 25 seconds. The final size of the data matrix was 1024 x 1024. The t1 time domain data was apodized with a squared sine bell function with a period (min to max) of 64 ms. The t2 time domain data was apodized with a squared sine bell function with a period of 128 ms. The gCOSY spectrum was

baseline corrected along the f1 frequency axis and then symmetrized.

The 1H-13C 2-D HSQC NMR spectrum was collected using 8.037 μs 1H RF pulses. The 1H sweep width and mid-point were the same as for the gCOSY spectrum described above. The 13C RF pulses were 17.8 μs in duration and the 13C spectral window was 18.85 kHz width, centered at 75 ppm. Each FID was the sum of 16 scans. Each FID consisted of 806 points (403 complex) - a small number required to keep 13C-decoupling times short - in this case the FID acquisition time was 100 ms. The number of FIDs collected was 512, giving 256 complex points in the t1 time domain. The total data collection time was 2 hours and 39 minutes. Linear prediction was used to double the number of t1 data points. The t1 time domain data was apodized with a Gaussian function with a period of 59 ms, and the t2 time domain data was apodized with a Gaussian function with a period of 13 ms. Following conversion of the t2 time domain to the f2 frequency domain, baseline correction was applied to the half-transformed interferogram. The size of the final matrix was 1024 x 512 (f1, f2).

The 1H-13C 2-D gHMBC NMR spectrum was collected using 8.037 μs 1H RF pulses. The 1H and 13C spectra windows were the same for this spectrum as for the HSQC spectrum described above. RF pulse durations were also the same as for the HSQC spectrum. Each FID was the sum of 64 transients. The number of data points in each FID was 512, giving a FID acquisition time of 127 ms. The number of points collected in the t1 time domain was 400. Linear prediction was used to extend the number of the t1 time domain points by 112 to 512. The total experiment time was 8 hours and 44 minutes. The t1 time domain data was apodized with a sine bell with a min to max period of 14 ms. The t2 time domain data was apodized with a sine bell with a period of 64 ms. The final size of the data matrix was 1024 x 512 (f1, f2).

3.12 Bibliography

[1] S. Edgar, K. Zhou, K. Qiao, J. R. King, J. H. Simpson, and G. Stephanopoulos, "Mechanistic insights into taxadiene epoxidation by taxadiene-5α-hydroxylase,"

Table 3.3: Strains Used in This Chapter

Strain	**Genotype**
TaxE0	MG1655 ΔrecA ΔendA DE3
TaxE1	MG1655 ΔrecA ΔendA DE3 ΔaraA::Trc-MEP ΔLacY::T7-TG
TaxE5	MG1655 ΔrecA ΔendA DE3 ΔaraA::Trc-MEP
TaxE9	MG1655 ΔrecA ΔendA DE3 p5-Trc-5ÎśCYP-CPR
TaxE12	MG1655 ΔrecA ΔendA DE3 ΔaraA::Trc-MEP ΔLacY::T7-TG p5-Trc-5ÎśCYP-CPR
TaxE14	MG1655 ΔrecA ΔendA DE3 ΔaraA::Trc-MEP ΔLacY::T7-TG p5-Trc-5ÎśCYP-CPR-Trc-TAT
TaxE15	MG1655 ΔrecA ΔendA DE3 ÎŤaraA::Trc-MEP ΔLacY::T7-TG p5-Trc-5ÎśCYP-o-CPR (operon)
TaxS1	MATa ura3ÎŤ0::URA-TEFp-5ÎśCYP-CPR-CYCt
TaxY1	MATa, leu2-270::LEU-TEFin-5ÎśCYP-CPR-XPR2t, ura3-302::URA3, xpr2-3

ACS chemical biology, vol. 11, no. 2, pp. 460–469, 2015.

[2] H. G. Floss and U. Mocek, "Biosynthesis of taxol," in *Taxol: science and applications*, pp. 191–208, CRC Press Boca Raton, FL, 1995.

[3] R. Croteau, J. Hefner, M. Hezari, and N. G. Lewis, "Taxol biosynthesis: cyclization and early hydroxylation steps of the pathway," 1995.

[4] D. Kingston, A. Molinero, and J. Rimoldi, "The taxane diterpenoids," in *Fortschritte der Chemie organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products*, pp. 1–165, Springer, 1993.

[5] J. Hefner, S. M. Rubenstein, R. E. Ketchum, D. M. Gibson, R. M. Williams, and R. Croteau, "Cytochrome p450-catalyzed hydroxylation of taxa-4 (5), 11 (12)-diene to taxa-4 (20), 11 (12)-dien-5a-o1: the first oxygenation step in taxol biosynthesis," *Chemistry & biology*, vol. 3, no. 6, pp. 479–489, 1996.

[6] S. Jennewein, R. M. Long, R. M. Williams, and R. Croteau, "Cytochrome p450 taxadiene 5α-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis," *Chemistry & biology*, vol. 11, no. 3, pp. 379–387, 2004.

[7] D. Pompon, B. Louerat, A. Bronine, and P. Urban, "Yeast expression of animal and plant p450s in optimized redox environments," *Methods in enzymology*, vol. 272, pp. 51–64, 1996.

[8] D. Rontein, S. Onillon, G. Herbette, A. Lesot, D. Werck-Reichhart, C. Sallaud, and A. Tissier, "Cyp725a4 from yew catalyzes complex structural rearrangement of taxa-4 (5), 11 (12)-diene into the cyclic ether 5 (12)-oxa-3 (11)-cyclotaxane," *Journal of Biological Chemistry*, vol. 283, no. 10, pp. 6067–6075, 2008.

[9] S. Jennewein, R. M. Long, R. M. Williams, and R. Croteau, "Cytochrome p450 taxadiene 5α-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis," *Chemistry & biology*, vol. 11, no. 3, pp. 379–387, 2004.

[10] O. Hayaishi, M. Katagiri, and S. Rothberg, "Mechanism of the pyrocatechase reaction," *Journal of the American Chemical Society*, vol. 77, no. 20, pp. 5450–5451, 1955.

[11] J. T. Groves and Y.-Z. Han, "Models and mechanisms of cytochrome p450 action," in *Cytochrome P450*, pp. 3–48, Springer, 1995.

[12] J. Z. Krevor, J. Davies, P. Watson, A. Greenberg, and J. Liebman, "Selective hydrocarbon activation principles and progress," *VCH, New York*, p. 19, 1990.

[13] D. R. Nelson, "Cytochrome p450: Structure, mechanism, and biochemistry," *Journal of the American Chemical Society*, vol. 127, no. 34, pp. 12147–12148, 2005.

[14] J. L. McLain, J. Lee, and J. T. Groves, "Biomimetic oxygenations related to cytochrome p450: metal-oxo and metal-peroxo intermediates," *Biomimetic oxidations catalyzed by transition metal complexes. Imperial College Press, London*, pp. 91–169, 2000.

[15] B. Meunier, S. P. De Visser, and S. Shaik, "Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes," *Chemical reviews*, vol. 104, no. 9, pp. 3947–3980, 2004.

[16] J. Rittle and M. T. Green, "Cytochrome p450 compound i: capture, characterization, and ch bond activation kinetics," *Science*, vol. 330, no. 6006, pp. 933–937, 2010.

[17] P. Meinhold, M. W. Peters, A. Hartwick, A. R. Hernandez, and F. H. Arnold, "Engineering cytochrome p450 bm3 for terminal alkane hydroxylation," *Advanced Synthesis & Catalysis*, vol. 348, no. 6, pp. 763–772, 2006.

[18] A. W. Munro, D. G. Leys, K. J. McLean, K. R. Marshall, T. W. Ost, S. Daff, C. S. Miles, S. K. Chapman, D. A. Lysek, C. C. Moser, *et al.*, "P450 bm3: the very model of a modern flavocytochrome," *Trends in biochemical sciences*, vol. 27, no. 5, pp. 250–257, 2002.

[19] C. J. Whitehouse, S. G. Bell, and L.-L. Wong, "P450 bm3 (cyp102a1): connecting the dots," *Chemical Society Reviews*, vol. 41, no. 3, pp. 1218–1260, 2012.

[20] M. A. Fischbach and J. Clardy, "One pathway, many products," *Nature chemical biology*, vol. 3, no. 7, pp. 353–355, 2007.

[21] R. D. Firn and C. G. Jones, "The evolution of secondary metabolism, a unifying model," *Molecular microbiology*, vol. 37, no. 5, pp. 989–994, 2000.

[22] R. D. Firn and C. G. Jones, "Natural products a simple model to explain chemical diversity," *Natural product reports*, vol. 20, no. 4, pp. 382–391, 2003.

[23] V. G. Yadav, "Unraveling the multispecificity and catalytic promiscuity of taxadiene monooxygenase," *Journal of Molecular Catalysis B: Enzymatic*, vol. 110, pp. 154–164, 2014.

[24] V. G. Yadav, "Combining metabolic pathway design and retrosynthetic planning for the design of a novel semisynthetic manufacturing scheme for paclitaxel," *Organic Process Research & Development*, vol. 18, no. 6, pp. 816–826, 2014.

[25] P. K. Ajikumar, W.-H. Xiao, K. E. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, and G. Stephanopoulos, "Isoprenoid path-

way optimization for taxol precursor overproduction in escherichia coli," *Science*, vol. 330, no. 6000, pp. 70–74, 2010.

[26] M. Girhard, K. Machida, M. Itoh, R. D. Schmid, A. Arisawa, and V. B. Urlacher, "Regioselective biooxidation of (+)-valencene by recombinant e. coli expressing cyp109b1 from bacillus subtilis in a two-liquid-phase system," *Microbial cell factories*, vol. 8, no. 1, p. 1, 2009.

[27] C. J. Paddon, P. J. Westfall, D. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. Leavell, A. Tai, A. Main, D. Eng, *et al.*, "High-level semi-synthetic production of the potent antimalarial artemisinin," *Nature*, vol. 496, no. 7446, pp. 528–532, 2013.

[28] B. W. Biggs, J. E. Rouck, A. Kambalyal, W. Arnold, C. G. Lim, M. De Mey, M. OâĂŹNeil-Johnson, C. M. Starks, A. Das, and P. K. Ajikumar, "Orthogonal assays clarify the oxidative biochemistry of taxol p450 cyp725a4," *ACS chemical biology*, vol. 11, no. 5, pp. 1445–1451, 2016.

[29] J. Hefner, S. M. Rubenstein, R. E. Ketchum, D. M. Gibson, R. M. Williams, and R. Croteau, "Cytochrome p450-catalyzed hydroxylation of taxa-4 (5), 11 (12)-diene to taxa-4 (20), 11 (12)-dien-5a-ol: the first oxygenation step in taxol biosynthesis," *Chemistry & biology*, vol. 3, no. 6, pp. 479–489, 1996.

[30] E. Corey, S. C. Virgil, and S. Sarshar, "New mechanistic and stereochemical insights on the biosynthesis of sterols from 2, 3-oxidosqualene," *Journal of the American Chemical Society*, vol. 113, no. 21, pp. 8171–8172, 1991.

[31] D. Kumar, B. Karamzadeh, G. N. Sastry, and S. P. de Visser, "What factors influence the rate constant of substrate epoxidation by compound i of cytochrome p450 and analogous iron (iv)-oxo oxidants?," *Journal of the American Chemical Society*, vol. 132, no. 22, pp. 7656–7667, 2010.

[32] A. D. Vaz, D. F. McGinnity, and M. J. Coon, "Epoxidation of olefins by cytochrome p450: evidence from site-specific mutagenesis for hydroperoxo-iron

as an electrophilic oxidant," *Proceedings of the National Academy of Sciences*, vol. 95, no. 7, pp. 3555–3560, 1998.

[33] J. Guerra-Bubb, R. Croteau, and R. M. Williams, "The early stages of taxol biosynthesis: an interim report on the synthesis and identification of early pathway metabolites," *Natural product reports*, vol. 29, no. 6, pp. 683–696, 2012.

[34] A. R. Modi and J. H. Dawson, "Oxidizing intermediates in p450 catalysis: A case for multiple oxidants," in *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*, pp. 63–81, Springer, 2015.

[35] S. Jin, T. M. Makris, T. A. Bryson, S. G. Sligar, and J. H. Dawson, "Epoxidation of olefins by hydroperoxo-ferric cytochrome p450," *Journal of the American Chemical Society*, vol. 125, no. 12, pp. 3406–3407, 2003.

[36] N. A. Barton, B. J. Marsh, W. Lewis, N. Narraidoo, G. B. Seymour, R. Fray, and C. J. Hayes, "Accessing low-oxidation state taxanes: is taxadiene-4 (5)-epoxide on the taxol biosynthetic pathway?," *Chemical Science*, vol. 7, no. 5, pp. 3102–3107, 2016.

[37] H. Itokawa, "3 taxoids occurring in the genus taxus," *Taxus: the genus Taxus*, p. 35, 2003.

[38] K. H. Hopmann, B. M. Hallberg, and F. Himo, "Catalytic mechanism of limonene epoxide hydrolase, a theoretical study," *Journal of the American Chemical Society*, vol. 127, no. 41, pp. 14339–14347, 2005.

[39] D. Brodkorb, M. Gottschall, R. Marmulla, F. Lüddeke, and J. Harder, "Linalool dehydratase-isomerase, a bifunctional enzyme in the anaerobic degradation of monoterpenes," *Journal of Biological Chemistry*, vol. 285, no. 40, pp. 30436–30442, 2010.

[40] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos, "Distributing a metabolic pathway among a microbial consortium enhances production of natural products," *Nature biotechnology*, vol. 33, no. 4, pp. 377–383, 2015.

[41] C. Kim, T. G. Traylor, and C. L. Perrin, "mcpba epoxidation of alkenes: Reinvestigation of correlation between rate and ionization potential," *Journal of the American Chemical Society*, vol. 120, no. 37, pp. 9513–9516, 1998.

[42] D. Ghigo, C. Riganti, E. Gazzano, C. Costamagna, and A. Bosia, "Cycling of nadph by glucose 6-phosphate dehydrogenase optimizes the spectrophotometric assay of nitric oxide synthase activity in cell lysates," *Nitric oxide*, vol. 15, no. 2, pp. 148–153, 2006.

Chapter 4

Engineering Early Paclitaxel Biosynthesis

4.1 Introduction

Portion of this Chapter are adapted from [1]

As demonstrated in Chapter 3 and in previous reports, taxadiene-5α-hydroxylase (CYP725A4), the cytochrome P450 catalyzing the first oxygenation of the taxadiene core [2], produces an array of nonspecific products in addition to the desired product taxadien-5α-ol when expressed in *E. coli* [3]. Indeed, less than 10% of taxadiene is converted to taxadien-5α-ol by taxadiene-5α-hydroxylase, presenting a significant bottleneck in practical production of early intermediates and downstream Taxol pathway elucidation, as only this product is able to be further transformed into Taxol [4, 5, 6]. As shown in Chapter 3, via both chemical and biological studies, this broad product profile is due to non-selective breakdown of an epoxide intermediate (**3**) rather than non-selective hydroxylation by taxadiene-5α-hydroxylase itself [4, 7, 8].

We thus set out to leverage chemical and mechanistic information to alleviate this bottleneck by improving the product specificity of the early paclitaxel biosynthetic pathway from the geranylgeranyl pyrophosphate (GGPP, **9**) precursor to the desired monohydroxylated diterpene product, taxadien-5α-ol (**2**). In this instance, differenti-

ating between a radical-rebound-based mechanism and an epoxidation-based one for taxadiene-5α-hydroxylase is essential as it informs engineering efforts. In the originally proposed radical-rebound mediated mechanism [9], poor selectivity for taxadien-5α-ol arises from aberrant catalytic cascades or incorrect substrate docking poses within the taxadiene-5α-hydroxylase active site [10, 11]. As a result, improvement of selectivity would only be correctable by engineering of the P450 itself, as discussed in Section 3.2.1 [12, 13]. In contrast, in the more recently proposed mechanism, taxadiene is first epoxidized to taxadien-4(5)-epoxide, **5**, and then undergoes non-specific degradation to a range of products, while an isomer of taxadiene, **8**, still undergoes the traditional radical rebound in a product-specific manner. This epoxidation-based mechanism for taxadiene (**1**) was supported by multiple lines of evidence including *in-vivo* enzyme behavior and similarity of enzymatic products to those obtained using chemical epoxidation [8]. Furthermore, chemical analogues of the active site heme moiety, such as FeIII(TPP)Cl, have been shown to catalyze both the epoxidation of **1**, as well as its nonselective decomposition to other products [7]. Under this mechanism, multiple locations for intervention are feasible, based on several plausible steps of chemical transformation from **1** towards **2** (Figure 4-1), described below. We thus investigated each of these three approaches to enable high-level production of early Taxol pathway intermediates:

1. Engineering the cyclization step which generates **1** and **8** from **9**.
2. Improving taxadiene-5α-hydroxylase specificity.
3. Controlling the epoxide breakdown pathway via introduction of additional enzymes.

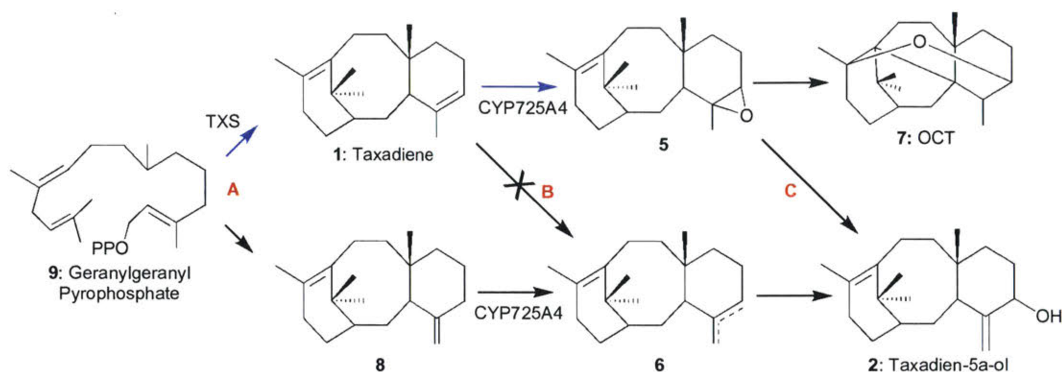

Figure 4-1: The known native pathway (blue) catalyzed by taxadiene synthase (TXS) and taxadiene-5α-hydroxylase (CYP725A4) terminates in an epoxide intermediate (**5**) which then undergoes nonspecific degradation to compounds such as OCT (**7**). Alternatively, three engineered pathways (black) are proposed originating from catalytic branch points (A, B, and C) in order to improve selectivity for the desired product, **2**.

4.2 Engineering Selectivity Through Addition of Plant Enzymes

It has been previously proposed that under an epoxidation-based regime, a second enzyme enabling the specific degradation of the epoxide would exist [9]. Thus, as a second enzyme would enable a plug-and-play solution and would also mimic the native solution, we set out to investigate this for our initial point of investigation by engineering the breakdown of the epoxide intermediate (**5**) for increased production of taxadien-5α-ol. We hypothesized that the epoxide intermediate (**5**) may under-go two alternative mechanistic trajectories to yield taxadien-5α-ol: a one-step concerted reaction or a two-step process of epoxide opening followed by dehydration. Indeed, both reaction mechanisms have precedent within terpenoid biosynthesis, exemplified by limonene epoxide hydrolase [14] and linalool dehydratase [15] respectively.

However, no such enzyme has been previously identified, and the ability to catalyze the general acid-base isomerization (Figure 4-2) from the epoxide to taxadien-5α-ol could have evolved from dozens of different enzymes classes, making the search difficult. Most feasible mechanisms leading from the epoxide to taxadien-5α-ol would

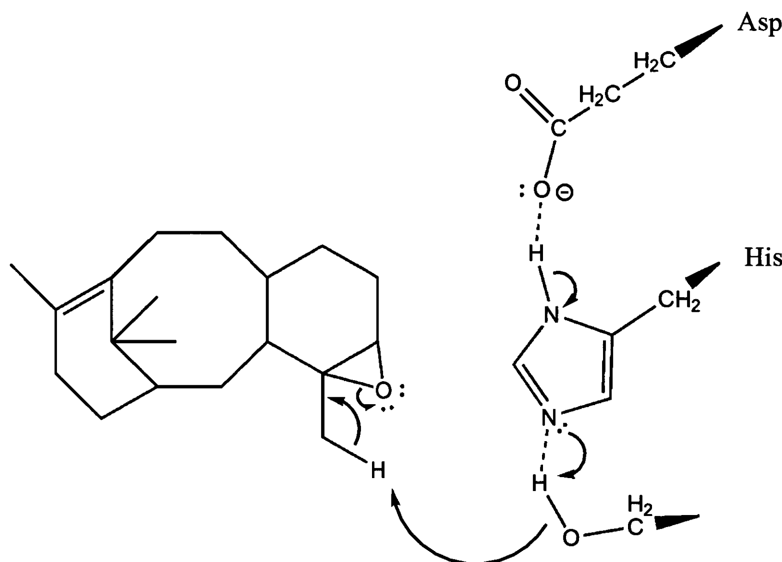

Figure 4-2: **Proposed Mechanism for Conversion of Epoxide to Taxadien-5α-ol.** Shown catalyzing the reaction is a catalytic triad, as present in many of the selected enzymes which were screened, although it is feasible this reaction may occur non-enzymatically.

likely require an enzymatic catalytic triad, found in a wide variety of degradative enzymes. Although no full genome sequence is available, de-novo transcriptomes [16] and differential expression data under methyl jasmonate induction [17] and between tissues [18] have been reported, and may act as a starting point for searching for such enzymes. Furthermore, Sun et al. previously identified a set of degradative enzymes upregulated in *T. cuspidata* upon methyl jasmonate induction, correlating with Taxol production, and we thus selected these enzymes (25 contigs with 2-fold upregulation or more) with sequences matching predicted degradative enzyme domains) for screening for hydrolase activity [17].

We obtained full-length sequences for 23 of these candidate genes by querying the initial partial transcripts against several other available *Taxus* transcriptomes via reciprocal B:AST. This allowed near complete transcripts to be obtained. We then derived RNA and cDNA from *Taxus Cuspidata* plant cell culture for both cloning of known sequences and for completing incomplete contigs, when required, via Rapid Amplification of cDNA Ends (RACE). Full length sequences are given in Appendix

A. Screening was performed by expressing candidate degradative genes in TaxE12 background (generating TaxE12-01 to TaxE12-25), and differences in observed selectivity were screened for. Expression of each of the candidate genes failed to yield significant alteration in product profiles, indicating that these genes are unlikely to guide epoxide degradation.

Enzyme discovery is further hindered as poor stability of epoxide **5** precludes activity-guided protein fractionation from *Taxus* extracts, and the necessary rearrangement may occur through simple acid-base catalysis. Furthermore, the broad product range not only occurs in heterologous systems such as in *E. coli*, but is also observed, at least to a degree, *in Taxus*, as demonstrated in the previous chapter, indicating such epoxide degradative enzymes may not exist. As a result, this approach was deemed infeasible as a route to overcome the Taxol biosynthetic bottleneck. This is also an important result as it reveals that in engineering efforts, the goal may not be to mimic nature, but rather to improve upon it.

4.3 Taxadiene-5α-Hydroxylase Engineering For Improved Selectivity

We next evaluated oxidation by taxadiene-5α-hydroxylase as a target for selectivity improvement. Although it is initially intuitive to attempt to alter selectivity by engineering the enzyme to preferentially perform hydroxylation rather than epoxidation, the epoxidation of taxadiene is proposed to be substrate driven [8, 19], and thus this approach is unlikely to prove sucessful. However, we noted taxadiene-5α-hydroxylase natively produces low levels of a di-hydroxylated product, proposed to be taxadien-5α-10β-diol [5, 11]. It is possible that a second hydroxylation is performed immediately before, or rapidly after epoxidation. In this scenario, the secondary hydroxylation acts essentially as a protecting group, preventing large-scale intramolecular rearrangements through either reduced nucleophilicity of the 12(13) alkene, steric hindrance, or other effects. A figure illustrating the possible path is found below in Figure 4-3.

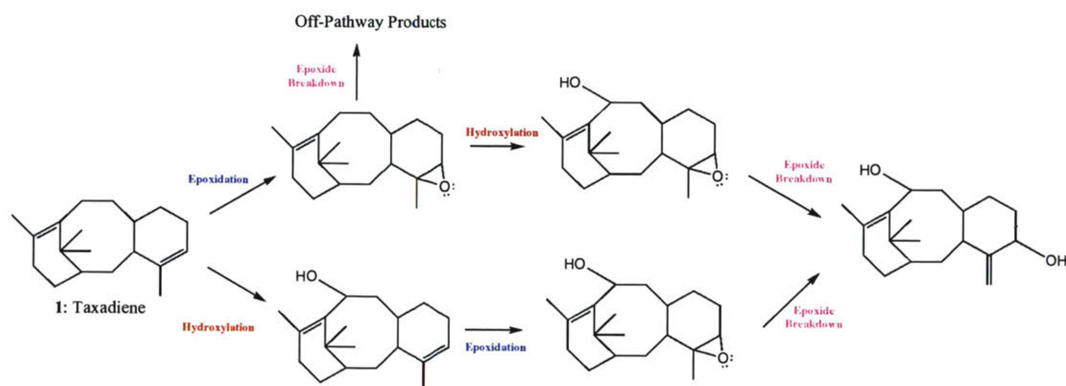

Figure 4-3: **Possible Mechanisms for Di-Hydroxylated Product Formation**

Thus, we set out to expand this di-hydroxylating capability of the enzyme through mutagenesis to circumvent the poor specificity of the first oxidation step. Due to the high frequency at which mutations disrupt P450 maturation, we employed a structural and phylogenetically-based approach to both conserve enzymatic activity and enhance this native di-hydroxylating ability (Figure 4-5 and 4-4). No crystal structure is available for the enzyme, thus we made use of homology modeling to generate an approximate enzyme model and determine active site residues which may influence catalysis. This was performed via computational docking and selection of all residues within a set radius. Furthermore, as P450s are easily rendered inactive via mutagenesis and no high-throughput screen is available we were required to focus our mutagenesis approach. To achieve this, we used a phylogenetically-guided methodology to reduce the screening and cloning burden. This approach was performed by aligning rationally-selected active site residues and comparing their sequence to that of enzymes downstream in the Taxol biosynthetic pathway, with the goal of generating active sites capable of holding more functionalized products. Following residue selection and multiple sequence alignment, a total of 30 mutants were selected, generated, and screened *in-vivo* in a taxadiene-production background (TaxE1).

While 18 of the 30 mutants retained catalytic activity, a majority of the mutants failed to alter product ratios. However, a single mutant, T380S, displayed dramatically altered selectivity and produced a di-hydroxylated product as its primary product with little decrease in total product titer (Figure 4-7). Although the retention

time and mass spectra resembled the proposed taxadien-5α-10β-diol produced by co-expression of the taxadiene-5α-hydroxylase and previously characterized taxane-10β-hydroxylase enzyme (Figure 4-6), this compound had not been fully characterized. We thus scaled-up bioreactor runs, and performed preparative chromatography and NMR data. Upon inspection of NMR, this was found to be a novel product, 5(12)-oxa-3(11)-cyclo-taxan-10-ol (**12**), which would arise via hydroxylation after epoxide degradation (Figure 4-8). Although this result is not the desired one, several valuable insights may be gained from it. First, this result highlights that dihydroxylated products are obtainable through this engineering approach. However, this also highlights that the proposal to exploit this di-hydroxylating ability must inherently compete with native epoxide breakdown, which is likely to be more rapid kinetically. Therefore, it is unlikely that an approach such as this will yield positive results.

4.4 Taxadiene Synthase Engineering for Increased Isomer Production

We next examined the first catalytic branch point as a point of investigation, and set out to engineer the terpene cyclase step, catalyzed by taxadiene synthase (TXS). Although it is not intuitive to focus efforts 'upstream' of the pathway bottleneck, this approach has many advantages. TXS natively produces two isomers, taxa-4(5)-11(12)-diene (**1**) and taxa-4(20)-11(12)-diene (**8**) at 92% and 8%, respectively [22], and taxadiene-5α-hydroxylase has been shown to act differentially upon the two isomers [8]. As shown in Chapter 3, while **1** is transformed into a range of products through an epoxide intermediate, **8** is hydroxylated with high specificity to the desired product, taxadien-5α-ol (**2**), exclusively. Thus, by engineering of TXS for increased production of **8**, the selectivity of the overall pathway can be improved as the P450 then behaves in a selective manner on the alternative substrate.

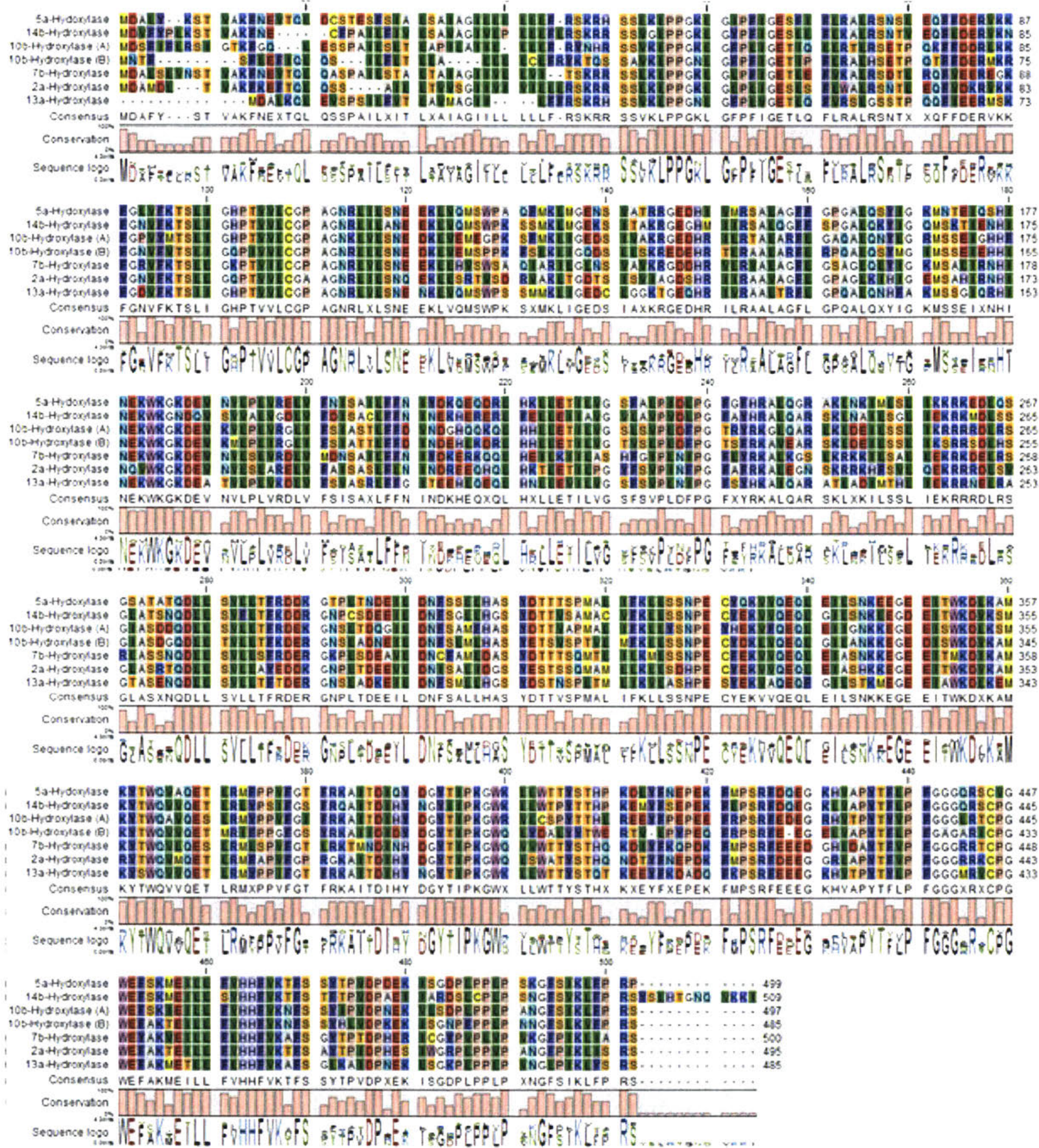

Figure 4-4: Sequence Alignment of Taxadine-5a-Hydroxylase to downstream Taxol P450s. Alignment to downstream P450s in the Taxol biosynthetic pathway including taxadien-10b-hydroxylase from two sources [19, 14], taxadien-13a-hydroxylaes [15], taxadiene-14b-hydroxylase [17], taxadiene-2a-hydroxylaes [18], and taxadiene-7b-hydroxylase [20].

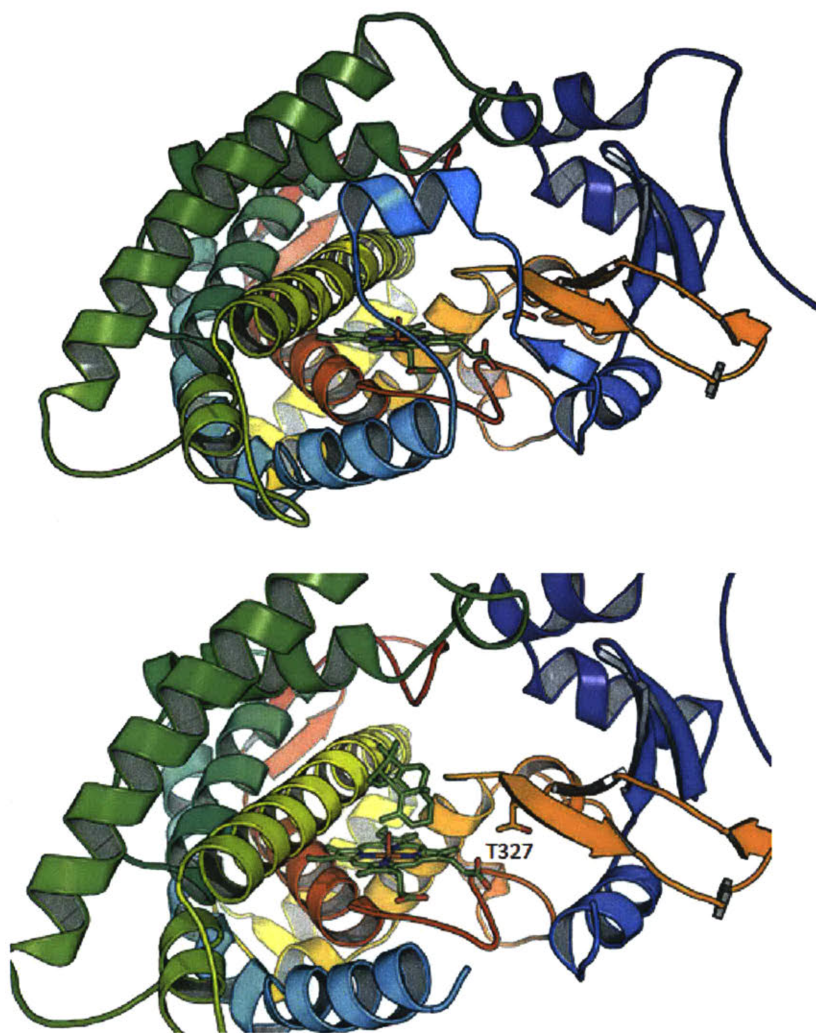

Figure 4-5: **Homology Model of Taxadien-5a-Hydoxylase.** Homology model obtained using Phyre2 [21], with active site and taxadiene docked (Green, bottom). The highly-conserved F helix can be seen in yellow adjacent to the heme cofactor and residue T380S is shown to the left of the substrate. Residues in proximity were selected subjected to phylogenetic analysis and mutagenesis.

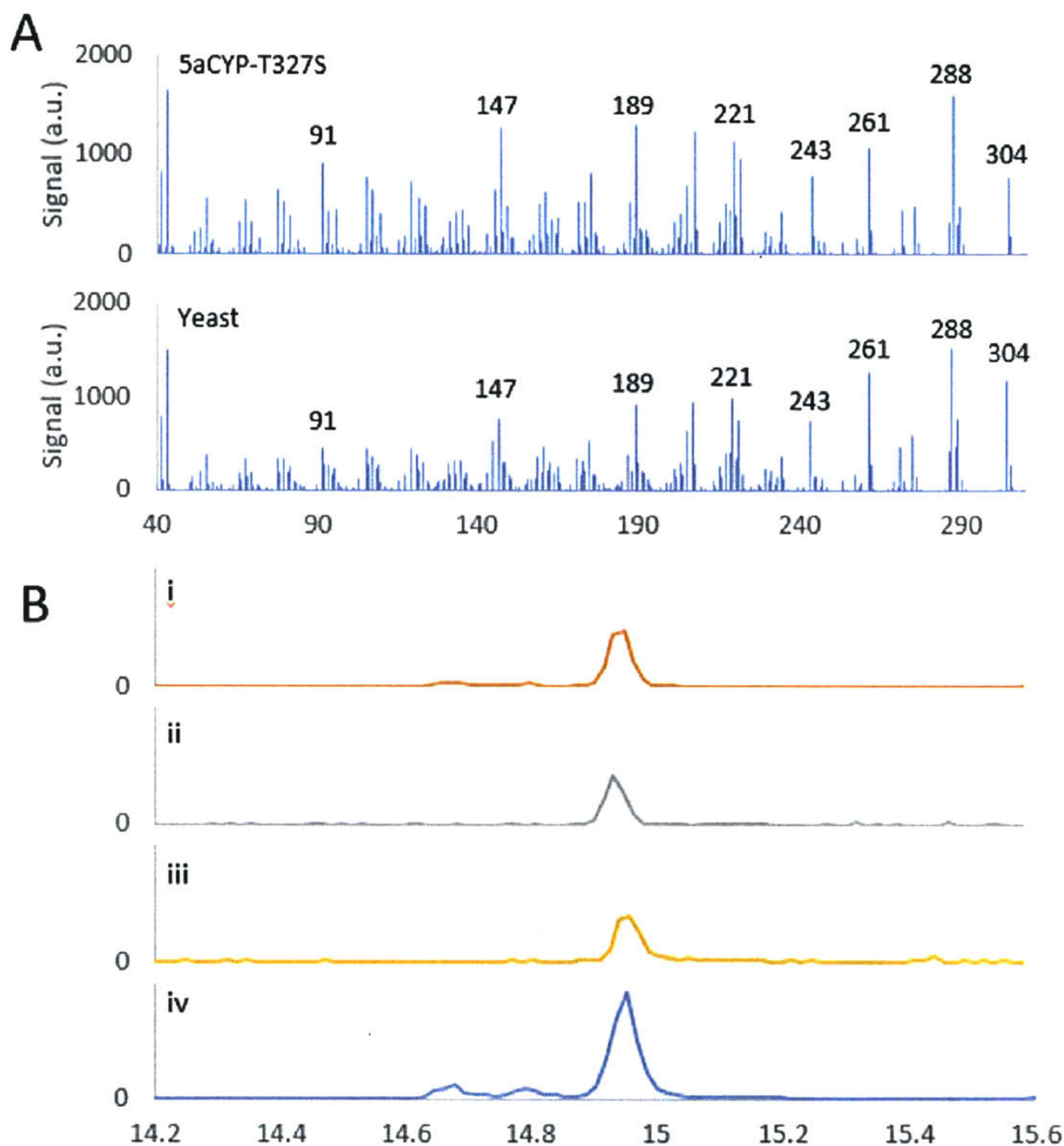

Figure 4-6: **Mass Spectra and Retention time of Observed Diol.** A) Mass spectra of the primary di-hydroxylated compound obtained (Top) to that obtained-from feeding taxadiene to yeast containing taxadiene-5α-hydroxylase and taxane-10β-hydroxylase (Bottom), showing high levels of similarity. (B) GCMS traces (shown with parent mass 304 m/z) obtained from feeding taxadiene (i) and taxadien-5a-ol (ii) to TaxE1, containing taxadiene-5α-hydroxylase. (iii) Trace obtained from T360S taxadiene-5α-hydroxylase mutant displaying diol production and (iv) trace obtained from feeding taxadiene to yeast containing taxadiene-5α-hydroxylase and taxane-10β-hydroxylase

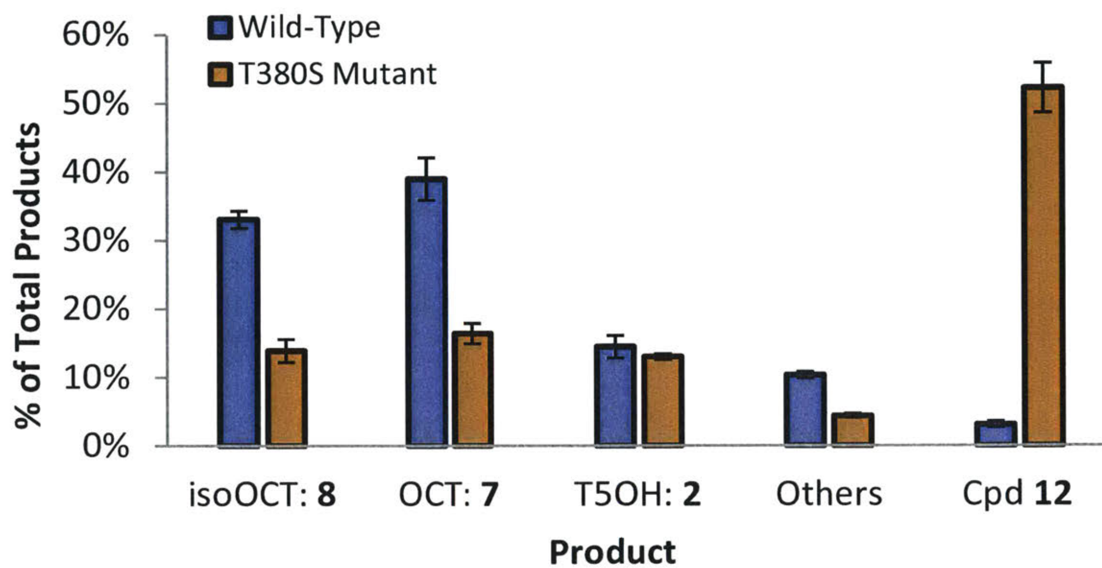

Figure 4-7: Observed Selectivity from T380S Mutant.

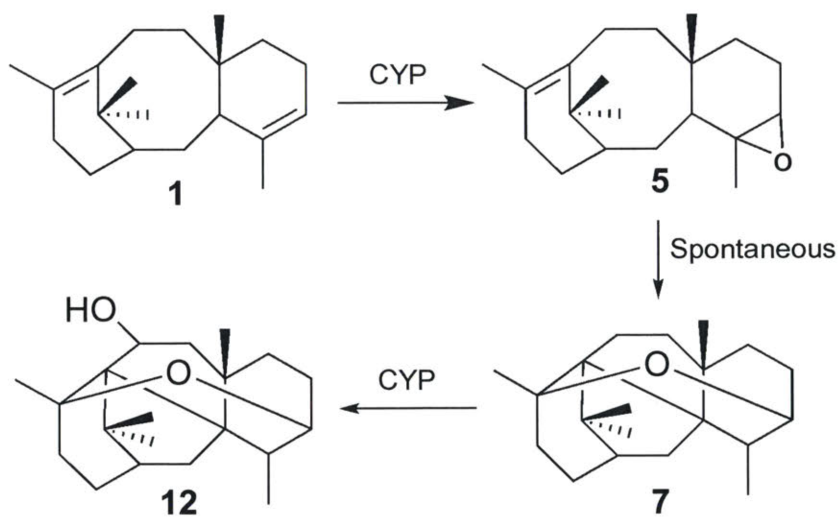

Figure 4-8: Proposed Pathway for Formation of Compound 12

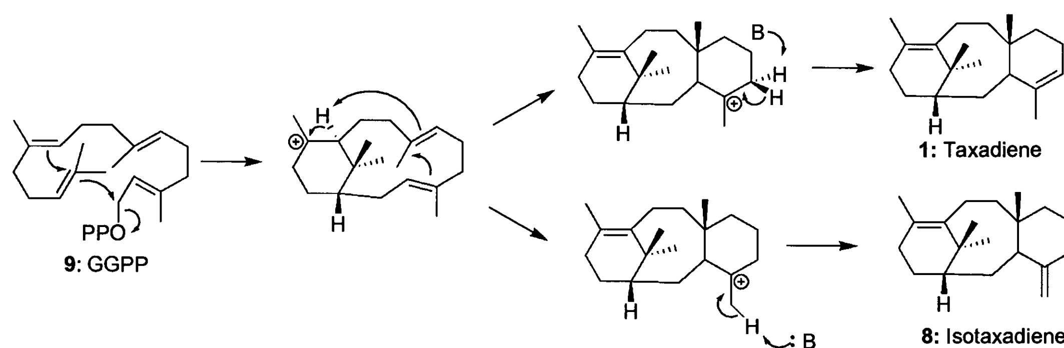

Figure 4-9: **Mechanism of Taxadiene Synthase** leading to the production of both the primary product, taxadiene, and a minor product, isotaxadiene

4.4.1 The Mechanism of Taxadiene Synthase

Our effort to selectively alter the product profile of taxadiene synthase was facilitated by the well-established taxadiene synthase catalytic mechanism (Figure 4-9 [23, 24, 25]. Close examination of this mechanism suggests that the final step of the catalytic cycle dictates which of the two isomers is generated; abstraction of H5β results in taxadiene, whereas abstraction of H20 gives rise to isotaxadiene. This aids engineering as fewer perturbations must be made than if it occurred early in catalysis. Furthermore, taxadiene synthase has a known crystal structure, from which we can gain structural insights [23].

4.4.2 Investigation of Stereospecific Base

We next employed computational docking of taxadiene into the taxadiene synthase active site. In agreement with previous literature, we obtained two poses which indicate the PP_i cleaved from GGPP in the intial reaction step (in a substrate-assisted catalytic regime) or residue Y688 may act as the base for abstraction (Figure 4-11). However, we also obtained a number of alternative poses with similar energetics and predicted poses varied significantly by choice of A or B chain and presence or absence of PP_i in the model. This is likely due to the highly hydrophobic nature of taxadiene and large active site volume, making computational docking unreliable and requiring a broader mutagenesis approach for improvement.

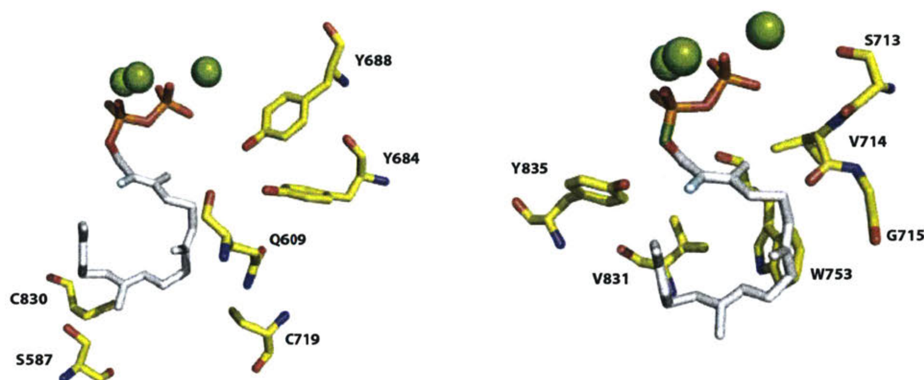

Figure 4-10: **Select Active Site Residue Locations.**

To determine which, if any, residues were directly responsible for the abstraction of H5β, we performed alanine-scanning across the six polar active site residues (S587, Q609, Y684, Y688, C719, and C830, as shown in Figure 4-10), and examined the biochemical function of the resulting mutant enzymes when expressed in a previously engineered *E. coli* strain with increased GGPP supply [4]. As seen in Figure 4-12, no mutant displayed abolished production of **1** while retaining production of **8**, indicating that no single residue appears to be solely responsible for abstraction of H5β. Alternatively, the PP_i may act as the abstracting base, as shown by our docking and proposed previously [23], or a single residue is responsible for the formation of both isomers.

Several other results can also be gleaned from this experiment. First, the C719A and C830A mutants display reduced in vivo cyclase activity, but retain the original selectivity (Figure 4-13), indicating these residues may affect enzyme turnover or stability. Furthermore, despite its distance from the magnesium cluster, the S587A mutant displays no activity and thus still appears to be essential for catalysis and/or overall protein stability. Interestingly, mutants of residues Q609, Y684, and Y688 yielded an unknown product with retention time and mass spectra resembling that of **1** and **8**. Following production via fermentation in bioreactors, purification, and

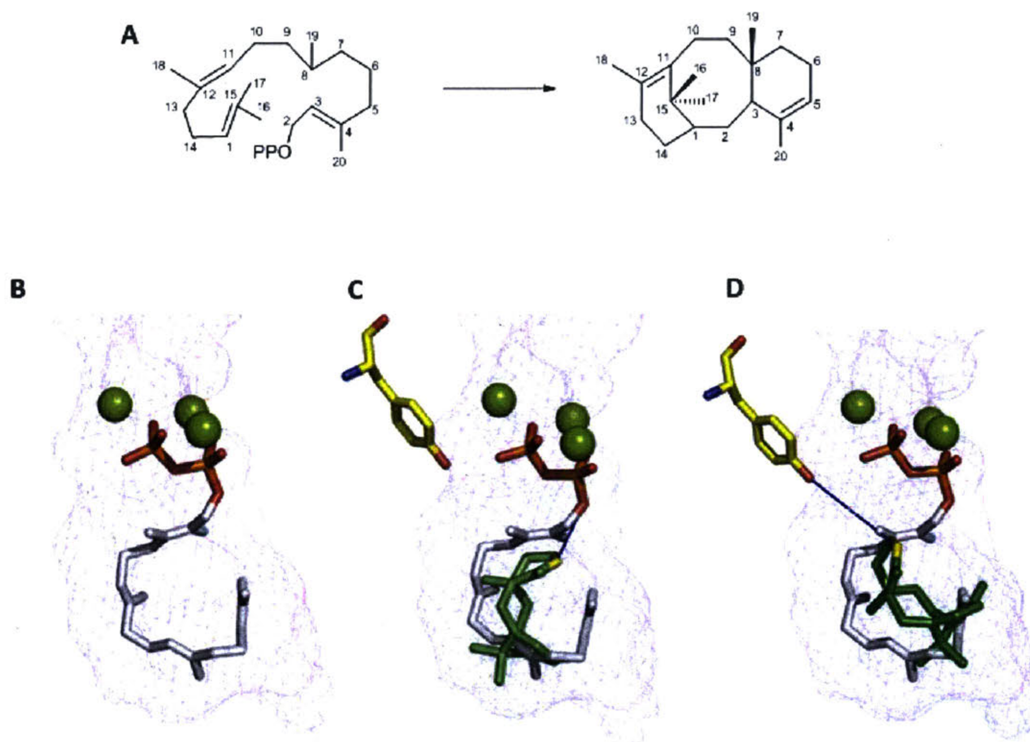

Figure 4-11: **Molecular Docking Results for Taxadiene.** A) Atom labeling configuration for the substrate, geranylgeranyl pyrophosphate (**9**), and the product, taxa-4(5),11(12)-diene (**1**). (B) Active site of taxadiene synthase with substrate analog docked, as obtained from available crystal structures (Accession 3P5R) [26]. (C) Docking result depicting taxadiene (green) bound in the known active site in an orientation in which PP_i of the substrate would perform the abstraction of H5β (blue dashes) (D) Docking result depicting taxadiene (green) bound in which the H5β abstraction would be performed by Y688.

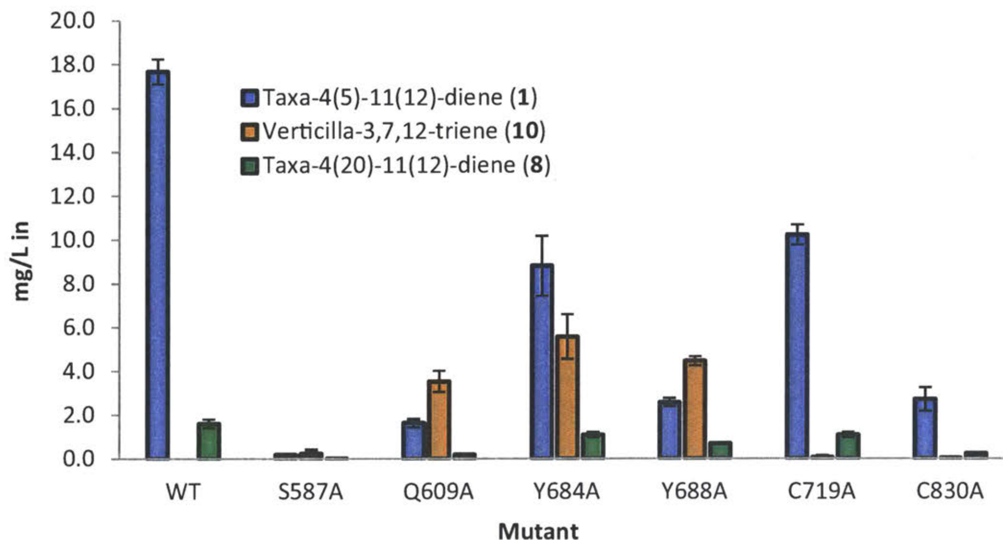

Figure 4-12: **Absolute Selectivity of Alanine-Scanning Mutants**

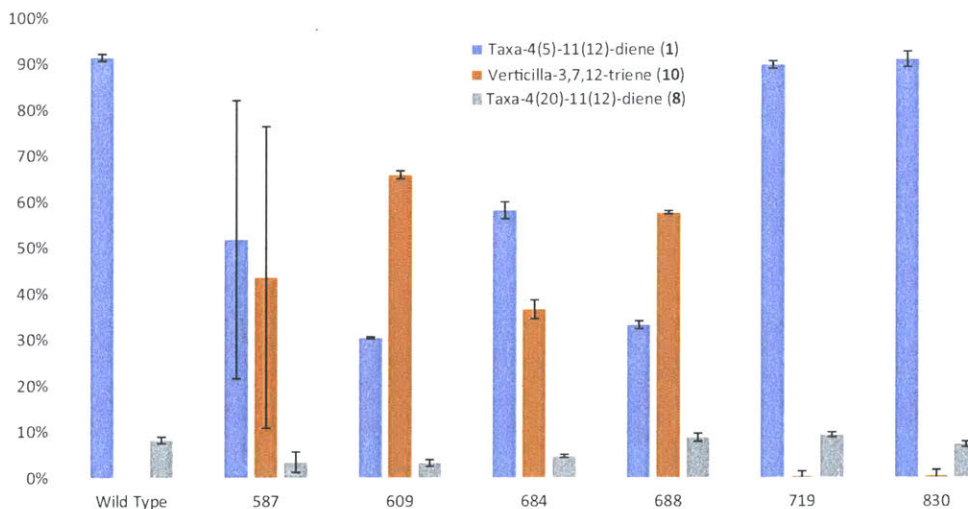

Figure 4-13: **Relative Selectivity of Alanine-Scanning Mutants.** Observed selectivity of mutants obtained through alanine scanning of active site polar residues. Large error bars observed on residue 587 are due to low observed titer, decreasing signal-to-noise ratio.

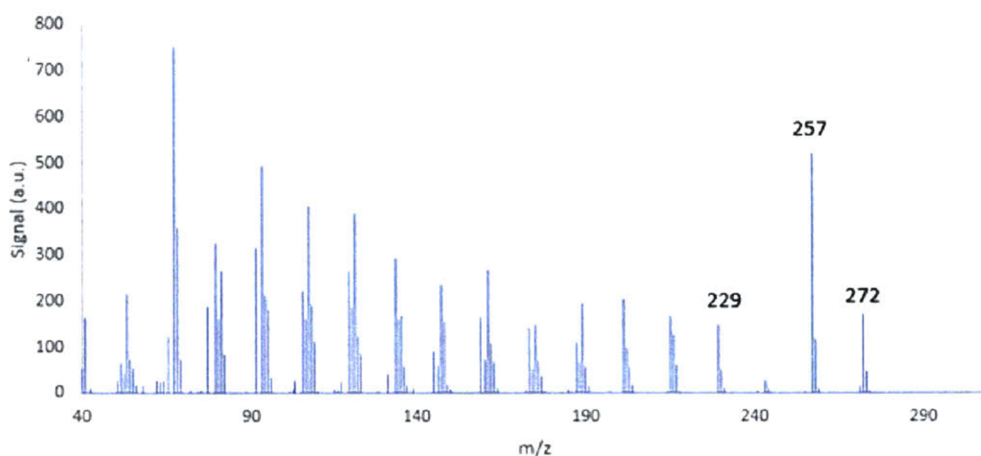

Figure 4-14: **Mass Spectra Obtained from Compound 10.** In contrast to taxadiene and isotaxadiene, the primary ion is 257.

characterization, the product was determined to be verticilla-3,7,12-triene (10, Figure 4-15. This product has been previously isolated from *Bursera suntui* plant tissue [27], and its derivatives have also been isolated from various Taxus species [26], although the terpene synthase responsible for its synthesis has not been identified to date. Based on the structure of this compound, we can make a reasonable mechanistic proposal for its production, as shown in Figure 4-15. For these mutants, it is likely that the change to a less bulky amino acid perturbs the active site geometry favoring alternative carbocation rearrangement to produce alternative terpene products.

4.4.3 Saturation Mutagenesis of Select Residues

To achieve a broader range of active-site perturbations, we performed saturation mutagenesis targeting select active-site residues of taxadiene synthase, which include the six polar residues as discussed above, as well as a number of additional residues in proximity to these polar residues or the cleaved PPi (S713, V714, G715, W753, V831, Y835). Since the taxadiene synthase crystal structure has an exposed active site due to N-terminal truncation, whereas the active form is likely capped by the N-terminus as in other terpene synthases [23], we selected 8 residues from the N-terminal region which we predicted to reside in proximity to the active site cavity (R84 to G91) for

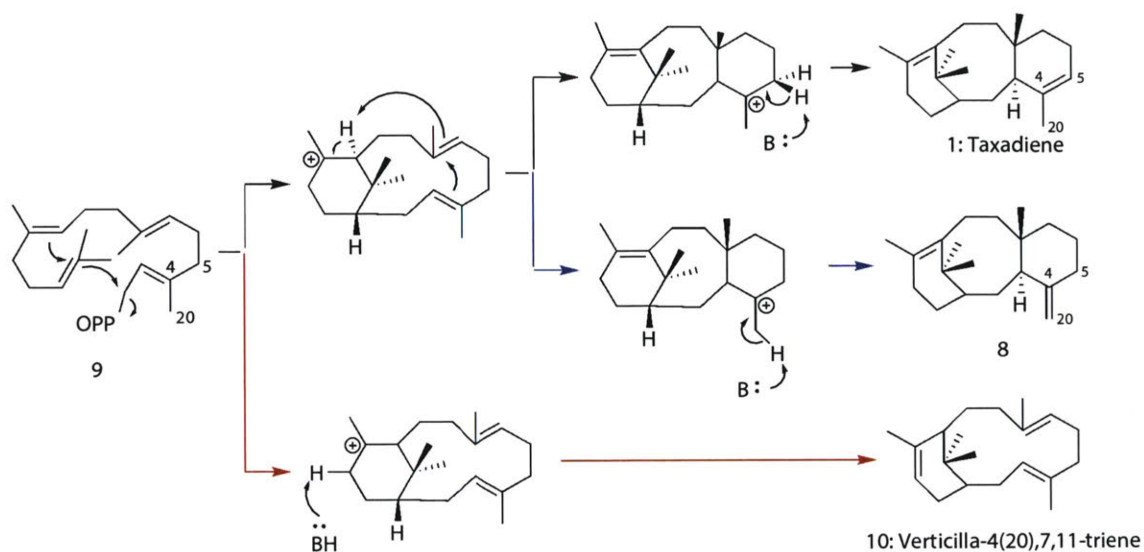

Figure 4-15: **Proposed Expanded Taxadiene Synthase Catalytic Cascade.** Black arrows are known functionality, blue arrows indicate proposed paths to iso-taxadiene, and red arrows indicate the pathways to the newly generated isomer.

mutagenesis as well. Enzymatic activity was then assayed using a whole-cell in-vivo system as above.

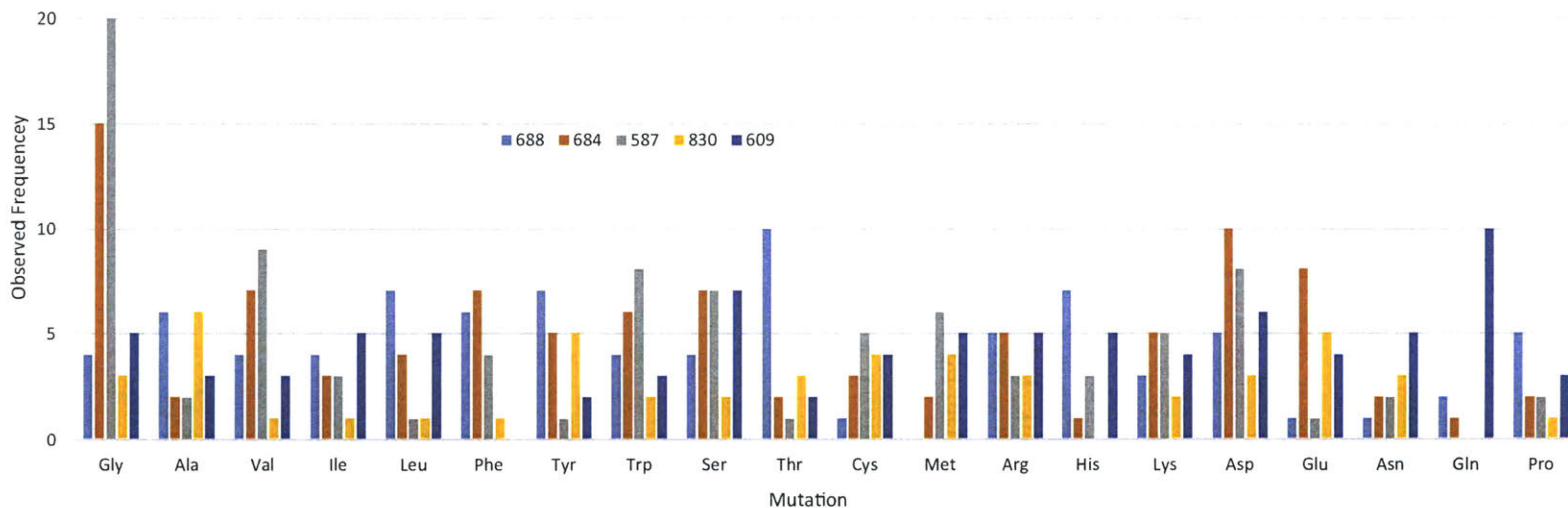

Figure 4-16: **Mutagenesis Coverage Observed From CLIVA Methodology.** Histogram depicting coverage of site-directed mutagenesis obtained by sequencing 100 colonies from mutagenesis libraries constructed using a 'small intelligent' approach and CLIVA. A total of 98 out of 100 desired mutants were obtained, with only two desired mutations (Q609F and Y688M) absent. This was interpreted as validation of our approach to mutagenesis approach and only five colonies were sequenced for subsequent target residues to ensure success of mutagenesis. Legend depicts residue numbers.

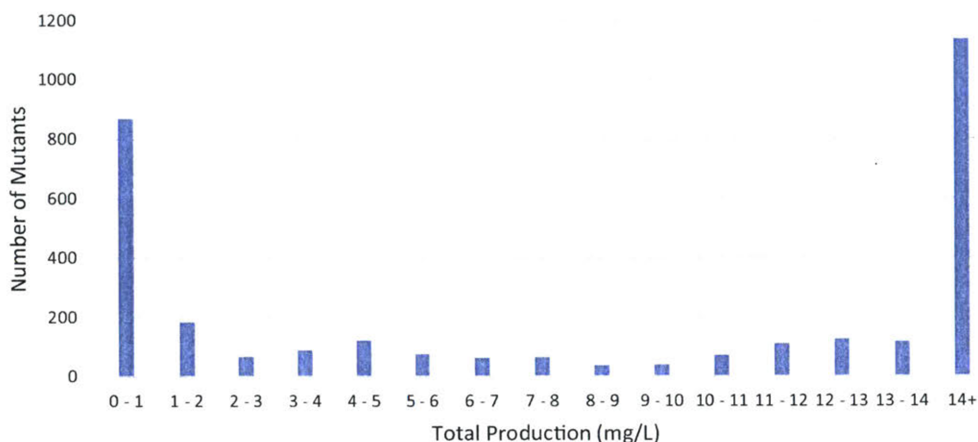

Figure 4-17: **Total Titer Observed from Taxadiene Synthase Mutants.** Histogram plot showing a bimodal pattern in which a majority of mutants either have no effect on total catalytic activity, or render inactive enzymes.

As shown in Figure 4-17, the selectivity and activity of the mutant library varied widely, typically consisting of a mixture of the three isomers previously obtained (**1, 8,** and **10**), with a majority of mutants clustered around the original selectivity with reduced activity. Although a large number of residues improved the apparent selectivity for **8**, this was achieved primarily through a reduction in the production of taxadiene with the titer of isotaxadiene remaining constant. Despite residue Y688 not appearing responsible for H5β abstraction from alanine-scanning, this residue appears to play an important role in selectivity, possibly due to its proximity to the PPi. The most successful mutant, Y688L, displayed markedly improved selectivity and yield of isotaxadiene (2.4-fold) with a concurrent decrease in production of taxadiene.

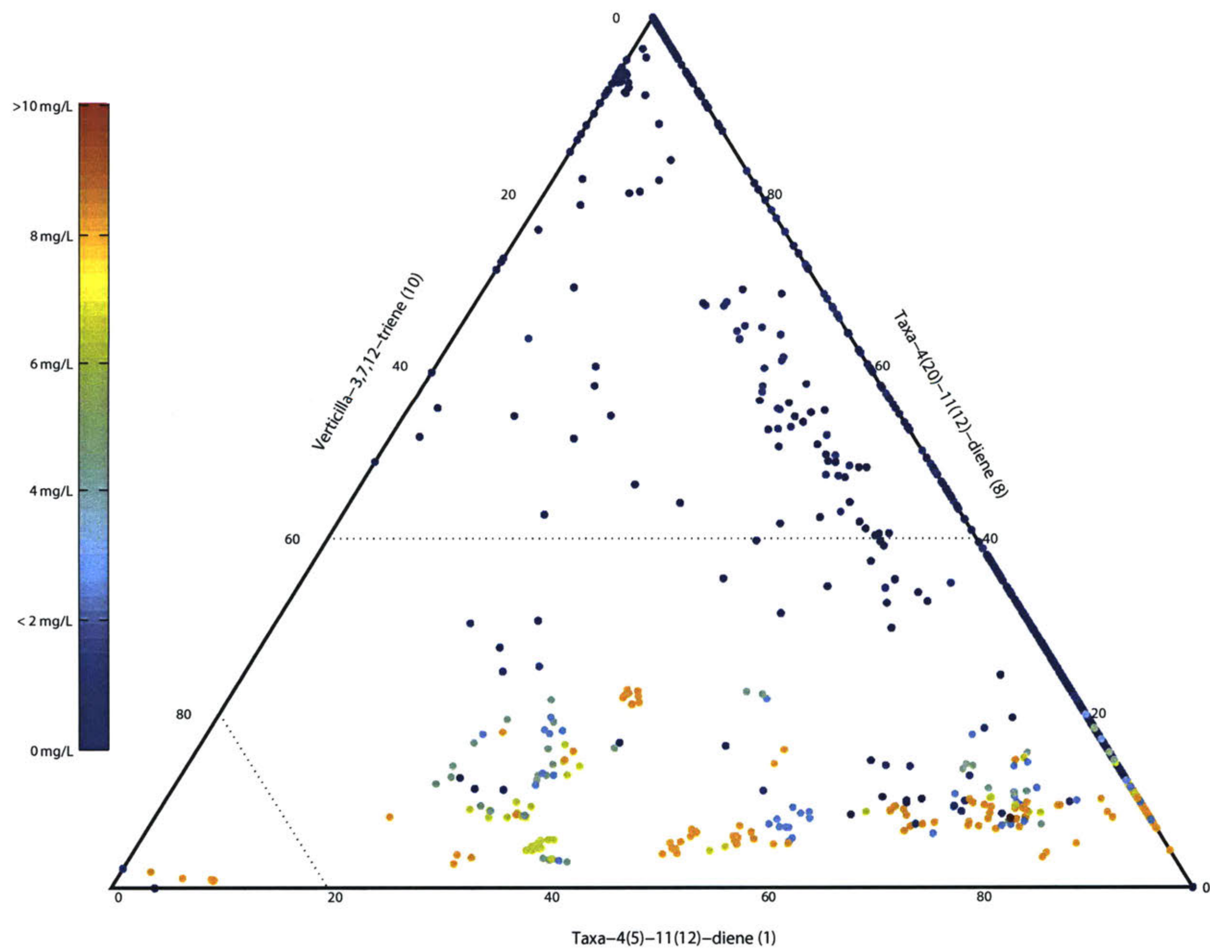

Figure 4-18: **Selectivity and Titer Observed From All Terpene Cyclase Mutants.**

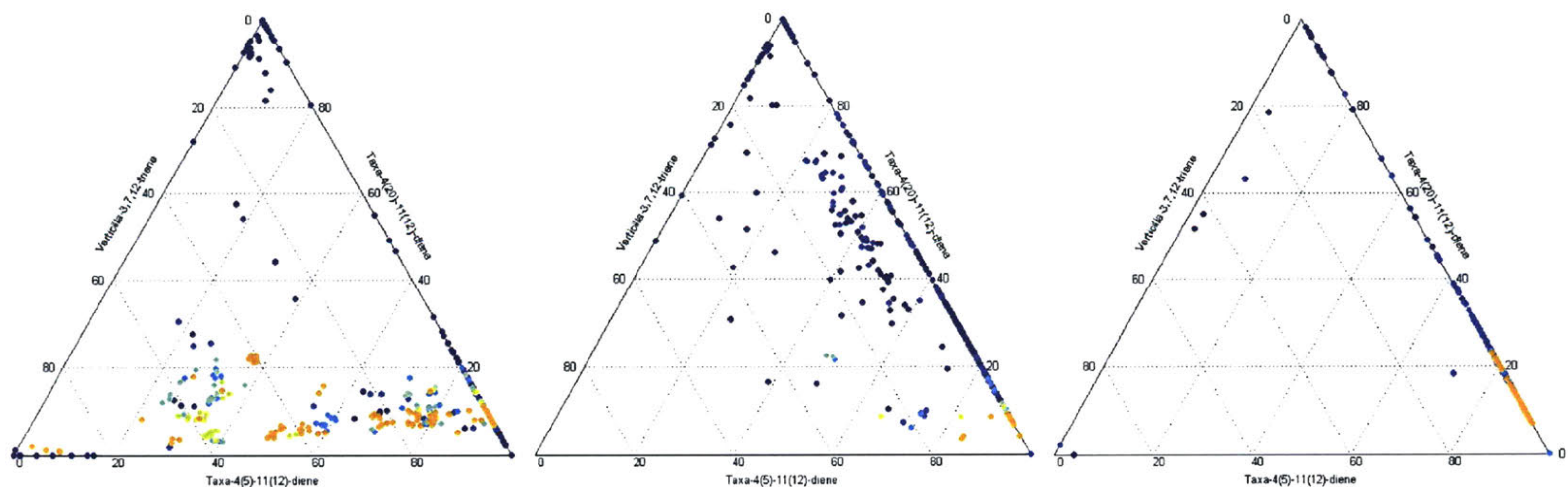

Figure 4-19: **Distribution Observed for Saturation Mutagenesis In Different Protein Regions**. Individual plots indicating the unique distribution of the three primary products obtained from mutagenesis of the three mutagenic groupings, active site polar residues (Left), additional active site residues (Center), and N-terminal residues (Right). Titer is indicated by color, as in Figure 4-18

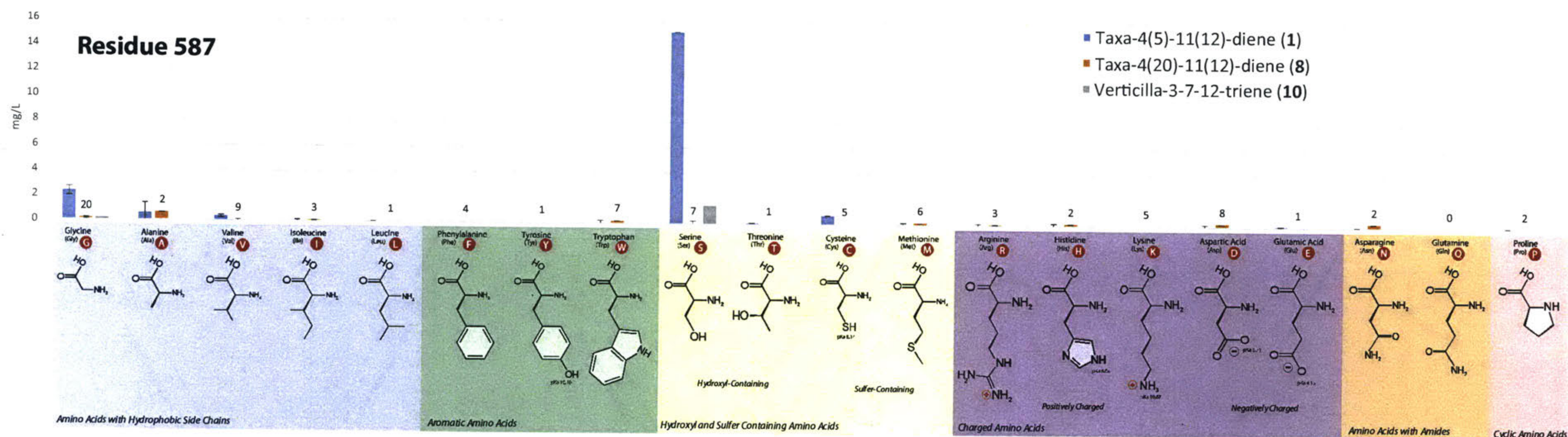

Figure 4-20: **Observed Selectivity for Individual Mutations at Residue S587.** Numbers above bar chart represent the number of mutants successfully obtained and screened for each residue.

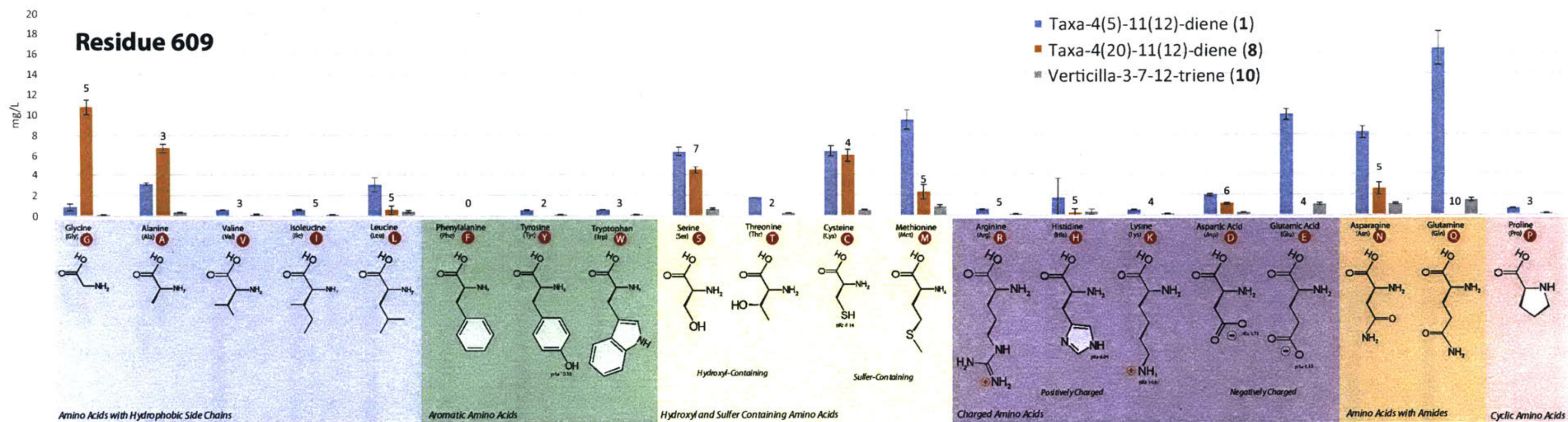

Figure 4-21: **Observed Selectivity for Individual Mutations at Residue Q609.** Numbers above bar chart represent the number of mutants successfully obtained and screened for each residue.

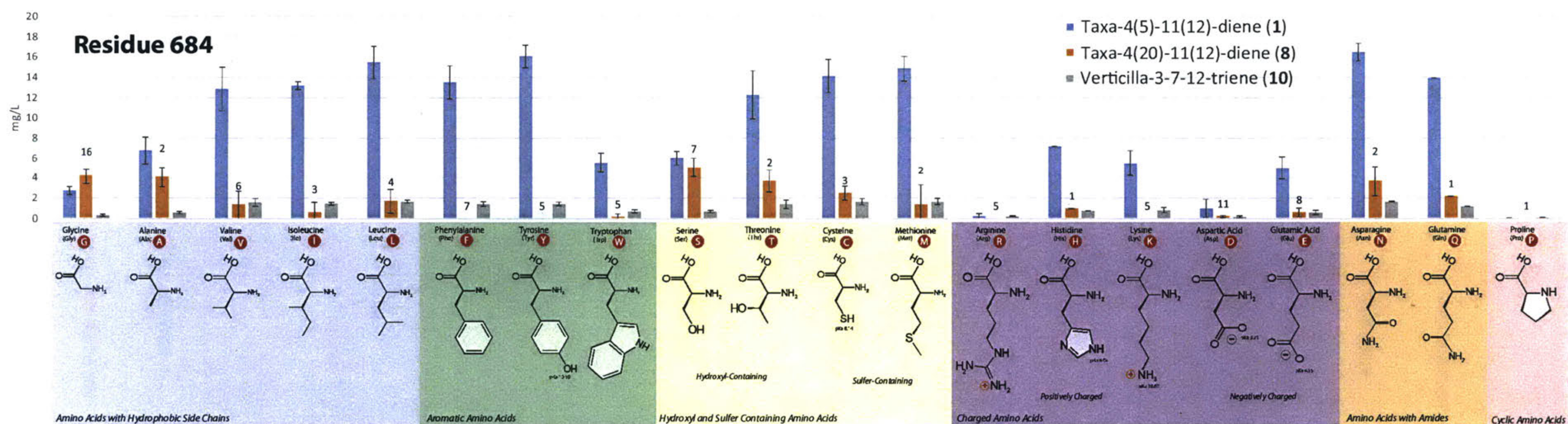

Figure 4-22: **Observed Selectivity for Individual Mutations at Residue Y684.** Numbers above bar chart represent the number of mutants successfully obtained and screened for each residue.

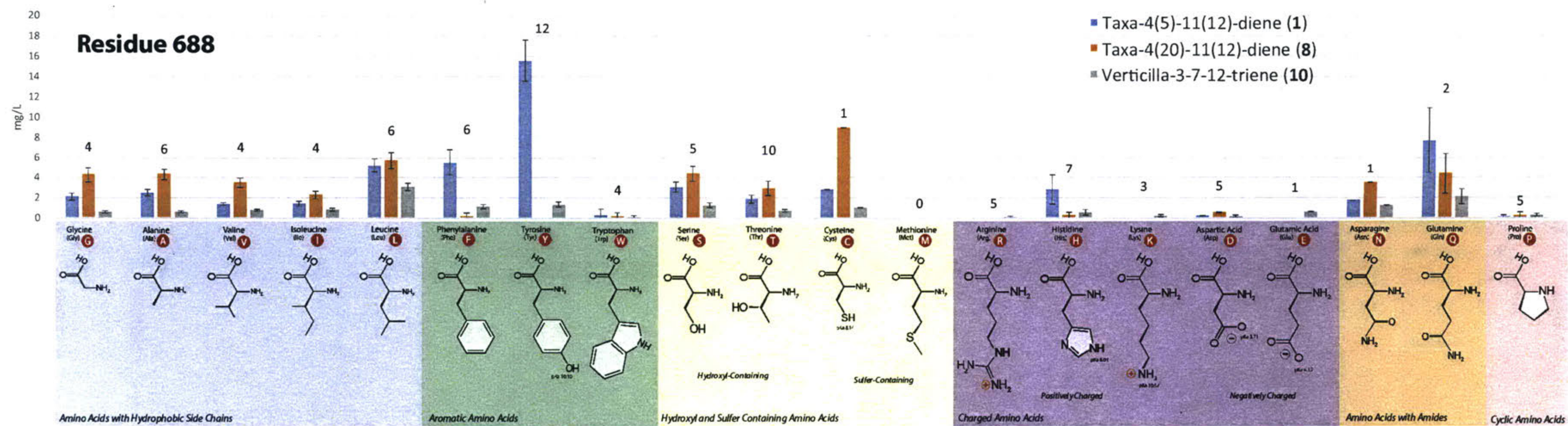

Figure 4-23: **Observed Selectivity for Individual Mutations at Residue Y688.** Numbers above bar chart represent the number of mutants successfully obtained and screened for each residue.

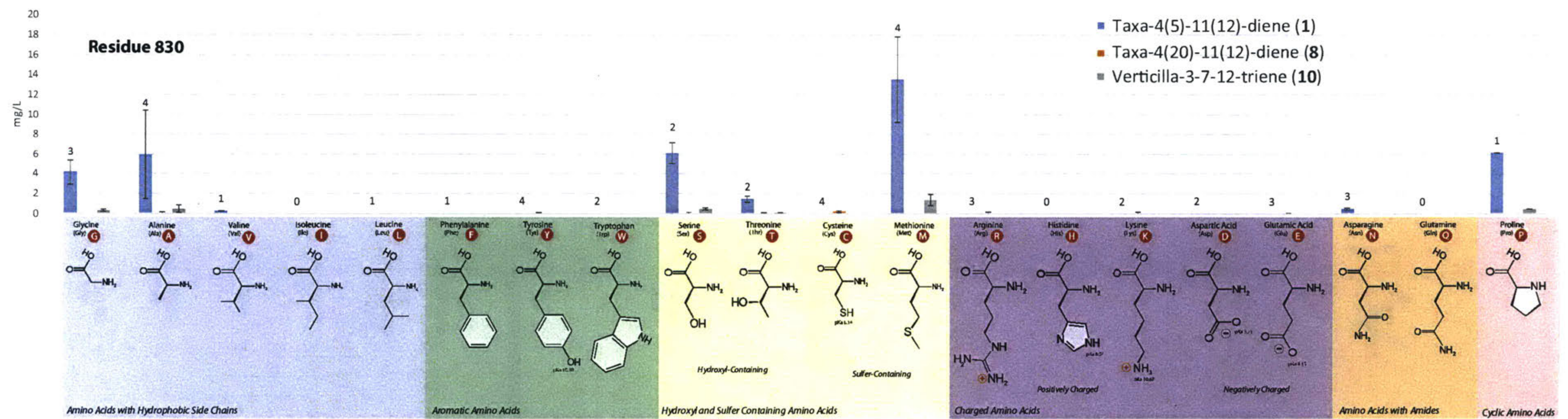

Figure 4-24: **Observed Selectivity for Individual Mutations at Residue M830.** Numbers above bar chart represent the number of mutants successfully obtained and screened for each residue.

Consistent with the alanine-scanning mutagenesis results, substitution of residue S587 with any other amino acid abolished production of the typically observed isomers (Figure 4-20). Additionally, mutant Q609G produced **10** as its sole product, yielding the first known terpene cyclase producing this product as its primary product. S587G also yielded a previously undetected product with a titer of less than 0.5 mg/L. Due to low titer and the fact that this was not a desired product, we did not pursue to resolve its structure. Furthermore, mutations in the N-terminal region primarily resulted in reduced enzymatic activity, and were less likely to generate **10** than core active site residues (Figure 4-19).

We next aimed to test whether taxadiene synthase mutants could be used to improve the production of the desired hydroxylated product, taxadien-5α-ol, when co-expressed with taxadiene-5α-hydroxylase. With this aim, we transformed four of the mutants obtained above (Y688F, Y688L, Y688Q and Q609G) into a strain containing a chromosomal copy of the upstream MEP pathway and CYP725A4 and its cytochrome P450 reductase partner [8]. As seen in Figures 4-25 and 4-26, taxadiene synthase mutants with improved selectivity for **8** displayed an increased production of **2** following co-expression with taxadiene-5α-hydroxylase due to the high selectivity observed for hydroxylation of **8**. The most successful taxadiene synthase mutant, Y688L yielded a 2.4-fold improvement in titer as predicted (Figure 4-26).

4.5 Conclusions

In summary, we investigated three possible approaches to improve the selectivity and throughput of early Taxol biosynthesis in *E. coli*. This was done by taking a holistic approach to selectivity improvement, focusing not only on the oxidation catalyzed by taxadiene-5α-hydroxylase, but also upstream and downstream of it. Our results demonstrate taxadiene synthase, acting upstream of the non-specific oxidation step, is a key enzyme in directing pathway flux toward Taxol precursors, such as taxadien-5α-ol (**2**). We further showed that engineering of taxadiene synthase is capable of improving the production of **2** through an unexpected metabolic route, whereas engi-

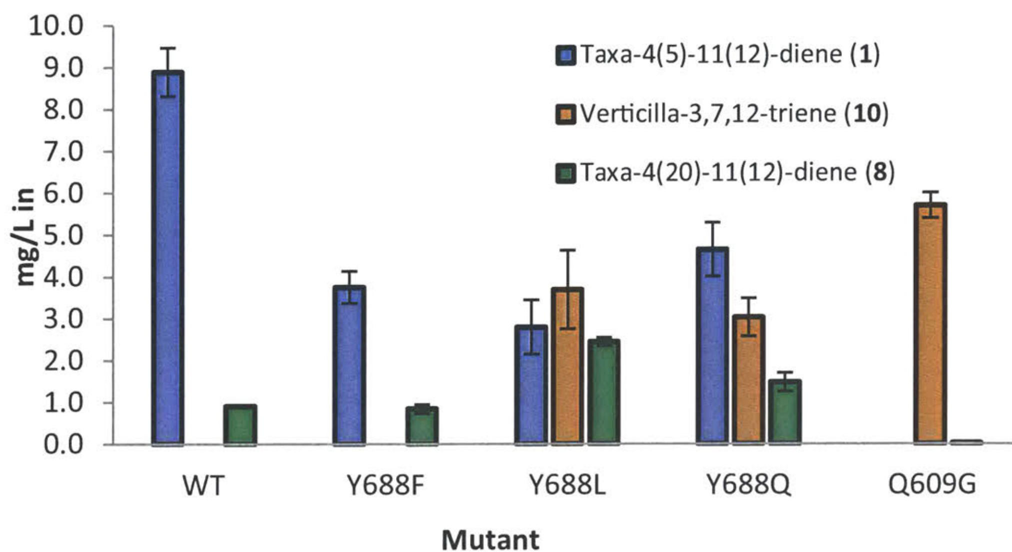

Figure 4-25: Cyclization Selectivity Observed From Select Mutants.

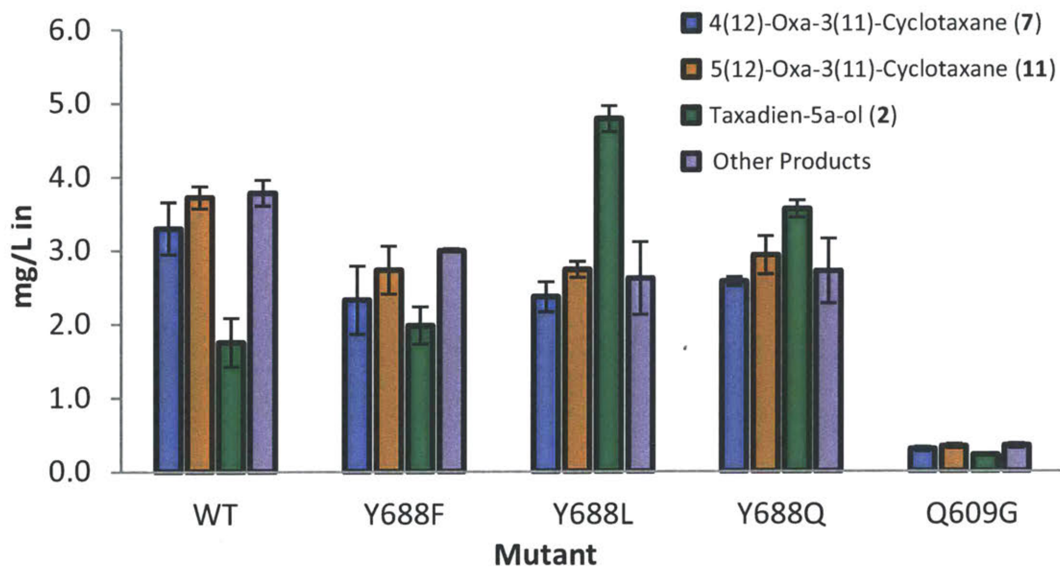

Figure 4-26: Oxidation Selectivity Observed From Select Mutants.

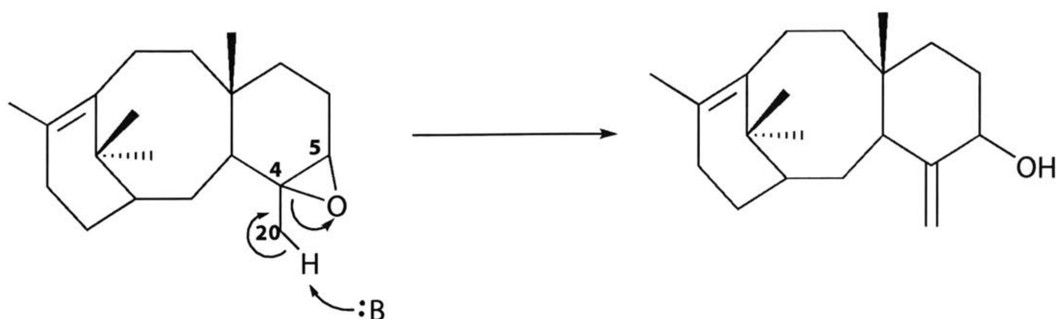

Figure 4-27: Acid-Base Catalysis of Epoxide Intermediate to Taxadien-5a-ol. Proposed non-enzymatic rearrangement of an epoxide intermediate is feasible via abstraction of H20 followed by epoxide ring opening. An SN2 reaction focusing on C4 may also act as a feasible route to taxadien-5a-ol.

neering of taxadiene-5α-hydroxylase and screening of hydrolase genes were ineffective at decreasing the production of undesirable byproducts. As only single-mutants were generated here, further engineering is likely to enable near complete shifts in selectivity for the desired product. Furthermore, application of the approaches utilized here to reduce the formation of byproducts and drive the carbon flux toward Taxol precursors will enable both heterologous production and pathway elucidation.

4.6 Materials and Methods

TXS Mutagenesis A combination of the 'small intelligent' approach and CLIVA was utilized in order to generate libraries saturating each selected residue. In brief, set of five primers consisting of one reverse primer, immediately upstream of the mutagenesis site, and four forward primers, each one containing one of the recommended degenerate primers (NDT, VMA, ATG, TGG), were designed and mixed as previously reported for PCR reactions. PCR products were gel purified, ligated using CLIVA [28], and transformed directly into the desired expression strain (TaxE5). Sequencing of five selected residues (S587, Q609, Y684, Y688, and C830) displayed that random selection of 100 colonies for each residue gave 98% coverage, thus sequencing was discontinued for remaining residues following initial verification, with 100 colonies screened for each residue. Taxadiene Synthase mutants were screened via fermenta-

tion in a previously constructed strain overexpressing the upstream MEP pathway for improved precursor supply denoted TaxE5. Docking of Taxadiene into the Active site cavity was performed using Autodock VINA [29]. Docking was performed using both A and B chains, as well as with and without the PP_i (cleaved from the substrate analogue) present.

Mutagenesis of Taxadiene-5α-hydroxylase. Homology models of taxadiene-5α-hydroxylase were generated using Phyre2 [21]. For modeling, taxadiene-5α-hydroxylase was truncated to the 'proline-hinge' region to remove any membrane binding domains. Comparison to those generated using MODELLER [30], showed similar conformations. The active site was determined via inspection and taxadiene was docked using AutoDock VINA to ensure adequate active site volume. Residues within the active site were then selected and compared to other known P450's in the downstream Taxol biosynthetic pathway via sequence alignment and residues were selected rationally for mutagenesis. Selection of downstream enzymes was done with the aim of expanding the active site of taxadiene-5α-hydroxylase to improve its ability to act on larger, more hydroxylated substrates. Mutagenesis of taxadiene-5α-hydroxylase was performed using the QuickChange II Site-Directed Mutagenesis kit (Agilent) and listed primers. Following PCR, products were DpnI digested and transformed into DH5α (NEB) for sequence confirmation followed by cloning into the expression strain (TaxE1) for screening.

Sequence Deduction and Screening of Degradative Enzyme Candidates. A previously published list of candidate degradative enzymes was generously provided by Dr. Qiu. Transcriptome information from Taxus tissues was also kindly supplied by Dr. Yang. Reciprocal blast of candidates to existing datasets, followed by alignment of matches enabled extension of most contigs to predicted full-length sequences. 3âĂŽ RACE was performed when necessary using established procedures [31], enabling sequence determination and cloning. Following alignment, duplicate sequences and sequences with no predicted degradative domain (predicted using PFAM) were removed and a set of 23 putative degradative enzymes were obtained.

RNA was then extracted from *Taxus cuspidata* plant cell culture and the 23 can-

didate genes candidate genes were cloned into the *E. coli* expression vector pET-28α via the Gibson assembly method. Full sequences of obtained clones are listed below in Appendix A. Constructs were then expressed in a TaxE12 background and screened for differences in observed selectivity using gas chromatography as described above.

Small scale fermentations. Small scale fermentations of taxadiene synthase mutants (with the TaxE5 background), taxadiene-5α-hydroxylase mutants (with the TaxE1 background), and degradative enzymes (in the TaxE12 background) were performed by inoculating a single colony of *E. coli* into LB media (10g/L tryptone, 5 g/L yeast extract, 10g/L NaCl, pH=7), followed by overnight incubation (37°,250 r.p.m.). 20μl of overnight culture were inoculated into 1 mL of small-scale culture media (5 g/L yeast extract, 10 g/L tryptone, 15g/L glycerol, 10g/L NaCl, 100 mM HEPES, pH 7.6), unless otherwise stated, in a 14 mL pyrex glass tube. Cultures were induced upon inoculation with 0.1 mM IPTG and shaken (22°/250) for 4 days prior to analysis of product distribution.

Quantification of Isoprenoids. Quantification of isoprenoids was performed as described in Chapter 3. Briefly, at the conclusion of fermentations, 200ÂţL of fermentation broth was sampled and added to 200μL ethyl acetate containing internal standard (caryophyllene; 0.90 mg/L) and 100μL of 0.1mm glass beads. This solution was then vortexed at 4°C for 30 minutes and centrifuged at 18,000g for 10 minutes. The top, organic, layer was extracted and 1μL was subjected to analysis via GCMS. The GCMS consisted of a Varian Saturn 3800 GC containing an HP-5ms (Agilent Technologies, USA) column, and Varian 2000 MS, with helium at a flow rate of 1mL/min as the carrier gas. The injector and transfer lines were maintained at 250°C. For oxygenated taxanes, the column oven was maintained at 100C for 1 minute and increased at a rate of 15°C/min to 175°C, then at 4°C/min to 220°C, then at 50°C/min to 290ÂřC, and held at this temperature for 1 minute. For analysis of taxadiene isomers the column oven was maintained at 100C for 1 minute and increased at a rate of 15ÂřC/min to 165°C, then at 2°C/min to 170°C, held at this temperature for 7 minutes, increased at 50°C/min to 290°C, and held at this temperature for 1 minute.

Large-Scale Fermentation for Production of Unknown Products. A 13-Liter Bioflow bioreactor (New Brunswick) was used for all semi-preparative scale production of desired compounds. Seed cultures, generated by inoculating a single colony into 100mL LB and shaking overnight (37°C,250 r.p.m.), were inoculated into 10 liters of fermentation media (5 g/L yeast extract, 20 g/L glucose, 13.3 g/L KH2PO4, 4 g/L (NH4)2HPO4, 1.7 g/L citric acid, 0.0084 g/L EDTA, 0.0025 g/L CoCl2, 0.015 g/L MnCl2, 0.0015 g/L CuCl2, 0.003 g/L H3BO3, 0.0025 g/L Na2MoO4, 0.008 g/L Zn(CH3COO)2), 0.06 g/L Fe(III) citrate, 0.0045 g/L thiamine, 1.3 g/L MgSO4, pH 7.0). Temperature was maintained at 22°C and pH was maintained at 7.0 via addition of NaOH. For characterization of the unknown cyclization product, cultures were performed under anaerobic conditions to reduce cell density, aiding in product recovery. For production of the unknown di-hydroxylated taxane, oxygen was supplied as sterile-filtered air at 1 L/minute with agitation adjusted to maintain DO at 10%. Induction with 0.1 mM IPTG was performed when OD reached 0.8. Taxanes were obtained from culture broth in a one-step extraction using a C18 chromatography resin and a 100% ethyl acetate for elution (100 mL). The effluent was then dried to completion, suspended in acetonitrile and used for HPLC purification.

Semi-preparative HPLC purification of Verticilla-3,7,12-triene was performed using a Supelco DiscoveryC18 (25cm x 10mm x 5μm) HPLC column under isocratic conditions (90% Acetonitrile in water, 6 mL/min) and had elution times of 19.1 and 20.2 minutes, respectively. Partially pure compound **11** was generated using a Supelco DiscoveryC18 (25cm x 10mm x 5μm) HPLC column using a ramp of 20-100% Acetonitrile in water over 30 minutes with a flow rate of 6 mL/min. A fraction ranging from 19 to 21 minutes was collected and then further chromatographed over Sephadex LH-20 eluting with CHCl3-MeOH (1:1) to yield 40 fractions. The 21st fraction was subjected to silica gel column eluting with Hexane-CHCl3 in gradient to afford compound 11 (Fracton 28 and 29).

4.7 Bibliography

[1] S. Edgar, F.-S. Li, K. Qiao, J.-K. Weng, and G. Stephanopoulos, "Engineering of taxadiene synthase for improved selectivity and yield of a key taxol biosynthetic intermediate," *ACS Synthetic Biology*, 2016.

[2] K. Nicolaou, Z. Yang, J. J. Liu, H. Ueno, P. Nantermet, R. Guy, C. Claiborne, J. Renaud, E. Couladouros, K. Paulvannan, *et al.*, "Total synthesis of taxol," *Nature*, vol. 367, no. 6464, pp. 630–634, 1994.

[3] N. C. Wilde, M. Isomura, A. Mendoza, and P. S. Baran, "Two-phase synthesis of (-)-taxuyunnanine d," *Journal of the American Chemical Society*, vol. 136, no. 13, pp. 4909–4912, 2014.

[4] P. K. Ajikumar, W.-H. Xiao, K. E. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, and G. Stephanopoulos, "Isoprenoid pathway optimization for taxol precursor overproduction in escherichia coli," *Science*, vol. 330, no. 6000, pp. 70–74, 2010.

[5] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos, "Distributing a metabolic pathway among a microbial consortium enhances production of natural products," *Nature biotechnology*, vol. 33, no. 4, pp. 377–383, 2015.

[6] D. Rontein, S. Onillon, G. Herbette, A. Lesot, D. Werck-Reichhart, C. Sallaud, and A. Tissier, "Cyp725a4 from yew catalyzes complex structural rearrangement of taxa-4 (5), 11 (12)-diene into the cyclic ether 5 (12)-oxa-3 (11)-cyclotaxane," *Journal of Biological Chemistry*, vol. 283, no. 10, pp. 6067–6075, 2008.

[7] N. A. Barton, B. J. Marsh, W. Lewis, N. Narraidoo, G. B. Seymour, R. Fray, and C. J. Hayes, "Accessing low-oxidation state taxanes: is taxadiene-4 (5)-epoxide on the taxol biosynthetic pathway?," *Chemical Science*, vol. 7, no. 5, pp. 3102–3107, 2016.

[8] S. Edgar, K. Zhou, K. Qiao, J. R. King, J. H. Simpson, and G. Stephanopoulos, "Mechanistic insights into taxadiene epoxidation by taxadiene-5α-hydroxylase," *ACS chemical biology*, vol. 11, no. 2, pp. 460–469, 2015.

[9] S. Jennewein, R. M. Long, R. M. Williams, and R. Croteau, "Cytochrome p450 taxadiene 5α-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis," *Chemistry & biology*, vol. 11, no. 3, pp. 379–387, 2004.

[10] V. G. Yadav, "Unraveling the multispecificity and catalytic promiscuity of taxadiene monooxygenase," *Journal of Molecular Catalysis B: Enzymatic*, vol. 110, pp. 154–164, 2014.

[11] B. W. Biggs, J. E. Rouck, A. Kambalyal, W. Arnold, C. G. Lim, M. De Mey, M. OâĂŹNeil-Johnson, C. M. Starks, A. Das, and P. K. Ajikumar, "Orthogonal assays clarify the oxidative biochemistry of taxol p450 cyp725a4," *ACS chemical biology*, vol. 11, no. 5, pp. 1445–1451, 2016.

[12] V. G. Yadav, "Unraveling the multispecificity and catalytic promiscuity of taxadiene monooxygenase," *Journal of Molecular Catalysis B: Enzymatic*, vol. 110, pp. 154–164, 2014.

[13] V. G. Yadav, "Combining metabolic pathway design and retrosynthetic planning for the design of a novel semisynthetic manufacturing scheme for paclitaxel," *Organic Process Research & Development*, vol. 18, no. 6, pp. 816–826, 2014.

[14] K. H. Hopmann, B. M. Hallberg, and F. Himo, "Catalytic mechanism of limonene epoxide hydrolase, a theoretical study," *Journal of the American Chemical Society*, vol. 127, no. 41, pp. 14339–14347, 2005.

[15] D. Brodkorb, M. Gottschall, R. Marmulla, F. Lüddeke, and J. Harder, "Linalool dehydratase-isomerase, a bifunctional enzyme in the anaerobic degradation of monoterpenes," *Journal of Biological Chemistry*, vol. 285, no. 40, pp. 30436–30442, 2010.

[16] E.-K. Lee, Y.-W. Jin, J. H. Park, Y. M. Yoo, S. M. Hong, R. Amir, Z. Yan, E. Kwon, A. Elfick, S. Tomlinson, *et al.*, "Cultured cambial meristematic cells as a source of plant natural products," *Nature biotechnology*, vol. 28, no. 11, pp. 1213–1217, 2010.

[17] G. Sun, Y. Yang, F. Xie, J.-F. Wen, J. Wu, I. W. Wilson, Q. Tang, H. Liu, and D. Qiu, "Deep sequencing reveals transcriptome re-programming of taxus× media cells to the elicitation with methyl jasmonate," *PloS one*, vol. 8, no. 4, p. e62865, 2013.

[18] G. Ge, P. Xiao, Y. Zhang, L. Yang, *et al.*, "The first insight into the tissue specific taxus transcriptome via illumina second generation sequencing," *PLoS one*, vol. 6, no. 6, p. e21220, 2011.

[19] D. Kumar, B. Karamzadeh, G. N. Sastry, and S. P. de Visser, "What factors influence the rate constant of substrate epoxidation by compound i of cytochrome p450 and analogous iron (iv)-oxo oxidants?," *Journal of the American Chemical Society*, vol. 132, no. 22, pp. 7656–7667, 2010.

[20] S.-t. Li, P. Zhang, M. Zhang, C.-h. Fu, C.-f. Zhao, Y.-s. Dong, A.-y. Guo, and L.-j. Yu, "Transcriptional profile of taxus chinensis cells in response to methyl jasmonate," *BMC genomics*, vol. 13, no. 1, p. 295, 2012.

[21] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, "The phyre2 web portal for protein modeling, prediction and analysis," *Nature protocols*, vol. 10, no. 6, pp. 845–858, 2015.

[22] A. E. Koepp, M. Hezari, J. Zajicek, B. S. Vogel, R. E. LaFever, N. G. Lewis, and R. Croteau, "Cyclization of geranylgeranyl diphosphate to taxa-4 (), 11 ()-diene is the committed step of taxol biosynthesis in pacific yew," *Journal of Biological Chemistry*, vol. 270, no. 15, pp. 8686–8690, 1995.

[23] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, and D. W. Christianson, "Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis," *Nature*, vol. 469, no. 7328, pp. 116–120, 2011.

[24] X. Lin, M. Hezari, A. E. Koepp, H. G. Floss, and R. Croteau, "Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of taxol biosynthesis in pacific yew," *Biochemistry*, vol. 35, no. 9, pp. 2968–2977, 1996.

[25] Q. Jin, D. C. Williams, M. Hezari, R. Croteau, and R. M. Coates, "Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labeling," *The Journal of organic chemistry*, vol. 70, no. 12, pp. 4667–4675, 2005.

[26] H. Itokawa, "3 taxoids occurring in the genus taxus," *Taxus: the genus Taxus*, p. 35, 2003.

[27] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, and D. W. Christianson, "Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis," *Nature*, vol. 469, no. 7328, pp. 116–120, 2011.

[28] R. Zou, K. Zhou, G. Stephanopoulos, and H. P. Too, "Combinatorial engineering of 1-deoxy-d-xylulose 5-phosphate pathway using cross-lapping in vitro assembly (cliva) method," *PloS one*, vol. 8, no. 11, p. e79557, 2013.

[29] O. Trott and A. J. Olson, "Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of computational chemistry*, vol. 31, no. 2, pp. 455–461, 2010.

[30] B. Webb and A. Sali, "Comparative protein structure modeling using modeller," *Current protocols in bioinformatics*, pp. 5–6, 2014.

[31] E. Scotto-Lavino, G. Du, and M. A. Frohman, "5âĂš end cdna amplification using classic race," *Nature protocols*, vol. 1, no. 6, pp. 2555–2562, 2006.

Chapter 5

Developing An Alternative Route to Isoprenoids

5.1 Introduction

Two separate biosynthetic routes are known to lead to the universal precursors to all isoprenoids, isopentenyl pyrophosphate and dimethylallyl pyrophosphate (IPP and DMAPP). These pathways (depicted in Figure 5-1 below) are the recently-elucidated methylerythritol phosphate (MEP) pathway, found largely in prokaryotes, apicomplexan protozoa, and in plant plastids; and the 'classical' mevalonate (MVA) pathway, which is universal to all higher eukaryotes as well as a majority of archea. Although both pathways have been applied within the field of metabolic engineering, each has unique characteristics which must be considered, as they influence their applicability to large-scale, commercial production.

The MEP pathway has the higher theoretical yield of the two pathways, largely as it derives from glyceraldehyde-3-phosphate and pyruvate, prior to carbon loss during tranformation from pyruvate to acetyl-CoA. Another advantage of this pathway is its redox-balanced nature in which no excess reducing equivalents are generated, thus oxygen requirements are lower and a fermentative route to these products is feasible (but has yet to be fully realized). However, this adaptation to a low-oxygen environment appears to extend beyond redox balance, as several pathway proteins are

maladapted to an oxygen-rich environment, such as the iron-sulfer cluster proteins ispG and ispH. As a result, buildup of metabolites immediately preceding these enzymes occurs in an oxygen-rich environment, preventing high pathway fluxes [1]. A deeper analysis of pathway yield under various environments is performed in section 5.2.2, below.

Alternatively, the MVA pathway is more commonly used industrially. Despite having lower theoretical yields, this pathway is not inhibited by oxygen in the same manner and has required less intensive optimization to achieve high titers. Additionally, this pathway is native to *Saccharomyces cerevisiae*, enabling production of food and pharmaceutical products, as this organism is classified as 'Generally Regarded As Safe' (GRAS). However, this pathway is also not redox balanced and generates excess reducing equivilants which must then be regenerated, hindering yield at oxygen-limited conditions. Furthermore, although high titers as high as 40 g/L of the sesquiterpene amorphadiene have been reported, reported yields have failed to exceed 1%, with the remainder of the carbon going to biomass [2].

5.2 Evaluating Alternative Pathways

Due to the unique limitations of each of these pathways, we began evaluating alternative routes to isoprenoids by enumerating possible alternative entry points to isoprenoid biosynthesis. As shown in Figure 5-1, the traditional pathways can be further decorated via the inclusion of two modifications to the MVA pathway which rely on alternative phosphoylation patterns (lower right). Several entrypoints to the MEP pathway are also feasible including through the methionine salvage pathways (passing though 5'-methylthioadenosine), deoxyxylulose using promisious enzymatic activity [3, 4], or a recently proposed pathway through ribulose 5-phosphate or xylulose 5-phosphate [5].

Although pathways circumventing deoxyxululose synthase (DXS) are attractive, downstream steps may still act as bottlenecks and flux must still proceed through the IspG and IspH iron-sulfer cluster proteins [1, 6]. Usage of modified archeal MVA

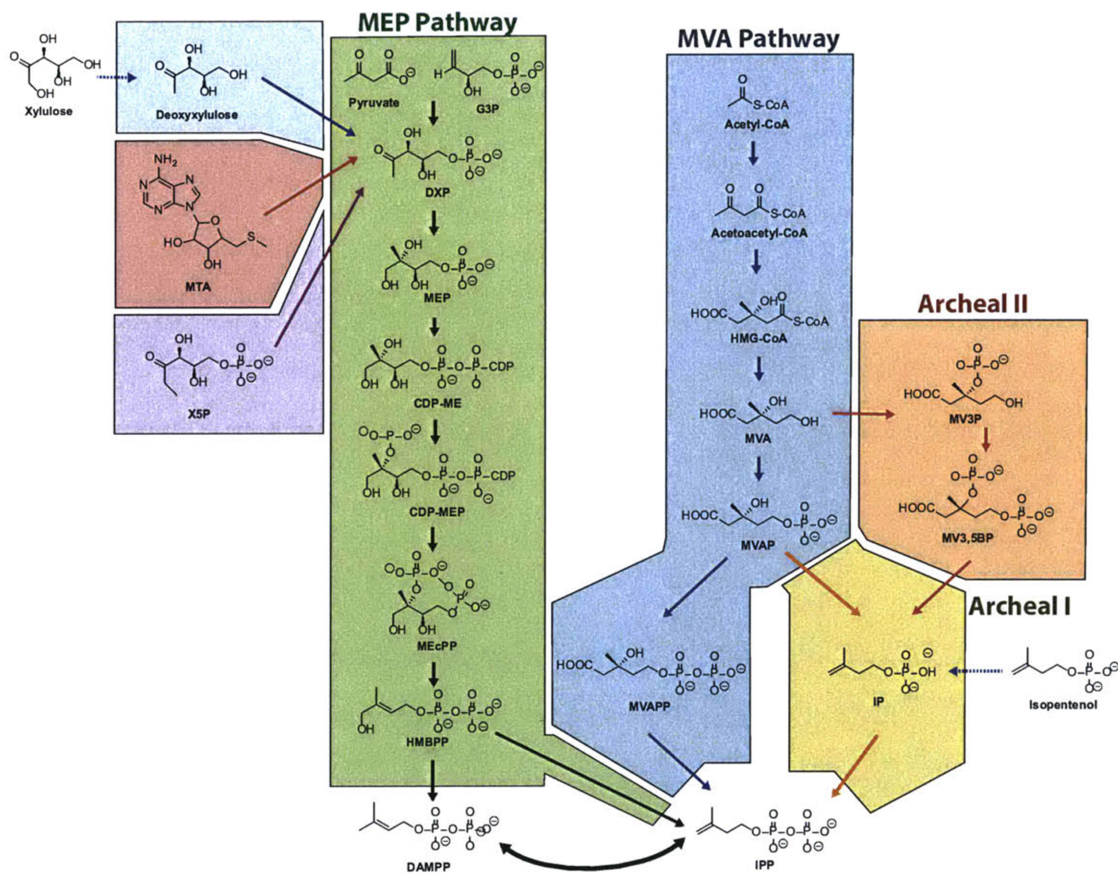

Figure 5-1: **Depiction of Pathways to IPP and DMAPP.** These pathways are depicted as they would be present in an engineering or wild-type *E. coli* cell. Within *S cerevisiae* the production of acetyl-CoA requires two additional ATP equivalents through what is known as the pyruvate-dehydrogeanse (PDH) bypass, this was not considered in order to conservatively estimate favorability of alternative pathways

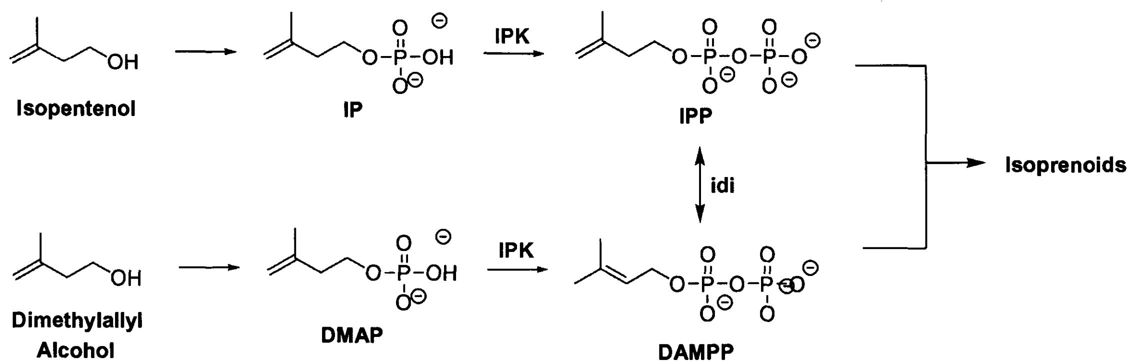

Figure 5-2: **Proposed Isopentenol Utilization Pathway** IPK catalyzes the reaction of IP or DMAP to IPP, while an undiscovered enzyme would catalyze the transformation of isopentenol or dimethylallyl alcohol to IP and DMAP, repsectively. Native idi activity results in isomerization and generation of both precursors for isoprenoid biosynthesis.

pathways also fail to yield improvements as these operate downstream of steps which result in carbon loss. One clever approach to overcome this has been implemented industrially by Amyris [7], in which phosphoketolase (PK) and phosophotransacetylase (PTA) are used to convert fructose-6-phosphate (F6P) or xyluose-5-phosphate (X5P) to acetyl-CoA to a theoretical 28.9% mass yield [7], although obtained yields were unreported. This modification decreases the excess reducing equivalents and utilizes the excess ATP generated with lower overall carbon loss. However, this increase in titer is still modest and other approaches may still be explored.

5.2.1 The Isopentenol Utilization Pathway

As seen in Figure 5-1, an alternative biocatalytic route beginning from bulk chemical isopentenol and proceeding through two phosphorylation reactions and an isomerization to IPP and DMAPP may be envisioned (Figure 5-2). This pathway is entirely orthogonal to the MEP and MVA pathways, using only a single gene from the archeal pathway, isopentenol phosphate kinase (IPK). As such, it overcomes the oxygen-sensitivity of the MEP pathway, and the oxygen-dependence of the MVA pathway, yielding a pathway which is oxygen agnostic. It is of value to note that the yield of ATP (which is required by the pathway) from sugar feedstocks is obviously higher for

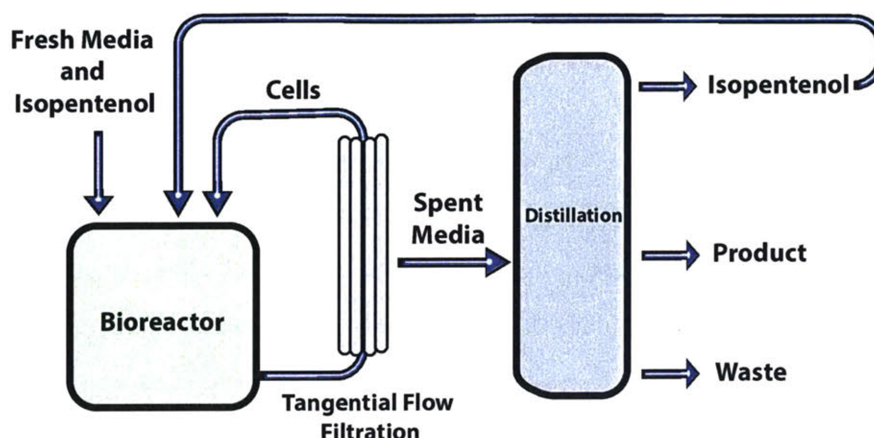

Figure 5-3: **Proposed Industrial Application of The Biocatalytic Process.** A large-scale high-density perfusion reaction system in which unused substrate (isopentenol) is recovered during product recovery and recycled is likely to yield highly economic processes.

a respiring cell, and thus oxygen availability influences yield through this mechanism.

It is important to make the distinction that this pathway is biocatalytic rather than fully biosynthetic. As a result, it offers unique advantages. In a traditional biosynthetic pathway, such as the MVA pathway, feedstock carbon not incorporated into the product or released as CO_2 is converted to biomass and cannot be recovered, lowering yield. In a biocatalytic process, such as the isopentenol utilization (IU) pathway, unconverted substrate can be recovered at the end of fermentation. For the isopentenol utilization (IU) pathway this is even more straightforward as products may require distillation for purification, and recovery of isopentenol would only require additional distillation column design. As a result, this process could easily be designed as a continuous or a perfusion-based system in which only isopentenol and minimal growth substrate are supplied, and products are removed continuously. This would enable high cell densities and decrease reactor size requirements and thus capital cost. A proposed set-up for this system is shown below in Figure 5-3.

5.2.2 Pathway Feasibility and Economic Analysis

To determine the practicality of such a pathway as an industrial tool, several parameters must be evaluated. Key among these parameters are yield, productivity, titer, scalability, and strain stability. We will discuss each of these and how the IU pathway excels in each category over traditional manufacturing using the MEP or MVA pathways. First, strain stability is an important characteristic as it has a large influence on process design parameters. A highly stable strain can be run in large reactors which require large numbers of generations from seed cultures, and can be run for long time periods, or even continuously. An unstable strain may produce quality results on bench-scale experiments, but fail to perform when scaled-up to industrial levels. Stability can be seen qualitatively to be lower for the total biosynthetic process. To understand why this is, we can enumerate the key burdens on a cell within a traditional bioprocess, these are:

1. Loss of Carbon from Biomass Production to Product Formation
2. ATP Cost
3. Redox Cofactor Requirements (Excess or Deficit)
4. Protein Expression Burden
5. Product Toxicity

Increasing cellular burden increases the probability of 'escape' to a non-production phenotype. With the MEP and MVA pathways, carbon flux not aimed toward product formation is lost to biomass production. However, with the IU pathway, product and biomass production are decoupled and thus this burden does not exist. Furthermore, although the IU pathway requires two ATP per IPP produced, the ATP requirement of the MEP and MVA pathways remains higher. Product toxicity is also reduced as the substrate itself is toxic and thus the differential is lower in the IU pathway, reducing the fitness differential between producers and non-producers. Toxicity of the substrate may be alleviated via process optimization and feeding. Thus,

the primary burden a cell experiences with the IU pathway is the protein expression burden from pathway expression. This burden may be minor relative to the other burdens discussed, and is also present in the MVA and MEP pathway expression strains.

To further compare, we can imagine a scenario in which a single nucleotide mutation occurs, rendering the first pathway step in each pathway non-functional. For a strain producing terpenoids via the MVA pathway, this mutated cell has highly increased fitness as the loss of carbon to production (a large burden) is now redirected to biomass, allowing this cell to outcompete producing cells. Furthermore, product toxicity, ATP cost, and redox factor requirements are now alleviated. In contrast, for the IU pathway, in a non-production mutant no carbon is redirected, product toxicity is not reduced (as substrate is of approximately equal toxicity), and only the minor ATP burden is alleviated, giving non-producing cells only a minor improvement in fitness, increasing the time before a non-producing population becomes large enough decrease bioreacor productivity

We can next evaluate pathway yield. This can be done first from a theoretical standpoint. As a point of reference for yield, we can compare the hypothetical IU pathway to the two natural pathways, the MEP and MVA pathways. We can begin by examining the molar equivalents of glucose required to generate one mole of IPP. This extends beyond loss to carbon dioxide and also includes cofactor requirements (shown in Figure 5-4). These cofactor requirements are of high importance when considering industrial fermentation as pathways which generate an of excess reducing equivalents will have higher oxygen requirements and thus impart design limitations on the finalized bioprocess.

From inspection, and as has been discussed in previous literature [8, 9, 10], the feedstock requirements of both glucose and isopentenol for the pathways are displayed in Table 5.1. The glucose requirement for the IU pathway arises as 2 ATP are required for biotransformation which are derived from glucose consumption, although these pathways are not directly coupled. From these results, we can now perform an economic analysis. At theoretical maximal yield we can then calculate a cost ratio

(isopentenol to glucose) needed for the IU pathway to be economically favorable over the MEP or MVA pathway. Based on the molar masses of the feedstocks and pathway substrate requirements, we obtain the values shown in Table 5.2,

		MVA	**MEP**	**IPP**
Aerobic	**Glucose**	1.48	1.24	0.07
	Isopentenol	-	-	1
Anaerobic	**Glucose**	10	1.74	1
	Isopentenol	-	-	1

Table 5.1: Feedstock Requirements Needed to Generate One Mole of IPP

	MVA	MEP
Aerobic	2.95	2.45
Anaerobic	18.84	1.55

Table 5.2: Maximum cost ratio ($ per kg isopentenol / $ per kg glucose) to allowed for industrial feasibility at theoretical yield. This is calculated by assuming a cost of 1 unit cost per kg of glucose.

From this economic analysis, it can be seen that at theoretical pathway yields, the minimum glucose to isopentenol cost ratio (per kg) for favorability over the widely used MVA pathway is approximately 3 under optimial conditions. However, typical cost ratios for these two commodity chemicals typically ranges from 4 to 6. This makes the proposed pathway slightly economically disadvantageous at theoretical pathway yields. However, due to the low practical yields from the MVA pathway and the ability to recover unused isopentenol, the economics are likely to shift in favor of our biocatalytic pathway over the traditional MVA pathway. This can be seen as the highest reported MVA pathway yield is 0.01% in literature [2], though it is likely that commercial producers have likely increased this significantly. Yet, yields of less than 50% of the theoretical maximum for the MVA pathway shift the economics in favor of the IU pathway.

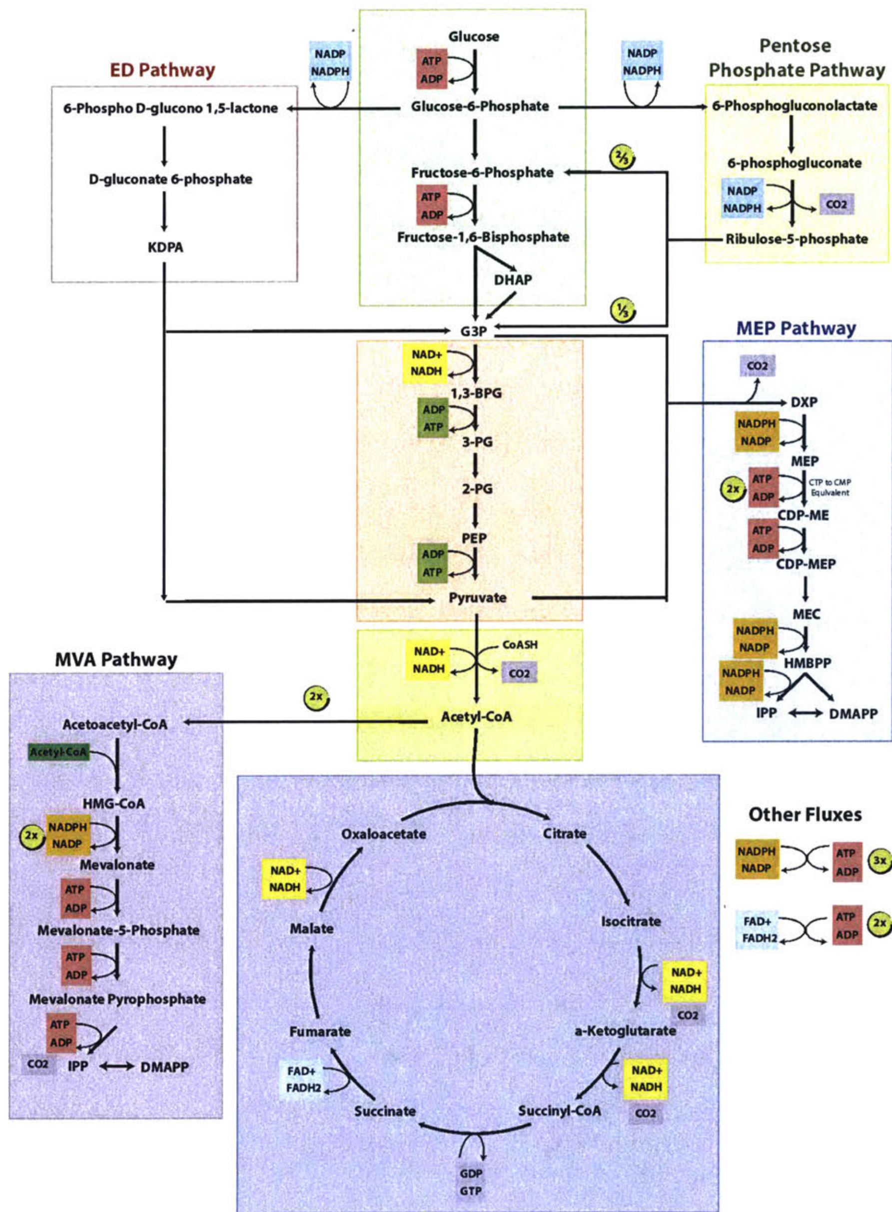

Figure 5-4: **Cofactor Requirements for MEV and MVA Pathways** Pathway modules are also depicted using colored boxes

5.3 Construction of an Isopentenol-based pathway

As shown in Figure 5-2, the IU pathway consists of three enzymatic steps, two phosphorylations and an isomerization. The final step, catalyzing the isomerization of IPP and DMAPP, is performed by *E. coli* natively, encoded by the gene *idi.* Additionally, the isopentenyl phosphate kinase (IPK) gene required for the second phosphorylation is present within many archea as it is an essential part of their MVA pathway (Figure 5-1), and has been found in several domains of life [11, 12, 13].

However, at the beginning of experiments, no enzyme catalyzing the first pathway step was known to exist and thus we set out to screen candidates. Following our initial work, another group published reports that some IPKs demonstrate promiscuous activity, and are capable of biotransformation of dimethylallyl alcohol to DMAPP [14]. This data was generated using ATP to ADP turnover, observed using a spectrophotomter and a coupled enzyme assay. However, during our own experiments, we observed (via LCMS) that the proposed enzyme catalyzed a degree of non-specific ATP hydrolysis without concomitant production of IP or IPP, which raises questions about the validity of using the coupled assay alone to obtain accurate activity measurements. Furthermore, the activities reported from an engineered enzyme, which we estimate would catalyze the production of nearly 1.5 grams of lycopene per gram of dry cell weight (DCW) per day (discussed below in Section 5.3.3), resulted in an actualized yield which was was less than one thousandth of that (at approximately 1 mg/L) in their cultures. Furthermore, statistical analysis given of the data display remarkably low standard deviations reported for yields and dry cell weights, indicative of systematic error.

However, it is important to note that although this previously published work is not complete in terms of generating a highly-functional pathway, a brief analysis of their work is still informative, especially of the *in-vitro* system. We may first compare this result (reported in mU/mg) to previous literature values of K_{cat} (with units of 1/s); to do this, we must perform a quick conversion. If we note that reported IPK molecular masses are in the range of 26 to 28 kDa, we can take an average value of

27 kDa for a quick approximation, yielding:

$$\frac{1\ mUnit}{mg\ protein} = \left(\frac{\frac{1}{1000}\ uMole\ Substrate}{min\ mg\ protein}\right) * \left(\frac{1\ min}{60\ s}\right) * \left(\frac{27{,}000\ mg}{mmole}\right) * \left(\frac{1\ mMole}{1000uMole}\right) \tag{5.1}$$

$$\frac{1\ mUnit}{mg\ protein} = \frac{4.5 * 10^{-4}\ uMole\ Substrate}{s\ *\ uMole\ protein} = 4.5 * 10^{-4}\left(\frac{1}{s}\right) = K_{cat} \tag{5.2}$$

We are now also able to compare this result to other literature. IPK from *Thermoplasma acidophilum* has a previously reported activity of 30 mU/mg protein on dimethylallyl alcohol, and thus a Kcat value of approximately 0.0135s^{-1}, compared to a value of 8.0s^{-1} for IP, (approximately 600-fold decreased) [15]. This degree of reduction of activity may be realistic to expect. Many characteristics of the IPK active site are specifically tuned to act upon phosphate moieties [16], and may thus be unlikely to act with this degree of activity on primary alcohols. Furthermore, for an enzyme to have high activity, it must typically also be highly specific for a single reaction type as the transition state of the reaction must be stabilized. Using a single enzyme for both phosphorylation reactions (one of an alcohol and one of a phosphoryl) appears to be unwise for this reason. We thus set out to characterize a number of genes and identify ones which enable efficient pathway construction.

5.3.1 Selection of Candidates

We first aimed to identify candidate enzymes catalyzing the phosphorylation of isopentenol and performed *in-vitro* screening. These candidates fell into two categories, archeal IPK enzymes, as utilized previously [14], and alcohol kinases acting on primary alcohol substrates with similar structures to isopentenol. This second group would rely on non-specific binding of the molecule only within the backbone rather than the portion which is acted upon. For this reason, we propose this set of enzymes are more likely to display enzymatic activity as the chemical transformation

and environment for catalysis remains the same. The genes we selected for screening include:

1. IPK from *T. acidophilum*
2. IPK from *M. jannaschii*
3. IPK from *M. thermautotrophicus*
4. IPK from *M. thermautotrophicus* (Triple mutant from [14])
5. IPK from *H. volcanii*
6. Mevalonate Kinase from *S. cerevisiae*
7. Glycerol Kinase from *E. coli*
8. Homoserine Kinase from *E. coli*
9. Choline Kinase from *S. cerevisiae*

5.3.2 Detection of IP

To ensure accuracy of our results, we selected initially for an LCMS-based assay. However, purchasable standards for IP and DMAP were not widely available. In order to demonstrate our capability to detect these metabolites we first utilized an enzymatic assay. This assay relied on the biotransformation of mevalonate-5-phosphate (MVAP) to IP utilizing the promiscious activity of mevalonate diphosphate decarboxylase from *S. cerevisiae*, as showin in Figure 5-5 [17]. Following enzyme purification and *in-vitro* assays, we were able to detect a peak with mass spectra matching the desired metabolite, as shown in Figure 5-6.

Although this result is sufficient to allow screening and comparison between enzymes, we next aimed to chemically synthesize IP and DMAP. This enables not only the generation of standard curves, allowing us to obtain absolute enzymatic activities, rather than relative ones, but also allows the testing of enzyme activity upon these intermediate substrates. We followed a previously reported procedure [18] and were

successful in producing IP and DMAP at milligram scales (confirmed via LCMS and NMR), enabling the generation of standard curves and for use in assays.

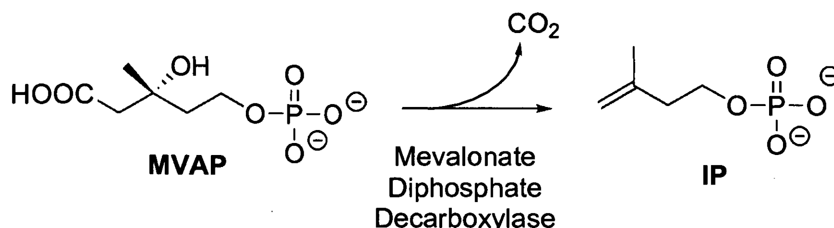

Figure 5-5: **Methodology for Enzymatic Detection of IP.** Mevalonate diphosphate decarboxylase typically catalyzes the biotransformation of mevalonate diphosphate to IPP, but also has promiscious activity allowing for the decarboxylation of the mevalonate monophosphate.

5.3.3 In-Vitro Screening of Candidates

We next screened candidate enzymes via LCMS. Initial experiments were performed using overnight enzymatic incubation; however, further experiments revealed this dramatically underestimated enzyme performance due to IPP and IP degradation. Thus, subsequent assays were performed over a 1-hour time interval. As shown in Figure 5-7, of the enzymes screened, choline kinase displayed the highest level of activity. This result was somewhat surprising as choline, as shown in Figure 5-9, is a quaternary amine alcohol. Thus, we expected its active site to be relatively polar, and therefore unlikely to accommodate isopentenol effectively due to its highly non-polar nature. However, inspection of choline kinase reveals much of the active site is defined by tryptohan and phenylananine residues, giving the substrate docking region a highly hydrophobic nature (Figure 5-10). Additionally, the size of choline is approximately that of isopentneol, likely aiding substrate binding. Further characterization of choline kinase revealed approximately 1.3 mU/mg activity under assay conditions (Figure 5-8).

Data regarding the enzymatic activity can also be directly applied to provide an estimate of applicability to industrial fermentation. To do this, we can relate enzyme activity measured *in-vitro* to approximate turnover by a mass of cells. This requires

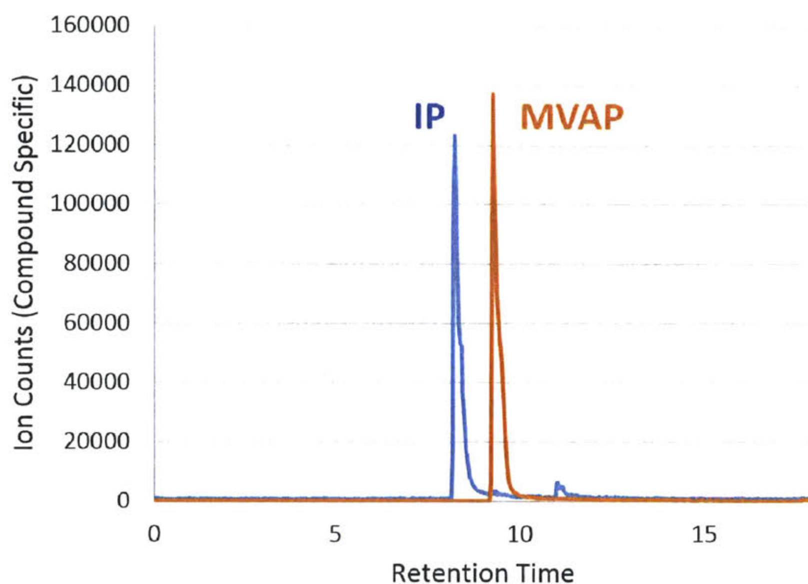

Figure 5-6: **Enzymatic Transformation of MVAP to IP.** LCMS trace depicting substrate (MVAP) and product (IP) obtained from *in-vitro* enzymatic transformation, displaying primary ions for each compound.

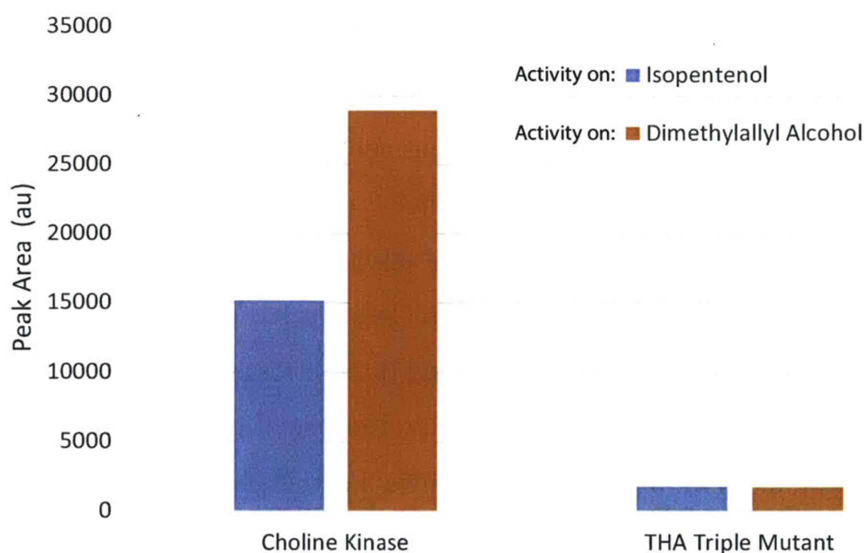

Figure 5-7: **Initial Screen of Isopentenol Kinase Candidates.** Peak heights shown are for total turnover to monophosphorylated compounds after 30 minutes. Only two genes are displayed as other screened genes displayed peaks heights indistinguishable from noise.

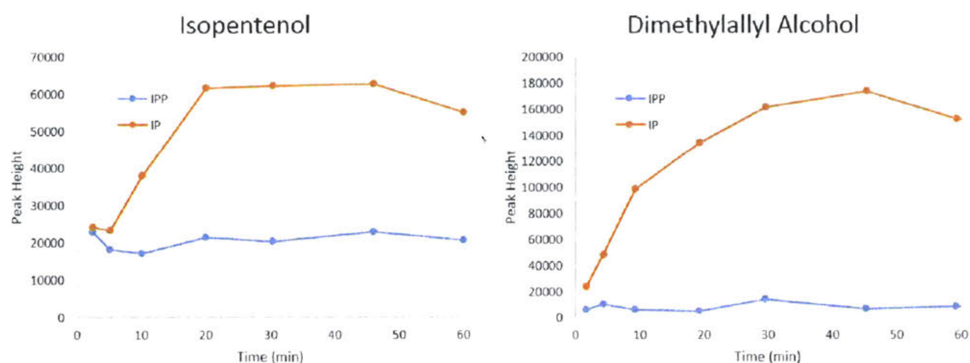

Figure 5-8: **Initial Screen of Isopentenol Kinase Candidates.** As can be seen, activity on dimethylallyl alcohol is higher than that on isopentneol and little IPP is produced from the reaction.

Figure 5-9: **Chemical Structure of Isopentenol and Choline.**

several assumptions:

- Isopentenol physphorylation is the rate-limiting step in the pathway as this is the sole step not performed by native enzymatic activity.
- The isopentneol phosphorylating enzyme comprises 5% of cell dry weight (DCW). This is aimed to be a conservative estimate as highly-expressed proteins can exceed 40%.
- No transport limitations of substrate in or product out of cells exists.

These assumptions are not perfect, as many terpene cyclases possess relatively slow kinetics, which would render them as rate-limiting when coupled with a highly-optimized IU pathway. Furthermore, transport of isopentenol across the membrane may become problematic; However, this may also be alleviated via efflux transporter knockouts and other engineering methodologies. Generally, these assumptions hold true at low enzyme turnovers, and begin to break down at extremely high enzymatic

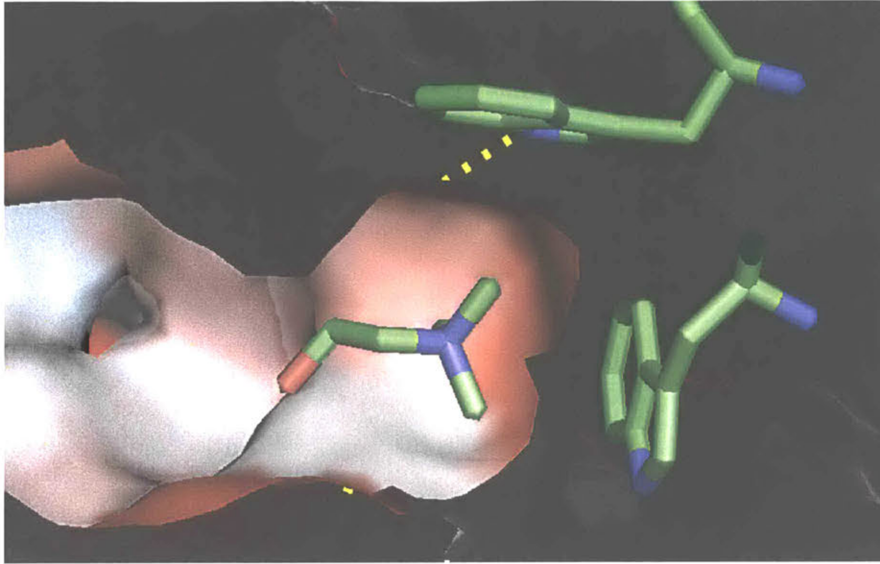

Figure 5-10: **Representation of Choline Kinase Active Site.**

activity. Applying these assumptions we can relate the activity of the protein in Units to productivity:

$$\frac{1\,Unit}{mg\ Protein} = \left(\frac{60\ min}{1\ hr}\right)\left(\frac{0.05\ mg\ protein}{1\ mg\ cellmass}\right)\left(\frac{1000\ mg\ Cells}{1\ g\ Cells}\right)\left(\frac{85\ ug\ Isopentenol}{1\ uMole\ Isopentenol}\right)\left(\frac{1\ g}{10to6\ ug}\right) \tag{5.3}$$

$$\frac{1\,Unit}{mg\ Protein} = 0.225\,\frac{g\ Isopentenol}{hr\ g\ DCW} \tag{5.4}$$

To continue, we can estimate 1 OD unit is equivalent to 0.39 g/liter DCW. Thus, we can estimate operation at an optical density value of 300, which is realisticaly obtainable under a perfusion, continuous, or fed-batch regime, and equates to a DCW value of 117 g/L. Thus, the productivity of an enzyme with 1 unit of activity is **29.8 g/L hr**. For reference, using our previously obtained relation between kcat and mU/mg protein, choline kinase posseses approximately 142 U/mg of activity on its native substrate at saturation conditions [19]. If this same level of activity were achieved for isopentenol, this would in turn yield a productivity of approximately 4 kg/L hr turnover of isoptenenol per hour, making it a highly effective for industrial use, although it is unlikely all the assumptions hold true at this high activity. To further illustrate the relationship, Figure 5-11 is provided below. This approximation

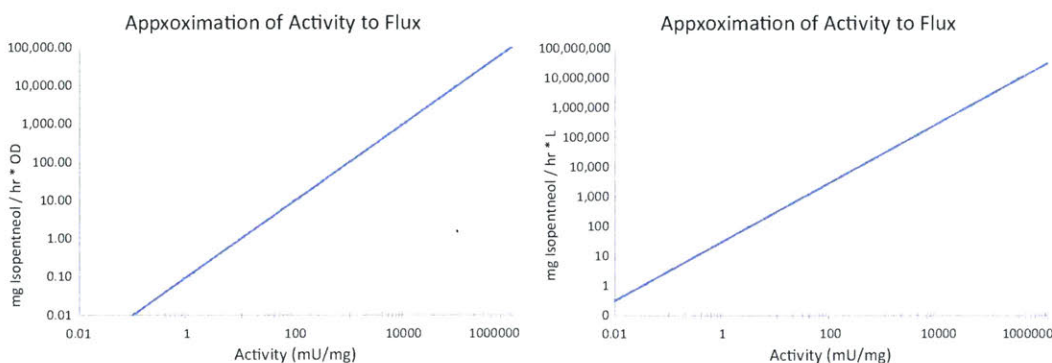

Figure 5-11: **Relationship of Enzymatic Activity to Productivity.** On left is turnover per liter for a single OD unit. On right proditivity is depicted for a OD 300 reactor.

demonstrates that although the current activity of choline kinase is too low to allow for industrial application, only moderate levels of optimization are required to begin seeing necessary improvements in activity.

5.3.4 *In-Vivo* Pathway Validation

Following the positive result yielded by choline kinase, we set out to demonstrate the functionality of this pathway *in-vivo*. This was performed via the construction of a plasmid containing the entire IU pathway, consisting of choline kinase, IPK, and isopentenyl diphosphate isomerase (*idi*). The IPK chosen was derived from *Arabidopsis thaliana* (atIPK), due to high catalytic effeciency reported (kcat/km)[12], while *idi* was derived from *E. coli*. Next, a strain containing this plasmid was grown on C13 glucose and unlabeled isopentenol. In this regime, pathway functionality can be demonstrated by a decrease in of isopentenol labeling. The result from this experiment is shown in in Figure 5-12, demonstrating a low level of pathway functionality *in-vivo*.

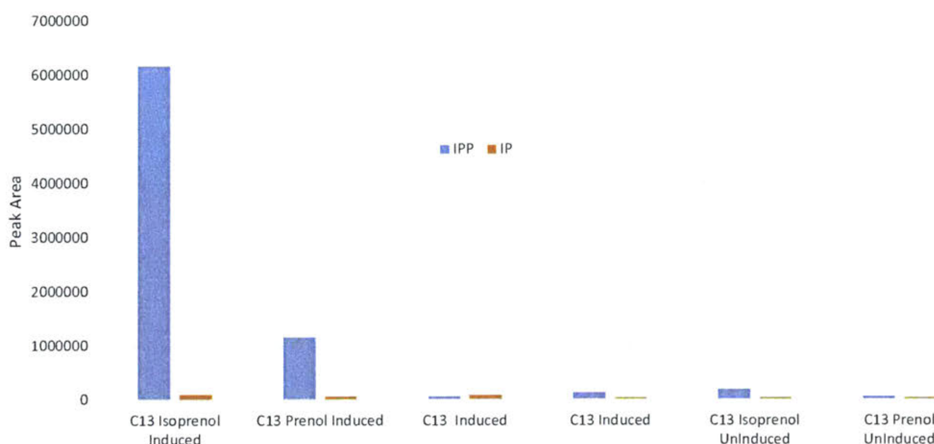

Figure 5-12: **Initial Screen of Isopentenol Kinase Candidates.**

5.4 The Need for High-Throughput Screening Methodologies

Although the generation of a functional pathway from isopentenol to isoprenoids is valuable in itself, optimization of the pathway is required before it can be utilized in any practical fashion. However, optimization of an enzyme, here choline kinase, for higher activity is a daunting task, one in which little information is available for rational, targeted mutagenesis. As illustrated in Figure 5-13, the types of mutagenesis one is able to apply to a system is directly related to the methodologies available for screening. Generally speaking, methods which enable screening of large libraries allow for approaches which are more combinatoral in nature and less rational, requiring less pathway information. Screening methodologies ranging from microfludic devices [20], to colorometric reporters [21] have seen vast phenotype improvements with little rational input. However, many of these methodologies are not immediately applicable to terpenoids.

Typical detection of terpenoids utilizes low-throughput gas or liquid chromatography. This method is labor intensive, requiring sample preparation and several minute long GCMS runs per sample, limiting the number of samples which can be run. A surrogate reporter molecule, lycopene, has been widely used within terpenoid biosyn-

thesis as it yields colonies with a reddish color, allowing for screening of colonies on agar plates. This methodology has yielded impressive results in applications ranging from rationally selected mutagenesis targets using MAGE [6], knockout mutagenesis [22], or even in engineering individual enzymes for higher activity [23]. However, despite the enormous body of work lycopene has enabled, it is far from an ideal surrogate molecule and has several drawbacks. The most obvious of these is the need to visually screen colonies. This visual screening, even when performed using image analysis software can be extremely noisy, leading to false positives or negatives. Furthermore, studies which have made use of combinatorial methodologies often result in the selection of genes involved in reactive oxygen stress. The reasons for this were recently highlighted [24]. An important learning from this is that although these mutations may improve the quantity of lycopene observed (due to a reduction in bleaching), they do not improve the true lycopene quantity, and thus are simply a result of systematic error built into the use of lycopene. Furthermore, lycopene is a highly hydrophobic molecule which typically partitions into either inclusion bodies, or within the cell membrane. As cells have a limited capacity for intracellular storage of hydrophobic molecules the maximum quantity of lycopene that can be produced prior to toxicity is relatively low.

We thus next focused our attention on developing novel methodologies for high-throughput screening of terpenoid-family compounds. As can be seen in Figure 5-13, while lycopene has been used, it still enables a relatively small number of colonies to be screened. FACS, or selection-based strategies are ideal as they allow for larger numbers.

5.5 Evolution-Based Pathway Optimization

Evolutionary-based pathway optimization is by far the most powerful tool in a biologists toolkit. Through this methodology, trillions of cells can be screened daily. A chemostat may be used with many experiments which allows for mutations to arise and be screened spontaneously. However, downsides also exist to this approach. If

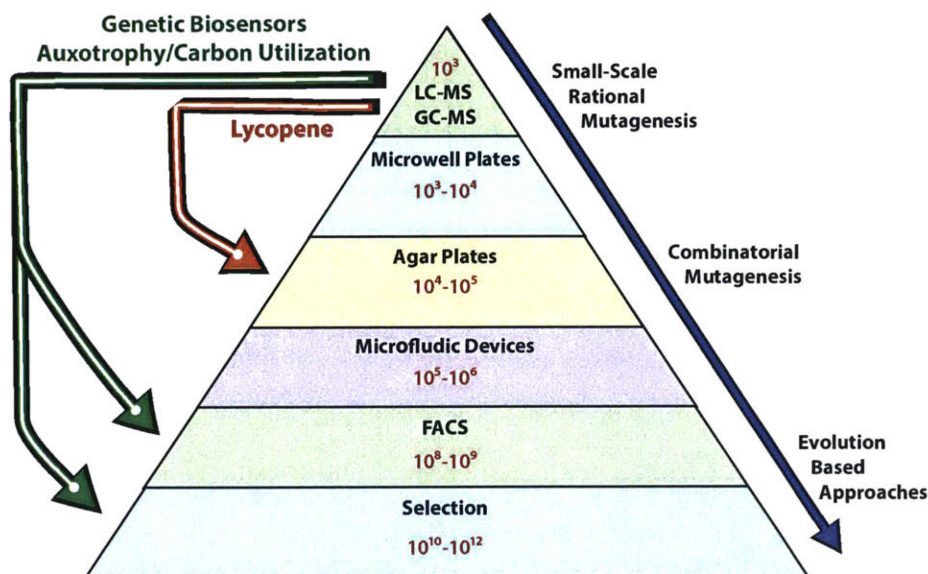

Figure 5-13: **Methodologies for Selection and Library Generation.**

run in an *in-vivo* rather than an *in-vitro* setting, it becomes increasingly likely that pathway mutations will occur which improve fitness, but are not within the objective genes or pathway. Some mutations are less likely than others, and the possibility that the population will reach a local maximal fitness, rather than a global one, during timescales examined is also possible. These limitations can largely be overcome through careful experimental design. We thus began investigation into developing two methodologies, one in *E. coli* and one in *Pseudomonas Citronellolis*, for this purpose. Although neither methodology is fully mature at the writing of this thesis, significant preliminary work has been devoted to build the necessary tools for these methodologies.

5.5.1 Applying Terpene Degradation in Pseudomonads

A potentially powerful methodology lies in expanding the substrate range of an organism. As this is potentially of use beyond terpenoids and the IU pathway, it will be discussed generally and then applied specifically to the IU pathway. The general concept consists of identifying an upstream production pathway to a compound which is to be optimized and used as the final production pathway. Next, an organism is

found which has the ability to consume this compound for growth as a sole carbon source. In these systems, carbon source range is then expanded via addition of the desired production pathway. In this way, the desired metabolic pathway (here the IU pathway) becomes an essential portion of the organism's carbon-utilization pathway. Thus, organismal fitness (and doubling time) become tied directly to enzyme and pathway optimality allowing for evolution-based approaches. After sufficient passaging, the IU pathway, contained on a plasmid, can be miniprepped and transformed directly into *E. coli*. Without the downstream degradation pathways present (as this is exclusive to the optimization stage), high pathway flux now results in the production of terpenes from isopentenol rather than growth. Although this approach is only implemented for a single pathway (the IU pathway), we hope this method will later be widely generalizable by using a wider array of organisms and biosynthetic processes.

In selecting a pathway by which to apply this methodology, we first turned to several *Pseudomonas* species due to their ability to consume a wide range of carbon sources. Following a survey of literature, four primary metabolites and their corresponding degradative pathways were chosen to complement the IU pathway. As shown in in Figure 5-14, while cymene and camphor both posses known degradative pathways with fully characterized metabolites and genetics, a single gene is missing from the biosynthetic route of each of these. Thus, these pathways are difficult to apply for our use. Pinene degradation is promising as it has a simple, known, biosynthetic route. However, while information on the metabolites involved in degradation are available, the genetics are unknown and thus we initially discounted the use of this pathway. The geraniol utilization pathway, however, has both a known degradation pathway with partially characterized genetics, and a simple biosynthetic route. For this reason, geraniol was initially selected as the target molecule for pathway optimization.

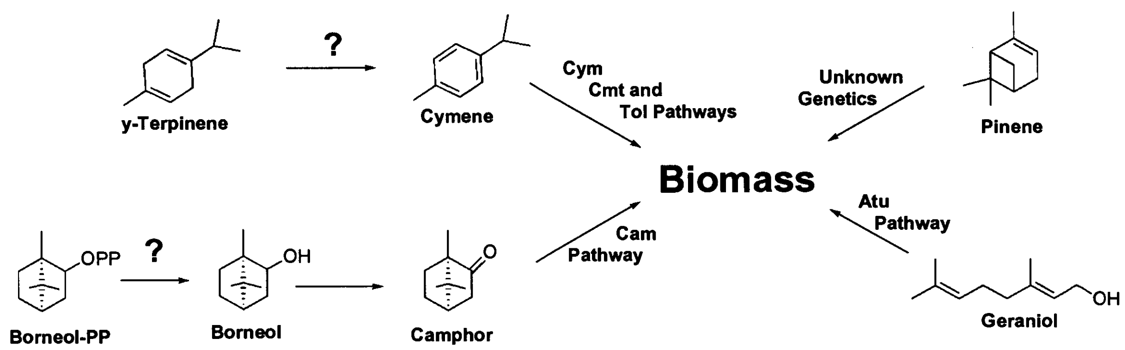

Figure 5-14: Pathways Evaluated for Optimization of IU Pathway.

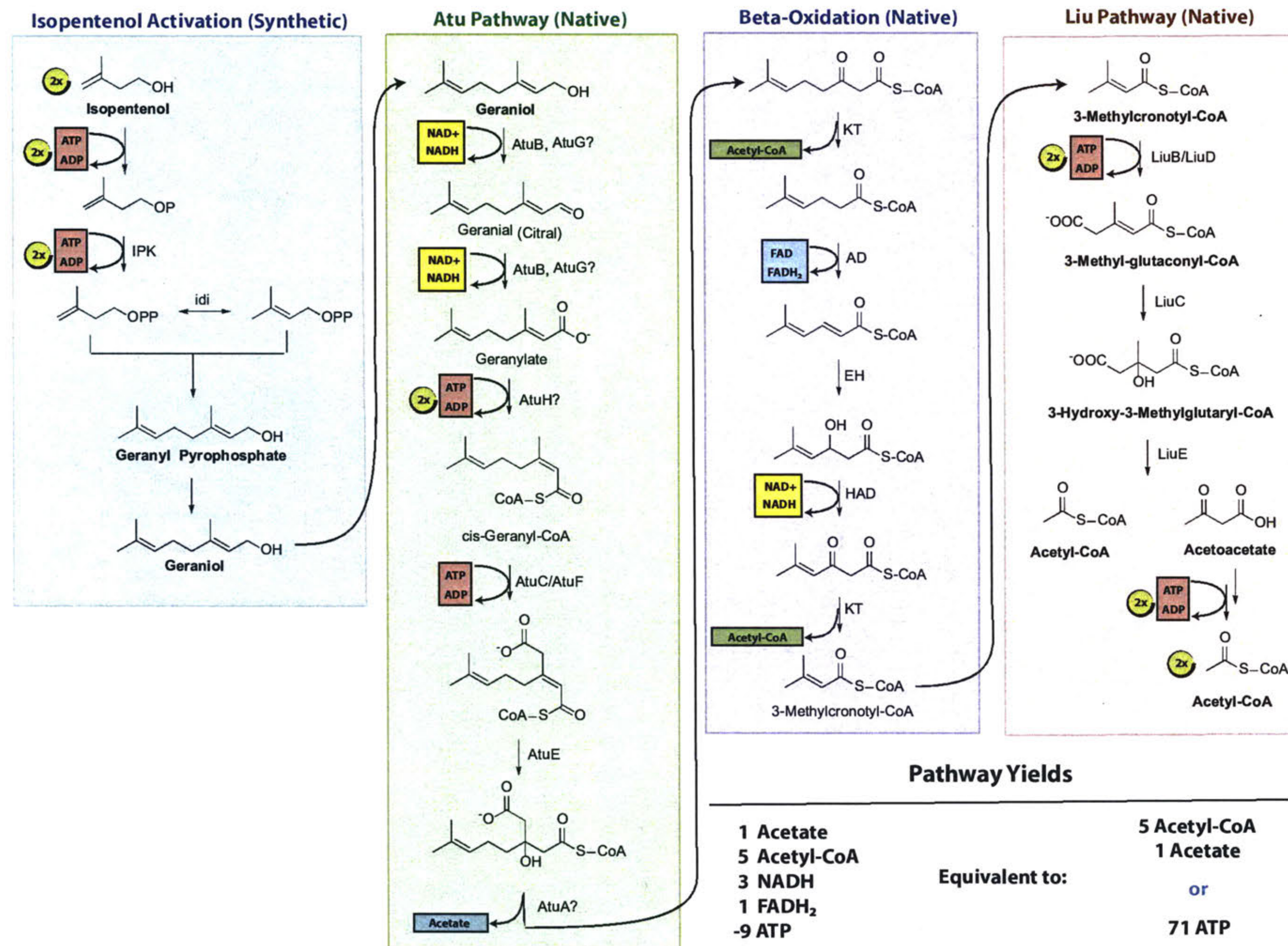

Figure 5-15: Full Pathway for Geraniol Production and Utilization from Isopentenol.

In order to understand the feasibility of such an evolutionary-based approach, it is informative to relate enzyme activity to a predicted doubling time. If doubling time is too long based on predicted activity, evolutionary based experiments will either take too long to complete, or carbon flux will be insufficient to satisfy maintenance requirements by the cell. On the other end, if the predicted doubling time is too short, other processes such as DNA replication may be rate-limiting and selective pressure to improve the IU pathway will be non-existent. To predict the doubling time we can express the change in cell mass over time as a function of biomass yield ($Y_{\frac{M}{S}}$), a fraction of cell mass which is the desired protein (Y_{FP}), and enzyme activity (A). As a conservative estimate, we set (Y_{FP}) at 10%, and ($Y_{\frac{M}{S}}$) to 30%. Thus the equation can be reduced written as:

$$\frac{dDCW}{dt} = \left(DCW\ \frac{g}{l}\right)\left(A\frac{uMoles\ iso}{min * mg\ Protein}\right)\left(MW\frac{ug\ Iso}{uMole}\right)\left(X_{FP}\frac{mg\ protein}{mg\ DCW}\right)\left(Y_{\frac{M}{S}}\frac{mg\ DCW}{mg\ Iso}\right)\left(\frac{1\ mg}{1000\ ug}\right) \tag{5.5}$$

Where activity is activity in units (U). Thus,

$$\frac{dDCW}{dt} = (DCW\frac{g}{l}) * (A * X_{FP} * Y_{\frac{M}{S}} * MW * 0.001) \tag{5.6}$$

Integration yields

$$DCW = C_1 e^{t*(A*X_{FP}*Y_{\frac{M}{S}}*MW*0.001)} \tag{5.7}$$

For the purpose of doubling-time calculations we can assume W = 1 at time zero. Thus

$$DCW = e^{t*(A*X_{FP}*Y_{\frac{M}{S}}*MW*0.001)} \tag{5.8}$$

Solving for the time at which W = 2W yields

$$t_d = \frac{ln(2)}{(t * (A * X_{FP} * Y_{\frac{M}{S}} * MW * 0.001))} \tag{5.9}$$

$$t_d = \frac{ln(2)}{(A * 0.00255)} \tag{5.10}$$

Thus we can directly estimate doubling time from activity (A, in U/mg protein). From this, we propose doubling times within the range of 2 hours and 7 days are relevant, and thus the required activity is between 136 units and 1.6 units. Additionally, this relationship may also be utilized during evolution as we can conservatively estimate the evolved activity of the enzyme from the doubling time of the culture.

5.5.2 Transformation, Expression, and Evolution in *P. citronellolis*

In order to perform evolution experiments in *P. citronellolis*, we must be able to stably transform and express enzymes within this system. We initially selected vectors from the Standard European Vector Architecture (SEVA) series. pSEVA-228 contains an oriT allowing for conjugative plasmid transfer, kanamycin resistance widely used in *Pseudomonas* literature, a broad host range vegetative origion of replication, and a methyl-toluate inducible pM promoter. In order to eliminate the need for inducers over long culture times, a constitutive proD promoter was used to replace the pM promoter. For proof-of-concept experiments we first aimed to express GFP. As initial attempts at electroporation failed to yield transformants, and the number of transformations required is relatively low, we opted to pursue conjugation as an alternative. Following conjugation colonies were obtained, and verified to be expressing via flow cytometry (Figure 5-16). Copy number variants were obtained and observed to have approximately a 10-fold improvement in GFP, but slower growth due to protein expression burden also existed.

We thus cloned and expressed our IU pathway coupled to a geraniol production pathway in *P. citronellolis.* However, following implementation, after extended culture times, growth on isopentenol alone was observed. It was hypothesized that this may occur through activation of isopentenol via acetylation and entry into the Liu pathway. In future work, knockouts will be constructed in which may remove this growth phenotype and allow for the application of this approach. One important note is that if the Liu pathway is indeed responsible for metabolism of isopentenol, a

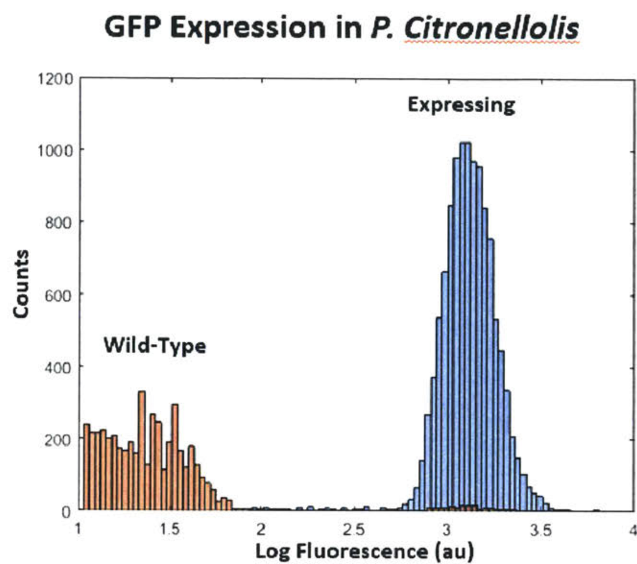

Figure 5-16: Flow Cytometry on Transformed P. Citronellois.

knockout will be unable to utilize geraniol, and thus an alternative metabolite (ideally pinene) will be required.

5.5.3 Isoprenoid Auxotrophic E. coli

An alternative approach in utilizing evolution for the improvement of pathway flux arises from the fact that terpenoids are essential to cellular function in nearly all organisms. Thus, *E. coli* bearing loss-of-function mutations in MEP pathway genes are unable to survive without supplementation of downstream intermediates. We thus envisioned an MEP-knockout strain in which the IU pathway has been added allowing for growth on isopentenol. Evolutionary approaches (*in-vivo* or *in-vitro*) can then be applied and cells with higer pathway flux will exhibit more rapid growth as their isoprenoid requirements are met. An important note is that only low levels (ug to mg/g DCW) of isoprenoids are required for normal cell growth. Thus, we proposed to use a promoter library to sequentially lower protein expression as protein activity increased in order to maintain selective pressure.

To apply this methodology, we first had to obtain MEP-pathway knockouts. This has been previously utilized, but relied on knockouts of upstream genes including

DXS and DXR. Previous literature has also demonstrated the appearance of escape mutants [25] in which growth is spontaneously restored. We thus sought to generate our own knockouts in which downstream steps or multiple genes are knocked out in order to prevent this escape phenotype. To achieve this we designed a 3 vector system to allow for plasmid compatibility based both on replative origin and on antibiotic resistance. In this way, we are able to utilize a rapid Cas9 inactivation system to achieve knockouts while retaining growth through a heterogeneously expressed lower mevalonate cassette restoring isoprenoid production and growth only in the presence of exogenously-supplied mevalonate. Importantly, this p15 cassette is constitutive as IPTG and arabinose inducible promoters are incompatable with the cas9 system widely used in *E. coli.*

- pBAD-proD-Lower Mevalonate Pathway (pACYC, Incompatibility group B, Chlorophenol)
- pCas9 (repA101, Incompatibility group C, Kanamycin)
- pTargetF (pMB1, Incompatibility Group A, Spectinomycin)

Through this method we were able to obtain knockouts in multiple MEP pathway genes including ispG, DXS, and DXR. In further experiments, we demonstrated that combination of these knockout strains with a plasmid containing the IU pathway restored growth in the presence of isopentenol and dimethylallyl alcohol. However, these mutants displayed delayed growth, with turbidity only appearing after 36 hours of culture time or longer. In contrast, supplementation with mevalonate yielded turbid cultures after less than 12 hours. This difference is reflective of the non-optimized nature of the IU pathway and dependence on promiscuous activity of choline kinase, rather than using an enzyme optimized for this reaction. Although not immediately obvious, this is a favorable result as this differential growth allows for rapid evolution experiments to be performed. Serial subculture and *in-vitro* evolution though error-prone PCR, both currently underway, will likely result in faster-growing strains which can be isolated with ease due to the long lag phase of cells containing wild-type enzymes.

5.6 Conclusions and Proposals for Future Work

In this chapter we enumerated obstacles to economic isoprenoid production through the MEP and MVA pathways. We then proposed, constructed and validated an alternative biocatalytic route from the bulk chemical isopentenol. Following this we evaluated and built tools for the use of evolution-based pathway optimization methodologies. Future work lies in the application of these evolutionary approaches to the optimization of the IU pathway.

5.7 Materials and Methods

Enzymatic Synthesis and Detection of Isopentenol Phosphates. For the purpose of this study isopentenyl monophosphate (IP) was enzymatically synthesized using mevalonate-5-phosphate (mev-5-P) as the starting compound. A codon-optimized gene encoding mevalonate pyrophosphated decarboxylase (PMD) was cloned in a pET28a vector and expressed with a C-terminal 6-His tag. The protein was subsequently purified and used in an in vitro assay. The assay was adapted from [26] . The assay was run in a volume of 150mL in which 625 mM of mev-5-P were incubated for 30min at 40°C, with 20uL of purified enzyme, in a buffer with 50 mM $(NH_4)HCO_3$, 10mM $MgCl_2$ and 2mM of ATP at pH 7.5. The sample was then quenched with 4 volumes of EtOH and subsequently lyophilized. The lyophilized product was then resuspended in 100iL of water and was analyzed through LC-MS-MS (C-18 resin column, aqueous phase: Acetic Acid and TBA, organic phase: MeOH). The reactant (mev-5-P) eluted at 8.28 min, whereas the product (IP) was detected based on its expected fragmentation pattern and eluted at 9.29 min.

Chemical Synthesis of Isopentenol Phosphates. For the purpose of this study isopentenyl monophosphate (IP) and dimethylallyl monophosphate (DMAP) were chemically synthesized using isoprenol or prenol respectively as the starting compounds. The general procedure was adapted from [18].

All commercially available chemicals and solvents were used as supplied without

further purification. All new compounds gave satisfactory spectroscopic and/or analytical data. 1H and 31P NMR spectra were recorded at 300 MHz and chemical shifts are reported in parts per million (ppm) downfield from the internal standard, tetramethylsilane (TMS). The products were also able to be detected by LC-MS-MS, where they exhibited identical characteristics (fragmentation patterns, elution times) to the equivalent products enzymatically syntesized.

General Procedures Trichloroacetonitrile (2.26 equiv) tetrabutylammonium phosphate (1.66 equiv) were added to a stirred solution of the appropriate isopentenol (1 equiv) in acetonitrile and the mixture was stirred for 4h. The solvent was evaporated and the sample was mixed with water and cooled for 6h at 4°C. This resulted in the formation of white precipitate (trichloroamide crystals) which was then separated by filtration. The resulting solution was purified by flash silica chromatography using a mixture of isopropanol/NH_4OH/H_2O as the eluent and the product was subsequently converted to the corresponding ammonium salt by percolation through a DOWEX 50WX8 ion-exchange column, using NH_4HCO_3 (0.025 M) as the buffer solution. Then the solution is lyophilized to produce the product in solid form.

isopentenyl monophosphate (IP) Prepared following the General Procedure, starting from isoprenol (80 uL, 0.53 mmol), trichloroacetonitrile (120 uL, 1.2 mmol) and tetrabutylammonium phosphate (300 mg, 0.88 mmol) in 2mL acetonitrile. After work-up, IP was obtained as a white solid. 1H NMR (300MHz, D2O): δ: 4.71 (s, 1H), 3.80 (q, 2H), 2.21 (t, 2H), 1.61 (s, 3H); 31P NMR (300MHz, D2O): δ 2.38

dimethylallyl monophosphate (DMAP) Prepared following the General Procedure, starting from prenol (80 ÎijL, 0.53 mmol), trichloroacetonitrile (120 uL, 1.2 mmol) and tetrabutylammonium phosphate (300 mg, 0.88 mmol) in 2mL acetonitrile. After work-up, DMAP was obtained as a white solid. 1H NMR (300MHz, D2O): δ 5.26 (t, 1H), 4.17 (t, 2H), 1.61 (s, 3H), 1.56 (s, 3H); 31P NMR (300MHz, D2O): δ 2.96

Cloning and Transformation. DNA constructs were via Gibson Assembly unless otherwise specified. All candidate kinases and IPK genes were codon optimized for E. coli and synthesized. E. coli was transformed using electroporation.

Transformation of Pseudomonads. Transformation of Pseudomonads was

performed via conjugation. This relied on the donor strain MFDpir, a Diaminopimelic acid (DAP) auxotroph. In brief, plasmids were transformed into MFDpir and grown in the presence of the appropriate antibiotic and DAP. Colonies of MFDpir and Pseudomonas were then isolated and precultured overnight. The following day, strains were subcultured and grown to exponential phase at 30°C and 250 RPM. When OD reached 0.4 for each culture, 1 mL was spun gently (3000 RCF, 5 minutes), the supernatant was removed, and the pellet was resuspended in 100 ul LB. 50 ul of each culture was then mixed on a 0.22 um filter and placed up on an LB plate complete with DAP. Following a 4 hours incubation at 30°C, the filter was placed into 10 mL LB, vortexed, supernatant decanted, and pellet plated on LB without DAP and with the appropriate antibiotic in order to remove MFDpir and select for positive transformants.

Characterization of Pseudomonas Transformanants Transformants were subjected to flow cytometry in order to confirm transformation and characterize heterologous protein expression level. Wild type pseudomonas and cells containing a vector encoding for constitutive GFP expression were cultured overnight, 2 ul of this culture was then resuspeneded in 200 ul LB and subjected to flow cytometry on a FACS Canto II.

Characterization of MEP-Pathway Knockouts MEP-pathway knockouts were grown in the presence of 1 mM mevalonate. To demonstrate growth on isopenenol, transformants were grown in LB supplemented with mevalonate overnight, and subultured into LB containing isopentenol or dimethylallyl alcohol, mevalonate, or a control with no supplimnetaion to demonstrate growth dependence on supplementation.

5.8 Bibliography

[1] K. Zhou, R. Zou, G. Stephanopoulos, and H.-P. Too, "Metabolite profiling identified methylerythritol cyclodiphosphate efflux as a limiting step in microbial isoprenoid production," *PLoS One*, vol. 7, no. 11, p. e47513, 2012.

[2] P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, *et al.*, "Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin," *Proceedings of the National Academy of Sciences*, vol. 109, no. 3, pp. E111–E118, 2012.

[3] J. Wungsintaweekul, S. Herz, S. Hecht, W. Eisenreich, R. Feicht, F. Rohdich, A. Bacher, and M. H. Zenk, "Phosphorylation of 1-deoxy-d-xylulose by d-xylulokinase of escherichia coli," *European Journal of Biochemistry*, vol. 268, no. 2, pp. 310–316, 2001.

[4] A. Hemmerlin, D. Tritsch, M. Hartmann, K. Pacaud, J.-F. Hoeffler, A. van Dorsselaer, M. Rohmer, and T. J. Bach, "A cytosolic arabidopsis d-xylulose kinase catalyzes the phosphorylation of 1-deoxy-d-xylulose into a precursor of the plastidial isoprenoid pathway," *Plant physiology*, vol. 142, no. 2, pp. 441–457, 2006.

[5] J. Kirby, M. Nishimoto, R. W. Chow, E. E. Baidoo, G. Wang, J. Martin, W. Schackwitz, R. Chan, J. L. Fortman, and J. D. Keasling, "Enhancing terpene yield from sugars via novel routes to 1-deoxy-d-xylulose 5-phosphate," *Applied and environmental microbiology*, vol. 81, no. 1, pp. 130–138, 2015.

[6] H. H. Wang, H. Kim, L. Cong, J. Jeong, D. Bang, and G. M. Church, "Genome-scale promoter engineering by coselection mage," *Nature methods*, vol. 9, no. 6, pp. 591–593, 2012.

[7] T. S. Gardner, K. M. Hawkins, A. L. Meadows, A. E. Tsong, and Y. Tsegaye, "Production of acetyl-coenzyme a derived isoprenoids," Dec. 10 2013. US Patent 8,603,800.

[8] P. K. Ajikumar, W.-H. Xiao, K. E. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, and G. Stephanopoulos, "Isoprenoid pathway optimization for taxol precursor overproduction in escherichia coli," *Science*, vol. 330, no. 6000, pp. 70–74, 2010.

[9] D. Dugar and G. Stephanopoulos, "Relative potential of biosynthetic pathways for biofuels and bio-based products," *Nature biotechnology*, vol. 29, no. 12, pp. 1074–1078, 2011.

[10] E. Gruchattka, O. Hädicke, S. Klamt, V. Schütz, and O. Kayser, "In silico profiling of escherichia coli and saccharomyces cerevisiae as terpenoid factories," *Microbial cell factories*, vol. 12, no. 1, p. 1, 2013.

[11] L. L. Grochowski, H. Xu, and R. H. White, "Methanocaldococcus jannaschii uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate," *Journal of bacteriology*, vol. 188, no. 9, pp. 3192–3198, 2006.

[12] L. K. Henry, M. Gutensohn, S. T. Thomas, J. P. Noel, and N. Dudareva, "Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants," *Proceedings of the National Academy of Sciences*, vol. 112, no. 32, pp. 10050–10055, 2015.

[13] N. Dellas, S. T. Thomas, G. Manning, and J. P. Noel, "Discovery of a metabolic alternative to the classical mevalonate pathway," *Elife*, vol. 2, p. e00672, 2013.

[14] Y. Liu, Z. Yan, X. Lu, D. Xiao, and H. Jiang, "Improving the catalytic activity of isopentenyl phosphate kinase through protein coevolution analysis," *Scientific reports*, vol. 6, 2016.

[15] M. F. Mabanglo, J.-J. Pan, B. Shakya, and C. D. Poulter, "Mutagenesis of isopentenyl phosphate kinase to enhance geranyl phosphate kinase activity," *ACS chemical biology*, vol. 7, no. 7, pp. 1241–1246, 2012.

[16] M. F. Mabanglo, H. L. Schubert, M. Chen, C. P. Hill, and C. D. Poulter, "X-ray structures of isopentenyl phosphate kinase," *ACS chemical biology*, vol. 5, no. 5, pp. 517–527, 2010.

[17] A. Kang, K. W. George, G. Wang, E. Baidoo, J. D. Keasling, and T. S. Lee, "Isopentenyl diphosphate (ipp)-bypass mevalonate pathways for isopentenol production," *Metabolic engineering*, vol. 34, pp. 25–35, 2016.

[18] R. K. Keller and R. Thompson, "Rapid synthesis of isoprenoid diphosphates and their isolation in one step using either thin layer or flash chromatography," *Journal of Chromatography A*, vol. 645, no. 1, pp. 161–167, 1993.

[19] C. S. Hudson, R. M. Knegtel, K. Brown, P. A. Charlton, and J. R. Pollard, "Kinetic and mechanistic characterisation of choline kinase-α," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1834, no. 6, pp. 1107–1116, 2013.

[20] B. L. Wang, A. Ghaderi, H. Zhou, J. Agresti, D. A. Weitz, G. R. Fink, and G. Stephanopoulos, "Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption," *Nature biotechnology*, vol. 32, no. 5, p. 473, 2014.

[21] W. C. DeLoache, Z. N. Russ, L. Narcross, A. M. Gonzales, V. J. Martin, and J. E. Dueber, "An enzyme-coupled biosensor enables (s)-reticuline production in yeast from glucose," *Nature chemical biology*, vol. 11, no. 7, pp. 465–471, 2015.

[22] H. Alper, Y.-S. Jin, J. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in escherichia coli," *Metabolic engineering*, vol. 7, no. 3, pp. 155–164, 2005.

[23] E. Leonard, P. K. Ajikumar, K. Thayer, W.-H. Xiao, J. D. Mo, B. Tidor, G. Stephanopoulos, and K. L. Prather, "Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control," *Proceedings of the National Academy of Sciences*, vol. 107, no. 31, pp. 13654–13659, 2010.

[24] M. Bongers, P. K. Chrysanthopoulos, J. B. Behrendorff, M. P. Hodson, C. E. Vickers, and L. K. Nielsen, "Systems analysis of methylerythritol-phosphate pathway flux in e. coli: insights into the role of oxidative stress and the validity of lycopene as an isoprenoid reporter metabolite," *Microbial cell factories*, vol. 14, no. 1, p. 1, 2015.

[25] J. Perez-Gil, E. M. Uros, S. Sauret-Güeto, L. M. Lois, J. Kirby, M. Nishimoto, E. E. Baidoo, J. D. Keasling, A. Boronat, and M. Rodriguez-Concepcion, "Mutations in escherichia coli acee and ribb genes allow survival of strains defective in the first step of the isoprenoid biosynthesis pathway," *PloS one*, vol. 7, no. 8, p. e43775, 2012.

[26] J. C. VanNice, D. A. Skaff, A. Keightley, J. K. Addo, G. J. Wyckoff, and H. M. Miziorko, "Identification in haloferax volcanii of phosphomevalonate decarboxylase and isopentenyl phosphate kinase as catalysts of the terminal enzyme reactions in an archaeal alternate mevalonate pathway," *Journal of bacteriology*, vol. 196, no. 5, pp. 1055–1063, 2014.

Chapter 6

Development of a Terpene-Responsive Biosensor

6.1 Introduction

As discussed in Chapter 5 (page 156), few tools are available for high-throughput screening of terpenoids. While many problems, such as optimization of the isopentneol utilization (IU) pathway are amenable to the previously discussed evolutionary approach (page 158), this approach is not suited to all optimization problems. We thus next set out to develop a broadly-applicable platform for high throughput screening and optimization of upstream terpene biosynthesis pathways.

To achieve this we focused on the use of genetically-encoded terpene-responsive biosensors and began by evaluating several possible systems. The first possibility, a system consisting of heterologously expressed G-protein olefactory receptors has the advantage that it is theoretically adaptable to a broad range of terpenoids. However, this system may experience ligand cross-talk, suffers from low reporter strengths, and likely requires a vast amount of engineering to be fully functional [1, 2, 3, 4, 5]. Another option, based on the yeast protein Cup9 has been demonstrated to be capable of detecting and responding to the acyclic terpene geraniol. However, the system depends on ubiquitination and protein degradation, has a homolog within *S. cerevisiae*, and would thus require complex work-arounds [6, 7]. Furthermore, both of

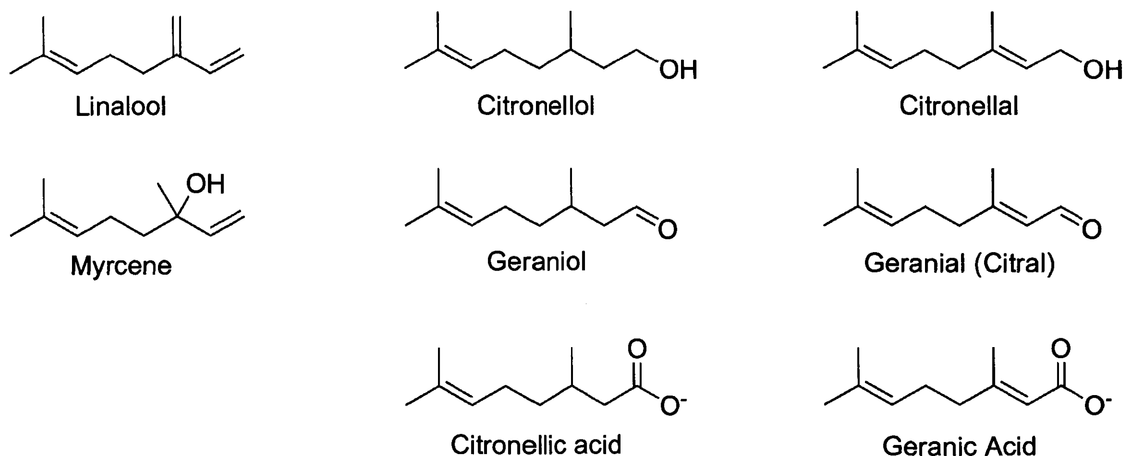

Figure 6-1: **Terpenes Capable of Acting as Sole-Carbon Sources in *Pseudomonas* Species.**

these options are, at present, only capable of being expressed in eukaryotic systems and could not be used within prokayotic cells for engineering.

Another possible system would rely on ligand-responsive transcriptional regulators, such as those belonging to the AraC, TetR, LacI, and MarR familes. These systems are typically prokaryotically derived and easy to heterolgously express in both prokaryotes and eukaryotes, allowing for broad applicability [8]. Furthermore, a wide variety of ligands have been characterized [9, 10] and they have been demonstrated to be engineerable [11]. We thus set out to implement this system for the sensing of isoprenoids.

6.2 AtuR and Genetic Basis of Regulation

A survey of literature revealed a proposed TetR family repressor and binding motif present in *Psuedomonas citronellois* [12]. As discussed in Chapter 5 (page 156), many pseudomonads including *P. citronellolis*, *P. fluorescens*, *P. aeruginosa*, and *P. mendocina*, are capable of degrading short-chain acyclic terpenes including citronellol, geraniol, their aldehydes and carboxylic acids, myrcene, and linalool [13, 14]. This ability extends to other bacterial species and may play a role in terpene recycling in plant species as well [15, 16]. Pathway metabolites were originally elucidated by

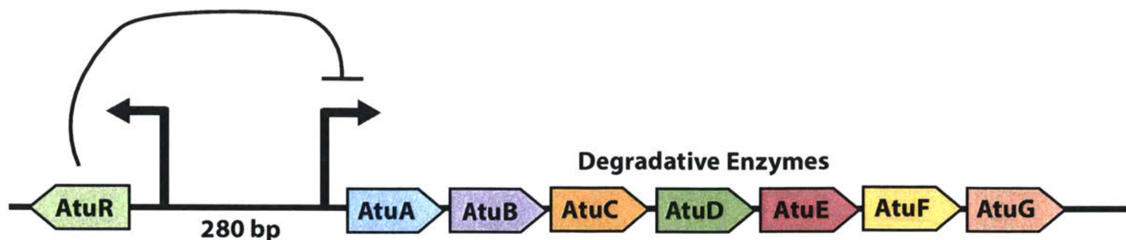

Figure 6-2: **Organization of Atu Gene Cluster in *Pseudomnoas citronellolis***

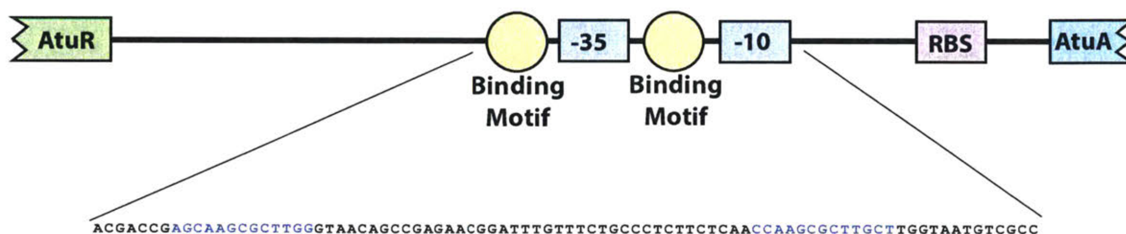

Figure 6-3: **Genetic Basis for Repression in *Pseudomonas*** The sequence from *P. aeruginosa* in which the cluster was studied is given. Binding motif is highlighted in blue and underlined.

Suebert et al. [13, 14, 17] (Discussed in more detail in Chapter 5, Page 156). While more recent work has uncovered the underlying genetics and enzymology [12, 18, 19].

A depiction of the gene cluster identified to be partially responsible for degradation of acyclic terpenes can be seen in Figure 6-2. As is typical for TetR family repressors, the repressor protein (AtuR) is oriented upstream and opposite in direction to the degradative enzymes AtuA:G which support utilization of acyclic terpenes. The intergenic region spans 280 base pairs (bp), and a binding motif consisting of a 13 bp palindromic sequence sits adjacent to the -35 and -10 regions of the promoter [12]. This is represented in Figure 6-3.

6.3 Initial Attempts at Promoter Design

In an initial attempt to generate an acyclic-terpene responsive biosensor, we first directly cloned the promoter sequence from *Pseduomoas arugosona* PAO1. The AtuR-AtuA intergenic region was synthesized and placed upstream of GFP, while AtuR was codon optimized and expressed in the location and orientation of the original AtuR

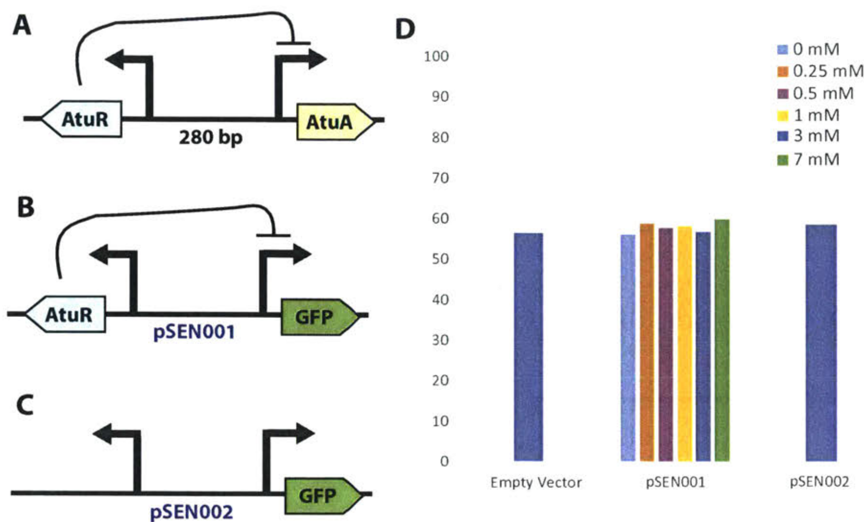

Figure 6-4: **Testing of Initial Synthetic Biosensor Construct.** A) Original Sequence from *Pseduomoas arugosona.* B) Sequence as cloned into *E. coli* with AtuR in a codon optimized form and the degradative cluster replaced with GFP, generating pSEN001. C) Construct with AtuR removed, generating pSEN002. D) Performance of constructs showing non-functionality of cloned system.

gene to generate the plasmid pSEN001 (Figure 6-4A, 6-4B). As desired, no basal GFP fluorescence was observed; however, the addition of acyclic terpenes did not increase fluorescence as desired. Thus, we concluded that two possibilities exist: AtuR repression is too strong, or the native *Pseudomonas* repressor may be too weak when used in *E. coli.* To differentiate between these options, we removed the AtuR gene, generating pSEN002 (Figure 6-1C). Following removal, no fluorescence was observed, thus the native promoter sequence was deemed non-functional in the heterologous host.

Following this result, we next set out to design a synthetic promoter sequence mirroring the elements of the native promoter, but within the context of a promoter known to function in *E. coli.* The strong constitutive promoter proD [20] was selected for this use. A plasmid was constructed from the parent plasmid pET28a, by replacing the T7 promoter region with the constitutive promoter and placing binding motifs both between and upstream of the -35 and -10 regions, as in the native host. Next, the LacI gene was replaced with AtuR, generating pSEN010. Ideally, this

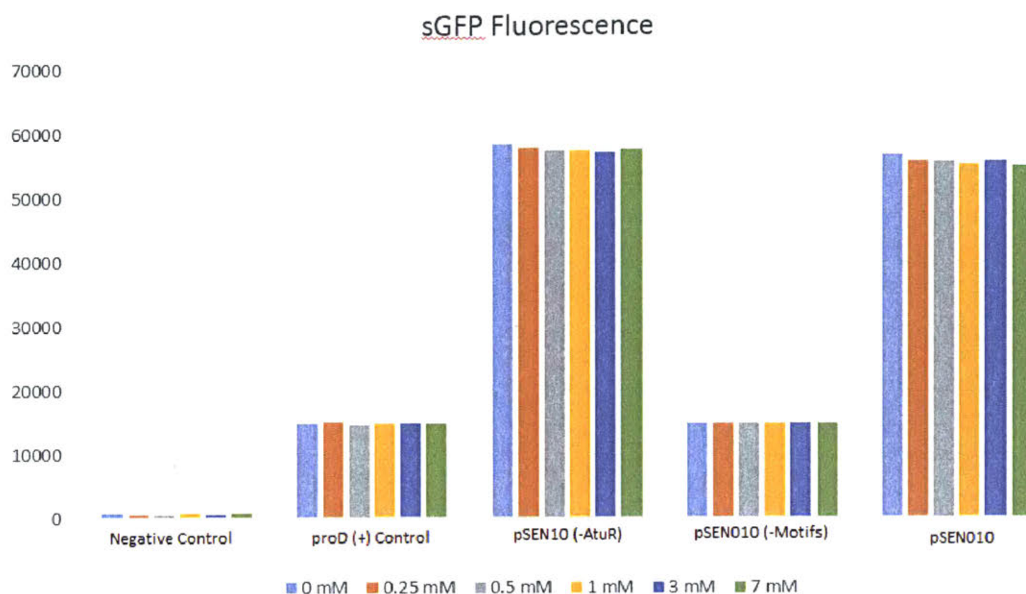

Figure 6-5: **Results from First Synthetic Promoter** Comparison of fluorescence obtained from various strains including a negative (empty vector) control, a positive control (proD), and pSEN010 without repressor binding motifs or AtuR expression, as well as the full pSEN010 construct.

plasmid should constitutively express AtuR, which would enable biosensor functionality. Upon construction and testing of this promoter, it was observed that addition of the binding motif increased expression by approximately four-fold as compared to the non-modified proD promoter and that the proD promoter itself was functional. However, no change was observed in the presence of acyclic terpenes.

The failure to obtain a clear result from this strategy demonstrated the need for a modular approach in which repression by AtuR can first be demonstrated, followed by derepression by acyclic terpenes, rather than approaching both tasks simultaneously.

6.3.1 Implementation of Two-Plasmid System

To achieve the proposed modularization, AtuR was expressed on a pBAD33 backbone, and expression of the repressor protein was verified (Figure 6-6). The synthetic proD promoter carrying a binding motif was expressed from a modified pET vector as before, but lacking the LacI and promoter sequence previously used to power AtuR ex-

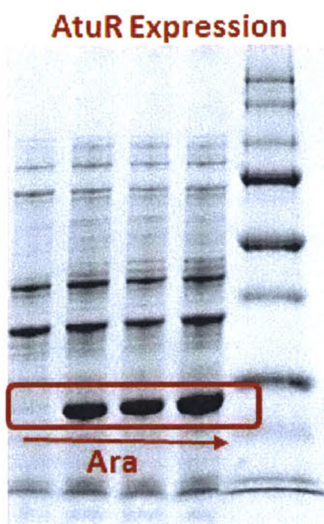

Figure 6-6: **Demonstration of AtuR Expression.** A band demonstrating AtuR expression is clearly visible with increasing arabinose titration

pression (yielding pSEN020). As shown in Figure 6-7, titration of arabinose to induce AtuR expression, and comparison to the pBAD33 empty vector, revealed a functional repression system. However, observed repression was low, with approximately 1.5-fold repression as compared to the control. We hypothesized that derepression (by acyclic terpenes) will only partially alleviate GFP repression, and thus a 1.5-fold shift was deemed insufficient for a functional biosensor. As previous reports demonstrated the importance of binding motif location [10](Shown in Figure 6-8), we investiated this design constraint as well via testing of four independent constructs (Figure 6-8), yet no improvement was observed .

6.3.2 Re-Investigating the Binding Motif

Following the low performance of the initial biosensor constructs, we began an investigation into the binding motif. We began by exploring binding behavior *in-vitro*. TetR-repressor family proteins typically respond to motifs which are internal palindromes as well as palindromic to eachother. Although these vary in length and in fidelity, as some motifs are highly divergent, length typically exceeds 15bp. Two examples are shown below in Figure 6-9. In contrast, the motif previously proposed for

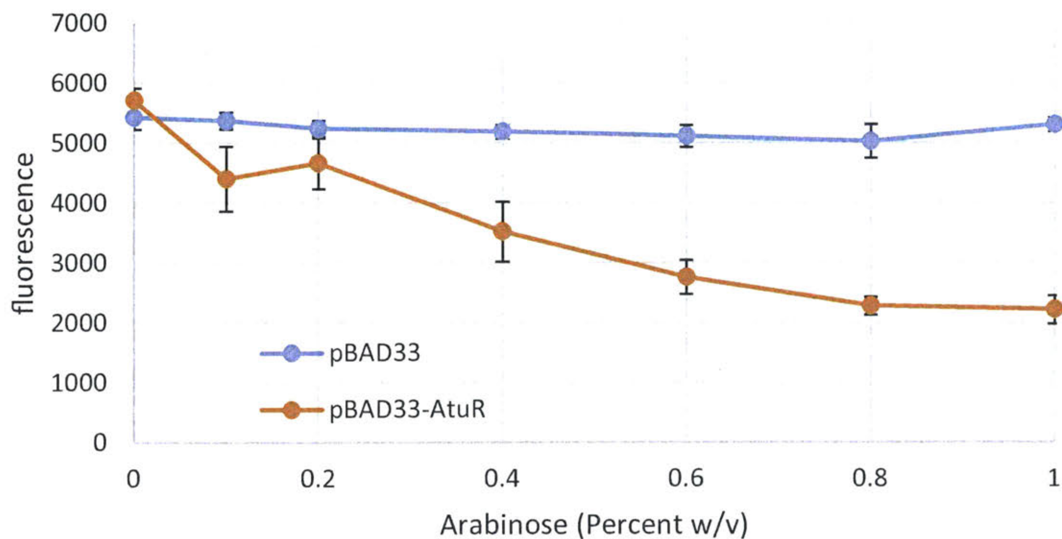

Figure 6-7: **Response obtained from pSEN020 Biosensor Construct.** Results for a vector containing a proD promoter flanked by binding motifis (pSEN020) in tandem with the empty pBAD33 vector (Blue), or a pBAD promoter expressing AtuR.

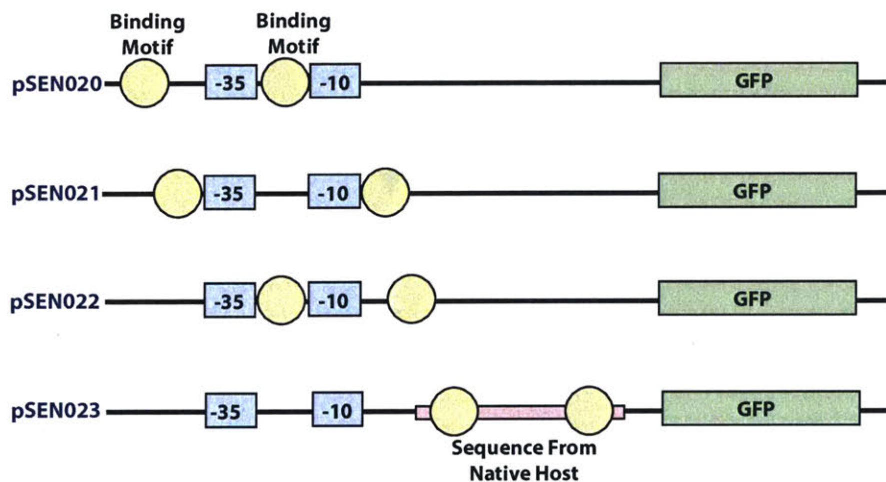

Figure 6-8: **Orientation of Binding Motifs in Tested Constructs**. The "road-bolock" methodology in pSEN023 utilizes 60 bp from the native host which includes both the proposed binding motif as well as internal sequence from the native host. This placed following the proD promoter, and prior to the ribosomal binding site.

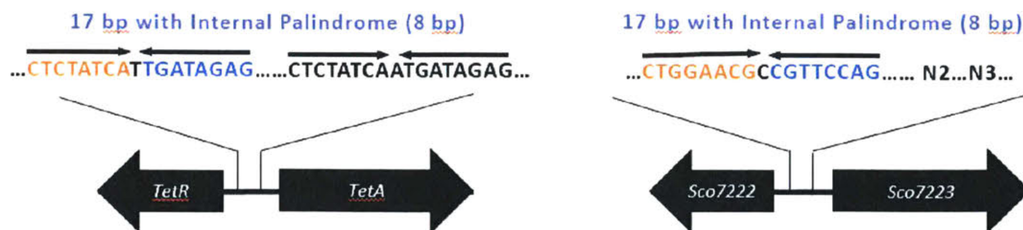

Figure 6-9: **Typical TetR Repressor Binding Motifs.** On the left, TetR is the most typical of the family with two sequences which are palindromic to each other, as well as containing internal palindromes. On the right is another member of the family which utilizes three binding sites.

AtuR binding [12] is only 13 bp with an internal palindrome length of 5 bp (Figure 6-12, Top). In order to gain further insight, a sequence alignment was performed between various *Pseudomonas* species containing Atu gene clusters, including both those which have and have not been experimentally confirmed to utilize acyclic terpenes as a sole carbon source, as well as other species with high sequence similarity to Atu genes (via protein BLAST). This alignment is shown in Figure 6-11. Interestingly, *P. aeruginosa* PAO1 possesses a single-nucleotide polymorphism (A to G), as compared to other species. Crucially, this SNP is immediately adjacent to the binding motif. If this mutation is ignored, the binding motif can be extended 4 bp upstream and 1 bp downstream, as shown in Figure 6-11. This allows the binding motif to now be redefined, yielding one with a total length of 18 bp with a 7 bp palindrome, more in line with the expected length (Figure 6-12, Bottom).

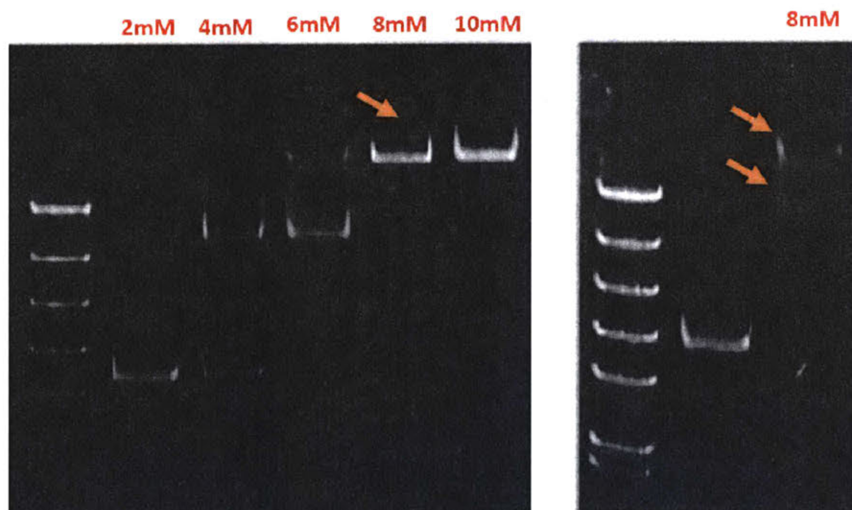

Figure 6-10: **Literature Results obtained from AtuR Titration.** On the left is a titration curve for AtuR in which full binding is achieved. On the left is a condition which is proposed to be identical, but only partial binding is achieved due to partial truncation of the binding region. Reproduced from [12].

```
                       ↓
Experimentally       ┌ P. aeruginosa PAO1    : ACGACCGAGCAAGCGCTTGGGTAACAGCCGAGAACG-GATTTG--TTTCTGCCCTCTTCTCAACCAAGCGCTTGCTTGGTAATGTCGCC
Verified Growth      │ P. citronellois SJTE  : GCAACCAAGCAAGCGCTTGGCCATCACTTGGTTGGGTTGTCTGGGCTTATGCCCTCTTTTCAACCAAGCGCTTGCTTGGTAATGTCGAA
on Terpenes          │ P. citronellois P3B5  : CTGACCAAGCAAGCGCTCGGCCGTACTTT--TGGCCGATTTGCCCTGCCCCAACGCTTTTCAACCAAGCGCTTGCTTGGTAATGTCGAC
                     └ P. fluorescens Pf5    : GTTACCAAGCAAGCGCTTGGCCAGAAGAATA--------------GGCCAACGTATTTACAAACCAAGCGCTTGCTTGGTAGCCTCCTC
Pseudomonas          ┌ P. denitrificans      : CCAACCAAGCAAGCGCTTGGCCATCCAATGGTTGGGGTTGTCGC-TCGATCGGGGCTTTTCAACCAAGCGCTTGCTTGGTAATGTCGAC
Strains with Atu     │ P. chlororaphis       : GCCACCAAGCAAGCGCTCGGGTGGAAAA-----------------CCCGCAGCCGTTTACAGACCAAGCGCTTGCTTGGTAGTCTCG
Gene Clusters        │ P. pseudoalcaligenes  : CTGACCAAGCAAGCGCTTGGGCGCATTCT--TGGCCGAATTGTCGTCTTTCGTCACTTTTCAACCAAGCGCTTGCTTGGTAGTGTGAGC
                     │ P. resinovorans       : GCGCCCAAGCAAGCGCTTGGGCAAAAAGAAC--------------TGCTTGGGCACTTTCAAACCAAGCGCTTGCTTGGTAGTGTTGAT
                     └ P. knackmussii        : CCAACCGAGCAAGCGCTTGGCCATTACTTGGTTGGGTGTTCTGGC-CGAATCGGCACTTTTCAACCAAGCGCTTGCTTGGTAATGTCCGA
Other Organims       ┌ H. chejuensis         : TTTACAAACCAAGCGCTTGCTTGGTATCGTTGGACAGGCTCGAATCAACAATAAGCGTTCAACAACAATAACTCTTCAAACGACATAC
with Atu             └ M. salarius           : TTTACAAACCAAGCGCTTGCTTGGTATTGTCGTTTCGATCGTTGGA---TAACCTCTCG-----------------------------
Gene Clusters
                                                      Original Motif                                  Original Motif
```

Figure 6-11: **Alignment of AtuR AtuA intergenic Regions for Motif Analysis.** Originally proposed motif is highlighted in yellow, while the extended motif is highlighted in blue. The SNP is indicated by an arrow and red text.

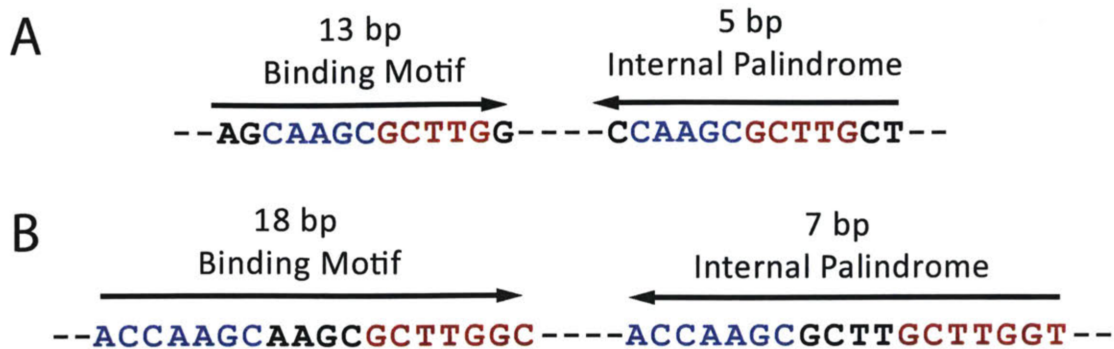

Figure 6-12: **Motif Sequence Comparison.** Comparison of previously-reported binding motif (top) to the revised binding motif (bottom).

Furthermore, this result allows us to reinterpret previous literature data. Förster-Fromme et al. reported of an electromobility shift assay (EMSA) experiment to confirm motif location [12]. In this experiment DNA is subjected to gel electrophoresis in the presence of purified protein. If the protein binds to DNA, an apparent shift in molecular mass is observed. Although diffusion and separation of DNA and protein is reasonable to expect, these in fact migrate together, due to a molecular caging effect imposed by the gel. In their result, two distinct shifts were observed, interpreted to correspond to binding to two motifs, which our data agrees with. As titration of AtuR was performed, full binding was observed at 8 nM AtuR (Reproduced in Figure 6-10). However, decreased binding affinity was observed when a shortened sequence, still proposed to contain two complete motifs, was used. We now can explain this result as, from our predictions, several residues were trimmed, leading to weakened binding affinity.

With this new information in hand, we next sought to implement our proposed extended binding motif. To achieve this, we designed two constructs. Due to the limitations imposed by the length of the motif (18 bp) which is larger than the distance between the -10 and -35 regions, use utilized a 'roadblock' methodology consisting of both the binding sequences and the sequence between them from the native host

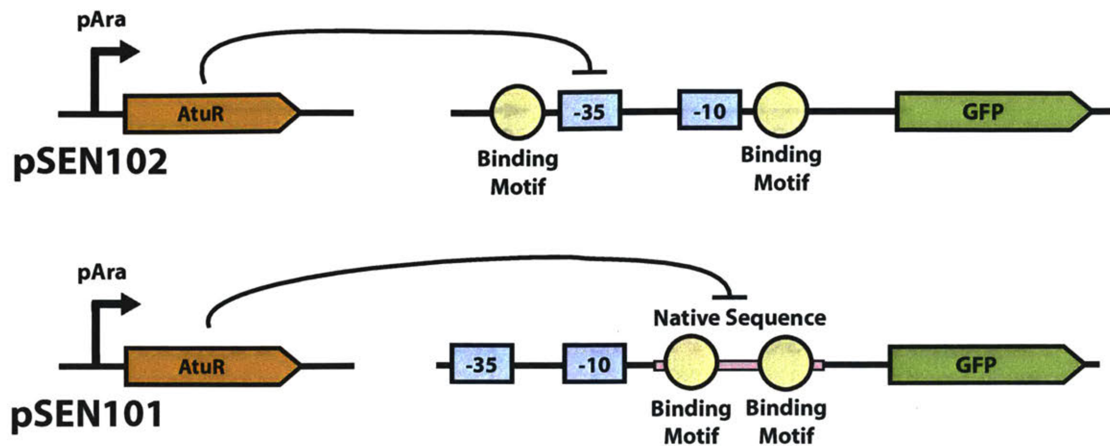

Figure 6-13: **Design of Extended Motif Constructs.** pSEN101 (bottom) utilized a 'roadblock' methodology, while pSEN102 utilized a methodology in which the binding motifs span across the promoter region.

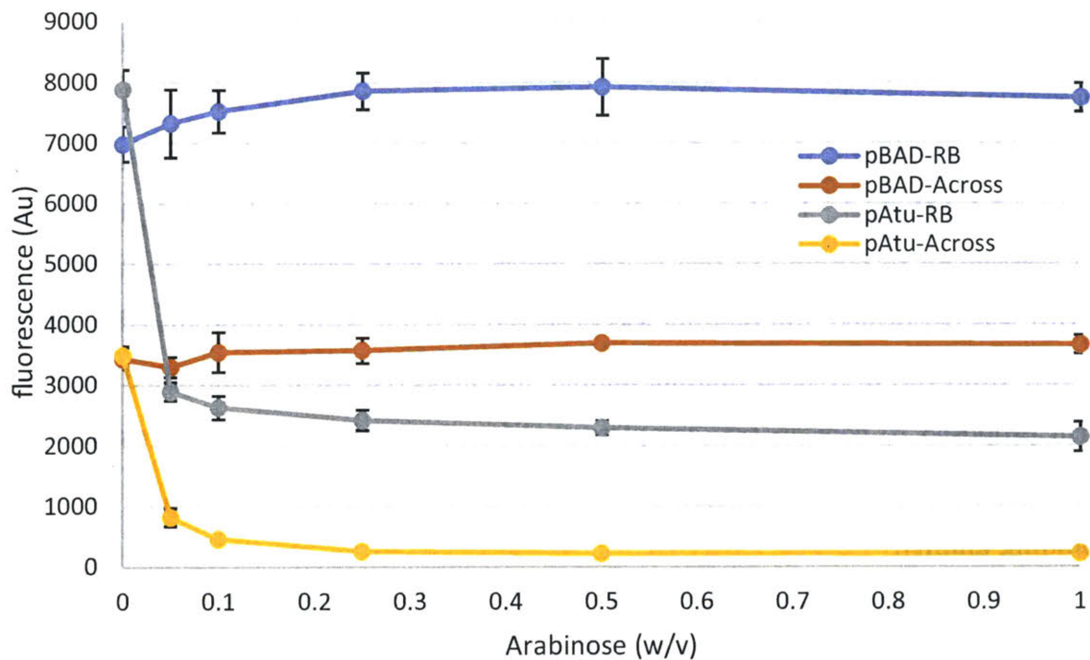

Figure 6-14: **Repressor Performance with Extended Binding Motif.** Use of the extended motif results in an 6.5-fold and 16.4-fold improvement in repression for the 'Roadblock' (pSEN101) and 'Across' methodology (pSEN102), respectively.

(pSEN101), as well as a configuration in which binding motifs were placed flanking the transcription factor binding site (pSEN102). These were expressed in the presence of the arabinose-inducible AtuR-expressing plasmid, or an empty-vector control.

As shown in Figure 6-14, use of these extended motifs improved repression significantly. A 6.5-fold change was observed for the roadblock methodology, whereas no change was observed previously, and a 16.4-fold change was improved for the flanking configuration. This result displays more than an order of magnitude change over the previous best result, yielding true repression via AtuR expression.

Despite this positive result, promoter leakiness arose as a primary concern from this experiment, as this was the biggest impact decreasing the dynamic range of the promoter. Fluorescence observed in the fully-repressed format was nearly 100-fold higher than the negative control. Thus, we next sought to address this.

6.3.3 T7 Promoter

Encouraged by the repression obtained in the section above, we next aimed to further extend the dynamic range of the repressor system via leakiness reduction. Many organisms utilize multiple carbon sources, while retaining a preferred source (such as glucose over xylose). Often, these systems rely both on a repressor system, and also large global transcriptional changes. Thus, we speculate that other factors beyond AtuR repression may influence gene expression in the native host. This is supported by the high level of conservation in the -35 and -10 regions observed in the multiple sequence alignment, which we postulate may indicate additional levels of expression control. This would impose important design implications as biosensors which respond to stressors, such as heat or the presence of antibiotics, must simply switch on in the presence of an inducer, but those for carbon source utilization may be 'leakier' as they have two modes for regulating transcription, as may be the case with AtuR.

We thus aimed to build on our system in a way that mimics this natural phenomona via the addition of an inducible, T7, promoter. As an added benefit, this promoter consists of only a single sequence, rather than two, as with bacterial promoters, allowing for the sequence to be tightly flanked by repressor binding motifs.

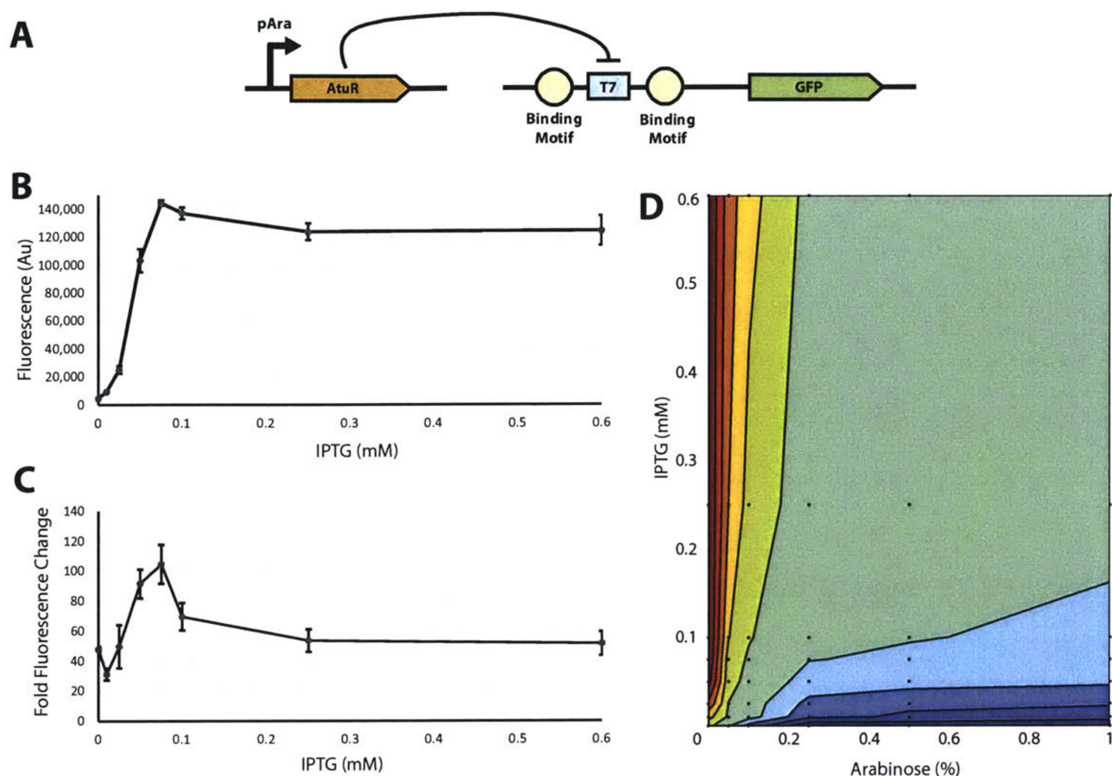

Figure 6-15: **Construction and Testing of an T7-AtuR Biosensor System.** A) Schematic of repressor system. B) Fluorescence observed in the presence of a control plasmid (pBAD33), showing full induction at 0.1 mM IPTG. C) Fold repression observed (no arabinose induction versus full arabinose induction) for each IPTG concentration, showing a maxima at 0.1 mM. D) Full dataset for repressor system for titrations of arabinose and IPTG.

As before, this plasmid was constructed (yielding pSEN201), and expressed in the presence of both an empty vector control, as well as an arabinose-inducible AtuR expression plasmid. We next characterized the repression observed across the two inducer concentrations, as shown in in Figure 6-15D. As can be seen, at low levels of IPTG, maximal expression is low, and at high IPTG levels, AtuR expression is too low to adequately repress GFP expression. As reflected in Figure 6-15B, 0.1 mM IPTG yields both full expression of the T7 promoter (in the absence of arabinose), and also the highest fold-repression change (approximately 120 fold). This change is nearly an order of magnitude better than previous constructs and thus we continued onward with this construct.

6.3.4 Determining Ligand Responsiveness

Following construction of a functional repressor system, we next aimed to characterize derepression behavior. No previous data existed on the ligand responsiveness of the AtuR protein. However, as this cluster is known to be essential to consumption of acyclic terpenes we investigated this group of molecules. The responsiveness was first evaluated *in-vitro* via an EMSA for citronellol, geraniol, their aldehydes and carboxylic acids as this is the most widely reported group of compounds. Surprisingly, molecular caging within the gel was sufficient to allow mobility shifts in the presence of acyclic terpenes. For this effect to be pronounced, EMSA experiments were run with a large molar excess of inducers, and higher initial voltages were required indicating that rapid absorption of bound (or unbound) DNA into the gel is essential to proper visualization. As shown in in Figure 6-16, we first demonstrated our ability to perform EMSA using a small (40bp) oligimer consisting of two annealed primers, rather than of the entire intergenic sequence previously used. Next, we demonstrated ligand responsiveness. As shown in in Figure 6-17, the AtuR repressor appeared to be most responsive to the aldehyde of geraniol (citral, a combination of isomers), but also responsive to geraniol and nerol, as well as its carboxylic acid, though to a lesser extent. A significantly weaker response was observed for citronellol-derived compounds.

We next sought to apply this information *in-vivo*. Following initial experiments necessary to optimize exposure to the volatile compounds, a dodecane overlay (10 µl in 200 µl total culture volume) was used for characterization in a 96-well plate format. Toxicitiy was observed beginning at 10 mM acyclic terpenes, and was more pronounced for the alcohols than the aldehydes or carboxylic acids. As a result, only concentrations up to 10 mM were considered. As can be seen in Figure 6-18, the results from the *in-vivo* assay mirror those of the *in-vitro* assay with citral eliciting the highest response.

Figure 6-16: **EMSA of AtuR Binding to Short Oligomer.** Lanes contain the motif or a control sequence added in its place; as well as AtuR *E. coli* crude lysate as a negative control

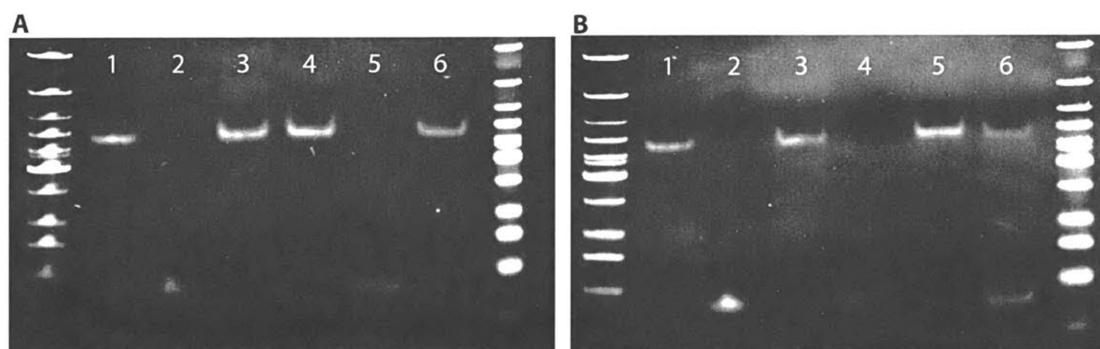

Figure 6-17: **EMSA Displaying Responsiveness to Acyclic Terpenes.** Lanes are as follows A1) AtuR and DNA control A2) DNA only control A3) Octanol Added A4) Citronellol Added A5) Geraniol Added A6) Nerol Added B1) AtuR and DNA Control B2) DNA Only B3) Citronellal B4) Citral B5) Citronellic Acid B6) Geranic Acid

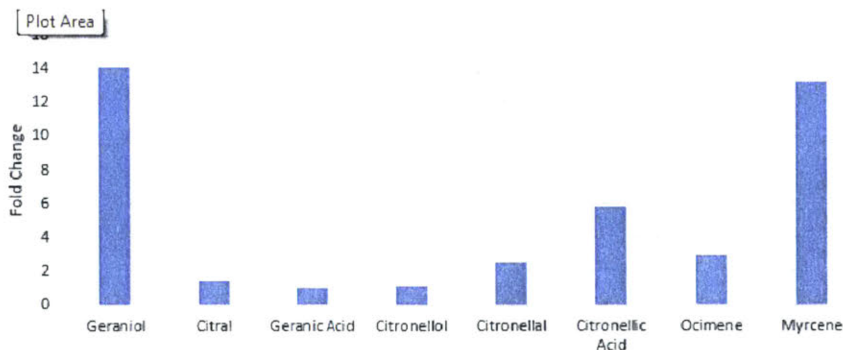

Figure 6-18: **Sensor Responsiveness to Acyclic Terpenes.**

6.3.5 Consolidation of Promoter

As not all cells are arabinose consumption knockouts, arabinose is a poor choice of inducer if one aims to have broad applicability of a biosensor. Additionally, time of repressor induction may influence sensor functionality. To remedy this, we replaced the pBAD promoter powering AtuR expression with a series of consitutive promoters [20]. As shown in Figure 6-19, below, promoters weaker than proB failed to provide adequate AtuR expression, allowing fluorescence in the absence of acyclic terpenes. Additionally, we reasoned that promoters that had excess expression may provide repression tighter than desired and thus may decrease dynamic range. As a result, we selected the weakest promoter giving near-full repression and generated a single plasmid containing the entire biosensor construct (proB). Following this, the biosensor construct was combined into a single plasmid and its performance was assayed in response to citral, as shown in in Figure 6-20. As can be seen, the biosensor responds over a range of 0 to 10 mM and has approximately a 6-fold dynamic range. Responses over a larger range may yield better dynamic ranges, but may be masked by toxicity.

6.4 Conclusions and Future work

In this chapter we have designed and constructed the first terpenoid-responsive biosensor. With geraniol as a surrogate molecule, optimization of upstream terpenoid biosynthetic processes are imaginable. These optimizations may be done on a far

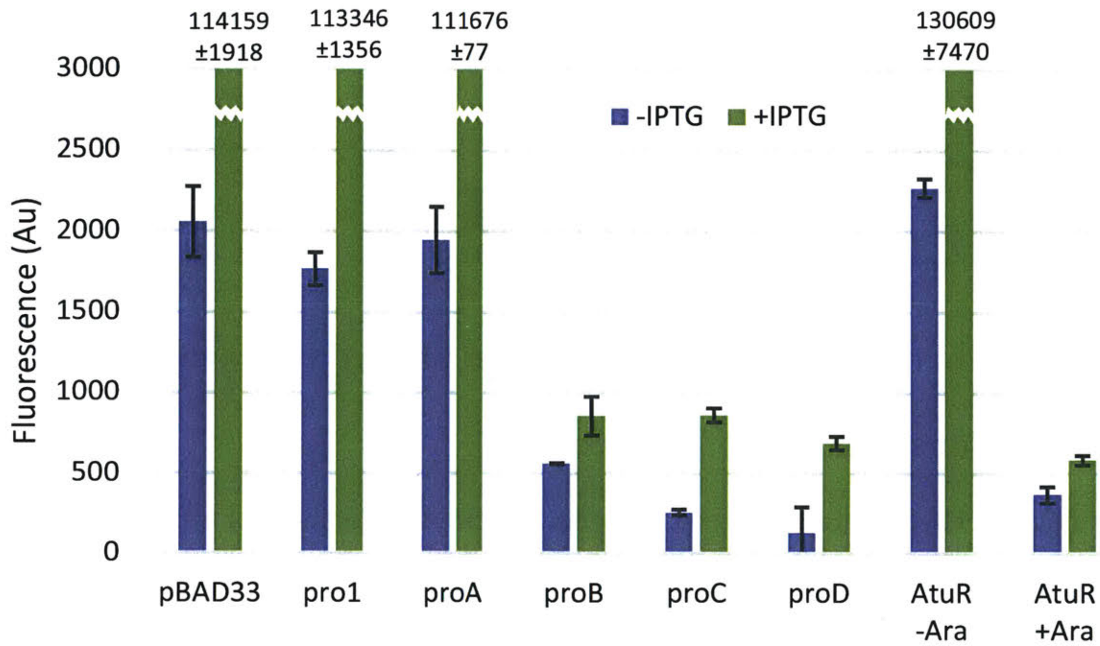

Figure 6-19: **Use of Constitutive Promoters for Repressor Expression.**

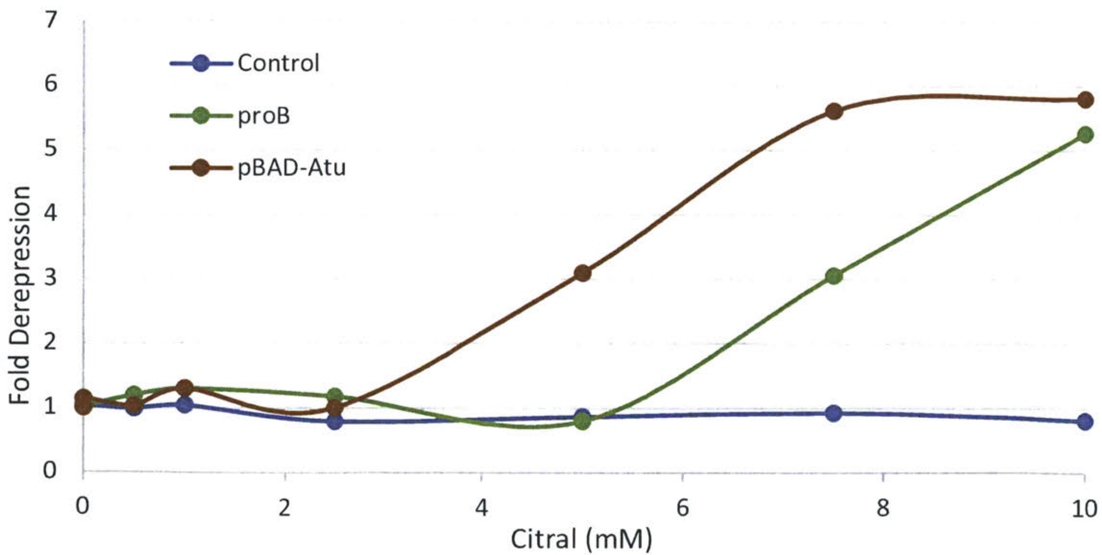

Figure 6-20: **Response of Promoter to Geraniol.**

larger number of cells than the previous lycopene-screening methods due to the ability to use flow cytometry and cell sorting to select positive phenotypes. Furthermore, this system may be used in a co-culture system as a "drop-in" solution, allowing rapid detection of geraniol within a culture produced by a cell without the biosensor construct, although this would be done in well plates at reduced throughput.

Future work in this area is directly related to application of the biosensor. Although in this chapter we demonstrated the ability of the biosensor to detect geraniol in the culture media, validation that this works when the geraniol is produced endogenously is essential. Furthermore, geraniol is a highly hydrophobic molecule and is capable of diffusion across cell membranes. Thus, flow cytometry of producer strains may result in "cheaters" appearing as positive producers due to geraniol concentration in bulk media from producers. To remedy this several solutions exist, including co-culture with *P. citronellois* to scavenge extracellular geraniol, fusion of geraniol synthase and AtuR via a linker domain to increase local geraniol concentration, and diverting only a portion of terpenoid flux to detectable terpenes such as gernaiol.

6.5 Materials and Methods

Cloning and Transformation. DNA constructs containing palindromic repeats due to the presence of the AtuR repressor could not be generated via Gibson Assembly due to the high frequency of recombination, and were thus constructed using CLIVA.

Culturing of Biosensor Constructs. Biosensor constructs were isolated as fresh colonies from plates and precultured in 3mL of LB media with the correct antibiotics overnight at 37°C and 250 RPM. Following overnight incubation, 50 ul was inoculated into 10 mL LB media in a 50 mL falcon tube containing LB and shaken at 37°C for 1.5 hours. At this time, cultures were poured into 25 mL reservoir and aliquoted to the designated wells of a 96-well microplate.

Cultures were induced at the conclusion of pre-culture and prior to distribution into wells. Arabinose, when used, was added to wells prior to addition of cells and was added as a mixture with LB with appropriate antibiotics for a 20 ul total volume.

IPTG was added similarly for a total of 10 ul to total well volume. In experiments in which all cultures were subjected to arabinose or IPTG induction, these components were added directly into pre-culture prior to distribution. Total well volume was consistent at 200 ul.

Plates were incubated in a Jitterbug Microplate Incubator/Shaker and shaken at maximum shaking speed at 37°C. Incubation time was 4.5 hours and plates were sealed with Breatheasy strips as excessive evaporation was observed from uncovered plates and high levels of anaerobicity were observed when fully capped. Exogenous acyclic terpenes were added as a portion of a dodecane overlay (10 ul in 200 ul total culture volume), to alleviate toxicity and prevent evaporation during experimental time course.

Measurement of Fluorescence. Fluorescence was measured on Spectramax II, with excitation and emission wavelengths set to 485 and 530, respectively. Plates were shaken for 5 seconds prior to reading. Plates used were falcon-brant flat bottom, non-culture treated, 96-well plates.

Final experiments were analyzed via flow cytometry in order to improve accuracy of results. Prior to flow cytometry, 2 ul of cells was added to 198 ul of cold (4°C) phosphate buffered saline (PBS) and mixed. Flow Cytometry was performed on a Canto II flow Cytomter with 10,000 measurements, gating was performed manually and was uniform samples within experiments.

Electromobility Shift Assays. Electrobolity Shift Assays (EMSA) were run as follows: Short oligonucleotides supplied as primers were mixed in a 1:1:2 ratio with annealing buffer (forward primer: reverse primer: buffer) to obtain a DNA solution which was 25 uM in each primer. This annealing buffer contained 70 mM Tris HCl (pH 8), 10 mM MgCl2, and 5 mM DTT. Primers were annealed via heating to 94°C and holding for two minutes prior to cooling to 4°C over 1 hour.

Reactions contained 0.5 uM oligonucleotides, 2.5 uM TetR protein, and 12 uM acyclic terpene inhibitor (when appropriate), 1.25 mM MgCl2, 2.5 mM NaCl2, 20 mM Tris-HCl (ph 8), 1 mM DTT, and 20 ng/ul nonspecific DNA. The reaction was then allowed to proceed for 20 minutes at room temperature.

The solution was then loaded on a 1% polyacralmide gel and subjected to electrophoresis under high voltage (240 V) for approximately 30 seconds in order to rapidly encopreate the DNA and Protein mixture into the gel, the voltage was then reduced to 60 V and the gel was run to completion. During experiments, it was determined that rapid loading of the gel and high initial voltages were essential for proper visualization. This was hypothesized to result from the dissolution of ligand from AtuR following dilution within the well if allowed to sit for too long.

6.6 Bibliography

[1] V. Radhika, T. Proikas-Cezanne, M. Jayaraman, D. Onesime, J. H. Ha, and D. N. Dhanasekaran, "Chemical sensing of dnt by engineered olfactory yeast strain," *Nature chemical biology*, vol. 3, no. 6, pp. 325–330, 2007.

[2] Y. Fukutani, J. Ishii, K. Noguchi, A. Kondo, and M. Yohda, "An improved bioluminescence-based signaling assay for odor sensing with a yeast expressing a chimeric olfactory receptor," *Biotechnology and bioengineering*, vol. 109, no. 12, pp. 3143–3151, 2012.

[3] Y. Fukutani, T. Nakamura, M. Yorozu, J. Ishii, A. Kondo, and M. Yohda, "The n-terminal replacement of an olfactory receptor for the development of a yeast-based biomimetic odor sensor," *Biotechnology and bioengineering*, vol. 109, no. 1, pp. 205–212, 2012.

[4] J. Minic, M.-A. Persuy, E. Godel, J. Aioun, I. Connerton, R. Salesse, and E. Pajot-Augy, "Functional expression of olfactory receptors in yeast and development of a bioassay for odorant screening," *Febs Journal*, vol. 272, no. 2, pp. 524–537, 2005.

[5] Y. Fukutani, A. Hori, S. Tsukada, R. Sato, J. Ishii, A. Kondo, H. Matsunami, and M. Yohda, "Improving the odorant sensitivity of olfactory receptor-expressing yeast with accessory proteins," *Analytical biochemistry*, vol. 471, pp. 1–8, 2015.

[6] C. Byrd, G. C. Turner, and A. Varshavsky, "The n-end rule pathway controls the import of peptides through degradation of a transcriptional repressor," *The EMBO Journal*, vol. 17, no. 1, pp. 269–277, 1998.

[7] Y. Lu, C. Su, O. Unoje, and H. Liu, "Quorum sensing controls hyphal initiation in candida albicans through ubr1-mediated protein degradation," *Proceedings of the National Academy of Sciences*, vol. 111, no. 5, pp. 1975–1980, 2014.

[8] B. C. Stanton, V. Siciliano, A. Ghodasara, L. Wroblewska, K. Clancy, A. C. Trefzer, J. D. Chesnut, R. Weiss, and C. A. Voigt, "Systematic transfer of prokaryotic sensors and circuits to mammalian cells," *ACS synthetic biology*, vol. 3, no. 12, pp. 880–891, 2014.

[9] S. P. Wilkinson and A. Grove, "Ligand-responsive transcriptional regulation by members of the marr family of winged helix proteins," *Current issues in molecular biology*, vol. 8, no. 1, p. 51, 2006.

[10] F. Zhang, J. M. Carothers, and J. D. Keasling, "Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids," *Nature biotechnology*, vol. 30, no. 4, pp. 354–359, 2012.

[11] W. Chen, S. Zhang, P. Jiang, J. Yao, Y. He, L. Chen, X. Gui, Z. Dong, and S.-Y. Tang, "Design of an ectoine-responsive arac mutant and its application in metabolic engineering of ectoine biosynthesis," *Metabolic engineering*, vol. 30, pp. 149–155, 2015.

[12] K. Förster-Fromme and D. Jendrossek, "Atur is a repressor of acyclic terpene utilization (atu) gene cluster expression and specifically binds to two 13 bp inverted repeat sequences of the atua-atur intergenic region," *FEMS microbiology letters*, vol. 308, no. 2, pp. 166–174, 2010.

[13] S. G. Cantwell, E. P. Lau, D. S. Watt, and R. R. Fall, "Biodegradation of acyclic isoprenoids by pseudomonas species.," *Journal of bacteriology*, vol. 135, no. 2, pp. 324–333, 1978.

[14] R. R. A. Y. Fall, J. L. Brown, and T. L. Schaeffer, "Enzyme recruitment allows the biodegradation of recalcitrant branched hydrocarbons by pseudomonas citronellolis," vol. 38, no. 4, pp. 715–722, 1979.

[15] S. Iurescia, A. M. Marconi, D. Tofani, A. Gambacorta, A. Paternò, C. Devirgiliis, M. J. van der Werf, and E. Zennaro, "Identification and sequencing of β-myrcene catabolism genes from pseudomonas sp. strain m1," *Applied and environmental microbiology*, vol. 65, no. 7, pp. 2871–2876, 1999.

[16] X. Guan, T. Diez, T. K. Prasad, B. J. Nikolau, and E. S. Wurtele, "Geranoyl-coa carboxylase: a novel biotin-containing enzyme in plants," *Archives of biochemistry and biophysics*, vol. 362, no. 1, pp. 12–21, 1999.

[17] A.-c. A. C. Homologous, R. R. Fall, and M. L. Hector, "Acyl-Coenzyme A Carboxylases. Homologous 3-Met hylcrotonyl-CoA and Geranyi-CoA Carboxylases from," vol. 16, pp. 4000–4005, 1977.

[18] B. Höschle, V. Gnau, and D. Jendrossek, "Methylcrotonyl-coa and geranyl-coa carboxylases are involved in leucine/isovalerate utilization (liu) and acyclic terpene utilization (atu), and are encoded by liub/liud and atuc/atuf, in pseudomonas aeruginosa," *Microbiology*, vol. 151, no. 11, pp. 3649–3656, 2005.

[19] K. Förster-Fromme and D. Jendrossek, "Identification and characterization of the acyclic terpene utilization gene cluster of pseudomonas citronellolis," *FEMS microbiology letters*, vol. 264, no. 2, pp. 220–225, 2006.

[20] J. H. Davis, A. J. Rubin, and R. T. Sauer, "Design, construction and characterization of a set of insulated bacterial promoters," *Nucleic acids research*, p. gkq810, 2010.

Chapter 7

Conclusions

7.1 Thesis Summary

In this thesis, methods for the optimization of terpenoid biosynthetic pathways were investigated. Initially, downstream biosynthetic pathways were studied, with a primary focus on the product Taxol. Taxadiene-5α-hydroxylase was identified as a bottleneck enzyme in Taxol biosynthesis, and several engineering strategies were employed in order to alleviate this bottleneck. Following this work, focus was shifted upstream toward production of the precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) through the biocatalytic transformation of the bulk chemical isopentenol. Finally, strategies for rapid improvement of upstream terpenoid production pathways were developed. This included a terpenoid-responsive biosensor and two evolutionary-based approaches. The key findings and accomplishments of this thesis work are summarized below.

7.1.1 Taxol Pathway Elucidation and Bottleneck Alleviation

Chapters 2, 3, and 4 focused on the chemotherapeutic drug Taxol. In Chapter 2, we developed a platform for the rapid elucidation of the Taxol biosynthetic pathway. This included utilizing an array of transcriptomic data to establish candidate pathway genes. We then developed rapid cloning and screening methodologies to evaluate

these genes. Following early successes, and *in-vitro* confirmation, we were able to reconstitute two additional Taxol biosynthetic steps in *E. coli* and *S. cerevisiae.* As we attempted to progress further down the Taxol pathway, a bottleneck in the early pathway became evident. We discovered this bottleneck arose from poor selectivity of the taxadiene-5α-hydroxylase enzyme. Although our pathway elucidation efforts focused primarily on Taxol, the methodologies developed and employed could feasibly be applied to a wide range of secondary metabolites.

In chapter 3, we investigated the mechanism of taxadiene-5α-hydroxylase in order to gain insight into the observed poor selectivity discovered in chapter 2. This was performed primarily by examining enzyme behavior in varying hosts and media, byproducts obtained from chemical epoxidation, and the products yielded from alternative taxadiene isomers. These insights were then applied in chapter 4 through three approaches. Two of the approaches, epoxide hydrolase enzyme screening and P450 engineering, were shown to be unlikely to yield results. The third approach, engineering of taxadiene synthase, proved to be a viable strategy toward correcting the selectivity of the early paclitaxel biosynthetic pathway. This approach was validated, yielding an improvement over 2.5-fold in titer for the desired product through shifts in selectivity.

7.1.2 Development of a Biocatalytic Route to Terpenoids

In chapters 5 and 6, we shifted focus toward the construction of an alternative biocatalytic route to terpenoid precursors. This began in chapter 5 with the conceptualization and a techno-economic analysis of a pathway leading from the bulk chemical isopentenol to the precursor molecule isopentneyl diphosphate (IPP). We then screened candidate enzymes *in-vitro* and constructed the desired pathway consisting of three pathway genes, choline kinase, isopentenol phosphate kinase, and IPP isomerase. Following reconstitution *in-vivo*, the pathway was shown to be functional via metabolite labeling. Next, a strain dependent on isopentenol for terpenoid production and thus cell viability was developed. Tools were then constructed for the use of an evolution-based platform for pathway optimization.

Chapter 6 then followed up by focusing on the development of a terpenoid-responsive biosensor which would have broad applicability beyond the isopentenol utilization (IU) pathway developed in chapter 5. Initially, implementation of a TetR-family repressor and binding motif previously reported in literature failed to yield a functional biosensor. However, following sequence alignment and evaluation of longer binding motifs, a functional sensor was obtained. This sensor was further optimized via alteration of binding motif location and replacement of the promoter with a phage T7 promoter. To conclude this chapter, ligand responsiveness was characterized and the biosensor was validated within an *E. coli* system.

7.2 Future Directions

7.2.1 Taxol Pathway Elucidation and Optimization

While significant strides were made in both construction of methodologies for Taxol pathway elucidation and in the study of the enzymes itself, a great deal of work remains. Pathway elucidation is, at present, limited by the acquisition of pure taxadiene and downstream metabolites. Further optimization of the upstream MEP pathway may feasibly enable production; however, due to the high cell masses required, extraction of intermediates for analysis would remain difficult. An alternative to biosynthesis is the chemical synthesis of taxadiene, as demonstrated by the Baran group [1], which would enable gram-scale quantities of pure taxadiene to be obtained. This taxadiene could then be utilized via the methods outlined in chapter 2 to enable pathway discovery. Furthermore, A promising route to production of large-scale quantities of taxadiene would arise from the utilization of the IU pathway described in chapter 5. This pathway is advantageous over chemical synthesis in that not only can taxadiene be made, but pathway steps may be added as they are discovered. This allows for production of more functionalized products than are chemically accessible. However, the IU pathway itself must first be optimized, as discussed below.

The improvement of selectivity via terpene synthase engineering discussed in chap-

ter 4 stands as an important proof of concept, yet a large portion of pathway flux is still diverted into off-pathway products. The most obvious next step for further improvement of the early pathway involves screening of larger numbers of mutants, or combinations of highly-productive single mutants from this work. Furthermore, screening of taxadiene synthases from other *Taxus* species may reveal differences in product profiles and may guide a more rational mutagenesis approach. Together, these approaches may enhance early pathway flux toward Taxol.

7.2.2 Optimization of the Isopentenol Utilization Pathway and Application of Evolutionary-Based Approaches

Future work on the IU pathway depends on transforming this pathway from a physiologically validated one to one which has industrial relevance. The development of an *E. coli* strain dependent upon isopentenol for survival, as performed in this work, stands as an important step in this direction. This strain can be leveraged in order to evolve the pathway for higher flux. This is made possible through the use of weak promoters which, when combined with the natively low enzymatic activity, causes this pathway to act as a bottleneck in cellular growth. Thus, cells which evolve more efficient isopentenol activation pathways will out compete other lineages, allowing for evolution-based pathway optimization. The vast space for improvement using only our existing strain can be seen as it takes nearly 3 days to achieve optical densities rivaling that obtained from overnight cultures of wild-type cells. Following optimization through the use of *E. coli*, it is feasible to imagine using a similar methodology in an archeal system. In these systems a larger quantity of terpenoids, and thus higher pathway flux, is required as they form the basis of cellular membranes in these organisms.

In addition to optimizing the IU pathway itself, a logical next step from the work in this thesis is to optimize the bioprocesses which would surround a functional pathway. Work toward optimal feeding regimes of isopentenol and product separation are essential in developing a continuous reactor system. Perfusion cultures may enable

high cell densities, and thus turnover, while retaining low growth rates as production is uncoupled from growth. Thus, these systems are of value for investigation as well.

7.2.3 Optimization of an Isoprenoid-Responsive Biosensor

In the construction of a terpenoid-responsive biosensor, we demonstrated up to 200-fold repression when using the phage T7 promoter. However, derepression when exposed to acyclic terpenes was much lower. This resulted in a biosensor capable of achieving only approximately a 10-fold change in fluorescence. While the reasons for this are unclear, there are many ways in which this issue may be corrected. Most obviously is error prone PCR (EP-PCR) of the AtuR protein, followed by cloning, and two successive rounds of flow cytometry. In this strategy, one would first select for sensor constructs still capable of repression, followed by ones with improved responsiveness to geraniol. This selects for constructs which both retain binding activity, and exhibit improved release kinetics. Following successive rounds of PCR and screening, it is feasible to imagine an improved sensor may be obtained.

In addition to optimizing the sensor for improved dynamic range and responsiveness, direct application of the sensor to engineering is an obvious next step. This would be performed via the creation of a geraniol-producing microorganism and co-transforming it with the biosensor plasmid. Following this, a library of cells containing the sensor construct can be used in conjunction with flow automated cell sorting (FACS). This would enable rapid screening of trillions of cells and genetic variants, allowing for random or semi-targeted mutagenesis as a practical method of pathway optimization. However, a primary concern in the application of FACS is the role of 'cheaters' within the system, or cells which report positively, through florescence, yet do not demonstrate improved geraniol production. Early experiments demonstrated that acyclic terpenes aggregate within cells and cell membranes following spiking (approximately 90% contained within the cell pellet after 1 hour), and thus high-producer strains may efflux a large portion of their production. This in turn reduces the difference between producers and non-producers. Two primary possibilities are feasible to remedy this. First, AtuR and geraniol synthase may be fused to locally increase

the geraniol concentration. Alternatively, co-culture with *pseudomonas* may be utilized, allowing for rapid scavenging of any extracellular geraniol, due to its ability to consume geraniol as a sole carbon source.

7.3 Concluding Remarks

By focusing on both the production of a single terpenoid, Taxol, as well as the production of precursor molecules, we have demonstrated methodologies to overcome obstacles to terpenoid production at any stage of their biosynthesis. For many terpenoids, including those that have simple or well characterized pathways, production of the precursor molecules IPP and DMAPP are often limiting. Considerable research has been levied to relieve this limitation both academically and industrially. Early work in this field, which still continues to a lesser extent, focuses on producing commodity chemicals. However, the economics of commodity chemical production from sugar feedstocks is highly dependent on competition from non-renewable resources. This is demonstrated, for example, by the shift of Amyris from a focus on the terpenoid based fuel farnesene, to specialty chemicals (primarily terpenes) such as lycopene and squalene. This shift raises an important question, namely that if it is economically infeasible to produce fuels, may one leverage cheap fuel-like molecules in the production of secondary metabolites. Biological systems excel at performing highly specific and chrial chemical transformations with ease. However, massive engineering efforts must be devoted in order to divert carbon to product production reliably during long culture times. By leveraging petrochemically-derived fuels in combination with the high specificity of biology, we determined it was possible to create an economically favorable pathway. We believe this work, while impactful on an academic and scientific level, truly stands to revolutionize the production of terpenoids in industry by decreasing cost, increasing scalability, and improving long-term strain stability.

While the production of IPP and DMAPP through upstream engineering is essential, it is not enough in isolation to produce the large array of terpenoids which are commercially relevant. Many terpenoids possess uncharacterized pathways, or

bottlenecks later in their biosynthesis. Fortunately, the study of secondary metabolites is a burgeoning field, one which is rapidly expanding. Two primary technologies: cheap and accurate sequencing, and increased computational power are rapidly enabling the discovery of candidate genes in a large number of biosynthetic pathways. While we were able to make use of these technologies to a degree in chapter 2, this data is not sufficient in itself. Rapid and efficient pathway elucidation requires not just computational power, but validated methodologies as well. Our development and demonstration of these methodologies rapidly accelerates pathway discovery and construction of finalized strains.

Furthermore, while the terpenoid-responsible biosensor we developed still possesses limited dynamic range, optimization of this sensor may yield a highly functional method for the detection of a number of isoprenoids. Optimization of upstream pathways has typically been limited by screening methods dependent upon GCMS, LCMS, or lycopene colorometric reporting. Development of this terpenoid-responsive biosensor enables screening of vastly more strains and opens up the ability to perform broader mutagenesis approaches. It is our hope that the optimization and utilization of this sensor will continue and greatly increase the methods available and the number of constructs which may be screened for terpenoid production.

7.4 Bibliography

[1] N. C. Wilde, M. Isomura, A. Mendoza, and P. S. Baran, "Two-phase synthesis of (-)-taxuyunnanine d," *Journal of the American Chemical Society*, vol. 136, no. 13, pp. 4909–4912, 2014.

Appendix A

P450 Sequences from Transcriptome Analysis

CYP01

ATGCCAACAAACCAAAAGGGCATATGGTATTTCAGCAGAGTCGCTAAGTAAARTAGGCTA
AGCACTCTGGAGCGCTTCATCTTGCAGCATATATATAGTCTCTRCYATTCATCTTGCAGC
AYATGTACTMCCATTCCTCTTTCCTATTCACTCCCTCCTCTCTCAGACCCACCTGCTCCA
AATGGATAGCTTCATTTTTCTGAGAAGCATAGGAACAAAATTTGGGCAGCTGGAGTCTTC
CCCTGCTATTMTTTCCCTTACCCTCGCACCTATTCTCGCCATTATTCTTCTCTTGCTCTT
CCGTTACAATCACCGATCCTCTGTTAAACTTCCCCCTGGAAAGTTAGGTTTTCCTCTCAT
CGGGGAGACCATACAATTATTGCGGACACTCCGATCAGAAACACCTCAAAAGTTTTTTGA
TGAYAGATTGAAGAAATTCGGTCCTGTTTACATGACTTCCCTAATTGGGCATYCCACAGT
TGTACTCTGCGGSCCTGCGGGAAACAAATTAGYTCTTTCGAACGAGGACAAGCTGGTAGA
GATGGAAGGGCCCAAGTCTTTCATGAAACTGATTGGGGAAGATTCCATTGTTGCTAAAAG
AGGCGAGGATCATCGCATCTTACGCACTGCACTTGCTCGGTTTTTGGGYGCTCAAGCYTT
ACAAAATTATCTGGGTAGAATGAGTTCAGAAATAGGACACCATTTCAATRAAAAATGGAA
GGGTAAAGATGAAGTGAAGGTGCTTCCTTTGGTAAGAGGGCTTATCTTCTCCATTGCAAG
CACCCTGTTTTTCGATGYAAATGATGGACACCAACAGAAGCAACTTCATCATCTTCTSGA
AACTATTCTTGTRGGRAGTTTGTCAGTCCCGCTGGACTTTCCAGGAACTCGTTATCGTAA
AGGGCTTCAGGCGCGGYTGAAGCTTGATGAAATTCTCTCCTCTCTAATAAAACGCAGAAG
AAGWGATCTGCGTTCAGGCATAGCTTCTGATGATCAAGATCTACTGTCGGTGTTGCTCAC
CTTCAGAGATGAAAAAGGGAACTCACTGACAGACCAGGGGATTCTGGACAACTTTTCTGC

TATGTTTCATGCTTCATATGACACCACTGTTGCACCAATGGCCTTGATATTTAAGCTTCT
ATACTCCAATCCTGAATACCATGAAAAAGTATTTCAAGAGCAGTTGGAAATAATTGGCAA
TAAAAAGRAAGGGGAAGAAATCAGTTGGAAGGATTTGAAATCTATGAARTATACATGGCA
AGCAGTTCAAGAATCACTACGAATGTACCCWCCAGTTTTTGGAATATTTCGTAAGGCTAT
CACTGATATTCATTATGATGGGTATACAATTCCAAAAGGATGGAGGGTTTYATGTTCACC
TTATACTACACATCTGAGAGAAGAGTACTTCCCTGAGCCTGAARAATTCAGGCCTTCAAG
ATTTGAGGATGAAGGCAGGCATGTGACTCCTTACACATATGTACCATTTGGAGGAGGCCT
GCGCACATGTCCAGGATGGGAATTTTCAAAGATTGARATAYTACTGTTTGTCCATCATTT
YGTTAAAAAYTTCAGCAGYTACATTCCAGTTGATCCCAATGAAAAAGTTTTATCAGATCC
ACTACCTCCTCTCCCTGCCAATGGATTTTCCATAAAACTTTTTCCGAGATCCTAATCCAT
GACAGAGCATTAGATCAAGATGTTGGAAATAGATGGKTTAATACGATAAAGATTGTGCCT
CAGTAGAGTTTGGCTTTTATTGCTGTCTTTTATATGTATGYTWGAAAAGTTTTCCTCTAT
GAGATTATCAACTAGCAATTACACATGAGTTGTAAATTATGTCCCAGAAGATAACATTAT
TAAGCTTGTTGAAGTTGTATTTTGTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

CYP02

CAGATTCCATGTTTTGTTTTTAGCAGACTTGATTTCACATATGTAGAAGAGATAATGGAG
AAAGTATTATATAAACTCGACCCTCTTTTATCAGTATTTCGACCACGGATAACCTCTGTT
TTCTGCTCCTGCTTTTCTTCTTCCAAAATGGACGCCCTGTATAAGAGCACAGTTGCAAAA
TTTAATGAGGTCACACAGCTGGACTGTTCCACTGAATCTTTTTCCATTGCYCTCTCAKCT
ATTGCTGGTATTCTTCTGCTTCTCCTGCTCTTCCGTTCTAAACGCCACTCCTCCCTTAAA
CTTCCTCCTGGGAAATTAGGCATCCCTTTCATTGGCGAGTCGTTTATHTTCCTKAGRGCT
CTTCGATCGAACTCGCTGGAGCAATTTTTTGACGAGAGAGTRAAGAAATTCGGYCTCGTG
TTCAAGACCTCCTTRATTGGRCATCCCACAGTAGTRCTCTGCGGYCCTGCGGGAAACCGG
CTTATTCTGTCCAACGAGGAGAAGCTGGTGCAGATGTCGTGGCCCGCTCARTTTATGAAG
CTCATGGGRGAGAATTCCGTTGCCACCAGGAGGGGTGAAGACCATATAGTTATGCGCTCT
GCTCTTGCAGGTTTTTTCGGCCCTGGTGCGCTGCAGAGTTACATTGGTAAAATGAATACA
GAGATCCAGARTCATATCAACGAAAAATGGAAGGGAAAAGATGAGGTGAATGTACTTCCT
TTGGTAAGAGAGCTCGTCTTCAACATTTCGGCCATCTTGTTTTTCAACATATATGATAAG
CAGGAACAGGATCGTCTGCATAAGCTTTTGGAAACTATTCTGGTCGGAAGTTTTGCTCTT
CCRATTGACTTGCCCGGATTTGGTTTCCATAGAGCACTCCAGGGACGGGCCAMGCTCAAC
AAAATTATGCTGTCTTTAATTAAAAAGAGAAAAGAAGATCTGCAGTCTGGATCGGCAACA
GCCACKCAGGATCTGCTCTCTGTTTTGCTCACTTTCAGAGATGACAAAGGACTCCACTC

ACCAATGAYGAGATACTCGACAACTTTTCTTCTCTGCTCCATGCCTCCTATGACACCACC
ACTTCGCCAATGGCTTTGATTTTCAAGCTCTTGTCTTCCAATCCAGAATGCTATCAAAAA
GTAGTTCAAGAGCAATTGGAGATACTTTCCAACAAAGAGGAGGGCGAAGAAATCACATGG
AAGGATCTSAAAGCCATGAAATACACATGGCAAGTAGCTCAGGAAACRCTGCGGATGTTT
CCTCCAGTTTTCGGAACATTTCGCAAGGCCATCACTGACATTCAGTATGATGGTTACACA
ATTCCAAAAGGGTGGAAGCTGTTGTGGACAACTTACAGTACACATCCCAAGGACTTGTAT
TTCARTGAACCAGAGAAATTCATGCCKTCAAGATTCGATCAGGAAGGAAAGCATGTAGCT
CCTTACACATTTTTRCCCTTCGGTGGAGGCCARCGGTCATGTGTGGGATGGGARTTTTCA
AAGATGGAGATWTTACTRTTCGTTCATCATTTTGTCAAAACTTTTAGCAGCTACACCCCA
GTTGATCCCGACGAAAAAATATCAGGGGATCCACTCCCTCCTCTTCCTTCCAARGGATTT
TCCATTAAACTGTTTCCCGAGACCATAGTCAATTGAAGGAGAAAACCACAGTGCAGAACT
GCTATTCTTGAATCCTCGCTCAAGAATAATACAAACATGCATCACCAACAATGTTTAATG
CACTCAATGCAAATTAACAGTGTGTCAGCATTGACAGTCCCCCTCTTAGCTCTAGTTGGC
TCATCTTGAACTACTATTGTATTATGGGAATGATTGTTGTTTTCCATTGTGTACTCTCCC
CTCTTTTCGTTTTTAAATTATCTTTTAACAGCTCCATGGTT

CYP03

ACGGATAATCTCTTCTGCTCTTGCATTTCTTGTACTCCAATCCAAAATGGATGTCTTTTA
TCCGTTAAAAAGTRCAGTARCAAAATTTAACGRATGTTYCCCTGCWATTMTTTYCATTGT
CCTCAGTGCTGTTGCTGGCATTGYTCTGCCCCTGCTGCTGTTCCTMCGTTCTAAACGCCG
TTCCTCTGTTGGACTMCCCCCAGGGAAATTAGGTTACCCTTTCATTGGCGAGTCGTTACT
GTTCCTGAAGGCTCTTCGATCAAACACAGTTGARCAATTTTTGGAYGAGAGAGTGAAGAA
WTTYGGGAATGTCTTCAAGACGTCATTAATTGGGCATCCRACRGTAGTTCTCTGCGGGCC
TGCRGGAAACCGGCTAATCCTGGCGAACGAGGAGAAGCTGGTGCAGATGYCGYGGCCCAA
ATCCTCTATGAAAMTCATGGGGGARAAGTCTATYACYGCCAAAAGGGGCGAAGGCCATAT
GATCATCCGYTCCGCACTGCAAGGCTTTTTCAGCCCTGGTGCTCTGCAGAAATACATAGG
CCAAATGAGTAAAACAATAGAAAATCATATTAATGAGAAATGGAAGGGMAACGACCAAGT
GAGTRTAGTTGCTTTGGTAGGAGATCTCGTCTTCGATATWTCGGCCTGTTTGTTCTTCAA
TATAAATGAGAAGCATGAACGGGAACGACTGTTTGAGCTTTTGGAGATTATAGCTGTCGG
AGTTTTGGCTGTTCCRGTGGATCTTCCCGGRTTTGCTTACCRTCGGGCACTTCAAGCACG
GTCGAAGCTTAATGCAATTCTCTCCGGTTTGATAGAAAAGAGAAAAATGGATCTGAGCTC
AGGATTAGCGASTAGCAATCASGATCTTCTTTCTGTGTTTCTCACCTTCAAAGATGACAG
AGGAAATCCAYKSAGCGATGAGGAAATCCTCGACAACTTTTCCGGGCTGCTTCATGGATC

CTATGACACCACTGTTTCAGCAATGGCCTGCGTTTTTAAGCTTTTGTCTTCCAATCCCGA
ATGCTATGAAAAAGTAGTTCAAGAGCAATTGGGGATACTTTCGAATAAATTGGAAGGAGA
CGAAATCACATGGAAAGATGTGAAATCCATGAAATATACATGGCAAGTCGTTCAGGAAAC
GYTACGATTGTATCCGTCAATTTTTGGATCATTTCGCCAGGCCATCACTGACATTCATTA
TAATGGTTACATAATTCCAAAAGGGTGGAAGCTTTTGTGGACACCATACACAACACATCC
CAAGGAAATGTATTTCAGTGAGCCGGAGAAATTCCTGCCTTCGAGGTTCGATCAGGAAGG
GAAACTTGTAGCTCCTTACACATTTTTACCCTTYGGTGGAGGCCAGCGTTCATGTCCAGG
ATGGGAATTTTCAAAGATGGAGATTTTACTGTCGGTGCATCATTTTGTTAAAACATTCAG
CAMCTTCACCCCASTTGACCCAGCAGAAATAATTGCWAGAGATTCCCTCTGCCCTCTCCC
TTCCAATGGGTTTTCTGTAAAACTTTTTCCTAGATCCTAKTCACTTYACAYAGGCAACCA
GGTCAAGAAAATATAATTATATATACCAACATTCAAAGAACTGGTATTGAAGAAATATGT
ACACGTGCCCATTAGTTTTTGTGTGTGAAAACATGTACAAAGTATTTACCTGACACCATT
AGAATCATGCATGGTAGATAAAAGAGCATGCATTTGTATTTGAAGCCTTACTCTTCATAT
AGGATTATATTATTTTTCAAAAAAAAAAA

CYP04

ATTCTCCAATTATTGAACGATATATGAAGTAAGTAGCTAAGTTCAACCTCTGCTATTTGC
TGGTTTTCATTTTTATAATGGACGCCATGGATCTCACAGTTGCAAAGTTTAAGGAATTYA
CGCAGCTACAGTCCTCTGCTATTCTTCTCACTGYTGTTTCTGGAATCATCGTCATCGTAA
TCCTGCTCCTCCGTTCTAAACGCCGCTCCTCTCTCAAACTTCCTCCGGGGAAATTAGGCC
TCCCTCTCATTGGGGAATCGTTATCATTCCWGTGGGCTCTTCGATCAAACACACTCGAAC
AGTTTGTGGACAAAAGAGTGAAGAAATACGGCAATGTCTTCAAGACATCGTTACTTGGGC
AACCYACAGTAGTACTSTGYGGCGCAGCCGGAAACCGCCTAATTCTGTCGAACCAGGAGA
AGCTGTTGAGCCGAACGGTGTCGGATCGAGTAGCGAAACTGACGGGTGATRMTTCTATTT
CGGTTATAGYGGGAGACAGTCATCGCATCATACGCGCAGCAGTTGCAGGRTTTTTGGGGC
CRGCAGGACTCMAGATTCACATTGGCGAAATGAGYGSACATATCCGAAATCATATCAACC
AAGTATGGAAGGGAAAAGATGAAGTGAACGTACTTAGTTTGGCAAGAGAGCTGGTCTTCG
CCAYCTCGGCCAGTTTGTTTTTAAATATAAATGATAGAGAGGAACAGCACCAATTGCATA
AGACTCYCGAAACTATTCTTSCCGGATATTTTTCTGTTCCTATAAACTTCCCCGGATTTG
CCTTTCGYAARGCACTGGAGGGAAACTCGAAGCGTAGGAAACATTTCTCTGTTTTACWAG
AAAAGAGAAGAAGGGATCTGAGCKTAGGGTTAGCATCCCGCAMTCAGGATCTGCTYTCTG
TTTTGCTCRCCTWCRAAGATGACAAAGGGAATCCACTCACSGATGAGGASGWCCTCGACA
ACATTTCTGCGCTCATTGATGGCTCCTACGAGAGCACCTCTTCACAAATGGCCRTGCTTT

TAAAGCTGTTGTCTGMCCATCCASAATGCTATGAAAAAGTAGTTCAAGAGCAATTGGAGA
TAGCTTCACATAAAAAGGAAGGAGAAGAAATCACATGGAAGGATGTGAAAGCCATGAGAT
ACACATGGCAAGTAATGCAGGAGACGCTGCGGATGTTYGCCCCTGTTTTTGGACCTCGTG
GGAAAGCTATAACTGACATTCATTATGACGGTTACACCATTCCAAAAGGATGGCAACTTT
CATGGGCAACTTATTCRACCCATCAGAATGATACATATTTCAATGAGCCGGACAAATTCA
TGCCGTCCAGATTCGACGAGGAAGGAGGGCGTTTGGCTCCTTACACATTCGTGCCATTTG
GAGGAGGGAGAMGGAMATGCCYAGGATGGGAATTCGCAAAGACTGAGATATTACTGTTCG
TCCATCATTTTGTTAAAACATTCAGTGCCTACACSCCAATCGAYCCTCACGAAAGTATTY
GGGGGCGTCCACTCCCTCCTGTCCCTGCCAATGGATTTCCTATTAAACTTATTTCTCGAT
CCTAAGATTAACTTCCGCTACAGGGGTCTGAACTTTTCCTTATAATGACAGAATAGTGAT
ATCGTCAATTTTAGTTTTATTGTCCATGTTTGTTCAGTTTTTTGTAAGTCTTTTTTTTTT
ATCGACACTAATGGATGTTTTACACGCTGTTTTGTTATTGTGGGTTTTGTTAATAAACAA
TACCTTCGCATT

CYP05

CGCATTATCACATGGAACTCGTTTATTCTACCGCTTCCCCCAATTGCAATTATTTTATTC
CTGTTATATATCACCGGTCCTCCTGTTATCCTCCAACCAAAAACTCTCGTCTMMGCTCTG
CAAACTCAGTCTCTCTCCCGTTGCTCTGTTTTCTGCTCAGCAACTCSTTTCTGAAACGAA
AATGGATGCCCTTAAGCAATTGGAAGTTTCCCCTTCCATTTCTTTTCGTTACCCTCGCAG
TAATGGYAGGCATTATCCTCTTCTTCCGCTACTAAACGCCATTCCTCTGTWAAACYYCCC
CCTGGAAATCTAGGCTTCCCTCTGGTTGGGGAGACAYTGCAGTTCGTGAGGWCACTYGGC
TCGAGCACTCCACAGCAGTTTATTGAAGAGAGAATGAGCAAATTTGGGGATGTGTTCAAG
ACTTCCATAATCGGGCATCCCACAGTAGYGCTSTGTGGACCTGCCGGAAACCGGTTGGTT
CTGTCGAACGAGAACAAGCTGGTGCAGATGTCATGGCCGAGCTCCATGATGAAACTCATY
GGCGAAGATTGTCTCGGCGGCAAAACGGGAGAGCAGCATCGGATCGTACGCGCTGCACTA
ACTCGGTTTTTGGGTCCTCAAGCATTGCAGAATCATTTCGCTAAAATGAGCTCGGGAATC
CAACGCCACATCAATGAAAAATGGARGGGAAAGGATGAGGYCACTGTACTTCCTTTGGTA
AAAGACCTCGTCTTCTCCGTCGCAAGCCGCTTGTTTTTTGGTATAACTGAGGAGCACCTG
CAGGARCAACTTCATAACTTGTTGGAAGTTATTCTTGTGGGATCTTTTTCTGTTCCACTC
AACATTCCCGGATTCAGTTACCATAAAGCGATKCAGGCAAGGGCYACCCTCGYTGACATC
ATGACCYMTTTGATAGAAAAGAGGAGAAATGAGCTGCGTGCAGGCACTGCATCTGAGAAT
CAAGATTTRCTCTCTGTTTTGCTCACTTTCACTGACGAAAGGGGGAATTCACTGGCGGAC
AAGGAGATCCTCGACAACTTTTCTATGTTACTTCATGGATCATATGACTCCACCAATTCC

CCACTTACCATGTTGATTAAAGTCTTGGCCTCCCATCCAGAAAGCTATGAAAAAGTGGCT
CAAGAGCAATTYGGAATACTCTCCACYAAAATGGAGGGAGAMGAAATTGCTTGGAARGAC
CTGAARGAGATGAAATATTCATGGCAAGTTGYTCAGGAAACATTGCGCATGTWTCCTCCC
ATTTTTGGAACATTTCGCAAAGCCATCACTGACATTCATTACAATGGTTATACAATTCCA
AAAGGATGGAAACTTTTATGGACAACTTACAGTACTCAWACCAAGGAAGRGTATTTCAAR
GRCGCCGATCAATTCAAGCCATCAAGATTTGAGGAGGAAGGGAAGCATGTAACCCCTTAC
ACATACTTACCTTTCGGAGGAGGCATGCGTGKTTGTCCAGGGTGGGAATTCGCCAAGATG
GAGACATTACTGTTTCTCCATCAYTTTGTTAAAGCCYTCTCTGGGTTGAAGGCAATTGAT
CCAAATGAAAAACTTTCAGGGAAACCACTTCCTCCTCTCCCTGTCAATGGGCTTCCCATT
AAACTCTATTCCAGATCTTAATCTTAAGCCAGTTCCATAAAAAAGACCTATATTATTATA
TGAATAGAAAGTGTAATACCTGTAACTGGCTTTACCATCGGAGTACCGTGACTTGTATTC
ACGAGGTATGTGAAATTCAAAACAGGAACACAAGATGATTGCTCATGCTAAACTTATCTG
TAACTGGGTTTTCTATATGTAATTTCTATTATAAATAAGATTACTGGTTTGGGTTCTTAA
AATAAAAAAAAAAAAAAAAAAAAAAA

CYP06

GTCTCCTCCCTCAAACTCTGCTGGTCGAAATGAATACCTTCAGTTTCCTCGAGTTCATTC
AACTCCAGTCTATTCTATTCCTTACCCTTCTTGCCCTTCTTCTTCTCTGCCTATTCCGAT
ATAAAACCCAGTCCTCTGCTGTTAAACTTCCCCCTGGAAACTTAGGGTTCCCTTTCATTG
GGGAGACAATACCATTCTTGAGGGCACTTCACTCAGAAACACCGCAAACCTTTTTCGATG
AGAGAATGAAGAGATTCGGTAATGTTTTCGTGACTTCGCTAATTGGACAGCCAATAGTTG
TACTCTGCGGGCCTGCCGGAAACCGCTTACTGCTCTCCAACGAGGACAAGCTGGTGGAAA
TGTCACCACCCAAGTTTTCGCTGAAGCTGATCGGCCAGGATTCCCTTCTGTCCAAAAGAG
AAGACGAGCATCGCACCCTACGCGCTGCACTCGCTCGCTTTCTGCGACCCCAAGCCCTTC
AGAGTTATATGGGTATAATGAGTTCAGAAATCGAACACCATATCAATGAAAAATGGAAGG
KAAAAGATGAAGTGAAGATGCTTCCTTTGATAAGAGGACTTATCTTCTCCATTGCAACCA
CCTTGTTTTTCGATATCAATGATGAACACCTCAAGGATCGCCTTCATCACCTTTTGGAAA
CTATTCTTGTGGGAACTGTGTCTCTTCCGCTCGACTTYCCAGGAACTTCTTTTCGWAAAG
CTGTTGAGGCGCGGTCGAAGCTCGATGAAATTCTTTCTTCTTTAATAAAAAGCAGAAGAA
GCGATCTGCATTCAGGCATAGYTTCTGATGGTCAAGATCTGCTCACAGTCTTGCTCACCT
TCAAAGACGAAAGAGGGAATTCACTCGCCGATAACGAGATCCTCGACAACTTTTCTCTTA
TGCTTCACGCCTCCTATGAGACCAGTGTTTCACCAACGGTTTTGATGTTTAAGCTGCTGT
CCTCCAATCCTGAATGTTATGACAAACTAGTTCAAGAGCAATTGGGAATACTTGCCAATA

AAAAGGASGGRGAAGACATCAGCTGGAAGGATGTGAAAKCTATGAAATATACATGGCAAG
TAGTTCAGGAAMCAATGAGGATRTTCCCTCCAGGTTTTGGATCATATCGCAAGGCGATCA
TCGATATTGATTATGATGGTTATACAATTCCAAAAGGATGGCAACTTTATGACGCCTTAT
ACTACACATGGGAAAGAACAGTACTTCCATATCCCGAACAATTCAGGCCTTCAAGATTTG
AAGAAGGAGAACTTGTAGCTCCTTACACATTCTTACCATTCGGAGCAGGCGCGCGCATAT
GTCCAGGATGGGAATTCGCAAAGACAGAGATATTACTGTTCGTCCATCATTTTGTTAAAA
ATTTCAGTAGCTACCACTTAGTTGACCCCAAGGAAAAAATATCAGGGAATCCATTCCCTC
CTCTCCCTAACAATGGATTTTCTCTGAAAGTATTTCCCAGATCGTAATCAATTAAAGGGG
GGAGACCAACATGTTGAAAATAGCGGTGTAAATACAATAAACATTGTGCTTTAAAATTAC
TCTAGCTTGATAGTTATGTTACACCTTGCCCAGATTTTTTATTTAATTCCTGGCGGCAG
AACAACTTTCCAAGGGCTTTGGACAGTTGTTTGTAGGAAAGTTGTCTGCATTAAACTAAA
ACTTTCCTTCTGTGTAGATGCTGATCAAGCTGTGGATTCTTGAGGAAATTTTTTATTTAT
GAATATTGCAACCATCATTTTTTGTASKACTTCAAAAAAAAAAAAAAA

CYP07

ATGGATACCTTAATTCAGATCCAGTCTTCCCCTGATTTCCTTTCCTTTACTCTCACAGCG
TTTCTGGGCGTTGTTGTGCTTTTGATATTCCGTTATAAACACCGATCCGCTCTCAAACTT
CCACCTGGAAATTTAGGCTTGCCCTTCATTGGGGAGACAATAACATTTGCATCTCAACCT
CCTCAGAAGTTTTTAAACGAGAGGGGGAAGAAATTTGGTCCTGTTTTCAAGACGTCGCTA
ATTGGGCATCCCACAGTTGTTCTCTGCGGCTCCTCGGGAAACCGTTTTCTCCTCTCCAAC
GAGGAAAAGCTGGTGCGGATGTCTTTGCCCAACTCATACATGAAACTCCTGGGGCAGGAT
TCCCTTCTGGGGAAAACGGGASAGGAACATCGGATTGTGCGTACCGCACTAGGACGTTTT
TTGGGCCCCCAAGAGTTGCAGAATCATGTGGCCAAGATGAGTTSAGACATTCAGCATCAC
ATCAACCAAAAATGGAAGGGGAATGATGAAGTGAAGGTGCTTCCTCTGATAAGGAACCTY
GTCTTCTCCATTGCAACCAGCTTGTTTTTCGGTATAAACGATGAGCACCAACAGGAGCGA
CTTCATCTTCTTTTGGAAACTATTGTAATGGGAGCTGTGTGTATTCCGMTCGCCTTTCCA
GGATCTGGTTTTCGTAAAGCGCTTCAGGCACGGTCGGAGCTCGATGGAATTCTCATTTCT
TTAATGAAAATCAGAAGAAGCGATCTGCGTTCAGGCGCAGCYTCAAGCAACCAAGATCTA
CTGTCGGTGTTGCTCACCTTCAAAGATGAAAGAGGRAATCCATTGACAGACAAGGAGATC
CTCGACAACTTCTCTGTTCTACTTCATGGYTTATATGACACCACAATTTCACCACTCACC
TTGATTTTTAAGCTCATGTCCTCCAATACTGAATGCTACGAGAATGTAGTCCAAGAGCAA
TTAGAAATACTTTCCMATARAGAGAAGGGAGAGGAGATCGGTTGGAAGGATCTGAAATCT
ATGAAATATACYTGGCAAGCCATTCAGGAAACCTTGAGAATGTTCCCTCCGGTTTACGGA

AATTTTCGYAAGGCTTTGACTGATATTCATTACGATGGCTATACAATCCCAAAAGGGTGG
AGGGTTTTATGTTCGCCTTTTACCACGCACAGCAATRAAGAATATTTTAATGAGCCAGAT
GAATTCAGACCTTCAAGATTCGAGGGGCAAGGAAAGAATGTGSCTTCTTACACATTCATA
CCGTTCGGAGGMGGCCTGCGCATATGTCCAGGATGGGAATTTGCAAAGACAGAGATGTTA
CTGTTTATACATTATTTTGTTAAAACTTTCAGCAGCTACGTCCCAGTTGACCCCAACGAA
AAGATTTCAGCAGATCCGCTCGCTTCTWTCCCTGTTAATGGATTYTCCGTAAAACTTTTT
CCAAGGTCCTAATCCAATTGAAGGGGGCGACCACAACCGACATGCTYGGAMRTGTTGAAA
ATTGACAGTTTAATAGAATAAACATTGTGCTTTGAATTCATTTTAGCTTCATAGTTGNNN
NNNNNNNNNNNTTTTGGAGCATTCTCATTTACTGACATTGGTTAATACTTGATCATTGAA
TTCGATGGAAATCAACCAGCAATATCTGATTCGCCA

CYP08

ATATATCTCGAGTAATAGAAGATCAAGATCCATGGAGGTGGCTCTTTCCCCAGTCAGAGA
AGTTTTACTGTGGGGGCTTTCATGGCTTTGTCTGTACATTGCATCTAGATATGTTCTGAG
TAAGAGTAAAAAGCTTCCTCCTGGTCCTTCGGGATGGCCTGTTTTAGGCTGCTTACCATT
GTTAGGCTCAATGCCTCATGTCGCTCTGACAAATTTGTCAAAGAAATATGGCCCAATTCT
TTATCTAAAATTGGGTACCTCTAATATGGTGGTGGCCAACACTCCTGCTGCTGCGAAGGC
TTTTTTGAAGACTCTGGACATTAATTTTTCCAACAGGCCTGGCAATGCTGGGGCTACTTA
TTTGGCATACAACTCTCAGGACATGGTGTGGGCACCCTACGGTGGAAGATGGAAGATGCT
CAGAAAGGTGTGTAATCTCCACATGCTAGGAGGCAAGGCCTTAGATGACTGGCAGCCTGC
GAGGGAAGCCGAGATGGGTCACATGTTGAGGTCAATCCTCGCTCGTTCCCGCAGTGGGCA
GACTGTGAACATCCCTGAACTGCTCAACCTGTTGGCAGCCAACGTTTTTGGGCAGATTAT
CCTGAGCAAGCGAGTTTTTGAGAGCGAGGGGGAGGATGCCAACGAGTTCAAGGAGATGGT
AGTCGAGCTCATGACTACCGCGGGGTATTTTAACATAGGCGATCATATCCCCTCCATTGC
ATGGATGGATTTGCAGGGAATACAGCGGGGTATGAAGAAATTACATAAGAGATTTGACGA
AATGCTAACCAGGATGATTGCAGAACACCAGGCGGCGGCCAAGAAGCGAGCTGCTCCAGA
CTTCTTGGACATTGTAATGTCCCAACGGGATAACTGTGATGGCCAGGGAGGAAGGTTGTC
TGACGTTCACATCAAGAGTCTACTTTTGAATTTGTTCACAGCGGGGACCGACACTTCAAG
TAGCGTTATAGAATGGACGCTTGCAGAGTTAATTCAAAACCCTAAGCTTCTTAAGCAGGC
ACATTTWGAGATGGACAAAGTGGTTGGTCGTGATAGGCGACTTAAAGAGTCAGATATTCC
AAAGCTCCCCTACCTRATAGCCATTTGCAAGGAAGGCTTCCGGAAGCATCCCTCCACTCC
TCTCAGCCTACCACGAGTCTCTTCTGAGCCTTGTCAAGTAGGTGGCTACTATGTGCCCAA
GGGAACTCGTCTCATGGTCAACATTTGGGGCATCGGTAGGGATCCTGAAATATGGGACAA

ACCCCTTGAATTCAATCCAAACAGGTTTGTGGGATCCAAAATTGATCCACGTGGAAATGA
CTTTCAACTCATACCTTTTGGTGCGGGWAGGCGTATTTGTGCTGGTACTAGGATGGGGAT
TACCATGGTGGAATACAATTTGGGGTCGTTGATCCATGCTTTTAATCTGGGAWATGCCAC
CAGGTAAAGAGAGCTTGAAACATGGATGAGTCTTATTGGACTTGCGCTACAAAAGAAAGA
GCCCCTTGTGGCCAAGGCTACCCCTCGCCTTGCCCTTACACCTATACCTAGATATGCACA
AAATTGTTCTAGGTTGGAATAGCTTATGAATATTGTCAAACCCATCTATCATTGCTGATA
TGCATGTTCATTTATCTTGCTAGATGTTCAAAGAATATCTCTACCTTGCTCAACCGCCTA
TATTAATAACATACAAAGAGTGATATCCTTCTNTATTGGACTGTTGTTATACTCCAAAGA
GACAGAAATACCTCGGCACTGCTGACCT

CYP09

TGTGCGGCAATGAGGGAAACCGGTTTATTTTCGCCAACGAGGAGAAGCTGATGCACCTGT
CGTGGTCCGGTCGATATGCGAAAATCCTTGGTGGGGAATCCGTTTCCATGAAGAGGGGAG
ATGATCATCGCAGTGTACGTGCCGCATTCGCAGGGTTTTTGAGCCCTTCGGCGCTGCCTC
TTTACGTAGCTAAAATGAGTGCGCAGATCCAAAATCATATCAACGAGAAATGGAAAGGAA
GAGATGTAATTGCTGTAGTTCCTTTGGTAAAAGAGCTGGTCTTCAACGTTTCCTCCAACT
TGTTTTTCGGCATGAATGATAGCGAGGAACTGCATCGATTGCATCAAATTTTCGAAACTA
TCGTAGATGGACATGTTTCCATGCCGATAGACCTTCCCGGATTCACCTTTCATAAAGCAC
TTCAGGGAAGGGCGAAGCTCAAGAAAGTTTACTCTTCTTTAATAGAAAGGAGAAGAAGCG
ATCTGAGCTCCGGATTGGCATCTGCTAATCAGGATCTCATCTCTGTTTTACTCACCTACA
AAGATGATAGGGGGTATACAATGACCCACAACGAGCTCCTCGACAATTTTCTTTCCCTTA
TTGAATCCTCCTATGATTCCATCAATTCACCAATGGCCTGCCTTTTTAAGCTTTTGTATG
CCAATCCAGAATGCTATGAAAAAGTAGTTCAAGAGCAATTGGGGATACTTTCTGGTAAGA
AGGAAGGASARGAAATCTCRTGGAAGGATCTGAGATCCATGAAATACACATGGCAAGYAM
TTCAGGAAACGCTACGAMTKTATACTCAAGTTGCTGGAATATTTCGCAAAGCCATGACTG
ACATTCATTATGATGGTCACACCATTCCCAAAGGGTGGCAACTTCTTTGGGCAACCCAAA
CTACACATCTGAACGACAAATATTTCAGTGAGCCKGAAAAATTCATGCCTTCCAGATTCG
ATGAAGAAGGAAACAATGTGATTCCTTACTCATTCGTACCATTTGGAGGAGGGCGGCGGA
TGTGTCCAGGTTGGGAATTCGGAAAGATGGAGMTCTTACTCTTYGTCCATCACTTTGTTA
AARCGTTCAGTGGCTTTACCCCAATYGATCCGRACGAAAAAATTACTGGGAATCCWTTTC
CTCATCTCCCTGCCAATGGATTTYYAATAAAACCTAYTCYCAGATCCTAATACAGACACT
ACTCGTTGTTGTTCACGTTTCATACAAATAATGAATCTGATCAGCGTGTGGAGTTATTGT
ACTTCAGCTTCTTGTTATGAATAAATAATGGATGTAGTCATGTGATCTCAAGTTATTCCA

ATCACAAAGTATAAATACAAGGTCCATATTATTCT

CYP10

CAAATGTTATGTGCAGAGAGACAGTAGAGACAGAGAGTCGTCTGGTATCTTGTTACGCTC
ATCCAAATGGACAGCTTCRCTTTTGTAACCACCAAAATGGGAAAAAWTTGGCAAGTCATT
CAGGTGGAGTACATTCTATCCCTTACCCTCACAGCTATTCTTCTCTTCTTCTTCCGTTAC
AGAAACAAATCCTCTCATAAACTTCCCCCTGGAAACTTGGGCTTCCCTTTTATTGGGGAG
ACCATACAATTCTTGCGGTCACTTCGATCACAAACACCTGAATTKTTTTTTGACGAGAGG
GTGAAGAAATTCGGTCCTGTTTTCAAGACCTCGCTAATTGGGGCTCCCACAGTGATATTC
TGCGGGGCGGCAGGGAGCCGATTAGTTCTGTCTAACGAGGACAAGCTSGTGCAGATGGAA
TCGCCMAGCTCTTTAAAGAAGCTAATGGGGGAGAATTCCATTCTGTATAAAAGAGAAGAG
GAACACCGCATCTTGCGTTCTGCATTATCCCGCTTTTTGGGTCCCCAAGCTTTGCAAACK
TACATTGCTAAAATGAGTACAGAAATCGAGCGTCATATCAACGAAAAAATGGAAGGGAAAA
GAWGAAGTGAAGACGCTTCCTTTGATAAGAGGGCTCGTCTTCTCCATTGCAAGCAGTCTG
TTTTTCGATATAAATGATGAGCCCCAACAGGAGCGACTTCATCATCATTTGGAAAGTCTT
GTTGCAGGAAGTATGGCTGTTCGCCTCGACTTTCCAGGAACTCGCTTTCGTAAAGCCGTT
GAGGCGCGTTCGAAGCTGGATGAAGCTCTCCATTCTTTAATAAAAAGCAGACGAAGCGAT
CTGCTTTCTGGCAAAGCTTCAAGTAATCAAGATCTTCTTTCGGTGCTGCTCAGCTTCAAA
GATGAAAGAGGAAATCCACTGAGAGACGAGGAGATCCTCGACAATTTTTCTCTTATACTT
CATGCCTCGTATGATACCACTATTTCACCAATGGTTTTGACATTGAAGCTGCTGTCCTCC
AATCCAGAMTGCTATGACAAAGTAGTTCAAGAGCAATTTGGAATACTTGCCAATAAAAAA
GAGGGAGAGGAAATCAGTTGGAAGGATCTGAAAGCTATGAAATATACATGGCAAGTAGTG
CAGGAAACACTGAGGATGTTCCCTCCACTTTTTGGATCATTCCGCAAGGCTATGGTTGAT
ATT

CYP11

ATCCTYTGCATCACTAATCCTTGCTCTTCAGCCCTTCCTCATCTCTGCTAATCCAAATGG
ATTCCTTMAGTTTTCTAAAAWGCATGGAAGCGAAATTCGGCCAAGTCATACACCGGGATC
AGTCTTCCAGTACTGCTCTTCTGTCCCTCGCATTCACAGCTGCTSTTGCCATTTTTCTTG
TGTTGCTCTTTCGATTTAAAAGCCGGYCCTCTACTAAKTTCCCTCCAGGAAATTTTGGCT
TCCCTTTCATTGGAGAGACGATACAGTTCTTGCGGGCACTTCGATCAGAATCGCCTCATA
TGTTTTTTGATGAGAGATTGAAGAAATTYGGGCGTGTATTCAAGACGTCATTAACTGGGC
ATCCCACAGCTGTGTTCTGCGGGCCTGCGGGAAACCGGTTTATTTACTCGAATGAGCACA
AGCTGGTGCAGTCGTCTGGGCCCAACTCCTTCGTCAAACTGGTTGGCAGCAATCCATCG

TGACCAARACAGGAGAGGAGCACCGCATCTTTCKTGGTGTCCTGAACGAGTTTCTGGGGC
CTCATGCCTTACAGAGTTATACGCCTAAAATGAGTTCCAAAATCCAGGAGAATATCAATA
AGCATTGGAAGGGTAAAGATGAAGTGAACATGCTTCCTTCGATAAGACAGCTCGTCTTCT
CCATTTCAAGCAGCTTGTTTTTTGATATTAATGATGAGGATCAACAGGAACAACTTAAAA
CTCTTTTAGAAACTATTCTTGTKGGAACTTTGTCGGTTCCCCTCGACATTCCAGGATCTA
ATTTTCGTAAAGCTCTTCGGGCGCGTTCCAAGCTGGATGAAATTCTGTCTCGTTTAATCG
AAAGCAGAAGAAAAGATATGCGTTCTGGGATAGCTTCTACCAGTAAAAATCTACTGTCGG
TGCTGCTCGCCTTCAAAGATGAAAGAGGGAATCCATTGACGGACACGGAGATCCTCGACA
ACTTTTCTTTTATGCTTCACGCCTCATACGACACCACCGTTTCGCCCACRGTTTGTATAT
TTAAGCTGCTCTCCGCCAATCCAGAATGCTATGAAAAAGTAGTTCAAGAACAATTGGGAA
TACTTGGCAATAAAAAGGACGGTGAAGAAATSTGTTGGAACGATCTGAAAGCTATGAAAT
ATACATGGCAAGCAGCTCAAGAAACAATGAGGCTTTTCCCTCCAGCGTTTGGATCATTTC
GCAAGGTCATCGCCGATATTCATCATGATGGCTATATAATTCCCAAAGGATGGAAAGCTA
TGGTGACAAATTACAGTACAAGTAGGAAAGAAGAGTACTTCGATGAACCAGACAAWTTCA
AGCCTTCAAGATTTGGGGATGGAAAGTATGTGGCTCCGTACACRTTCTTACCTTTCGGGG
CAGGCATACGCATATGCCCAGGATGGGAGTTCGCTAAGTTGGAGATGTTACTGTTCATCC
ATCATTTTGTCAAAAATTTCAGCGGATACCTCCCACTTGACACCAAGGAAAAGATTTCCG
GAGATCCATTCCCTCCTCTCCCCAAAAATGGATTTCCCATTAAACTATTTCCGAGAACCT
AACTATTTCTACAGCTAGATTCCTCTTCTAACTTTCCGCTATGCTATTATCAATAAAAAC
AACATGTTACTTAATGTGCCAMATTTGATMTCATTAGAGAATTTATAAAACAAGGGCTGC
GTTGCAGCAGTAGTGCCTGCTGTAAAAGCTTGAAAGCAAATTGTTTATATGCGAAGCAAG
GTATTGGTAAGGAATTGAGAACTGAGTGAGATTACTCCTGGTTGCGCTGTTTTCATTTGT
TAAATCTTAACTGAAATTGACAATGCCTCTGCAGAAGCCTTCCGTAGC

CYP12

ATGCTTATCGAAATGGATACCTTCGTTCAGCTCGAGTCYTCCCCTGTTCTTCTTTCCCTT
RCCCTCACACTTATTCTTCTTTTTATATTCTGTAGTAAACAATACAGATCCTCTCTTAAA
CTTCCCCCTGGAAACATGGGCTTCCCTCTCATTGGGGAGACGATAGCACTGGCATCACAG
ACACCYGATAAATTTTTCGGCGATAGAATGAAGAAATTCGGCAAGGTTTTCAAGACTTCG
TTAATTGGGCATCCCACAATTGTGCTCTGYGGTTCCTCCGGAAACCGTTTTCTCCTCTCC
AACGAGGAAAAACTGGTRCGGATGTTTCCGCCCAACTCAWCCAGCAAACTCCTGGGGCAG
GATTCTGTTCTGGGKAAAATAGGAGAGGAGCATCGGATTGTACGTACCGCACTAGCCCGC
TGTTTGGGCCCCCAAGCGCTGCAGAATTACGTGTCCAAAATGAGTTCAGAGATCCAACGT

CATATCAACCAAAAATGGAAGGGAAAAGGTGAAGTGAAGATGCTTCCTCTGATAAGAAGC
CTTGTCTTCTCCATCGCAACCAGCTTATTTTTGGTATTACCGATGAGCAACAACAAGAA
CGACTTCATCATCTTCTGGAAACAGTTGTTACGGGACTTTTGTGTATTCCGCTCGACTTT
CCAGGAACTACATTTCGTAAAGCACTTCACGCGCGGTCGAAGCTCGATGAGATTATGTCT
TCTGTAATAGAAAGGAGAAGAAACGATCTGCGTTTAGGCGCAGCTTCAAGCGACCAAGAT
CTATTGTCGGTGTTGCTCACCTTCAAAGATGAAAGAGGGAATCCATTCGCTGACAAGGAG
ATCCTGGATAACTTCTCTTTTCTACTTCATGCCTTATACGACACCACAATTTCACCACTC
ACGTTGGTATTTAAGCTGGTGTCCTCCAATCCTGAATGCTACGAAAATATAGCTCAAGAG
CAATTGGAAATTCTTCGCAATAAAAAGGATGGAGAAGATATCAGCTGGGCGGATCTGAAG
GATATGAAATATACGTGGCAAGCAGTTCAGGAAACCTTGAGGATGTGTCCTCCAGTTTAC
GGAAATTTTCGCAAGGCTTTGACAGATATTCATTATGATGGCTATACAATCCCAAAAGGA
TGGAGGATTTTATGTTCACCTTATACTACACATAGTAAAGAGGAGTATTTTGACGACCCG
GAGAAATTCAGACCTTCAAGATTCGAAGAGCAAGGAAGGGATGTGGCTCCTTACACATTC
ATACCATTCGGAGGAGGCCTGCGCATCTGTCCAGGCCGGGAATTTGCGAAGATGGAGATA
TTAGTGTTTATGCATCATTTTGTTAAAGCTTTCAGCAGCTTCATTCCAGTTGACCCTAAC
GAAAAAATTTCAACAGATCCGCTTCCTTCCATCCCTGTCAATGGATTTTCCATAAACCTT
GTTCCCAGATCCTAATCCAGTTGAAAAGTGCCCGGGCGACCACAACTAATTAGAAGCTTT
GAGATGTTGAAAACATTGTGTTTGGAAAAAGTTTTAGCTTCAATATAGTGCCCAGCTATA
GTTTGTGCATTGTGTAAAGTTTNNNNNNNNNNNGTTTCCTTAAAATAGATTATCCATAAA
AAATTAGTTTGGGTTTCCATTTCTTAAATAAACAAAGCCATGATATATTAAATCC

CYP13

CCGAACACTAACTTATACAATCTCTGTTCTCCTTATCATTTAGTCTGTCCTGTCTCTGAA
ATGGATAGCTTCAATTTCTTGAGAGGCATTGGAGCAGATTTTGGGGGATTCATTCAGTTC
CAGTCTTCCCCTGCTGTTCTTTCCCTTTCCCTGATCACAACTATTCTTGGCGTTCTACTT
CTCTGGTTCTTCCTTCATAAAAACGGTTCCTCTGTTACTCTCCCCCCTGGAAATTTAGGC
TTCCCTTTCATTGGGGAGACCATACCATTCTTGAGGGCACTTCGATCAGAAACACCTCAG
ACGTTTTTTGATGAGAGRGTGAAGAAATTCGGTGTTGTATTCAAGACTCGGATAGTTGGG
CATCCCACAGTTGTACTCTGCGGGCCTGAGGGAAACCGCTTTCTTCTCTCCAACGAGGAC
AAACTGGTGCAGGCGTCATTGCCCAACTCTTCCGAGAAACTAATTGGGAAATATTCCATT
CTGTCCAAAAGAGGGGAGGAGCATCGCATATTACGTGCTGCACTTGCCCGCTTTTTGCGA
CCCCAAGCTTTGCAGGGTTATGTTGCTAAAATGAGTTCAGAAATCCAACATCATATCAAG
CAAAAATGGAAGGGAAATGATGAAGTGAAGGTGCTTCCTCTGRTAAGAACCCTGATCTTC

AACATTGCAAGCAGCCTGTTTTTCGGCATAAATGATGAACACCAACAGGAACAGCTTCAT

CATCTTTTGGAAGCCATTGTTCTGGGAAGTCTGTCTGTTCCGCTCGACTTTCCAGGAACT

CGTTTTCGTAAAGCTCTTGATGCGCGGTCTAAGCTGGATGAGATTCTTTCTTCTTTAATG

GAGAGCAGAAGAAGGGATCTGCGTTTGGGCACGGCTTCTGAGAATCAAGATCTTCTTTCT

GTGTTGCTCACCTTCAAAGATGAAAGAGGGAATCCACTCACAGACAAGGAAATCTTCGAC

AATTTTTCATTTATGCTTCATGCCTCGTATGATACCACTGTTTCACCAACGGKTTTGATG

CTTAAGCTTCTCTTCTCTAGTCCTGATTGCTATGAAAAACTAGTTCAAGAACAATTGGGA

ATAGTTGGCAATAAAAAGGAGGGAGAAGAAATCAGCTGGAACGATCTGAAAGCTATGAAA

TATACATGSMARGTTGTGCAGGAAASTATGAGGATGCTCCCTCCAGTTTTTGGATCGTAT

CGCAAGGCTANCACCKATATCCATTATGATGGGTATACAATTCCAAAAGGATGGAATATA

TTCTGGTCACCTTATACTACACACGGGAAAGAAGAATACTTCAATGAAGCGGACAAGTTC

ATGCCTTCGAGATTCGAGGAAGGCAAATATGTTGCTCCTTACACATTCTTGCCATTCGGA

GCAGGTCTGCGCGTATGTCCAGGATGGGAATTTGCAAAGACCGAGATATTACTGTTCGTC

CATCATTTTATTACAACTTTCAGCAGCTACATCCCAATTGACCCCAAAGATAAAATTTCA GGGGATCCATTTCCTC

TAA

CYP14

GAGGGACTGAGATTATCTTCCTATTATTGAATAGAGAAGGTATTTAAGGTCTGTCATGGA

TCTATCCTTCCATGCATTTCTGTGTTAATGCAAATGGACGCTTTTAATATTTTAAAGGGC

CCTGCTGCAAAACTTAATGGAGTCGTGCAGCTCGGCTCTTACAYCGATCRYATTCTTTCC

ATTACAGTCGTTGCCTTCATTACTATTCTCCTGTTGCTCATGCTCCGTTGGAAAAGCSAG

TCTTCTGTGAAGCTTCCCCCGGGGAACTTTGGCTTCCCTTTGATYGGCGAAACATTACAA

TTGTTRAGGGCATTTCGATCTAACACKACTCAACAGTTTTTTGATGAGAGGCAAAAAAAA

TTTGGTKGTGTTTTCAAGACATCACTAGTCGGAGAACGCACGGTAGTACTCTGCGGTCCG

TCTGGAAACCGTTTAGTGCTCGCCAACCAGAACAAGGTGGTGGAGTCATCGTGGCCGAGC

GCTTTCATCAAACTCATCGGAGAGGATTCCATTGCCAACACAAACGGRGAGAAGCATCGG

ATCTTACGCGCCGCACTGCTTAGATATCTTGGTCCCGGGTCGTTACAGAATTATGTGGSG

AAAATGAGGTCAGAAATCGAACATCATATSAATGAGAAATGGAAGGGRAAAGAKSAAGWG

AAGGTGCTYSMTTTGGTAARARAGAATGTCTTCTCCRTYGCAACCRSCTTGTTTTTCGGT

GTNAATGAYGAGGAAAGAAAAAGGATCCGACCTCCATCAATCTTGCGGAAACTGCACTTT

GCGGGCAGTTTTTCTATTCCGCTGGACTTTCCAGGAACTAGTTATCGGAGAGCTCTGGAG

GCACGGTTGAAGCTGGATAAAATCCTCTCTTCTCTGATAGAAAGGAGAAGAAGCGATCTG

CGCTCGGGCTTGGCATCTGGTAATGAGGATCTGGTCTCCGTGTTGCTCACCTTCAAAGAC

GAAGGAGGAAATCCTCTGACAGACAAGGAGATCCTCGATAATTTCTCCGGGCTACTTCAC
GCATCGTATGACACCACAACTTCAGCACTCACCTTGACATTCAAGCTCATGTCCTCCTCT
GCTGAATGCTATGACAAAGTAGTTCAAGAGCAACTGAGAATAGTTTCCAATAAAAAGGAG
GGAGAAGAAATCAGCTTGAAAGATCTGAAAGACATGAAATATACATGGCAAGTGGTGCAG
GAAACTCTGAGGATGTTCCCTCCGCTTTTCGGATCATTTCGGAAGACCATCGCCGACATT
CAGTACGATGGCTATACAATTCCAAAAGGATGGAAAGTTTTATGGGCAACTTATACCACA
CATGGGAGAGATGAGTATTTCAGTGAGCCCCAAAAATTCAGGCCTTCGAGATTCGAAGAG
GGAGGAAAGCATGTGGCTCCTTACACATTCTTGCCCTTCGAAGGAGGGGAGCGCACCTGT
CCAGGATATGAATTTTCAAAGACTCATATATTACTGTTCATCCACCAATTTGTTAAAACT
TTCACTGGTTACATCCCGCTTGATCCAAACGAAAGCATTTCGGCGAATCCGCTCCCCCCT
CTACCTGCCAATGGATTTCCTGTAAAACTTTTTCAAAGGTCCTAA

CYP15

ATGGATACCTTCATTCAGCACGAGTCTTCCCCACTTCTTCTTTCTCTTACTCTCGCTGTT
ATTCTTGGCACAATTCTTCTTTTGATATTAAGTGGTAAACAGTACAGATCTTCTCGTAAA
CTTCCCCCTGGAAACATGGGCTTCCCTCTCATTGGGGAGACTATAGCACTTATATCAGAT
ACACCTCGGAAGTTTATCGACGACAGAGTGAAGAAATTCGGCCTGGTTTTCAAGACTTCG
CTAATTGGTCATCCCGCAGTTGTAATATGCGGCTCCTCCGCAAACCGTTTCCTCCTCTCC
AACGAGGAAAAGCTGGTGCGGATGTCTTTGCCCAACGCAGTACTGAAACTCTTGGGGCAG
GATTGCGTTATGGGGAAAACCGGAGTGGAGCATGGGATTGTACGTACCGCACTAGCCCGC
GCCTTGGGCCCCCAGGCGTTGCAGAATTATGTGGCCAAAATGAGTTCAGAGATCGAACAC
CATATCAACCAAAAATGGAAGGGGAAAGATGAGGTGAAGGTGCTTCCTCTGATAAGAAGC
CTCGTCTTCTCCATTTCAACCAGCTTGTTTTTCGGTATAAACGATGAGCACCAACAGAAG
CGACTTCATCATCTTTTGGAAACTGTAGCTATGGGACTTGTGAGTATTCCCCTAGACTTT
CCAGGAACTCGTTTTCGTAAAGCACTTTACGCGCGGTCGAAGCTCGATGAAATTATGTCT
TCTGTAATAGAAAGGAGAAGAAGCGATCTTCGTTCAGGAGCAGCTTCAAGCGACCAAGAT
CTACTGTCGGTGTTGGTCACCTTCAAAGATGAAAGAGGGAATTCATTCGCAGACAAGGAG
ATACTGGATAACTTCTCTTTTCTACTTCACGCCTTATACGACACCACAATTTCACCACTC
WCCTTGATATTTAAGCTGCTCTCCTCTAGTCCTGAATGCTATGAGAATATAGCTCAAGAG
CAGCTGGRAATACTTGGCAATAAAAAGGATAGAGAGGAAATCAGCTGGAAGGATCTGAAG
GATATGAAATATACATGGCAAGCAGTTCAGGAAACTTTGAGGATGTTCCCTCCAGTTTAY
GGATATATTCGCGAGGCTTTGACAGATATTGACTATGATGGCTATACAATACCAAAAGGA
TGGAGAATATTATGTTCACCTCATACTACGCATAGTAAAGAGGAGTATTTCGATGAGCCG

GAMGAATTCAGACCTTCAAGATTCGAGGATCAAGGAAGGCATGTGGCTCCTTACACATTC
ATACCATTTGGAGGAGGCCTGCGCATCTGTGCAGGCTGGGAATTTGCAAAGATGGAGATA
TTACTGTTTATGCATCATTTTGTTAAAACTTTCAGTCACTTCATTCCAGTTGACCCCAAC
GAAAAGATTTCGAGAGATCCACTGCCTCCCATCCCTGTCAAAGGATTTTCCATAAAGCCT
TTTCCTAGATCATAA

CYP16

CTTCCTTCACAATCAGCCCATTCAGTCTACAAACACGTGCAAATACCAAGTTATAAAGCC
TTTGCACTCCATCGATTTTCTTCACCATTCTGCTCTTCCATTTCAGTTGTGCTAATTGAA
ATGGATACCATASGAGCAARTTTTGGCGAAGTKATTCAGCCAGAGTATTCTCCTSYCATC
ATTTCCSNCGCYCTSGCAGCTTTTCTTGGYATTGTWATTTTCTCGATCTTCAGTTMCAMT
CGACGATCCTMTGTGAATCTCCCCCCTGGAAATTTAGGTTTACCTTTCATTGGCGAGACG
ATACARTTCTTGGGGGCACTTCARTCAGAAAAACCCCAYACATTTTTCGATGAGAGAGTG
AAGAAATTCGGTAAGGTCTTCAAGACTTCTCTAATTGGGGATCCCACGGTGGTACTCTGC
GGGCCGGCGGGAAACCGCTTAGTTCTGTCGAACGARGACAAGCTGGTGCAGTCCGCAGGG
CCCAAGTCTTTCCTGAAACTGTTTGGGGAGGATTCCGTTGCGGCCAAAAGAGAAGAGRGY
CATCGCATCTTRCGTTCGGCTCTGGGTCGATTTCTGGGTCCCCAWGCTTTACAGAATTAT
ATTGGGAAAATGAATTCAGAAATGCAACGNCATTTYGATGAMAAATGGAAGGGAAAAGAT
GARGTGAAGGTGCTTCCTTTGGTTAGAGGCCTCATTTTCTCCATTGCTACCTCCCTGTTC
TTCAATATAAATGATGACAGACAACGTGAGCAACTCCATGGTCTGCTGGATACAATACTT
GTGGGAAGTATGACTATTCCTCTGAACATTCCWGGAACTCTTTTTCGTAAAGCTGTCAAG
GCACGGGCGAAGCTGGACGARATTCTTTTTGCTTTRATAGAGARCAGAAGAAGAGAGCTG
AGATCRGGCCTAAATTCTGGTAATCAAGATCTTCTGTCGTCCTTGCTCACCTTCAAAGAT
GAAAAAGGGAWTCCACTKACAGACAAGGAGATCCTCGACAAYTTCTCTGTTATGCTTCAT
GCCTCSTATGACACTACTGTYTCACCAACGGTCTTGATATTGAAGCTYCTCGCCTCCAAT
CCTGAATSCTAYGAAAAAGTTGTTCAAGAGCARTTGGGAATACTTGCYRGTAAAAAGGAG
GGAGAAGAAGTCAATTGGAAGGATYTGAAAGCTATGCCATATACATGGCAAGCAATTCAG
GAAMCCCTAAGNATGYACCCNCCAGCKTTTGGAATGTTTCGAAGAGCTWTCMCYGATATT
CADTTSGARGGYTATACAATTCCRAAAGGATGGGCMATTGTGTGGMCANCTTATAGTACA
AATGGGAGAGAAGAGTWYTTCAATGAACCAGACAAATTCAAGCCTTCCAGATTYGAGGAA
GGMAAGCCMCTGGATCCTTACACATTCATACCATTYGGAGCAGGSGTACGCATATGTGCA
GGATGGGAATTTGCAAAGRCWGAACTATTACTSTTTGTCCATCMCTTTGTTAAAAACTTC
AGCGGTTGCATTRTAATTGATCCKAATGAAAAAATTTCAGGGGATCCATTCCCTCCWCTC

CCTACCAGTGGACAACKCATGAAACTTATTCCGAGATCATAAATAAATTGAATGGGGGAC
ATCACAACTACTTGATGAGATGGATCGGCATGTCGAACAATAAATTTAATGGCCTGAACA
ATACTAGCTTGCCCTGATACAGTTAATAGTTCTGTTCGTCTTATATATGTTAATGTTTCG
AGCTGGTTATAAAAATAAAGATTAGGTTAT

CYP17

ATGGAAACTAAATTTGGGCAACTTATGCAGCTNGAGTTTCTTCCCYTTATCCTCACACCT
ATTCTCGGCGCCCTTGTTCTTCTCCATCTCTTCCGTCATAGAAACCGATCCTCTGTTAAA
CTTCCACCTGGAAAGTTAGGTTTCCCCGTCATTGGGGAGACGATACAGTTCCTGAGGGCA
CTTCGATCACAAACACCTCAAAAGTTTTTCGATGATAGAGTGCAGAAATTTGGTGGTGTT
TTCAAGACTTCRCTAATTGGAAATCCCCTAGTGGTCATGTGCGGGCCTGCGGGAAACCGG
TTAGTTCTGTCCAACGAGGACAAGCTTGTGCAGTTGGAAGCRCCCAATTCCTTGATGAAA
CTGATGGGGCAGGACTCCCTYCTGGCCAAAAGACAAGAGGACCACCGCACCTTACGTGCT
GCACTAGCCCGGTTTTTAGGCCCCCAAGCTCTACANAATTATATGACTAAAATCAGTTCA
AGAACCGAACATCATATGAATGAAAAATGGAAGGGAAAAGATGAAGTGAGGACGCTTCCT
TTGATAAGAGAGCTCATCTTCTCCAATGCAAGCAGCTTGTTTTTCGATATCAATGATGAG
CACCAACAGGAGCGACTTCATCATCTTTTGGAAGCTGTTGTTGTTGGAAGTATGTCTATT
CCGCTGGACTTTCCAGGAACTCGCTTACGTAAAGCCCTTCAGGCGCGATCTAAGCTGGAT
GAAATTCTCTCCTCTTTAATAAAAAGCAGAAGAAAAGATCTTGTTTCAGGGATAGCTTCT
GATGATCAAGATCTACTGTCGGTGTTGCTCACCTTCAAAGACGAGAGAGGAAATCCACTG
ACCGACAAAGAGATCCTCGACAACTTTTCTCTTCTGCTTCATGCCTCGTATGACACCACT
GTTTCCCCAATGGTTTTGACATTGAAGCTCCTCTCCTCCAATCCAGAATGCTATGAAAAA
GTAGTTCAAGAGCAATTGGGAATAGTTGCCAATAAAAGGATAGGAGAAGAAATCAGCTGG
AAGGATTTGAAAGCCATGAAATACACATGGCAAGTAGTTCAGGAAACACTGAGAATGTTC
CCTCCACTTTTTGGATCATTTCGCAAGGCTATGGTTGATATTGATTATGATGGCTACACA
ATTCCGAAAGGATGGATGATTTTATGGACAACTTACGGTACACACCTGAGAGAAGAGTAC
TTCAATGAACCGTTGAAATTTAGGCCTTCAAGATTTGAAGAAGACGGGCGTGTGACTCCT
TACACATTCATACCATTCGGAGGAGGCGCGCGCACATGCCCAGGATGGGAATTTTCAAAG
ACGGAGATATTACTGTTCATCCATCATTTTGTTAGAACTTTCAGCAGCTACCTCCCAGTT
GACTCCAACGAAAAAATTTCAGCAGATCCATTCCCTCCCCTCCCTGCCAATGGGTTCTCC
ATAAAACTTTCAGCAGATCCATTCCCTCCCCTCCCTGCCAATGGGTTCTCCATAAAACTT
TTTCCCAGATCTCAATCCAATTGA

CYP18

GGGGGATTCAACATGGCTTATCCGGAGTTGCTCGAAAATTTATCGGGAGACCGAGCTCAA

TCTCCAGCAATAGCCGCGGTGCTTACAATTTTATTCTTGTTGGGGATTTTCTACATACTG

CGCGGGCTGAGAAACAATGGAAGAAGATTGCCCCCCGGCCCAATTCCATGGCCGATCGTG

GGAAATCTCCACCAGTTGGGAAAGCTTCCCAACCGTAATCTGGAAGAGCTCGCAAAGAAA

CACGGACCCATCATGCTCATGAAATTGGGTTCCGTTCCTGCCGTTATCGTTTCTTCCTCT

GCCATGGCAAAAGAAGTTCTGAAAACTCATGATCTGGTTTTCGCCAGCCGACCCGAAAGC

GCCGCAGGAAAATACATAGCGTATAATTACAAGGATATAGTTTTCTCTCCCTACGGACCT

TACTGGAGACAGATGAAGAAAATATGCGTGGTGGAATTGTTGAATGCCAGAAGAATCGAG

TCGTTGAGATCCGTAAGAGAGGAAGAGGTGTCTGTTATAATTCGTTCGGTGTGGGAGAAG

AGCAAGCAGGGTGCGGTCGCCGTCAATCTGAGCAAGACGCTGTCATCCCTTACACAGGGA

CTCATGTTGCAGATCTTTTCCAGTAACGATGACGGCGGGAATAGCAGCGTCACCGCCATT

AAAGAAATGATGTCSGAGGTGTCTGAGACGGCSGGAGCTTTTAACATTGGAGATTATTTT

CCATGGATGGACTGGATGGATTTGCAGGGTATACAGCGGCGCATGACGAAGGCACACGAT

TATTTCGACCAGGTCATTACGAAAATTATAGAGCAACACCAGAGGACGAGAGCGATGGAG

GACACTCAACAACCAAAAGACATAATTGACGCCCTGTTGCAGATGGAGAACACCGATGGC

GTCACCATCACAATGGAAAATATCAAAGCCGTCGTTTTGGGTATTTTTCTGGGCGGAGCG

GAGACRACGTCCACTACGTTGGAATGGGCGATGAGCGCGATGCTTGAAAACCCTGAGGTG

GCCAAGAAAGTGCAAGAAGAGATCGAATCCGTTGTGGGAAGARAGAGGGYGGTGAARGAR

AGTGATCTRGGAAAGTATGGAATACCTSCAATGTGTGGTGAAARAGACGATGAGATTATA

TCCGGCGGTGCCTTTGCTTATCCCGCAYGAATCGACMCAAGATTGCACTGTCAATGGATA

CTTCATTCCTGARAGAACCAGAATTCTCGTTAACGCGTGGGCRATAGGAARAGATCCAAM

CGTGTGGGATGATGCRCTGGCATTCAARCCARARAGATTTTTGGGYANAAAYGTGGACTT

GCARAAAGGAAAAGAGTTTTTCGATATGGTTCCCTTTGGTGCGGGAAGGARAGGATGCCC

AGGGGCAAGCATGGCCGTTGTGACGATGGAACWTGCGTTGGCACAACTCATGCACTGCTT

CCAGTGGCGCATTGAAGGAGAGTTGGATATGAGTGAACGCTTSGCAGCCTCCKTGCAAAR

AAAAGTYGATCTTTGTGTTSTTCCYCAATGGAGGCTAACTAGTAGCCCTTGAACTAAAAA

WATATATTTTCATCCTRTTTCTRCTCTTGAAAWCAAGTTTAAAAGTTTGAACCTTGAATC

CCTTAGAAATAGAATARTGTYGAGTGAGTTAGTTTGTTGAAACAACTCTCTCATAATAAT

CTGAATAGTTGTAGGAGATAGTTGACCTCTCCTATATTTCAGGAGTGATTTAGTGTATAA

ATATCTCCGATACAATGAATAAACGTTACAAAGAGAATTATTATTGTGTTT

CYP19

GTCCCTAATTCACAAGATGGAACTGTGGAATATGTTTCTGCCATGGATCTCCATTGCAAC

AGCAACATCCCTTACAGTGGGCATGGCATCYGCATTCGCTATAATTTATGCTCCTCTTTT
GCTGTCATTCCTGAGATTGAAGACGGCCAGAAGAAATATGCCTCCGATTCCTCCAGGAAG
CATGGGAATGCCATTCTGGGGAGAATCTCTGGGYTATCTCGGCTCATGGAATAACCAGAG
CAACCCTGACGTGTGGTACGACACACGGAAGGCCAAACACGGCAAAATTTTCACAACCCA
CATTCTGGGCAGCCCCACTGTGGTCATGTTGGGTCCGGATGCCAACAGGTTCATCCTCAY
TAACGAAAACAAGCTTTTTCTCAACAGTTGGCCCAAATCTCTCAACGCTCTCATCGGAAA
GCABGCCCTCATCAMTTCGCAGGGCGCAGAACACAAAAGGATGCGGCGAATTATACATTC
CGTGCTCGGCCMAAGAAACCCTGAAACTAGCCGTGGGAAGATTCGAAGGACTGGTGTTGC
ATCATCTCGATTCCGACTGGCATGGCGGCCAAATCATCCAAGCCTACCGCCAAGTTAAGG
ACATGGCGCTCTGTTTGGCTGCCGATTTTTTCATGGGGTTAAAGCCMGGAAAAGAATTGG
AGACTTTCAGGCGGCATTTCAGTGATTTCAGCGCGGGGCTTTTATCTCACCCTCTCGATC
TTCCCTGGACTGTGTTTGGGAAGGCSAAACGAGCGCGCGCCGCCATGGTCRCTCAGATTT
TTTCACAAATTCGGCTGCATAGGACTTCCATGCACAAAAGTGGAGAGGAGGGGGGAAATT
TCTTGGACATGGTGTTGGGTTCGCAGGAGAAGGGAGGCGATTTGAGGCTGAGTGAGGAGG
AGATTGCAGACAATCTTATGGGTCTTTTAACTGGCGGACAGGACACGACAGCCTCGGCAT
TAGCCACCATTCTGAAGCACCTCTCTCTCTCCCCACATCTATTACWAAGGCTTCGCAAAG
AGTGTGAAAAACTTAGAGATARCAAGGAGGYAGGGGRGCCTCTTACATGGAGTGAAMTAA
AAAGTGTGGGCTATTTACACAATGTAATCTCAGAAGGACTACGRATGGTAGCCCCCATAA
ATGGAGGATTTAAGAAAGCAAAAGTAGACGTTGTATATGGAGGTTATACTATTCCCAAAG
GATGGAAGGTTCATTACTCCGTGAGACAGACAAACAACAAAGAAGAGTATTTTCCTAGTC
CAGAGAGATTTGATCCAGATCGCTTCAATGAGAGACATGAGCCTTTTTCTTTCATCCCCT
TCGGCCAGGGTAATCGGATGTGCCCCGGAAATGAATTCGCAAGGTTGGAAATGGAATTAT
TTCTATATCATTTGGTTTTGAGATATGATTGGGAATTAATGGAGGYGGATGAACGCACCA
ACATGTACTTCATTCCTCACCCTGTGCACAGTTTGCCTTTACTACTTAAACACGTTCCTC
CTACATGAATCTCCTTCTCCAAAYGTGTTTTAATACTTTGAATTTGTTTGGAACTCAATG
CGAGTGACATTGTTGTACACAATACRAATGTTTTAGGCTTAAATTTAGTTTAACAAAAAG
ATAATAATATCCCCATCTAGGTGTGGATGACATGGGATTCAATGGTTAAAGGGTGATTCA
AATCCATTATTTTCAAAACTTAAAGAGGAACCCCTCACTTCAATGTTGGGGTGATTATAT
AATTGGAAGCAAAATAGAGATATTTTTGTATGTTAAAAGTGATATTATATTAATAATATA
AATATGCA

CYP20

ATGGAGCAGCTAATCTATAGTATTGTCTATTCCAATTGGTATTTATGGGTTTTGGGACTG

TTTATCTGTGTAATTTTACTGTTATTAAGACGGAGTAATGACAGACAAGGGAATGGATCC
GCCAATAAACCCAAACTTCCACCTGGATCAGCTGGATTGCCATTTATTGGAGAGACTATC
CGTTTTCTTAGAGACGCTAAATCGCCTGGACGGCGAAAGTTCTTTGATGAACATGAGCTC
AGGTATGGGCCGATTTTCAGATGTAGTTTGTTTGGAAGAACACGTGCAGTTGTGTCGGTG
GATCCCGAGTTCAATAAGTACGTCTTGCAAAATGAGGGAAGGCTGTTCGAATCCAACGCA
CTCGCGCCYTTCAGAAATCTTATCGGCAAATATGGATTGTCGGCGGTACAGGGGGAACTT
CAAAGGAAGCTCCATGCAACTGCTGTCAATTTGTTGAAGCATGAGACGCTCAGCTCTGAC
TTCATGGAAGATATACAAGACATCTTTCAGGCTGGAATGAGAAAATGGGAGGAGGAGGGA
GACATCCCTATTCAACACAAGTGCAATCAGATTGTTCTGAACTTGATGGCGAAGARATTG
CTGGACTTACCTCCATCAGAAGAAATGGGASATATTTATAAAGCTTTCGACGATTTCGTG
GGAGCTGTCCTCTCTTTCCCCCTCAATATCCCTGGAACCACTTATGCGAGAGGAATTCGG
GCCAGGGGAATTCTGTTAAAAAGAATTCACAAGTGTATAAAGGAGAGGAGAGAACATCCA
GAGGTGCTCCGCAATGACTTGTTGACCAAACTTGTGAGGGAGGGCACATTTTCGGACGAA
ATTATTGCAGATACAATAATCTTTTTTGTGTTTGCTGGTGTCGAAACTTCAGCAATGGCC
ATGACGTTTGCTGTAAAGTACCTCGCTGAGAATCCACGAGCACTGGAGGAGTTGAGGGCT
GAGCATGACGCTCTTTTGAAGGCCAAAGGGAAAGGCAATGAAAAGCTGACGTGGAATGAC
TACCAATCAATGAAATTCGTTCATTGTGTAATAAATGAAACACTTCGTYTGGGTGGTGCA
ACCGTGGTTCTTTTCAGGGAAGCCAAACAAGATATTAAAGTGAAAGATTTTGTTATTCCC
AAAGGATGGACCGTTTCTGTTTTCTTGAGCGCCACACATGTTGATGGAAAATACCATTAT
GAAGCTGACAAATTCCTCCCTTGGCGCTGGCAAAATGAGGGTCAAGAAACGTTGGAGGAG
CCATGTTATATGCCATTTGGAAGAGGTGGCAGGCTCTGTCCAGGACTCCATTTGGCAAGA
TTTGAAATTGCTCTCTTTCTTCACAACTTTGTCACTAAATTCAGATGGGAGCAGCTGGAA
ATTGATCGTGCGACTTACTTTCCTCTTCCTTCCACAGAAAATGGTTTTCCAATCCGTCTC
TATTCTCGAGTACAYGAATGAGGAAGATTTTATGTACACGAATAAGGAAGATTTTATTTT
CTAATTAAGTAGCAATTGTATGTGTACAATGCAATGTGAACGCCGAATGGTATATGTTAT
TTGATTTAAAGGGATAGTTATTACTTTTGTAGTTGATGTTGATAAGGGGTGTGATTGAAC
ATTTTGATCATAATCCCAAAAGGAGAAAAAGTAATGTCATTGTCCAACTCCTGCTTTAAG
TTAAATTGACATGCTTCAGTTG

CYP21

ATGGCATTCGAAGCAGCTACTGTTATTCTTTTCACTCTGGCTGCCCTGTTGCTAGTCGTC
ATACAGCGCCGTAGAATTAGAAGGCACAAATTGCAGGGGAAGGTGAAGGCACCACAACCT
CCTTCATGGCCCGTTATTGGGAATCTTCATCTGCTTACACAGAAAGTGCCTATTCACCGA

ATTCTATCTTCGCTTTCGGAGAGCTATGGACCAATCATGCATCTTCAACTCGGTCTCCGA

CCAGCTTTGGTTATTGCCTCTTCAGATCTGGCGAAAGAATGCTTCACAACAAATGACAAA

GCCTTCGCTTCTCGCCCACGTCTGTCTGCAGGAAAGCATGTAGGATATGACTACAAAATC

TTCAGTATGGCTCCTTACGGTTCCTACTGGCGAAACCTTCGGAAAATGTGCACGATCCAG

ATCCTCTCTGCAACCAGAATTGACTCCTTCAGACACATCCGCGTAGAGGAAGTTTCTGCT

CTCATTCGTTCGTTGTTTGACAGTTGCCAGCGAGAGGACACTCCAGTCAACATGAAAGCG

AGGCTCTCTGATCTCACGTTTAGTATCATCCTCCGTATGGTTGCCAACAAGAAATTATCA

GGACCTGTTTATTCCGAGGAATACGAAGAAGCGGATCATTTTAACCAGATGATAAAACAG

TCTGTGTTCTTACTTGGAGCATTTGAGGTTGGAGATTTCCTGCCGTTTCTCAAGTGGCTT

GATCTTCAGGGTTTCATAGCTGCTATGAAAAAACTGCAGCAGAAAAGAGATGTCTTTATG

CAGAAATTGGTGATTGATCACCGTGAGAAGAGAGGGAGAGTCGATGCAAATGCACAAGAC

TTAATTGATGTTCTCATCTCTGCAACAGACAACCATGAAATTCAGTCCGATAGTAACGAC

GATGTTGTGAAAGCCACCGCCCTTACAATGCTGAACGCAGGTACAGATACATCCTCGGTG

ASCATCGAATGGGCATTGGCGGCTCTGATGCAGCACCCTCATATTTTGAGCAAAGCCCAG

CAGGAGCTCGACACGCATATCGGACGCAGCCGATTACTAGAGGAAGCAGATCTGCACGAG

CTGAAATATTTGCAGGCAATTGTGAAAGAAACGTTGAGGCTATATCCAGCCGCACCTCTC

TTAGTTCCTCACGAAGCCATTGAGGATTGCACTGTTGGAGGGTACCATGTCTCCGCAGGA

ACGCGACTGATTGTGAATGCTTGGGCAATTCACAGAGACCCGGCAGTGTGGGAACGGCCG

ACCGTGTTCGATCCTGAACGGTTTTTGAAGAGCGGAAAAGAGGTTGACGTAAAAGGGCGG

GAKTTTGAATTGATTCCGTTTGGTTCAGGGAGAAGAATGTGTCCGGGCATGAGTCTGGCA

TTGAGTGTTGTTACGTATACGCTGGGGAGGCTGCTGCAGAGCTTCGAGTGGTCTGTTCCA

GAAGGTRTGATAATTGACATGACGGAAGGTTTGGGACTCACAATGCCCAAAGCAGTTCCG

TTGGAGACCATTATCAAACCTCGCCTTCCCTTCCATCTCTACTGAGATCTCTCATATATT

GCTGTCAACGAGTACAAAGTTTTAGTTCATGTGCAGGTACTCCAGCACACTGGCGGCCGT

ATAACTAGTCGGATCCGAGCCTCAGTGTATCCANATGCGTTAGTGCAGTAATTCATGGTC

ACTATGCTGTTTCCATGTGTGAAATTGTGTGATCCAGCTCACACAGTTCCACACAAATCT

ATATCGAGCCTGGAAGCATTAC

CYP22

CTTCAAACCAGTCCGAAGATCAATTTAATTCAACGTTGGACTTAGATTGGCAGGGTGACT

TGTATTGCAGGGTGACTYGTGTAGGTGTTTAGATTGGGTGTCGAATTTCTATCTTGCYTA

TTTGRATTTCGAYTTCTACAATTAAAACYATTTTAAGGATCGGATCTTCACCATTAAGYC

AAYTCAGGGGATYGATCCATTCATMTGGCCATGGAKTTGAATGCTCAGGAAACAATAGCS

AAAGCTCTCACCGATCTTCAGAACCCCTCTAGTACACTGTTTGCCATATTATCCACTGCA
CTTGCTTCAACACTGCTTGTGTTCATCATTTCCAAGATTGTCAAGACCGGAAATGGGAAA
AAACTTCCTCCAGGAAGTCTGGGCTTGCCTCTGATCGGTGAGACCCTTCARTTTCTGAGT
GCTTACAGATCTCACAAATCCAAGGATTGGGTTAAAAAAAGGGTTGCAAAATATGGYCCT
GTTTTCAAGACCTCCCTCATGGGGAGCCCAACAGTCATCGTCACTGGTCAGGCAGGAAAC
CGGTTTCTCTTCCACAACGACGGCCACAGAATACTGAACAAGCAGCCCAAGAATTTCACC
AGGATTTTTGGSGAGAGTGGTATTGGTGAACTYGCCGGGGAAGAGCACAAGCGAATAAGA
GGMGCCATCATGCAGTTCATGAAGCCGGAAGCCTTGCAGAAATTCGTGGGCAGAATGGAG
TCTGTCATTMTGCAACATTTTGCTNGATWGTTGGGAAGGAAAAGACATCATTACAGTGAT
GCCATTGATGAAGAAGCTTACKTTTCAAGTTGCSTGTGATCTTTTTTTCAGTCTGAGTGA
CAGTRCAGAGAGGGAAACTTTGGCTAATGAATTTAATACTGCTGTGAARGGTATTTGGTC
TGTTCCTCTGGATTTACCAGGGACTGCCTTTCGGAAGGCTTTTGAAGCTCGTTCTCGGGT
ATCTCAGAGGTTATCTTTGCTGTTGGACGTGAGGAGAAAGGAAATAGAGCAAGGCAAGGC
CTCACCTCGCCAAGAYTTAATGTCATGTTTGCTGACAATGAAGGATGAGAATGACAAGCC
CCTGACTGAGGAGGARATCTTAGAYATTGTYGTAGTGGTCATGATAGCAGGCCATGACAC
CACTTACAATCTTMTAACTCATCTTGTGAGAGGATTGGCACTCAACCCAGATATCTACCA
AAATATTCTACAAGAGCAAACTTTRGTATTAGAAGGAAAACAACCAAATGAACCTCTTAC
ATGGGAAGATCTACGTAAAATGAAGTATACRTGGAAAGTGGCTCAAGAAACACTACGCCT
ATTCCCTCCCATATTTGGAGGATTCAGGAAAGCAAACGAAGATGTTGAATATCAAGGTTA
TACAATTCCCAAAGGCTGGCAGTTATTTGGGAGACGAGCTCCACACATTGGAATGAGGA
GATTTTCAAGGAGCCRCATAAATTTGATCCATCTCATTTTGAYACTCAAGTTTCACGCTA
TGTGTTCACACCATTTGGAGGAGGACCTCGAATATGTCCTGGSTATGATTTTGCAAAGAT
AGAGACCATGATATTTTTGCATCATTTAGTAACCAAATGGAAGTGGACYCTGGTTGACCC
CAATGAAAAAATAATGTGTGACCCTACACCTACTCCTGTTATGGGACTGCCAATTAATCT
TCAAGCCAAAGGGAAGTGATTGAATATATTATTAATGGTGTTCTTTTATGTATAAACACT
ATTCATAACATTATGAGTATTATTGAATGTTAGCGTKCTTCTTATAATAGTTTTCAAGAA
GTGGGTGAAGGCTTAGAAGAAATYGAGAATTAGTAAGGCRATGGCTAGKTTATCTCGCCT
TGTATGGAGGCGSCYTGATCTATCTCCTATAGTTCWTTGTGTAATAAAACTCATGTTCCT
AATTTGACTTCAAWAATAGWGAGCATTTTTTTTTTAATATAAATAGTCAATTAATGTTT
G

CYP23

ATTGTAATTAAGTGCTTKGCAGTGAGATGGAGTTCGGAGAGCAAGCCATAGCTAAAACTG

TCATGGATTTCTTGCATTCAACCCCTCCCTCCCTTCTTATCTTGTCCACTGTCGCACTTG
CTTCAATAATATTTGTGCTTGTAGCCAAGATTTTCAGCCCTGGAAGAGGGAAAAACCTTC
YTCCTGGAAATCTGGGGCTGCCTGTGATAGGAGAGACCATTGGATTCTTGAATGCATTGA
AAGTTGATCAGTGCGACAAGTGGGTTGCAGACAAGGTTGCCAMATATGGCCCTGTCTTCA
AAACCTCCTTRATGGGCTGCCCMACYGCCGTSCTCACYGGCCAAGCGGGCAACCGMTTTC
TCYTMCAGAATGAYTACAACACCATCATCAACAAGCAGCCCACCACTTTTGTGAGGATTT
TYGGSTCTAAAAATATTAATGAGCTKTTYGMAGAWGAGCACAAGMGGATCAGAAATGCCA
TCATGRCATTCATGAAGCCRGAAGCCYTGCAGAAATGTGTGGGRAGRATGGAGGGCGTTG
TTMAGGAACATTTTGYTGAGYATTGGGAAGGGAGAGACACTGTTTYWGCKTATCCCTTGA
TGAARCAGCTCACTTTCCAAGTTGCTTCTGAYCTGCTCTTYGGTYTGAAGGAYCCYRWAG
AGAGGCAGATTCTGAACACAGAATCCATCACACTGATGAAGGCTATCTGGTCTGTTCCAT
TGAATTTGCCCGGAACCACTTTCAACAAGGGCTTAAGAGCCCGTGGTCGTGTTGTTAAGA
GACTCTCTTCCTTGATTAACTYRAGGAAGAGAGAAATAGCAGAGGGAAAGGTTGCCCCTG
WCCAGGATTTAATGTCATGGCTGATAACTTTGAAGGATGAAAACAACAATTCCCTTACTG
AGAATGAGATCATAGATAACCTGATTGTGCTCATGATAGGAGGTCACGATACCACTGCTG
TTCTTTTAACCCAGCTCGTCAGATTTCTGTGCTTACACCCGGATATTTAYAAAAACGTTC
TGCAAGAACAAACAGAGGTTCTTGCAGGAAGGCAACGGAATGGAGCCCTAGCTTGGGAGG
ACGTTAAAAAGATGGAGTACACACACAAGGTTGTTTATGAAGCATTGCGCATGGTACCTC
CAGTTTTTGGTGGATTCAGGGTCACAGTTCAAGATGTACAGTATAAGGGCTATACTATTC
CTAAAGGCTGGCAGTTATTTTGGGAGACCAGTACCACACATTCCGATGAAAGRGTGTTCA
GGGARCCAACAAAATTTGATCCCRGTCATTTTGATGACAAAACTCCGCCCTATACATTTA
TACCATTTGGAGGAGGAGCCAGACTTTGTCCAGGGTATGATTTCGCCAAGATGCAGGCAC
TCATATTTGTTCACCATTTGGTATCTACTTATGATTGGTCTATGYTGATTCCTGATGAGA
AAATGGTGTGTGAGCCTCTTCCTGTTCCGAGCAAAGGATTGCCGATAAAGATCCGAGCCA
GGAAGAATTGATCATGTTTTCCACTATATTGTCATGTTTGSTTTCATGGGWTTCAGATGC
TTTGCTTGTTTTATAGSCTTTAGCCRTTGTAGCCAGTTAAGCAGTACGATAAATTTCCTT
TGTGGTTCTAAGAAAAACTTGGKGATCTTYGGGCAAAGCATAATTTGCTGTTTTGTAAAG
TGAAGAATAAAAAATGTGTGTGAGTTCTTACTATTTTCTCTATTTTATTTTGGGGTTATG
TATAAAAAATAATATTTGTAATTTTTTTAATGATTAATAATAGTATTTCAC

CYP24

ATTCATAGAGATTATTGTGCAAGATGTCTGTCACAATTGTTGTGTGGGTAGTTGTAATAT
GCATTTTATTGGTGAAGCRGCCCTACAGGAAAGGGCATAATTTTTATGGACCCGAAATCW

TATCCCTTGCTTGGCTGCCTCATCTCCTTCTGGAAAAACCAAAGCAGGCTGTGCGAGTGG
TACACAGATCTGCTCAGAGAATCCCCCACGCAGACCATCACAGTCCAACGCSTGGGTGCC
ATTACCACTGTGGTCACAGCAAATCCAGAGAACGTGGAGCACATTCTCAAGACTCGATTC
GACAATTATCCCAAGGGGAAGATTTTCAAGGCAATTCTTCATGACCTGCTGGGCAATGGC
ATCTTCAACGCAGACGGAGWWMTSTGGAGAATGCAGAGAAAGATTGCGGGCGACGAGTTC
AATACAAAGTCGCTGCGGAATTTCATAGTTGAYTCAGTGCAGACGGAGATTACAGAACGC
CTGCTGCCTCTTCTTCTCCAGGCTTGTCAAAAAGGGAGTTTCGTAGATTTGCAAGACGTT
CTGCGTCGATTTTCGTTCGACAACATCTGCAAGGTGGCGTTCGGAATGGATCCAAATTGC
CTGGACCTCTCCCTTCGGCAATCCGATTTTGCAAACGCCTTCGATCTCGCCTCTCATCTG
AGCTCCAAAAGAGCGCTGTCTCTCACAAGCTTGGAATGGAGAGTGAAGCGGGCATTGAAT
CTGGGCTCCGAAAGACTTCTGGCCGATGCCCTGCGTGTTGTTCACGACTTCGCCACTGAA
ATTATTCGTGGAAGAAGTGCGGAAGTTAATAATCTGTCGCCCCAAGACAAACCAGACTTG
TTGTCGAGATTCGTGGCTCTCGCTAAAGACGACGTACTTGACGAACAATATTTACGAGAT
ATCGTAGTGAGCTTTGTATTGGCAGGCAGAGWCACCACAGCTTCTGCSTTGACGTGGTTC
TTTTGGCTGCTTTCTTCTCATCCMAAAGTCGAGGAAGCCATTCTCAAGGAAATCAATGAC
GAGTACGACAGTGGGTTGCAGGACTCGTACGAGAACATGAAAGGGATGCAATACCTGCAC
GCTGCACTGTGCGAATCRATGAGGCTCTAYCCTCCTGTTCGGTATGACTCAAAATATGCA
CTGGAAGACGACTTCCTTCCGGATGGCACGTTTGTGAGGAAGGGCAGCTGGGTGACTTAT
CATCCCTACGCGATGGGCAGGATGGACAGCATTTGGGGTGCAGACTGCATGGAGTTTGAT
CCAGACAGGTGGATGGAAGACGGTGTTYTTGTTCCGCARAGTCCATTCAAGTTCGCCGTG
TTTCAGGCTGGCGCGCGCATCTGTTTGGGAAGAGAAATGGCGTTTATCCAGATGAAATAT
ATTGCTATCTCTGTTATTCGTAGGTTTCGCTTAAGAKTGGATCCGCAGCACAAGGCGGAC
TCGATTCCAAGYCTGAGTGCGTCTATGAAGAATGGGCTGCCTGTTCGAGTGGAGAGGCGA
GAGGYCTCTTCATCGCCTCCCGCTCTGTGATTYGCAATGTCTACTGTGTATCTTCATAAA
ATCATACTTCACGTATATGTACACTTGTGGAAACTATACATTTATTAAATATATGTAAAT
GTAATTATATATTTATTGAATATGAAATGCAATAGAATTCTCAATATATGGATTTCAAAA
TCTGTAAGGAATGGGTGATGCCTTGCACAATGTAAACTCCATATTTCGCCGCAGCGTCGT
TTGCGATATGAGAAAACGAGAAGAATAATTGTATGTCTACAGGAGAGGGAGGAA

CYP25

GTTTTTCTAATAGGTCACCATGGAGTTCATAGCCATAAGAATCCTCCAGGACCCAGAGCA
CTTCCCATAATTGGCCACTTTCATCTCCTGATTGACAAGAAAAAACCCTTTCACAGAATT
CTTGATTCACTGTCCAAAGTGTATGGACCAATTATGAATCTCAGATTTGGTAGTCACCCA

GTTTTGATTATCAGTTGCTCAGAGCTTGCCAAAGAATGTTTTACGGTAAATGATTTGTCT

TTTGCTTCCAGGCCTCAGCTTGCTCAAGGAGAGCATTTGGGTTATAACTTTTCTTTCATG

GGTTGGGCTCCCTATGGTCTTTACTGGCGAACCACTAGAAAAATATGTGCTTTAGAGCTT

CTGTCTTCTAAACAGATACAAGCATTCCAATCCAAGAGGATGAAGGAGATTTGTAAAGCT

CTAAATTCTCTGTTTCTTCAGTCTCAGAAGGAAAACAGAGCTATAAGCATGAGAAATGTT

TTGTCCCAATTGACTCTGAATATTCTGATGACCATGATAATAGATGACAAATATTTTGGA

GAGGAAGCAGGGATTTCATTTGCAGAGGTTGCACATCTGATCGAAGAGTCTTTTGTGCTC

CATGGGGTTGTCAGTATTGGGGATTACATTCCCTGGTTGAAGTGGCTTGATTTGCAAGGG

TATGAGAAGGCTATGAAGGAGGTACAGCAGAAACTGAATCCTTACATGCAGAGAATCGTT

GAGAAGCACCGTGAAAACGCACAAAAAGCAAAGGAAGACATGGACTTCATTGATGTGCTC

ATATTGCAGGCGGAGGAGAATGGTGAAGCCATTCCTGATAAAGATACCTTCATCAAAGCC

ACTGCCGTGCAAATGTTTAGTGCAGGCTCAGATACTTCTTCACTTTCAATAGAATGGGCA

TTGTCAATATTAATTCATCATCCCAACATAATGAAGAAAGCCCAAGATGAACTCGATTCA

AAAATCGGACAGAACAGAGTGGTAGAAGAATCAGATATACCCCAGTTAAAATATCTGCAG

GCAGTGGTCAAAGAGACTTTGCGCCTTTTCCCAGCAGGGCCTTTACTTGTGCCACACAAA

TCCATTAAGGCTTGCATGGTTGGAGGCTTTCATATTCCTGAAGGAACAACACTGATGGTA

AATGCATGGGCGATTCAGAGAGACCCAAGAGTGTGGAATCAGCCTCTGGAATTCATGCCA

GAGCGTTTCTTTGAAAAAAATATAGAGATGGATAATATACAAAGCCGAGGGAATGAATTT

GAGATGATTCCGTTTGGGGCTGGAAGAAGAGGATGTCCTGGTGCTTCCCTAGCAATGTGT

ATGGTCCATATAACTATTGCAAGGTTGTTGCATAGCTTTGACTGGTTTGTTCCAAATGGT

AATGTGCTTGATATGGAGGAAGGAATTGGTCTGACAATGCCAAGAGCAGTGCCTTTGGAG

GTTGTTATTAAACCTCGTCTTCCCCCTCATCTATACCTGAATTCTATTCAGTAACAAATG

CCTGGAATTAGTTAAACTTCTACA

CYP26

TAGGACACACAGAAATGGAGCAGCAGATGATGTTAGCAGTAATTGTTTCTTGCAGCCTAT

CAGTGGCAGTTTTAGTGGTATTGGGGAGAAGAATAATATGGAGTCTAAGTAATACTCAGA

AAATAAAACTACCACCCGGATCCATTGGATGGCCCTTTATTGGCGAGTCAGTCGGCTTCT

TCAAGGCTATCAATTCCTCTCAACATCAACACCGTCGCAAGTTCATTGATGAACATGAGC

GCATGTATGGGCCGATTTTCCGATGTAGTTTGTTTGGAAGAGCACAGACAATAGTATCTG

TRGATCCAGAGTTCAATAAATATGTCCTGCAAAATGAGGGAAGGCTGTTTCAAGCCAAAT

ATCCTAGCTCTACAGTAAAAATTATAGGAAAATATGGTTTGTTGGATGTTCACGGAGAAC

TTCAAAGAAAACTTCATGGAACTGCTGTGAACTTGTTAAAGTATGAGAAGCTGAGCTCTG

ACTTTATGGATGACATACAWSATGTATTCTTGGATGGCATGAAGAAATGGGAACAGCAGA
GAGACATCCCTCTCCAGCACAAGTGCCATCAGCTTGTTCTGAATTTGATGGGGAAAAAAT
TACTAGACCTGCCCCCTGCTGAGRAGATGGAAGAAATATATAAGGCTTTTGACCAATACG
TGGGAGCKKCCCTCWGYTTGCCCCTCAATATTCCCGGAACTGCTTTTGCAAGAGGAATWA
AGGCAAGGAAGACTCTGATAAAGAAGATTTACAAGTGCATGGAGGAGAGAAGGCAGCATC
CAGAAGTAGCCCATAATGATCTTTTGTCCAAGCTTCTGAAAGAAGGAACATTTTCTGATG
AGATTATTGCAGATCTTATACTGTTTTTGCTCTTTGCGGGTTTTGAGACATCATCGACAA
CGATGGCATTTGCTGTAAAGTTTCTCACCGAGCATCCGCGGGTATTAGAGGATCTCAGGG
AAGAACATGATGTCATCTTAAAGGCAAAAGGAAAAGACAATCAAAAGCTGACATGGGATG
ACTATAAGTCAATGAAATTAACTCAGTGCGTGATAAAGGAAACACTTCGTTTGGTTAGTG
CTGCTCCATGTGTTTTTAGAGAACCCAAACAAGATATCAAATTCAAAGATTTTATTATTC
CAAAAGGTTGGACAGTGTTTGTATTCTTAAGTGGCCAACATTTGGATGAGAAGTACCATT
TCGAAGCTCACAAGTTCAATCCATGGCGTTGGAAAGATGAGGGTCAAGAGCTTTCTGACA
ACCCAAGGTATATGCCATTTGGAAAAGGTGCCAGGCTTTGTCCTGGATATCATTTGGCAA
GATTTCAAATTTGCCTCTTCCTTCACCACTTCATCACTAAGTTCAGATGGGAGCCAATAG
AAGYTGTTAGATCGACATATTTCCCTCTTCCTTCCATGGAAAAAGGATTTCCAATTCGTC
TCCATTCCCGACCACTACAATGATGATGTGTCAATCTAYTATCTTACTAAGATGTTTATG
ATTAATTTTCACTGCATCTMCAATAACCATGTATCTCGTGGTCAAATTTGACAATGGTGG
ATCGATTCATGAAGTAGATCATTTTCATCCAAACATGGCTCTTTGAATGTTACATTAATA
GAGCATTTATTTCATTATGAATAATTGTACAAATGGTGTTATTAACATTTCATGTTATCT
TTCATGGCATGGCTCTTTCAACACCACTTCAACAATGTTATTGTTCATAATGACACTTAT
ATTTCCAAGAAGATTA

CYP27

TTTGGAATCACATTCCTCATTGCCTMAACATGGATGTCCTAAATGTTCCCTGGTTCTCTG
TTGTGGTTGGCAGTGTGTTTGTTTACTTCTTGTWCAGAAGACTTAGGAAGGTTCAGTCAG
GGCTTCCACTTCCTCCTGGACCTCCACCTCTTCCYATTGTAGGAAACCTCCTTCAATTGG
GAGGAGGTAAGGCCACAGAGGTTATGCTTGCTCTTTCTCAGCAATATGGTCCTCTCATGA
CCCTCCGCCTTGGRACGAAAACYGTCATGGTGGTCTCTTCRCCGGCCATGGCGAAAGAGG
TTYTGAAAACTCAGGATCATAACCTCGCSGGACGATGGGTTTTGCAAGCAGCTAAGACGC
TTGACCAGCACAAGTCCTCAATTGGGTGGGCTGAGTATGGACCTCACTGGAAGAAACTCA
GGCGCATTTCTAGCACTGAGCTCATGACCTCCAGGAGACTCCAAGCTCTGCAGCACCTCA
GAAGTGAGGAAATCTTCCGAACTGTTCGCTTGATCTTTGAGGAGARGGGRAWGAGTGTAA

ATATTTCGCATACAGTGTTCTACACCGGTCTGAATCTGTTAGGCAAAATGATCTTCAGTG
CAAGTGTGTTTGATCCCAACAATCCTGAGTCCATGGAGTTCAAGGATGCATTTTCCCTGT
CGATGAAGCTTGCAGGCACGCCAAATGTTGCCGACTTCTTTCCRTTTCTGGCGTTCCTGG
ATCCCCAGCGCCTATGCCGTCAGATGGCCATTAATTTAGGAAAGTTCATAAGTTTTTAG
ATGTATATATACAGGATCGTGTGGCCAAGAGAGGCAAAAACGTTGAGCAGAACAAAGACG
CMGACAAAGATTTTCTAGACGTCTTGCTGGATTTCAGTGGCCCTGATTTGACCATCAAAG
ATATTCGAGCTCTAATYTTTGAATTGTTCATGGCTGGTAGCGACACCAGTGCAACCACAA
TAGAATGGGCAWTGGCAGAACTGATTCGCAAYCCCCACACAATGAAACGAGTCCAAGAAG
AATTGGATCAAGTAGTCGGTCACAACAGAAGAGTGGAGGAATCCGACATAGACCAGTTGC
CTTACCTTCATGCAGCAGTGAAAGAAGTGTTCCGTTTGCACCCCCCAATTCCTCTTCTGG
TACCACACCAGGCTGAAAACTCYTGCGAGGTTGCAGGGTACTCCATACCCAAAAACAGCA
AGATCATGGTGAACGTGTGGGCAATTGGAAGGGATCCTACAGTGTGGGAAGAGCCCTTAA
AATTCATTCCTGAGAGATTTCTGGAGGGTGACAAGAGCAAGAYAGAGTACAAGGGACAGG
ATTTTGAACTGTTGCCATTTGGGGCAGGAAGAAGAATATGCATGGGACTCCCTCTGGCAC
ATAAGATGGTGCATCTGGTRTTGGCAGCACTGGCTCACTCCTTTGACTGGGCACTTCCTG
ATGGATTGCCTTCGGAGAAACTGGACATGAGCGATGAGCTTGGAGCTGCATTGAAGAAGG
CCCAAGAGTTGACTGCCATTCCCACACCGCGCTTGCCCCATCACCTATATTAGTACCTAY
AATTATGAGTTCTTTCWTTTCAGTAAAGCTCCACATAGATATTGCCAAATGGTTTCTGGT
TTAATTTTAGACTTTTTTTTWAAATCTCAATTTGGAAGGCACCAGAGAAACCGTTGTTTT
ATGTAAATTGATTTTGCAATCATGAATTAATAAATATTTTGGCAACTTAAATTTACATAA
AACAACGGTTTCTCTGGT

CYP28

ATGGACGCAAATTCGTGCAGCCTGGTTTTTATTTCCTGCTCGCGGGCATAGCCCTGTCA
TATCTGATATTCAGCTGGAAGCTGTCCAGCAAGGCATCGAAGCGGCTGCCACCTGGACCC
CGACCCCTGCCTCTGATAGGCAACATACTCCAGCTGGGCAAGAGTCCCCACGAGTCCCTC
TTCCACCTGGCAAAGCTCCACGGGCCCTTGATGACCCTCCACCTGGGCTTCAGGACCACC
ATCGTCGCCTCCTCCCCTGCAATGGCGAAACAGGTATTGAAAACGCAGGACCCGGCCCTG
TCAGCACGCACAGTCATCGATGCAGCCACGTGTCTGTCCTACTGCGACCACTCGCTCGTC
TGGAGCGATTGCGTGCCACGCTGGCGCACGCTCCGGCGCATATGCACCACAGAGCTCTTC
TCCGCCAAGCGCCTCGATGCCCTCCAGGATCTCCGGAGAGTCCAGATCTGCAACATGATA
AGGACGATCTATGTGGATTCCCTGGCGGCGTCCAACGTGGACATCGGCCACGCCACGTTT
TTGACGAGCCTCAATTTGTTGGGCAACATGATATTTTCAAAGGATATGTTCGACAGGGGC

TCGGAAAAATCCCACGAATTCAAAGACACCATCAGGAAGATGATGGTAGTTGGCGGGAAG
CCAAATTTAGCGGACTTTTTCCCGTTTCTGCGGTTTTTGGATCCTCAGGGCKTGAGGAAA
GAGACGATTGCATATTTTAGGGTTATGTTTGGGATTTTTGGTGAGCATATAGAGGAGCGG
CTTCATAGCGGAATAAATGAGGATAGCGAAAAAGATTTTCTTGATATTTTGCTGGAGAGT
AAAACAGAGACAGGAGAGAATCTCACAAAACTTGATTTAACGTGCTTTTTTTATGAYTTG
TTTGCTGCGGGGAGCGAAACAACAAGTACTACAATTGAATGGGCAATGGCTGAAATWWTA
CGTAACCCAAAGATAATGGAAAGAGYAAAGRAAGAATTAGACAAAATAGTGGGGCGGGAA
AGAATAGCACAAGAAAATGACATCGACAATCTTCCTTATCTTCATGCAGTGGTTAAAGAA
ACCTTTCGTMTGCACCCTGCCGCACCTCTTCTAATCCATCATARAGCCCAAAGCTCTTGT
GAAATAGATGGATATGTTATACCCAAGGATGCCCAAGTGTTTGTGAAYGCTTGGGCAATT
GGAAGGGACCCGATAACATGGAAGGACCCTGAAAGGTTCTTTCCAGAAAGATTCATGGAA
TCAGAGGTGGAATATAAGGGGCAMCATTTTGAAYTGATTCCATTCGGGTCAGGGAGGAGA
ATATGCCCGGGGTTACCTTTGGCTCATCRAATGATTCACATAGGGGTGGCCTCTCTCATT
CAATGCTTTGATTGGTGTCTTCCAAATGGACAAAATCCRGAAGATTTAGATATGRGAGGA
GACTTYGGAATCACATTGCAAAAGTCACAACACTTGGTTGCAACACCCACACCCCGCTTG
CCATGTCATGTCTATAACTAATCACCAATCATTGCAAGACATCTTATAGGTAATTACTTG
AATAAATCAAGGTTACAATACTTTGTGTAGGTGTTATTATCTCCAACTAGAGAAATTTAC
CAGAAGTGTCATATAGTGGTCAATACCTCTCTTTACTATACAATTTTTGAACAATTATGC
TTTAAACAATAACACTGACGAGTGCTTAATGAAATTCAATGATGATCTACA

CYP29

CCCGAAGCTCTGCAGAAATTTGTGGGCAGAATGGAATCRGYTATTSTRCAGCATTTTTCY
GATTGTTGGSAAGGARAASAGAGTATTACAGTGTGGCCACTGATGAAGAAGCTTACTTTC
CATGTTGCCTGTGATCTGCTCTTCGGTTTGAAGGACGATGAAGAGAGGGAAATATTGGGC
AAGGAATTCGATACAGCTTCCAAGGGTCTGTATTCTCTCCCCCTGGATTTTCCAGGGACT
GCTTTTAACACGGCCTTAAAAGCTCGTTCTCGCATATATAAGCGGTTATCATTGCTGTTG
GAAGAGAGGAGGAGGGAAATAGAACAAGGTCAGGCTTCTCATTCCCAGGACTTAATGTCA
TGCATGCTCACTATGAAGGATGAAAATGGGGAATCCCTCACAGAAGAAGAGATCATAGAT
AACATGATGCTAGTTATGATGGCAGGCCATGAGACTACCTCTGCACTTCTCACAAATCTC
ATAAGAATGTTAGCATTACACCCAAATGTCTACCAAAAAGTTCTTCAAGAACATGAAAAA
ATCTTAAGAGACAGAGAATCACATGGCCCCTTGAGATGGGAAGACATTCAAAAGATGAAA
TATACGTGGAAGGTGGCCCAAGAAACACTTCGTTTAGTACCTCCACTTTTTGGTGCTTTT
AGAAAGACAATCAAAGATGTAGAATATAAAGGTTACACAATTCCCAAAGGCTGGCAATTT

TACTGGGTCACAAGTACACATTGGAATGAGGAGATTTTC

CYP30

CCGTTCCACTTACAGTGTTCTGTAACACTACAGAGTACAGCCCGTTTTTAACAGGCAACT
ATGGATATTTCCCCGATGGTATTAACCTTAGGGTTAGGGTTTCTGTTGCTTGCTGTTTTT
CTACTCAAGAAAAAGCCCTCCTCCGGCATTAATCTCCCCCCTGGCAGCTTCGGCTGGCCG
CTGCTCGGAGAATCTCTAGAGTTTCTGGCCTCTCAGAGGGATAACGCATGTTACGATTTT
ATGGAGAAGAGAGTTAAAAAATATGGAGATACTTTTAAAACCCACTTTCTGGGAGCCCCC
ATGGCCGTCTTCTACAGCCCTGCAGGTAACCGCTTTCTCTTCTCCAACGAAAACAAACTA
GTTCAGACCTCTTGGCCTTCTTCTCTTGCCAAGCTATTTGGAAATTCCATTCTTACAAAG
GTTGGGGACGAAGCGAAAGATCTGAGGAAGACGCTGCTTACTTTCTTGAGGCCTGAAGCT
CTGCAAAACTTCGTARGCAGAGCCGACACCATTATCCRTGACCACCTCACCCGTTACTGG
GTTGGGAAGGAAGAAGTAAAGACMTTTCCTCTCGCTAAGCGTTGCTTGTTTACAGTGGCT
TGCGCTCTTTTTGTCAGCCTCGGCGAAGGACCGGAACARGACGAGTTGTATGATCATTTT
ATGGACTTGATTAACGGAATGCTACAGATTCCGGTGGAYTTGCCGGGAACTCTGTACCGY
AAATCCAGAATCGGCAGCAATCACATACGCCGAATACTACAAAGTGTGATTGACAAGAGG
AGAGMGGACCTGGCAAGCGGTTTGGCTTCCTCGGATCAGGATCTGCTATCCTACCTGCTC
TGCAACGTAGATGGACGWGAGAATCCATTGAGTGATAKTGACATTAASGACAACATCATG
CAGCTTCTAATGGCSGGCCACGACACCAACGCTGTTRCTGTAACACTCCTGTTGARGTAT
CTGGCTCTCAATCCACACTGTTACCGCCAGGTCCTCCAAGAGCAACTGAATATAACGAGA
GAGAATGGGGAAAAGCAGCTACAATGGGATGACATTCAAAGGATGAAGTATACATGGAAC
GCCGCACAAGAGACTCTCAGACTCAATCCTCCTGCAGGAGGAACATGGCGAAAGGCCATT
GTTGAAATTTCCTATGGTGGTTTCACCATCCCAAAAGGCTGGAAGCTATTTTGGACAGCG
AACTCGACACACACAAGAATCCAAAGTACTTCCCTGATCCAGACAAATTCGAACCCGCAAGG
TTTGAGGGGAACGGACCCGCCCCCTACACATTCGTGCCATTTGGAGGAGGGCCTCGAATG
TGTCCCGGCAATGAATTCGCGCGTATGGTGATTCTGGCGTTCCTCCACAACATTGTGAAA
ATGTTCGAATATGATTTGGTAGATGTGAATGAGAAGGTTATTGTGGATCCGTTTCCTACT
CCGGTGGATGGGCAACCCATTAAGCTCCGCCCGCAATCAAAGAAATAAGTTAACCCCGTC
TTAACTGCAGTTTGTTTCAGACAACGAGCTAT

CYP31

ATGGCATTATCTGATATGCTTCAACTTCTTCAAGAAAGAGCAGCTGAATTTTCTACCCAA
ACAACAGCTGTGGGAATCACCATTTTAGTCCTCCTGTGGGCTTTCTACTGCTTCAAGAAA
AATCCAAAAYTGCCTCCGGGGCCATCTCCATTGCCYATCATAGGAAACCTCCATCAATTA

GGAAAGCTTCCTCACCGTATWCTGTACGATCTTTCTAAACAATATGGCCCAATTATGTTT
TTGAGATTTGGGTCTCTGCCTTGTGTRGTGGTTTCCTCTTCTGACATGGCGAAAGAGTTT
TTAAAAACTCATGATTTGGTATTTGCCAACAGACCATCTTCTGCTGCTGGAGAACACATT
GCATACAATTACAGAAACCTGGGTATGTCTCCCTATGGACCYTACTGGAGACACATGAGG
AAAATATGTGTGATGGAATTGTTGAGTGCAAAGAGAATAGAAAGCTTCAGGTYTATAAGA
GAGGAAGAGCTTTCTTTGGCTGTGAAGTCTGTGTGGGAAAAGAGCAGTAAAGGAACTGTT
TCTGTGAATGTCAGCAAGAGTTTTTCCTCCCTAACTACTGCTATMATCTGGAGGATGTTG
ACTGGGACTAAATATTCYGATGATGATGTGGGTGGCAGTGGAAATGAGCTCAGAGATATG
GTCAGGGTGGTTACCAGTACCATTGCTGCTGTTAATATTGGGGACTTCTTTCCTTATATA
GATTGGATGGATTTGCAAGGTATCAAGAAAAAGATGAAGAAGGCACACARATTTTTCGAC
AGAATTGTTGAGAAAATTATTGATGACCATGTTGAACAGCGGAGGAGGGCTTTGGATGAG
CACAAGGAGATGACTAAGGATATTGTGGATGTTCTTCTTGATATGGCGGAGTCTCAGACG
ATGGAGATGGAGATCACACGGGAAAAYATCAAAGCAATTATTTTTGATTTGTTCATTGGT
GGGATGGAGACAATGGCTACCACAATGGAATGGGCAATGACTGCACTCCTCAGAAACCCC
CGTGTAGCCAAGAAACTACAGGAAGAAATTGATTCAATAGTAGGCAAACACCGCATGGTA
ACTGAATCTGATCTACCCACCATGGACTACCTGCAATGTGTGCTGAAAGAGTCTCTGCGA
TTATACCCACCAGGRCCATTGCTGCTCCCCCACGAATCTAGAGAAGCTTGCACTGTAGGC
CCACAGGGCTACTTTATACCTGCAAAAACAAAGCTTTTTGTTAATGCYTGGGCCATTGCC
AGAGACCCAACTGTGTGGGAAGATGCCTTGGAATTCAAGCCTGAGAGATTTATGGGTAAT
AAGACTTTTGATTACCAAGGGCAAGARTTTGACATGCTTCCTTTTGGAGCAGGAAGGAGA
AGATGTCCTGGTGGTCACATGGCTATGGGGCAGATGTCCCTTGTGTTGGCTCAGCTGTGG
CATTGCTTTGACTGGAGACTTGAAGGAAATCCAGATGATTTGGATATGACTGAAGAATTT
GGAGCCACTGTGCCAAGGAAATTTAATCTCTTTGCCTTTCCTTCCTTGAAATTGTCCATA
AACTTGTAGGTTAGATTTTATAATCAAYCTTATGGTTAAACCTTKTGGTTTGCAGTGATG
AAAATCTTTATGCATAGGACCTCTTACTTCTTGATAARCTCATATTTGTTGAGTTTTGTA
ATAGGATGTAGTAGAAGGCTTGATGAGTTCCTGGTTTTATGTAAAGAGATGCAATGGGTA
TCTTAAAAAAGAATGGGTACTGTGAAGAAATGGTCTCCTCTTTTAGGTCAATGAGGTTTC
CTTTTCAGCTAGTCTTGGTTTTGGCTGTATATTGCTCTGTACAGCTGCTGGTTATGTGTA
TGGTCAAGGGTATTTTTTCATTRTATTTGTGCCAAGTGCCTTAGTGATTTTAGCTTTGTT
AAGTGTGATGTAATTGTTCTTTTAATGTATTATTCACTAGTTACATTC

CYP32

AGAGAGAGAGAGAGAGAGAGAGGTGGAATGGAGGCGAATATGCAGTTGGCAGTGGCAGTG

GCAGTGGTGGTRGTGGTGGTGGCTTGGAGGATCATTTCTGGAAAGCGCAATGGAAACTAC
AAAGATTACAATCTGCCTCTGCCACGTGGGGCGACTGGATGGCCACTGATCGGAGAGACC
ATCCCTTTCCTTCGAGCCATCAATAACCCTCGCAACCCCCGAAAGTTTGTTGAAGAACAT
GAGCGCAAGTATGGSCCGGTATTCCGAGCAAATTTGTTTGGAACATCAAAGACTGTGGTA
ACAGTCGATCCTGAGTTCAGTAAATAYGTGCTGCAAAACGAGGGAMGGCTATTCCACAGT
AAATACCTTTTGTCGATGAGGACTATTATTGGGAAATACAGTATCCTGGCCGTGCGCGGA
GAACTTCAGAGGAAGCTCCATGGCACYGTTAACAAATTCTTGAGCGCCGAGAAGCTTAGC
TCGGAATCCATCGCCTATATTCAGAATATTTTTGARTTGGGTATGMAGAAATGGGAGGGG
CACAATAGAGAATTGCATCTTCAAGAGGAGATTCATAAGGTTGTTTTGAATTTGATGGCA
AAGAAACTTATCGACTTAAATCCAGAGGAGGATATGAAAGAAATCAATAAACTCAGGGCT
CTCGATGACTTGGCRGGAGCGATAGTAACCGTTCCCCTCAACATCCCCGGCTCTCCTTTC
TCAAGGGGAATTAAGGCTAGGAAAACTGTGCTACAGAAGATTCACGAGTGCATAAAGGAG
AGGAAGGAGCATCCGGAAGTGATYCGCAATGAYATGTTGACGCATCTRGTAAAAGACGGC
AGCTTTCCTGAGGAGATCATTGCAGACTTGATAATGTCTCTCTTCTTTGCATCTTACGAG
ACTTCCTCCAAAGCCCTCACTTTCGCTGTTAAATTCCTCACAGACACCCCGCGCGCRTTG
GAGGAAATGAGGGCAGAGCACGACGCCCTCGTGAGCTCCAAAGGCGATGGAAAACTCACA
TGGGACGACTACGCCTCAATGAAATTCACTCACTGGGTGATAAAAGAAACTCTTCGGTTG
GGCAACGTTGCCCCCGCAGTATTCAGAGAAGCGGCACAAGACATCAAAACTAAAGATTTT
GTGATTCCCAAGGGATGGGCAATCTTGGTGTTGCTGAATGGCACGCATTTGGATGACAAG
AATTATTCGGATCCTCTCACATTCAACCCATGGCGCTGGCAAGACGAGCCTGCTCTCGAR
CTTTCCAAGRACCCATGGTTTCTGTGTTTCGGAGGAGGCGCCAGGTTTTGCCCGGGWTAC
CATTTTGCCAAACTTGAAATGGCTCTCTTTCTGCATTCCTTCATTAGTAAATTCAGGTGG
GAGGCGGTGGAAGAAGATGAGCTTCATTACTTYCCCTACACTTACATGAAAAAGGGGCTG
CCCCTTCGTCTCTCTCCTCGAGATTCCACCTAGARTATTAKATTTYAYRCCAAATAACAA
TGTACGGAAAGATGACGTGGCGATGAGCTCCACATCTTATATTGCCCTTTTACAAGTTTG
GAATTCAACGCCATTTCAATCTTACATTCATAAATAGCTCAACACACTTTAATGCTCAAC
GTCAATATAAATTTAATGTTTCCATTACAAATTACT

CYP33

CGCACTCGCGCCTTTCAGAAATCTTATCGGCAAATATGGATTGTCGGCGGTACAGGGGGA
ACTTCAAAGGAAGCTCCATGCAACTGCTGTCAATTTGTTGAAGCATGAGACGCTCAGCTC
TGACTTCATGGAAGATATACAAGACATCTTTCAGGCTGGAATGAGAAAATGGGAGGAGGA
GGGAGACATCCCTATTCAACACAAGTGCAATCAGATTGTTCTGAACTTGATGGCGAAGAA

ATTGCTGGACTTACCTCCATCAGAAGAAATGGGAGATATTTATAAAGCTTTCGACGATTT
CGTSGGAGCTGTCCTCTCTTTSCCCCTCAATATCCCTGGAACCACTTATGCSAGAGGAAT
TCRGGCCAGGGGAATTCTGTTAAAAAGAATTCACAAGTGTATAAAGGAGAGGAGAGAACA
TCCAGAGGTGCTCCGCAATGACTTKTTGACCAAACTTGTGAGGGAGGGCACATTTTCGGA
CGAAATTATTGCAGATACARTAATCTTTTTTGTGTTTGCTGGTGTCGAAACTTCAGCAAT
GGCCATGACGTTTGCTGTAAAGTACCTCRCTGAGAATCCACGAGCACTGGAGGAGTTGAG
GGCTGAGCATGACGCTCTTTTGAAGGCCAAAGGGAAAGGCAATGAAAAGCTGACGTGGAA
TGACTACCAATCAATGAAATTCGTTCATTGTGTAATAAATGAAACACTTCGTTTGGGTGG
TGCAACCGTGGTTCTTTTCAGGGAAGCCAAACAAGATATTAAAGTGAAAGATTTTGTTAT
TCCCAAAGGATGGACCGTTTCTGTTTCTTGAGCGCCACACATGTTGATGGAAAATACCA
TTATGAAGCTGACAAATTCCTCCCTTGGCGCTGGCAAAATGAGGGTCAAGAAACGTTGGA
GGAGCCATGTTATATGCCATTTGGAAGAGGTGGCAGGCTCTGTCCAGGACTCCATTTGGC
AAGATTTGAAATTGCTCTCTTTCTTCACAACTTTGTCACTAAATTCAGATGGGAGCAGCT
GGAAATTGATCGTGCGACTTACTTTCCTCTTCCTTCCACAGAAAATGGTTTTCCAATCCG
TCTCTATTCTCGAGTACATGAATGAGGAAGATTTTATGTACACGAATAAGGAAGATTTTA
TTTTCTAATTAASTAGCAATTGYATGTGTACAATGCAATGTGAACGYYGAATGGTATATG
TTATTYGATTTARAGGGATAGTTATTACTTTTGTAGTTGATGTTGATAAGGGGTGTGATT
GAACATTTTGATCATAATCCCAAAAGGAGAAAAAGTAATGTCATTGTCCAACTCCTGCTT
TAAGTTAAATTGACATGCTTCAGTTG

CYP34

AAGCATTTCATGGATCCTCAACCAGATTTTTTTTCATCAGCTCTGGCACTATTCACTCTG
CTTGTCTGCCTACTGCCTCTGATCATAATAGCCAAGTTTGTTAAGAAAGGTACAGGGAAA
AACATCCCTCCAGGAAGCCTTGGCCTGCCTGTTATAGGCCAGACCCTTCAATTTCTAAGA
GCAAAAAAATATAACAAAAGCCAGGAATGGATTCAAGAGAAGGTTGCAAAATATGGCCCC
GTTTTCAAAACATCCTTATTGGGTTGCCCCACAGTCGTCCTCACCGGCCAAGAGGGTAAC
AGATTTCTGTTGCAGAATGACTACAGCATGGTTATTAACAAGAAACCCATCACACTTGCC
AGGGTTTTGGGAGAGAAAAACTTGTTTGAACTTCGCGTTGAAGATCACAAGCGGATGAGA
GCTGCCATCATGCAATTCATGAAGCCCGA

CYP35

GCAATCACAATTATTAACTTAAACATATATARRAGTGGCAGAAGATCCACAATGGATACC
GTGGCCGAATGGACAAGCAGTTCWTGGATTCTTATAGCCGCCATGATTATTGGGTTCCTG
TTATTTTATCAACATATCACACATACAAACAATGGCAGACTCAATCTTCCACCGGGCAAT

CTGGGGCTTCCTGTAATTGGAGAAACATTACAGTTTTACAGAKCTCAGAAATCGAAYCGA
CTCTTCGAGGATTTCTTCGGGAAGAGAASCCACAAGTATGGGAGCGTTTACAAGACCGGG
TTGATGGGCTCTCCCACCGTAATTTTCACCGGCGCCGCCGGTAACCGTTTCCTGCTCTCA
AACGAATTCAAACTGGTTACGAGCTCGTGGCCGCAGAGCACCATAAATCTGATCGGCACG
GGTTCCATTATGGAAAAACAGGGCGCGGAACACAAGCGGCTTCGAGGCGCCGTCATGGCC
TGTTTCAGAGCAGATGTTCTGCAGAGCTTCGTGGGAAAACTCAGCAACGTAATTGAATAC
CACTTTGAAAGACATTGGAAAGGAAACCGCGTGGTTCGCGTCTATCCTCTCACGAAACTG
CTCACTTTCACTTCTGTGTGCAGTCTYTTTCTGGGACTGGAAGACGAGGGCGAGATCGAG
TATTTTCTGAAATTGTTTCAGATGGTGTTGCTCGGCGTTCTGTCTTTGCCGGTGGATTTC
CCAGGGAGTCGGTGCCGGCGAGCGAAACGTGCGAGGCGAGAAATCGACGAGTTTCTGTCT
GCGTTAATTGACAGGAGGCGGAGGGATTTGGCGGAGGGRAGGGCTGGCGGTGAGCAGGAT
TTGTTTTCCAGGTTGCTGACGTTAAATGACGAAGATGGTAACAGGCTTGGGGATAAGGAG
GTGATGGATAATATTGTCTTGTTTATGTTTGCTGCGCATGATACCACTTCTTTGGTCATT
ACTATGATGTGTAAATTCCTGTGCCACCATAAAGATTGCTATCATGAAATCTGCCAAGAA
CACAAGAAAATTCTAGAAAGCAAGGAACCAGGAGAGAAGCTCAATTGGGAAGACGCAAAA
AAACATGACATACACATGGAGAGTAGCCCAAGAACAATGCGGTTGTTGCCTCCAATATTT
GGTTCATTTAAGAAGGCAATTGTTGATTTTGAATACGAAGGTTACACCATACCAAAAGGA
TGGAAGGTTTTATGGACAGCATCAACTCACTATGATCCAATTTTTTTTGAGGAGCCACAA
AAATTCAAGCCGTCACGGTTTCAAGATACGATTCCACCATACGTGTTTGTACCATTTGGA
GGAGGAGTCCGTTTATGTGCAGGTTTAGAGTTTGCAAAATTACAAATTATGGTGTTCATG
CACCACCTTGTAACACAATATATTGGTCACTGGTGGATCCCGCTGAAGGAATACAGATGG
ATCCTCTTCCTTCTCCAATCAAGGGAATGCCTGTAAGGCTTGTGAAAAAGAGTTATTAAA
CAAATTCTCAATTTAGTAGTAAGGAATGTAGAGATTACTTGCCATGTATAAATTGATGGA
GCATGTAATGTAAGTGCCGAATTGGTTGATTGACAAATTGATCCTGCAAATATATGTGAG
TTTTTAGGTTGAACGATGTAAAATTAGTAGAGTTAAAGGTGGCTCACATACTTATTTTGG
CTTTATAATGTTAGATGGTGAGAAGA

CYP36

GATCTATATAATCTTCTGCATCATCCCTTCTCCATTTCATCTCTGGTTTCTGCTCTGCCT
TTTCTTACAAGAGTTTTCATAATGGATGCCCTTTCTCTTGTAAACAGCACAGTTGCAAAA
TTTAATGAGGTAACGCAGCTACAGGCTTCCCCTGCTATTCTGTCCAYTGCCCTCACTGCT
ATTGCAGGCATTATTGTGCTCCTCGTCATCACWTCTAAACGCCGTTCCTCTCTTAAACTT
CCTCCTGGAAAACTAGGCCTCCCTTTCATTGGCGAGACTTTAGAGTTCGTGAAGGCTCTT

CGATCAGACACACTTCGACAATTTGTGGAGGAAAGGGAGGGGAAATTTGGACGTGTGTTC
AAGACTTCATTGCTTGGGAAGCCCACTGTAATACTCTGCGGCCCTGCGGGAAACCGCTTA
GTTCTTTCCAACGAGGAAAAACTGTTGCACGTGTCGTGGTCCGCCCAAATTGCCAGAATY
CTGGGTCTCAATTCTGTTGCAGTGAAAAGGGGAGATGATCACCGCGTTCTGCGTGTCGCA
CTAGCAGGTTTTTTGGGCTCTGCAGGGCTACAGCTTTACATAGGTAAAATGAGTGCACTT
ATCAGAAATCATATCAATGAAAAATGGAAGGGAAAAGATGAAGTGAATGTACTGAGTTTG
GTAAGAGATCTTGTCATGGACAATTCAGCTATCTTGTTTTTCAATATATACGATAAAGAG
CGAAAGCAACAACTGCATGAAATATTGAAAATCATTCTTGCCTCACATTTCGGCATACCT
TTAAACATTCCCGGATTTCTGTATCGCAAAGCACTCAAGGGGAGYTTGAAGCGGAAAAAA
ATTCTCTCCGCTTTACTGGAAAAGAGAAAAGACGAACTGCGCTCARGATTAGCGTCTAGC
AATCAAGATCTTCTCTCTGTTTTGCTCAGCTTCAGAGATGAAAGAGGGAAACCACTGAGC
GACGAGGCAGTCTTAGACAACTGTTTTGCAATGCTGGATGCCTCCTATGACACCACCACT
TCACAAATGACTCTGATTTTAAAGATGTTGTCTTCCAATCCAGAATGCTTTGAAAAAGTA
GTTCAAGAGCAATTGGAGATAGCKTCAAATAAAAAGGAGGGAGAAGAAATCACAATGAAG
GATATCAAAGCCATGAAATACACATGGCAAGTGCTCCAGGAAAGTCTACGGATGCTTTCT
CCAGTATTTGGAACACTTCGTAAGACCATGAATGACATTAATCACGATGGTTACACAATT
CCAAAAGGATGGCAGGTTGTATGGACAACTTATTCTACACATCAGAAAGACATATATTTC
AAGCAGCCAGATAAMTTCATGCCTTCGAGATTCGARGAGGAAGATGGGCATTTGGATGCT
TATACATTCGTACCATTTGGAGGAGGACGGCGGACATGTCCAGGATGGGAATACGCAAAA
GTGGAAATATTACTGTTCCTCCATCATTTTGTGAAAGCATTCAGTGGTTACACCCCAACT
GACCCTCATGAAAGGATTTGTGGGTATCCAGTCCCTCTTGTCCCTRTCAAGGGATTTCCA
ATAAAACTTATCGCCAGATCCTGAGGCATGGACTACATTGGCATCAACATCCACAGAATA
AACAATATCGTCATAATTTGGTTTTATTRTTTGTGTTCCTTCAACATGATCTGTAAGCCT
GTTTTTTCATGAACGAATACTAATGAATGTTTCGCACCCTGTTCCATCATCGTTCGATTT
TTAATTTACATTGTTTTACAGTGGCC

CYP37

ATTCTTTTCCTATCCAAACTCTCCTCTAATCAATACATGATTATGACCGTATTTCCCTAA
ACAAACTCTCCCCTGCTCTATAGAAATTCATGAACTGCCTGATCATTCACACAACAAAGG
AAGGAAAGAAATGGAATCCATGATGTTGGAAGAGTTTGGAGGAGTCGTGACGGATATGTA
TCCTGCAACATGGATCGCCCCTCTAGTGATTCCTTCTGTCCTCTGTTTCTGTACTATTGC
GGTGGGTACTGTAGCATATGCTTGTGCTGTCATTTATGGTCCTCTGTTGTTGACACGTTT
AAGAATGATGAGCCGCAATATGCCTCCTCTTCCGCCAGGGAATATGGGGTTGCCCTTCTG

GGGAGAATCTCTGGACTATCTCAAGTCCATGAATAACCAGCTCAATCCGGATGTGTGGTT
AGATACGCGAAAGGCTAAGCATGGGAATATCTTCGCAACCCATATTCTGGGCAGCCCCAC
AGTGGTCATGCTGGGTGCTGCCGCCAACAAGTTCATCCTCATCAACGAAAACAAACTGTT
CCGCAGCAGTTGGAACAAATCTTTAACTGTTCTGATCGGAAGTAACGCCATGGTTACGTC
GCAAGGCGCCGACCACAAAAGGATGCGGCGCATCATCCACAGTGTTTTGGGGCAGGAAGC
GTTGAAGAATGCCGTGGGAAAATTTGAAGGGCTAGTTGTAGAGCATCTCGACTCGTATTG
GCACCCCGGCGGAATTATACTCGCACACCGCCTTGTTAAGGACATTACACTGCGCGTGTC
GGCGGAAATTTTTATGGGATTGAAGCCTAGGAAAGAGTTGGAGAGTCTCCAGACGCACTT
GCATGATTTCAGCGCGGGGCTCTTATCTCACCCTCTCGACCTTCCCTGGACGCCCTTTGG
CAAGGCAAAACGAGCGCGCGCTGCTATGCTCGCTCAAATTCACCAACTAATTCGGCTGCA
GAGAAGCAAAATGGACGGAGACGAGAAAGGAAACTTCATTGGCATGGTTTTGTCGGCGCA
CAAGTCCGGAAGCGATTCAGTTCTGAGCGACGAGGAGATTGCAGACAACCTGCTCGGCCT
TTTAACTGGCGGCTACGACACGACATCCTCCTGTTTGGCCACAATTTTGAAGCACCTCGC
TCTCTCCCCGCTTGTAAAACAAAGGCTTCGCACAGACTGTGAAAAGCTTAGAGAAAGTAA
GAAGGCAGGAGAGCCTGATCTCACCTGGAGCGACATAAAATGTTTAGACTATTTACGCAA
CGTAATCTCAGAAGGGCTGCGCATCGTGGCTCCTATCAATGGAGGATTCAAGCAAACAAA
AATAGACGTAATATACGGAGGTTATACTATTCCCAAGGGATGGAAGGTTCATTATTCTAT
CAGACACACAAACAACAAAGAAGAGTACTTTTCCAGTCCAGAGATATTTGATCCGGATCG
TTTTAATGAGCGACACGAGCCGTTTTCGTTCATTCCTTTCGGGCAGGGCAATCGAATGTG
CCCTGGAAATGAATTCGCAAAGTTAATAATGGAAGTTTTTCTCTATCATTTAGTGGTGAG
ATATGATTGGGATTTGGTGGACGCAAATGAACGGATCGATATGATGCGGTTGAGCCCTTC
CCCTCTCCAGGGTTTGCCTTTGCTACTTAAACAAGTTTATTAATTTGTATTAATCTATGA
GGCTCACAGAGAACTAACTATAGATGAGATGTATACTATCCCCGATGGCCTACCTTACCC
TTACTCTATAAGGGCCCAGTTATGGATAAGGCGGGGAAGTGTCTGCTTTCCCTGCAGATA
AAGATATATAGTTTAGATAGAAATATAGAGTTCAGAAATAGTGCCGTAATTATCTATGGG
CACAATATATGGTAAATAGGAGCCATCCCCATCTGCATCACATTCTAATTATATCATAAA
T

CYP38

GAATCACAATTATTAACTTAAACATATATAGGAGTGGCAGAAGATCCACAATGGATACCG
TGGCCGAATGGACAAGCAGTTCTTGGATTCTTATAGCCGCCATGATTATTGGGTTCCTGT
TATTTTATCAACATATCACACATACAAACAATGGCAGACTCAATCTTCCACCGGGCAATC
TGGGGCTTCCTGTAATTGGAGAAACATTACAGTTTTACAGAGCTCAGAAATCGAATCGAC

TCTTCGAGGATTTCTTCGGGAAGAGAAGCCACAAGTATGGGAGCGTTTACAAGACCGGGT
TGATGGGCTCTCCCACCGTAATTTTCACCGGCGCCGCCGGTAACCGTTTCCTGCTCTCAA
ACGAATTCAAACTGGTTACGAGCTCGTGGCCGCAGAGCACCATAAATCTGATCGGCACGG
GTTCCATTATGGAAAAACAGGGCGCGGAACACAAGCGGCTTCGAGGCGCCGTCATGGCCT
GTTTCAGAGCAGATGTTCTGCAGAGCTTCGTGGGAAAACTCAGCAACGTAATTGAATACC
ACTTTGAAAGACATTGGAAAGGAAACCGCGTGGTTCGCGTCTATCCTCTCACGAAACTGC
TCACTTTCACTTCTGTGTGCAGTCTTTTTCTGGGACTGGAAGACGAGGGCGAGATCGAGT
ATTTTCTGAAATTGTTTCAGATGGTGTTGCTCGGCGTTCTGTCTTTGCCGGTGGATTTCC
CAGGGAGTCGGTGCCGGCGAGCGAAACGTGCGAGGCGAGAAATCGACGAGTTTCTGTCTG
CGTTAATTGACAGGAGGCGGAGGGATTTGGCGGAGGGRAGGGCTGGCGGTGAGCAGGATT
TGTTTTCCAGRTTGCTGACGTTAAATGACGAAGATGGTAACAGGCTTGGGGATAAGGAGG
TGATGGATAATATTGTCTTGTTTATGTTTGCTGCGCATGATACCACTTCTTTGGTCATTA
CTATGATGTGTAAATTCCTGTGCCASCATAAAGATTGCTATCATGAAATCTKSCAAGAAC
ACAAGMARATTCTAGAAAGCAAGGAACCAGGAGAGAAGCTCAATTGGGAAGACGCAAAAA
AAATGACATACACATGGAGAGYAGCCCAAGAAACAATGCGGTTGTTGCCTCCAATMTTTG
GTTCATTTAAGAAGGCAATTGTTGATTTTGAATACGAAGGTTACACMATACCAAAAGGAT
GGAAGGTTTTATGGACAGCATCAACTCACTATGATCCAAWTTTTTTGAGGAGCCACAAA
AATTCAAGCCRTCACGGTTTCAAGATACGATTCCACCATACGTGTTTGTACCATTTGGAG
GAGGAGTCCGTTTATGTGCAGGTTTAGAGTTTGCAAAATTACAAATTATGGTGTTCATGC
ACCACCTTGTAACACAATATGATTGGTCACTGGTGGATCCCRCTGAAGGAATACAGATGG
ATCCTCTTCCKTCTCCAAYCAAGGGAATGCCTGTAAGGCTTGTGAAAAAGAGYTATTAAA
CAAATTCTCAATTTAGTAGTAAGGAATGTAGAGATTACTTGCCATGTATAAATTGATGGA
GCATGTAATGTAAGTGCCGAATTGGTTGATTGACAAATTGATCCTGCAAATATATGTGAG
TTTTTAGGTTGAACGATGTAAAATTAGTAGAGTTAAAGGTGGCTCACATACTTATTTTGG
CTTTATAATGTTAGATGGTGAGAAGA

CYP39

TGTCCCCGCGTCGTTCAGAAATATTATAGGCAAATATGGTTTGGCGGGGGTACATGGGGA
ACTTCAAAGGAAAATCCATGGAACTGCTGTCAATTTGTTTAAGTATGAGASGCTGAGCTC
TGATTTCATGGAAGACATACAAGACATCTTCCTGGCTGGAATGAGGAAATGGGAGGAGCA
GGGAGASATTCCTCTTCAACACAAGTGCCATCAGATGGTTCTGAACTTGATGGCGAAAAA
ATTGCTGGACTTACCTCCAGCAGAAGAAATGAAAGATRTTTATAAAGCTTTCGACGATTT
CGTAGGAGCTRTYYTCTYYTTGCCCCTCAATATCCCTGGAACCWCTTATGCBAGAGGAAT

TCAGGCCAGGGGAATTCTGTTAAAAAGAATTCACRAGTGTATAAAGGAGARGAGACAGCA
WCCAGASGTGGTCCGCAATGAYTTGTTSACCAAACTTTTGAGGGAGGGCACATTTTCGGA
TGAAATAATTGCAGATACAATACTCTTTTTGCTGTTTGCCGGTTTCGAAACTTCATCAAC
GGTCATGGCGTTTGCTGTGAAGTATCTCACTGAGAATCCGCGAGCTCTGGAGGAGTTGAG
GGCTGAGCATGATGCTCTCTTGAAGGCCAAAGGGAAAGACAATGAAAAGTTAACGTGGAA
CGACTACCAGTCAATGAAATTCATTCATTGTGTAATAAATGAAACATTTCGTTTGGCTAG
CATAGTCCCGTTTGTTAGCAGGGAAGCCAAACAAGATATTAAAGTGAAAGATTTTGTTAT
TCCSAAAGGATGGACAGTGTATGTTTTCTTSAGTGCCACTCATTTGGATGAAAAATACCA
TTACGAGGCTAACAAATTCCTCCCTTGGCGCTGGCAAAATGGGGGTCAAGAAAAAYTGGA
TGATCCATGTTATATGCCATTTGGAAGAGGTGGCAGGCTCTGTCCAGGACTTCATTTGGC
MAGATTTGAAATTGCTCTCTTTCTTCACAACTTTGTCACTAAATTCAGATGGGAGCAGCT
GGAAGCTGATCGTGCCACTTATTTTCTTCTCCCTTCCATGGAAAAAGGTTTTCCAATTCG
TCTATACTCTCGAGCGGACGAATGAGGAATGGTCGATTTAGTTATTAAAAAATAAAAACT
ATTATCTTATTGGAATGTGTTTTCTTTCTATCTAAAAGTTTGGTTCGATTTTTGGATTGA
AAAATATTTTCTATTTACCATATGTAATATTATCGTACATCATATTTATTTATATGAGCA
AAATTTTTATTTTGTAACAAAAAAAAAAAA

CYP40

CTTGACAATTGAAGGCAAACACAACAGCACAATCTATTTAGCGCAGATACATTTTTCCTA
TGTAGCATTCCAAGCTGTGCAGCAGTAATGGAGCAGCTAATATACAGTATTGCATATTCC
AATTGGTTTTTATGGGTTTTGGGACTGTTTACTTGTGCAATTTTACTGTTATTAAGATGG
GGTAAAGACGAACAACGAGATGGATTCGAAATTAAACCCAAACTTCCTCCTGGATCATCT
GGATTGCCATTTATTGGAGAGACTATCCGTTTTCTCCGAGAAGTTAAATCGCTCCCACGG
CGAAAATTCTTTGATGAACATGAACTCCAGTATGGGCCGATTTTCAGATGTAGTTTGTTT
GGAAGAACACGTGCTGTTGTGTCAGTGGATCCCGAGTTCAATAAGTACGTCTTGCAAAAT
GAGGGAAGGCTGTTCGAATCCAATGTCCCCGCGTCGTTCAGAAATATTATAGGCAAATAT
GGTTTGGCGGGGGTACATGGGGAACTTCAAAGGAAAATCCATGGAACTGCTGTCAATTTG
TTGAAGCATGAGACGCTCAGCTCTGACTTCATGGAAGACATACAAGACATCTTCCTGGCT
GGAATGAGGAAATGGGAGGAGCAGGGAGACATTCCTCTTCAACACAAGTGCCATCAGATG
GTTCTGAACTTGATGGCGAAAAAATTGCTGGACTTACCTCCAGCAGAAGAAATGAAAGAT
GTTTATAAAGCTTTCGACGATTTCGTAGGAGCT

CYP41

GTGTTTCTATGGATATCTATCCCATCCTTCCCACTGCAGTACTGATTCTCATCGTATTCT

ATTTTCTATGGAGATCTAAGCACAGCAGGGCCACAGTATTGTTTCGCCTTCCTCCTGGTC
CTCCTGCTTGGCCTATTGTGGGAAACCTGTTTCAGCTGGGGAAAAAACCCCACGAATCTC
TCCATGCGCTTTCCCTGCAATACGGCCCTCTCATCACCCTCCGTCTCGGCATGAGAACAA
CAGTGGTGGCCTCATCTCCTGCCATGGCAAAGGAGATTCTCAAAACCCACGACCACGCCT
TGGCCGGGCGGACAGTTATCGAAACAGCCAAGGCCCTTTCCCATCACAAGMCCTCCATGA
TTTGGGGYGAATATGGACCTCACTGGCGGCATCTCAGACGCATTTCAGCCGTTGAGCTTT
TCAGTCCCACAAGGCTACAAGCTCTGCAACCTCTGAGRAGAGACCAAGTCTTTCACACCA
TTCGACTCATTTTTGAGGAGAGGGGAAAAGCTGTAAATATTGGGAACATGGCGTTCTACA
CCTCTGTGAATCTGTTGGGCAACATGGTCTTCAGCASCAACCTSTTCGATCCACACAAAC
CAGCTTCGRTGGGGTTCAAAGACACTATCAGGAAAATAATGAAGCTTGGCGGAACGCCCA
ATTTGGCCGACTATTTYCCGTTTCTTGGGTTCTTGGATCCTCAGGGCCTGCGACGTCAGA
CGGCCAAGTATATGAAGGGAATCTATGATTTTTTCGATTTGCTCATACAGGAAAGGCTCG
CCGCGAGAAGCAAAAACCTGSAGAAGAAAGACTCCGAGAAGGATTTTCTCGACRTTCTGC
TGGATTTTAGAAGCGAGGATTTCACTCTTGTTGGAGTTAGGGCCTTAATCGCTGAACTGT
TTATTGCCGGTACAGAAACCACCACCACTACAATAGAATGGGTCATGGCAGAATTCATTC
GCAATCCGSAGAAGATGCAGAAAGTGCAACGAGAATYGGATGAAATGGTTRGTGGAAATC
GAAGAGTTGAAGAATCTGATTTAGATCGTCTGAGTTATCTACACGCAGCGGTGAAAGAAA
TATTCCGATTGCATCCTACWGCTCCTCTGCTGCTGCCCCACAAAGCTGAGAGCRSCTGYG
AGATSGAGGGATTYGTGATACCCAAAGACAGCCAGGTGATGGTGAAYGTATGGGCAATGG
GAAGGGACCCTGCAATCTGGAAGAATCCATTGGAATTTATTCCAGAGAGATTCTTGGAAG
GCGAGAATGGGAAGATGGATTATAAGGGACAAAACTTYGAGCTGATTCCATTCGGAGCAG
GGAGGAGAATATGCGTGGGGCTTCCTTTGGCGARTCGTATGGTGCATCTGGTTGTGGCTT
CGTTGCTCCATTCATTCCAGTGGACGCTTCCGAATGAGATGAGGTGCGAGGAGATGGACA
TGAGCGACGAGTTTGGAATCACATTAAAGAAGGTCATAGAGTTGAATGCAATCCCAATGC
CTCGTCTACCACATCACATATATTAGTTCATTGGTTATAATCTGCCGACTCTTCTTATCT
TGAGGAGTGGACACAGGGAGTGGCTAGTCTTGTCTTTGATATCTAATTCTCGCCTAGCAT
AGGAAATTGCATTTGATTTCAAATAAAACGAGTTCAATGGATGTTGTTTTACTATCAAAT
CTCGCCTCGCACAACGAAATGCTGGTTTTCCAAATCACTATTTGTTACTATGTTCAATAC
TTTGTTTGATTTTGGAGTCTTACATTAAAATATGTGGTCATATATATGCATAAAGGGTGG
ATTTTCCC

CYP42

AGAGCTATTGCATACAACTTTGTGTGCAATCAACTCTTTTGGGGTTTCTTCTTCCATGGC

CACTCCCATTCTGCTCGGATTCCTCTTCTTAATTCTTGTCATATACAAGATTTATTGCCA
CTTAAGTCCTAAGCAGATGCAGCAGATTATGCAGAGCCAAGGCATACAAGGTCCCAATCC
GAARTTTCTGYTAGGAAATCTGCTTGATATCAAGAAAATAACTGAGAAAGCTTCCAAGAA
GCAAAAATCTTCTGCAACTCACAACTTGGAGCATCGTTTCTTCCCATACTTTGTAGAATG
GAAGAAAGCATATGGGAAGAATTTCGTATACTGGGTTGGATGTGAGGCAGTGTTGTATGT
GCAAGATCCAGAGTTGATACAAGAGATAAGCTCATCAGGGTCYTTCAACTGGGGGAGGCC
AGCCTTTCTGAAAAAAGACAGATTTCCTCTTTTCGGCAATGGRCTTATCATGGCAGAAGA
CCAAGATTGGCTTCACCAACGTCGAATTATTGGGCAGGCTCTCACAGCTGAGAAARTTAA
GTTGGTGTTTGGCTCAGTGATGGAGGCTTCTCTGCCCATTCTAAATGGCTGGGCTGTGAG
AATAAAGAAGTGTGGCCAAGGTAGCATAGAGATTGAAGTTGATGGAGATCTGAATAAAAT
CACTGGACATGTTATCTCCAAACTGTTATTTGGKAGTGACTACAAAAAAGGCTTTGAGAT
RCTTGATAAGCTCAAAGTTCTTCAACACACCTTATTCAAAGCAACCAGAATTGTGGGCAT
CCCTGGAAGCAGGTTTATACCTTGGGCTGGGAATAAAAAGGTTTGSAGGTTGGGCAAGGA
AGCTGATTTTCTAATCATGGAAATCATCAATGCACGCAAAGGATTTGCAGATGATTTGTT
GGGGCTCCTGTTGCAAGAAGTCAATCATGACAAGGGATTAACAACTCAGGATGTGGTGGA
TGAATGCAAGACTTTGTTGATAGCAGGGCAGGAAACAACCAAACTGTCCTTAACTTGGAC
ACTCATGTTATTAGCACTGAATCCCCATTGGCAAGACAGAGTTAGGGCCGAGGTCATGGA
GATCACCAATGGAGCCTATCCTGACATCAATATGCTCTCTCAAATGAAAATTATGACGAT
GGTTCTGAACGAGTCCATGAGGCTGTATCCCCCAGTGTCTTACACAGTGAGACAGGCCAA
GCGAGATATACGGGTGGGAAATCTAGAGATTCCCAAAGGGATGTCGGTGCTGATAAATAT
ATTGGGAATGCACGAAGATCCTCATTTATGGGGAGGTCCAGAGGATGTTTATGAATTCAA
TCCAGAGCGCTTCTCTAATGTGGAATTTAGCAGGAAAAACAGTGGATACATTCCATTTGG
ATTCGGAGGAAGAGTTTGTCCAGGGGAGAAGCTGGCTCGCATGGAGCAGAAGGCAATACT
GAGCATGATAGTGTCCAGATTCTCCTTCACCATATCTCCAAATTACWGTCATTCTCCAAC
ATCTCTGTTGAGCATCATGCCCTCTAATGGACTGCAGCTTCTTTTCACCCTTCTCTAGCT
TTTACAACACATTGTCATGCCCTCTAAAACTTAGTTTGCAGAGGG

CYP43

TCACCTTCGCTGTTGGGAGGAAATACATTCATTCTCACACATCTATCAACTGCTAGATGG
AGTTGTTTTGGCCAGATAATCTTGTAGTATTACCGATGTATTCTGCGATTTTGTTGTGTG
TATGGGCATTCCTATTGCACACTGCAAATATTTGCTGGTTTCGACCCAGAAAAATCAACA
GTATGTTACACAAACAAGGCATCGCAGGCCCTCCGCCACGGTTCCTGTTGGGAAATCTTC
TAGATATGAAGAAGATAAAGGAAGAATTTATGGAAAAATCACCGCATAATACTGGAATTA

CAAGTGTTAGACATGATATTGGGCCTTTGCTATTCCCCCACTTTGTCCAATGGAGCAGAG
TTTATGGGAAAACGTTTGTGTTTTGGCTGGGAACAGAGCCCTTTCTTTATGTGCGAGATC
CTGAATTGTTCAAAAAGACGAGCTTGTCGAGATCTCTTGATTGGGGAAAGCCTGATGTTC
TGAAGAATGATAAAGTTTCGCTGTTTGGGAATGGATTGATCATGGCAGATGGAAAAGATT
GGGCTCATCACAGACGAATTGTTTCCCCAGCTTTTCATATGGATAAAATCAAGGGTATGA
TAGGTGACATGGTGGAATCAACGATACCCATGCTTGAGCGTTGGAGAAACACCGTTATAG
GAGGAGGAGGATCTGCAGACATGGATATAGATGAGGATATAACGAAAGTCACAGCAGATA
TCATAGCAAAGACATCATTTGGGACTAGCTATGAGAAGGGTAAACAAGTATTTGCCAGCC
TAAGAACTCTTCAATATTTTCTCTACAAACCTAATCGCTTTCTTGGGGTTCCAGGAAGCA
GATTCATAAATGGTAAAGTGCATTCTACAGCCAAAAAGCTGGGAGAGGAGACTGATAAAC
TGCTCATGTCAATCATAACWGCGCGAAGAAAACKTCAAAGCTCRGGCGACTTTGGAAACG
ATTTGTTGGGACTTATTTTGGCTGAAAACGAGAGTGATGAAGGAAARCAGAAGCTCACTA
CACGGGATTTGATGGACGAATGCAAAACCTTTTTCGTGGCAGGGCATGAATCAACTGCAA
TGTCTGCAATGTGGACCATGATGCTTCTTTCCTTGAATCCAAACTGGCAAGAACGCGTTC
GAACAGAAGTGATGGAGATTACACAAGGCAGAATTCCTGATGCAGAGATGCTCTTCAAAA
TGAAAAATATGAGTATGGTGTTGAGCGAGTCCATGCGACTGTATCCAGCAGCCTCGTACA
CGTTAAGGCAGGCCAAACAAGACATGCAAATGGGGGACATTCTTGTTCCCAAAGGAACCT
CAGTCTGGTTTGACATCCTTGGCTTACACCACGATCCAGACTTATGGGGGGCAGATGTGA
ATGAGTTTAAACCTGAGCGATTCGAACTCGGAATATCCAACGCATGCAACCATTCAATGA
GTTTTCTTCCCTTTGGATTTGGACAGAGAGTGTGTGTAGCCCAGAATATGGCAAATATTG
AAGTAAGGTTGATCATGAGCATGGTGTTGTCAAAATTTCGGTTTTCACTGTCTCCCACAT
ACTGCCATTCTCCCATCAWTGTGCCGAATCTCAAACCTTCTCAAGGAATGCCTCTTGTAT
TTGAGTGCTTGTGATTACTCTCCATAATTTTGCAGCTCTAAGGTGTCTTACACCTGAACC
AATAGASTATTTATCAAAATGCTCTAATAATCWGAGGACAAGGATTCTTGAAGTAGACAA
GCAGTTTTAAGTTTTTTTGGTACAAAAGAAGGATAAGAGTATAGAGGTTTTAATTTGAG
GCTTGTAGAGCTTGCCATGTTGTTGATGCAAGATCGTTGTAATTGAATATAATGTATGTT
TCTATGACTGTTATACGTATGAAAGTTGTTT
)

Appendix B

NMR Spectra

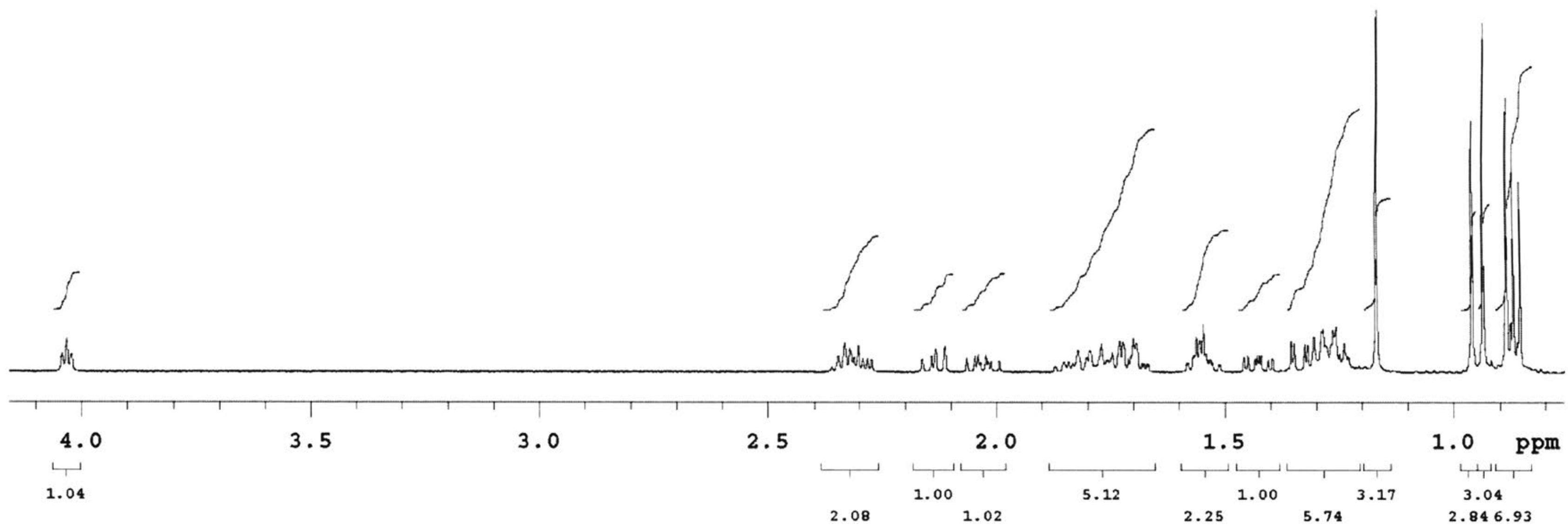

Figure B-1: 1H NMR of Compound 11

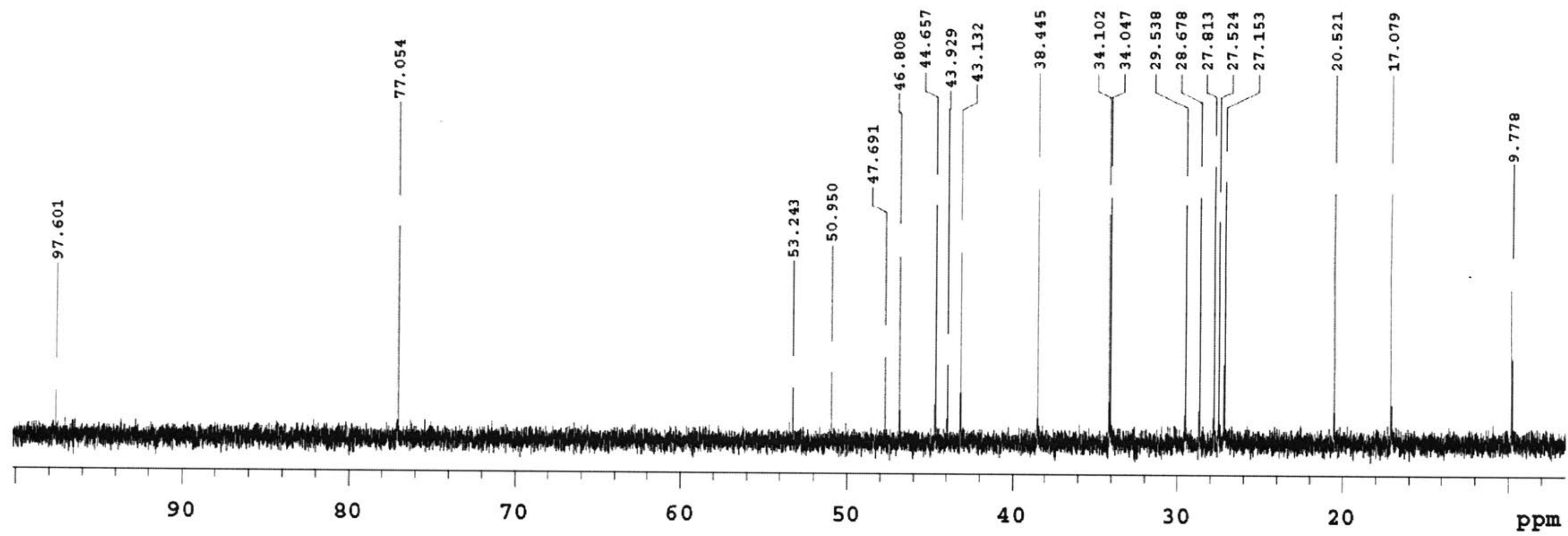

Figure B-2: 13C NMR of Compound 11

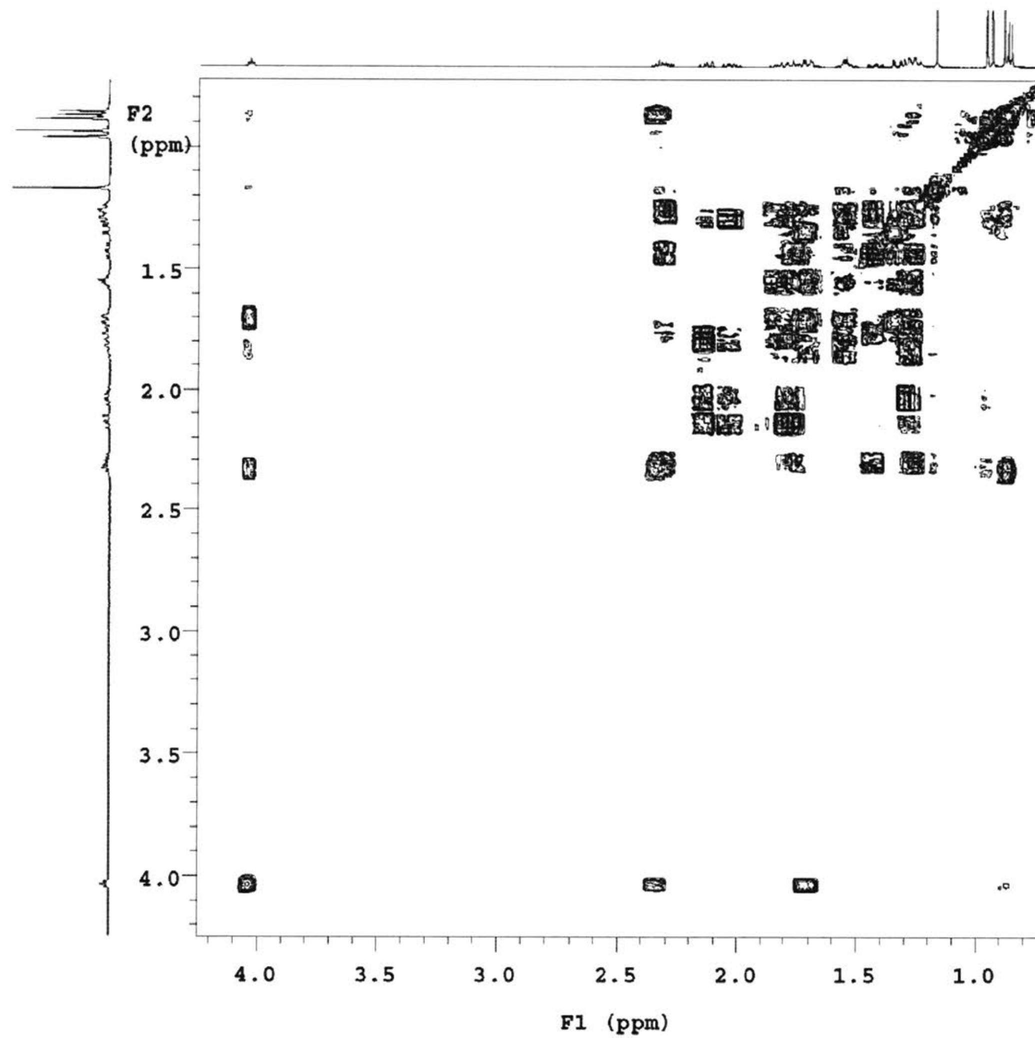

Figure B-3: gCOSY NMR of Compound 11

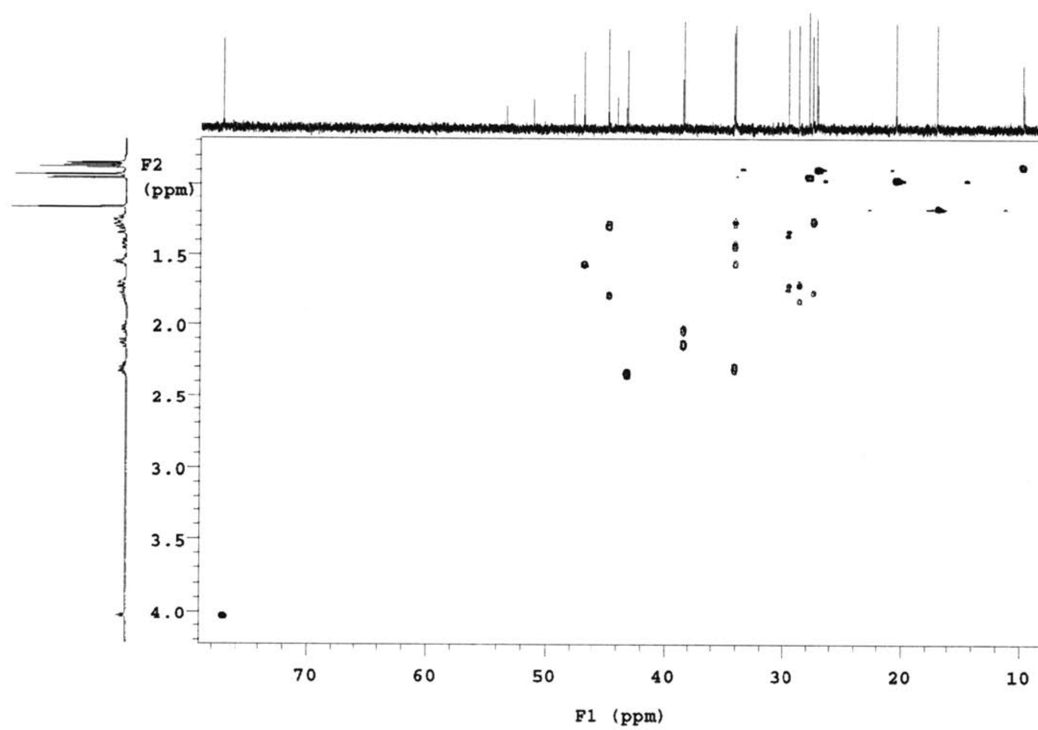

Figure B-4: HMBC NMR of Compound 11

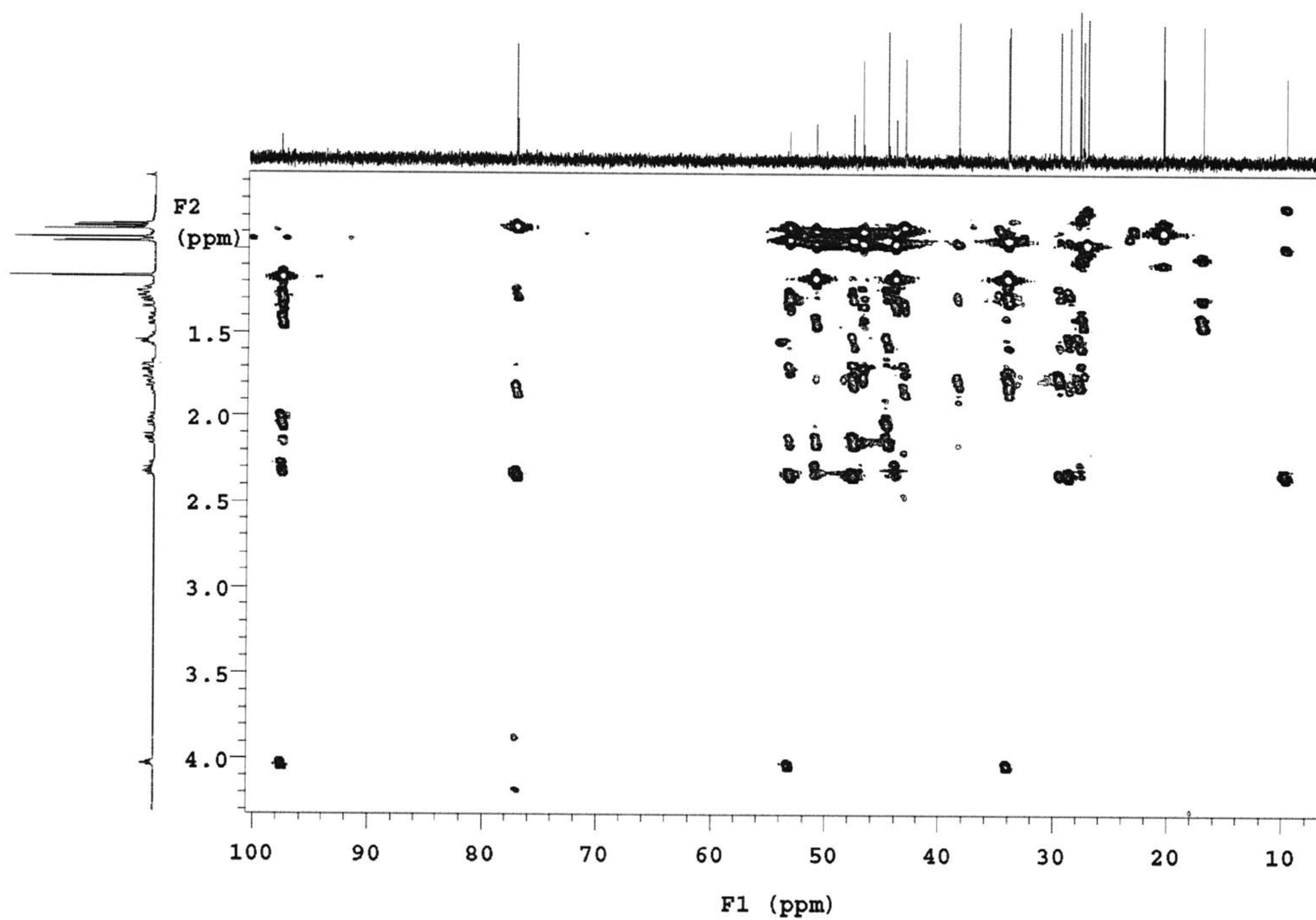

Figure B-5: HSQC NMR of Compound 11

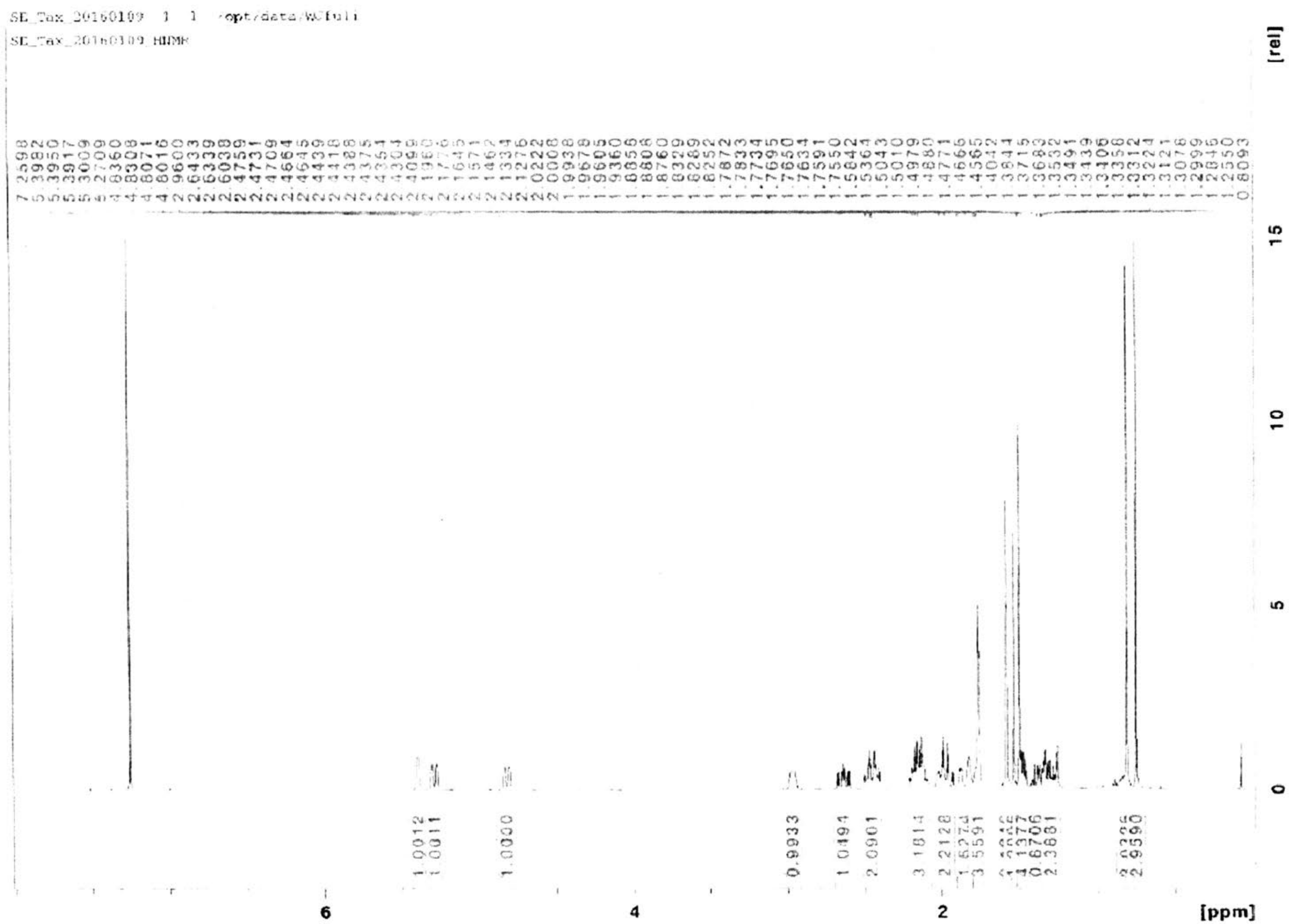

Figure B-6: 1H NMR of Verticilla-3,7,12-triene

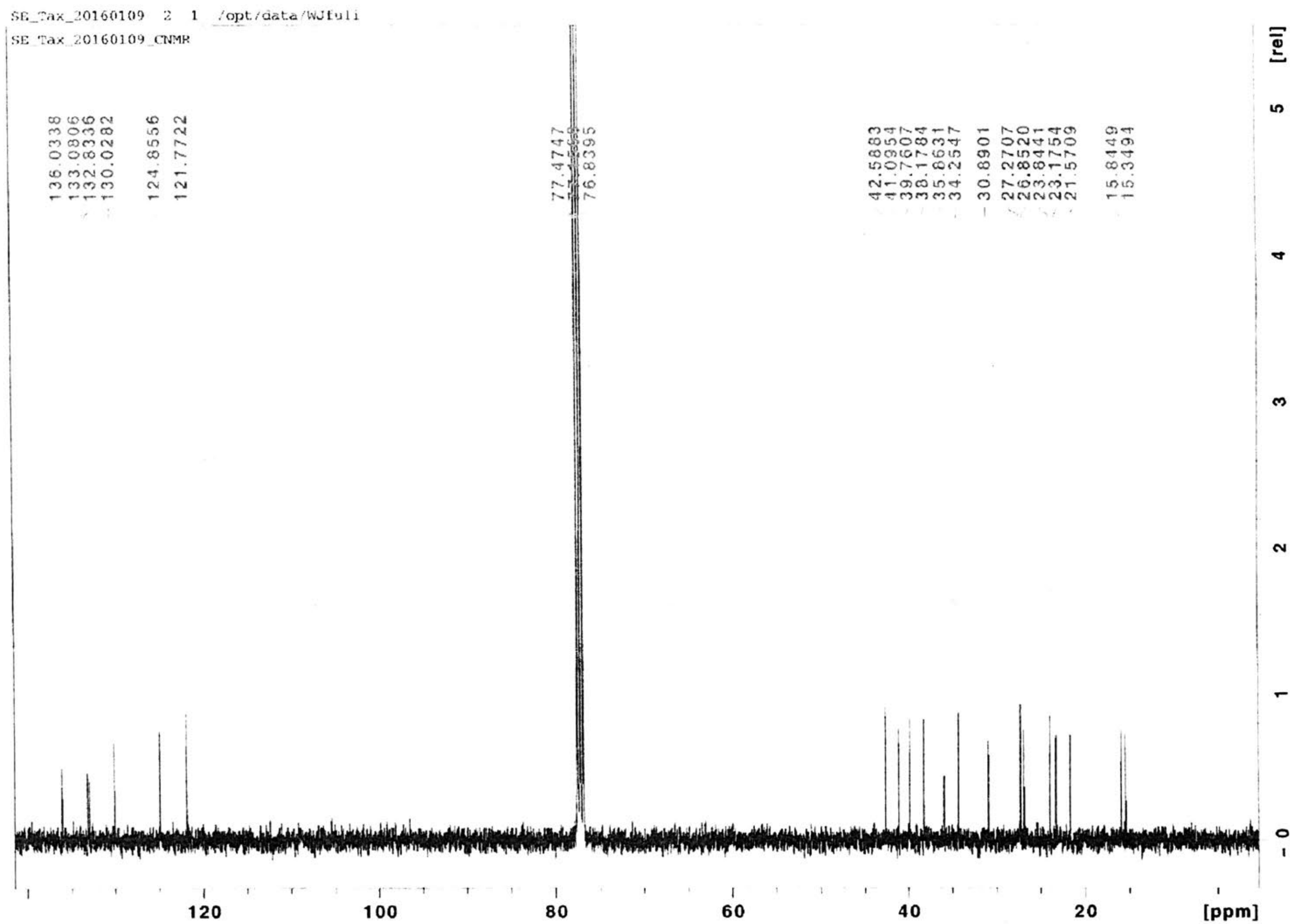

Figure B-7: 13C NMR of Verticilla-3,7,12-triene

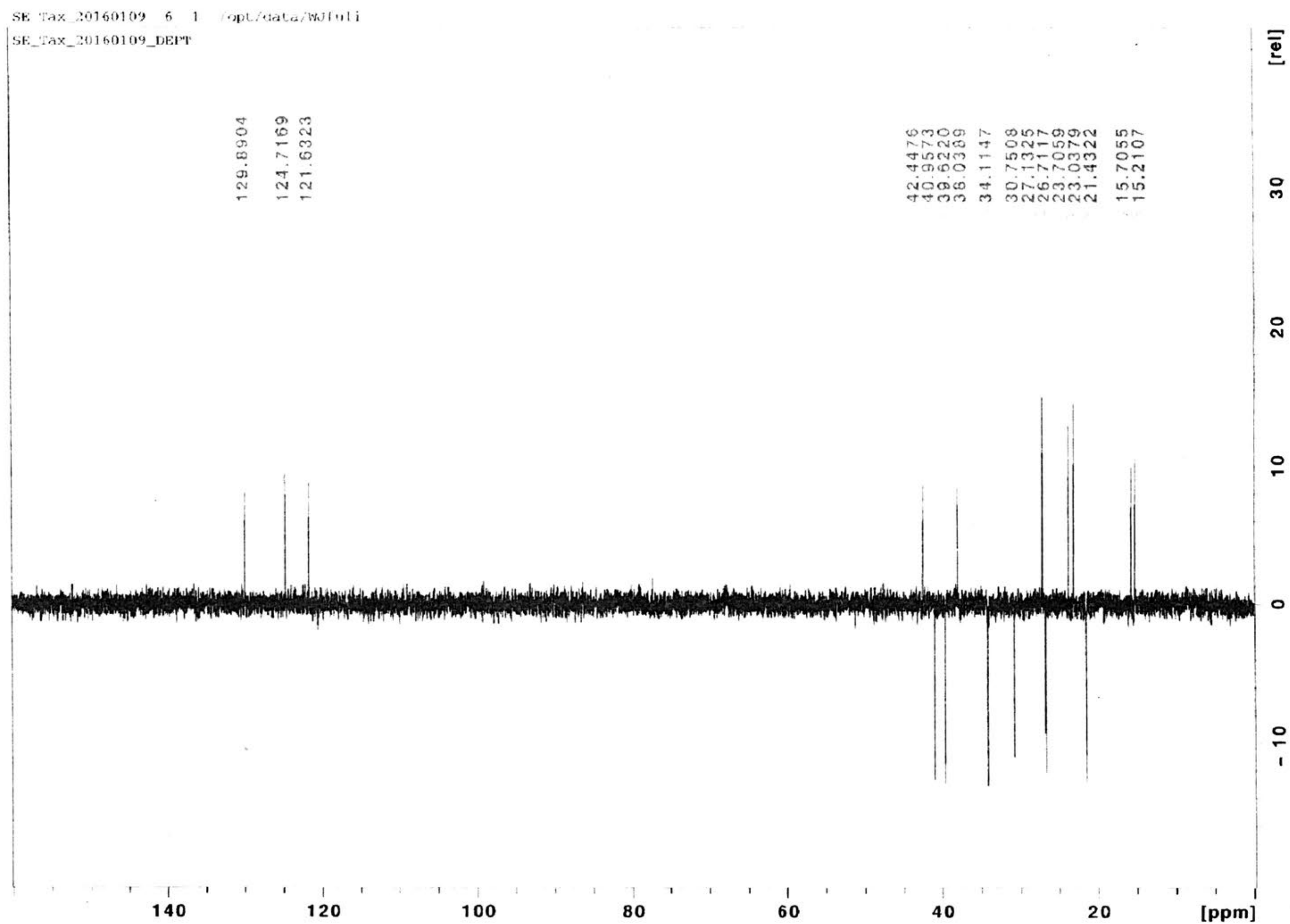

Figure B-8: DEPT NMR of Verticilla-3,7,12-triene

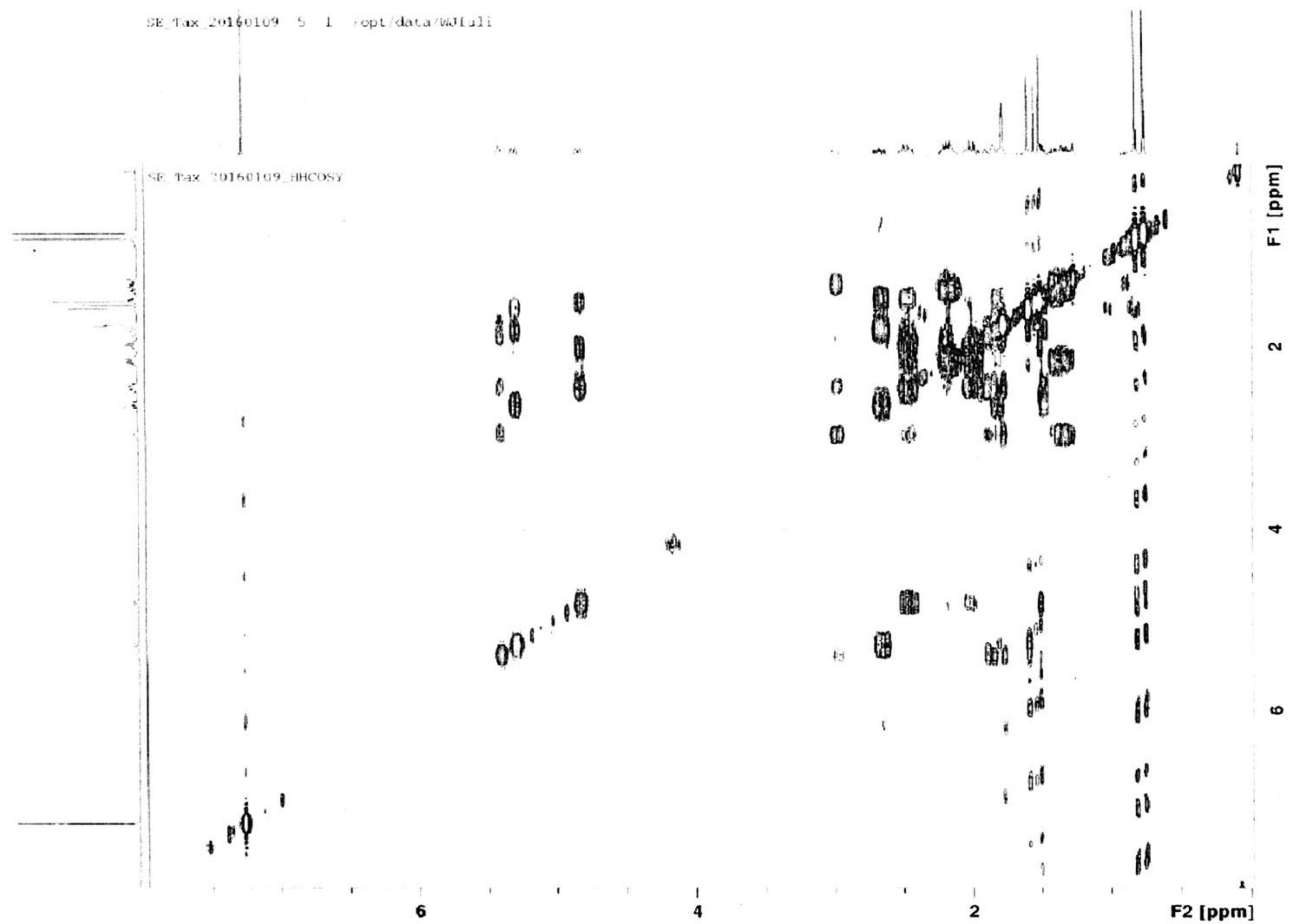

Figure B-9: HHCOSY NMR of Verticilla-3,7,12-triene

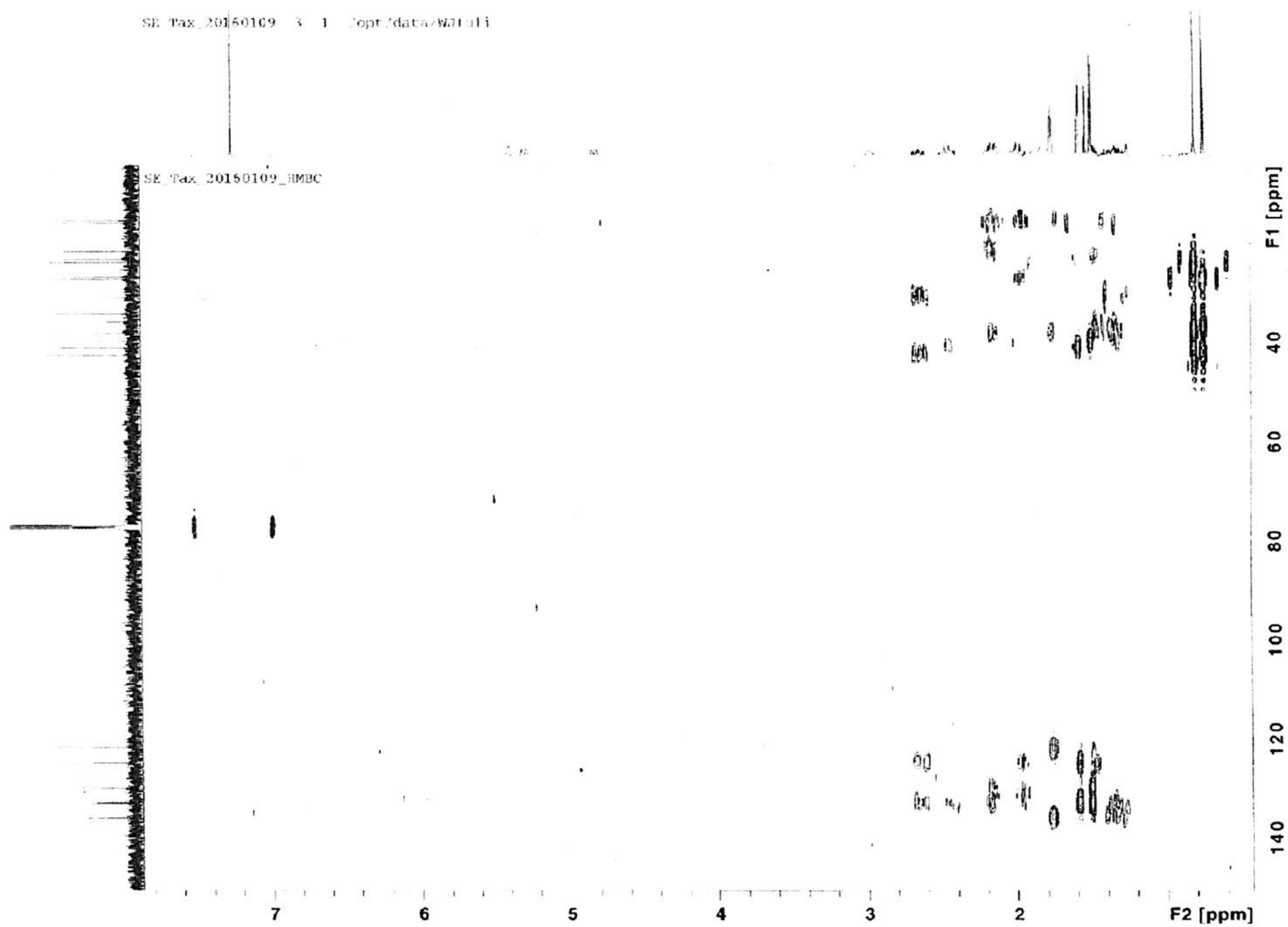

Figure B-10: HMBC NMR of Verticilla-3,7,12-triene

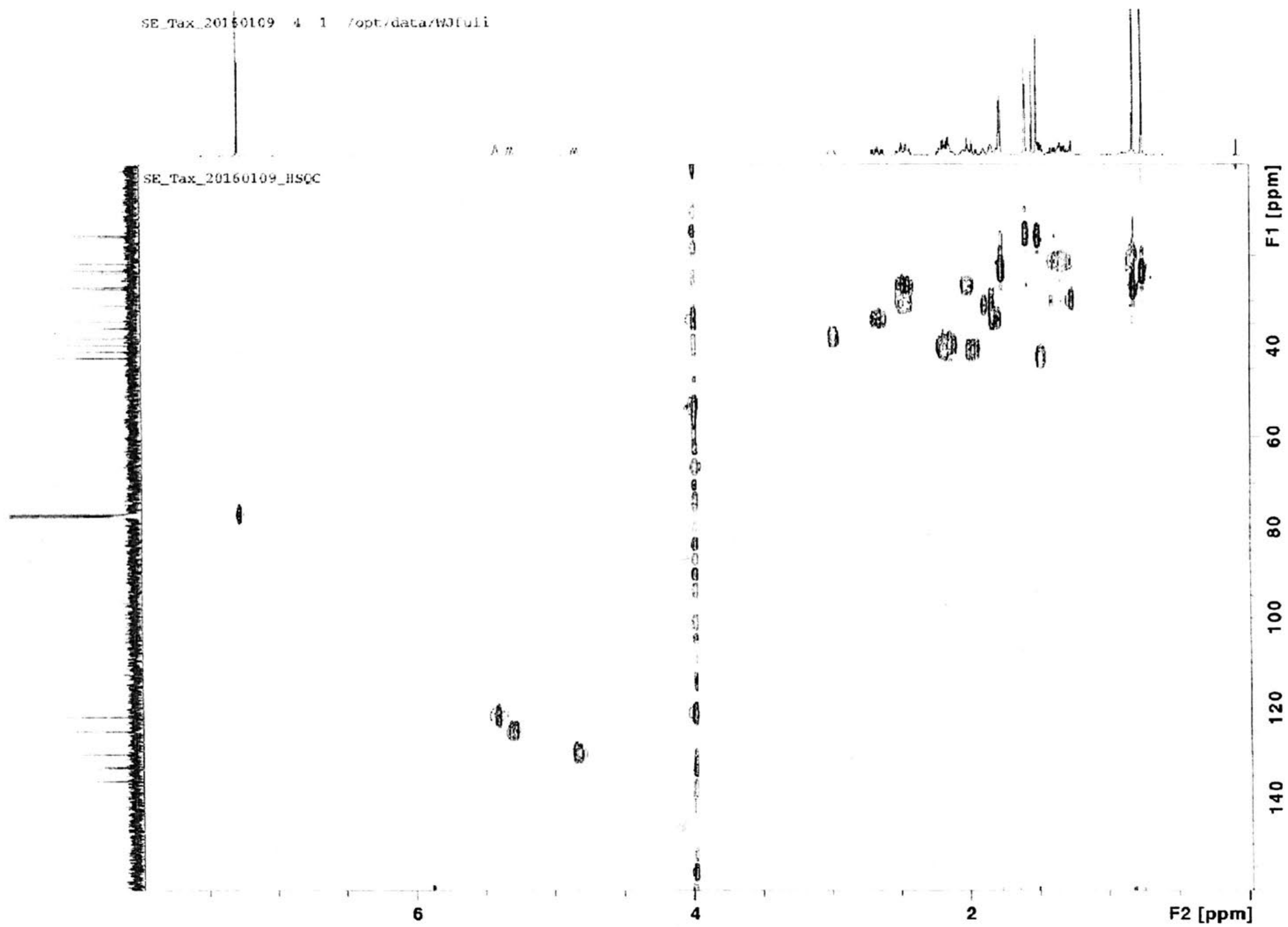

Figure B-11: HSQC NMR of Verticilla-3,7,12-triene

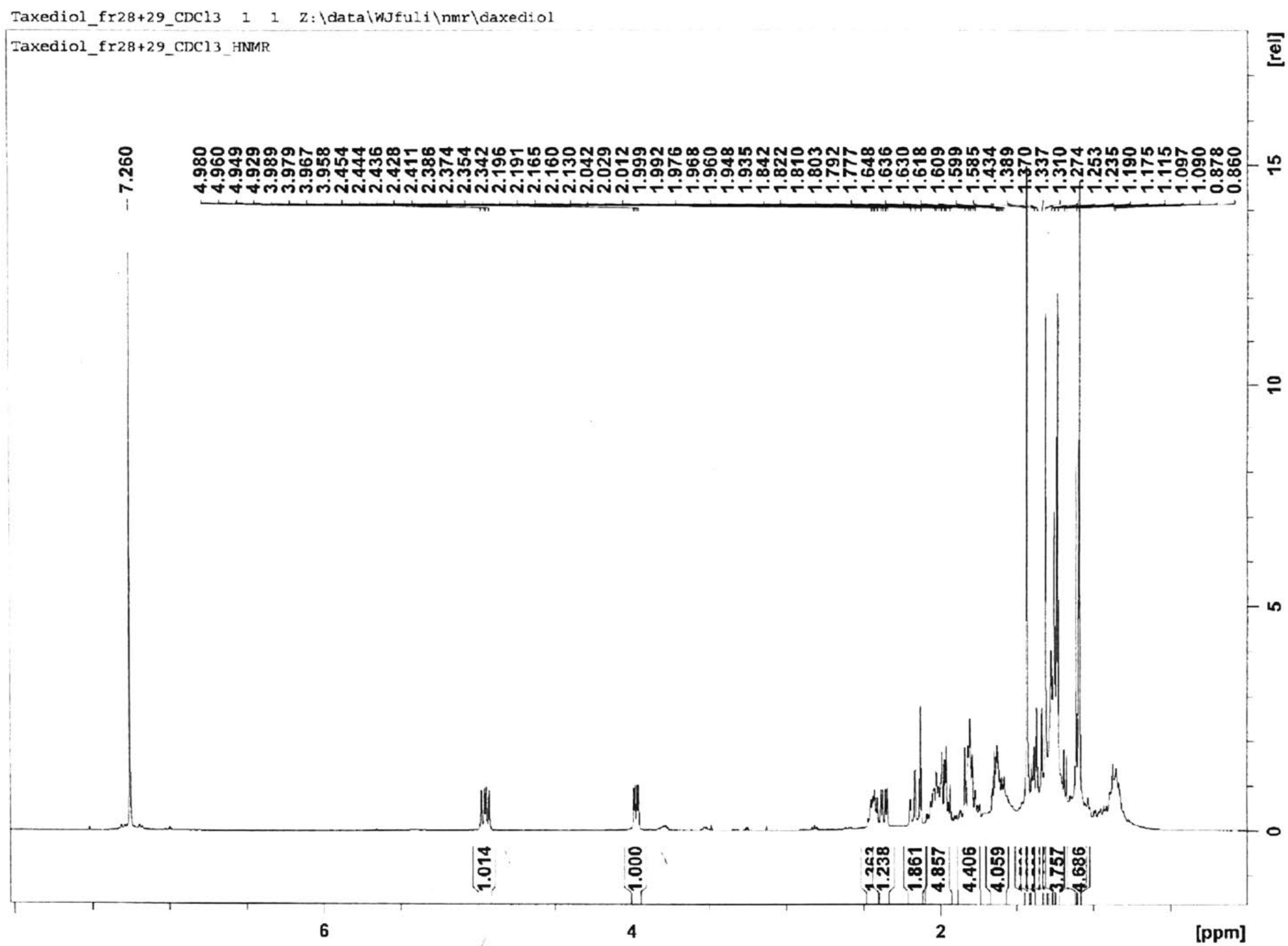

Figure B-12: 1H NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol

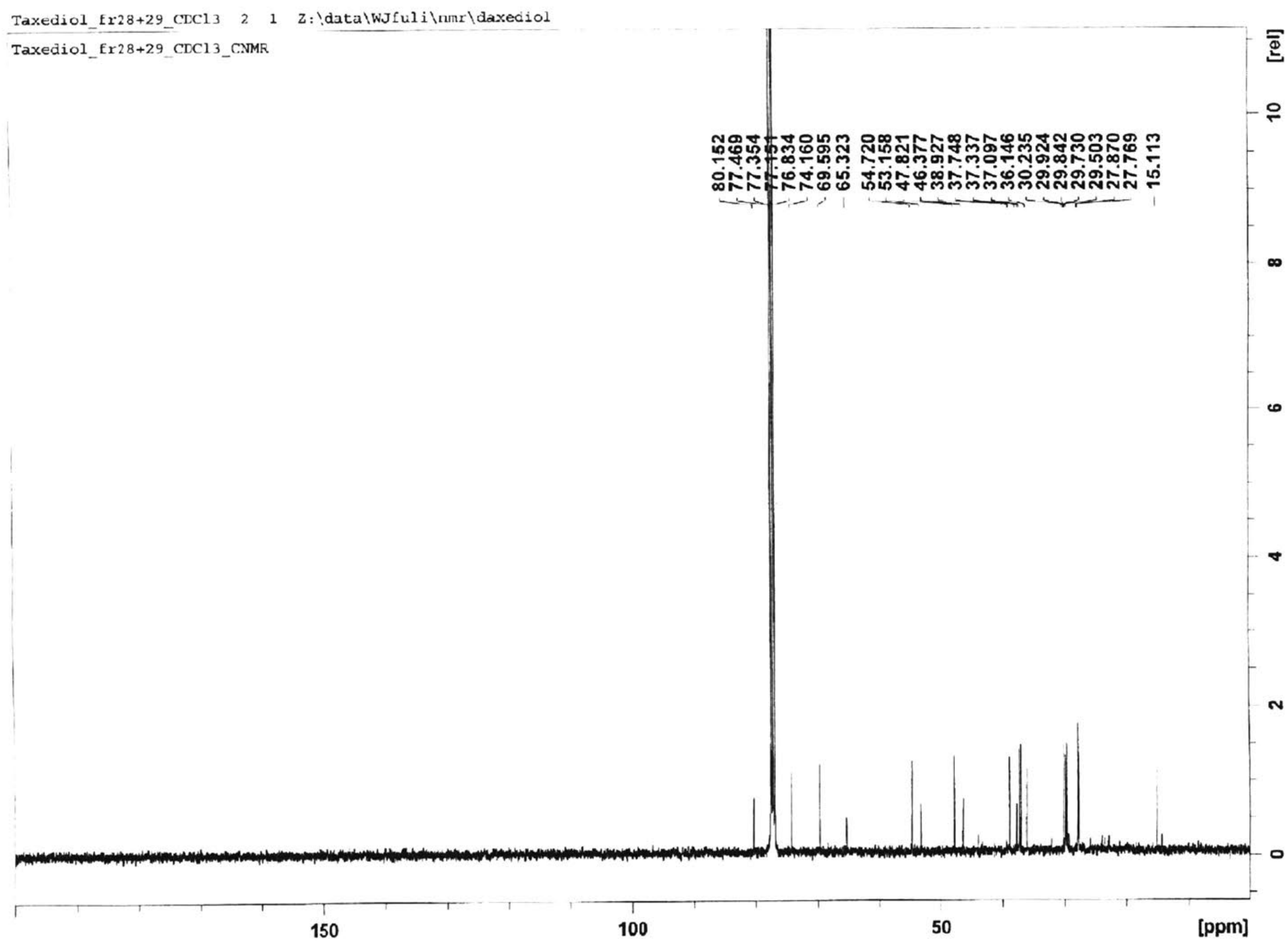

Figure B-13: 13C NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol

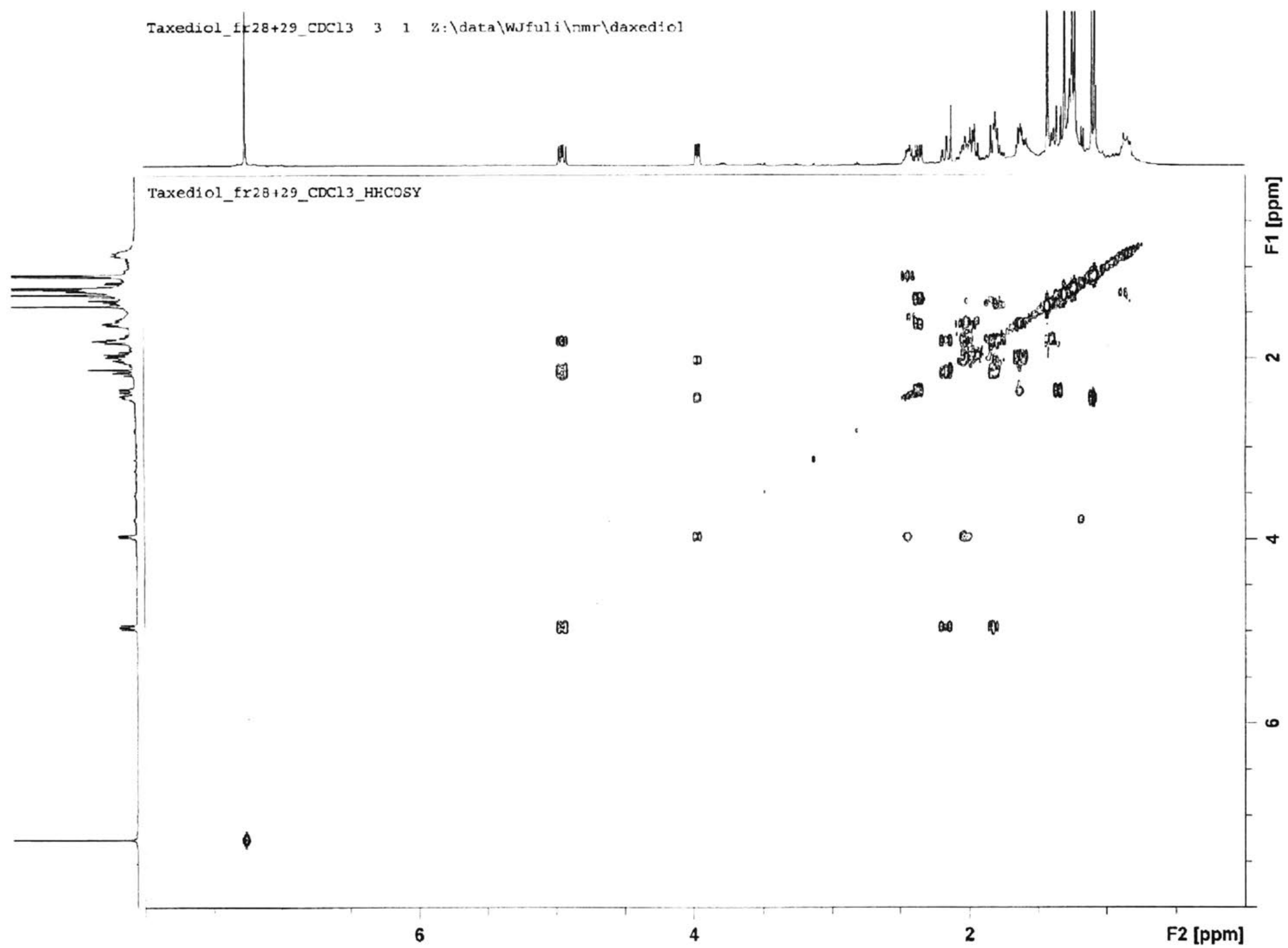

Figure B-14: HHCOSY NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol

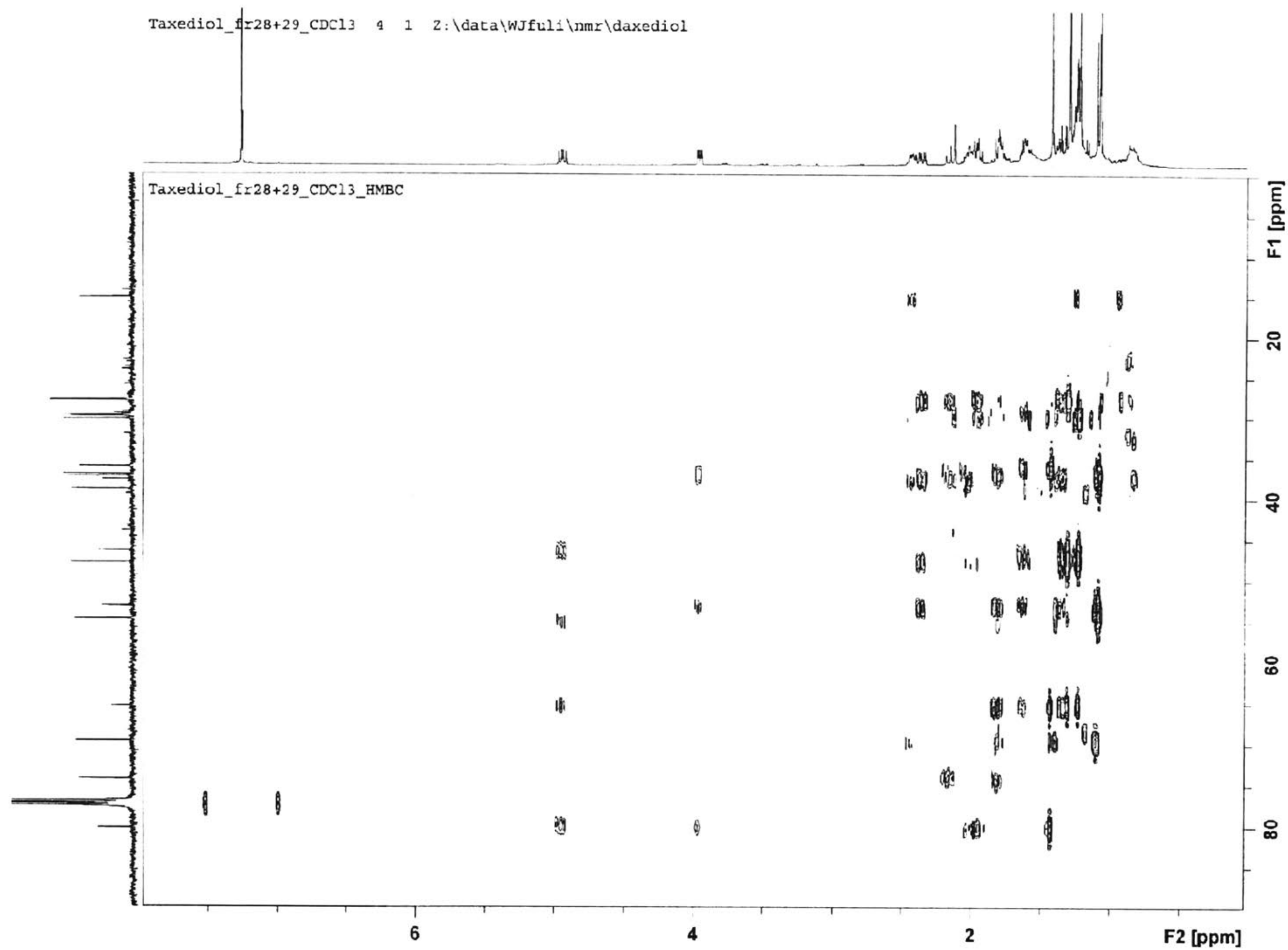

Figure B-15: HMBC NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol

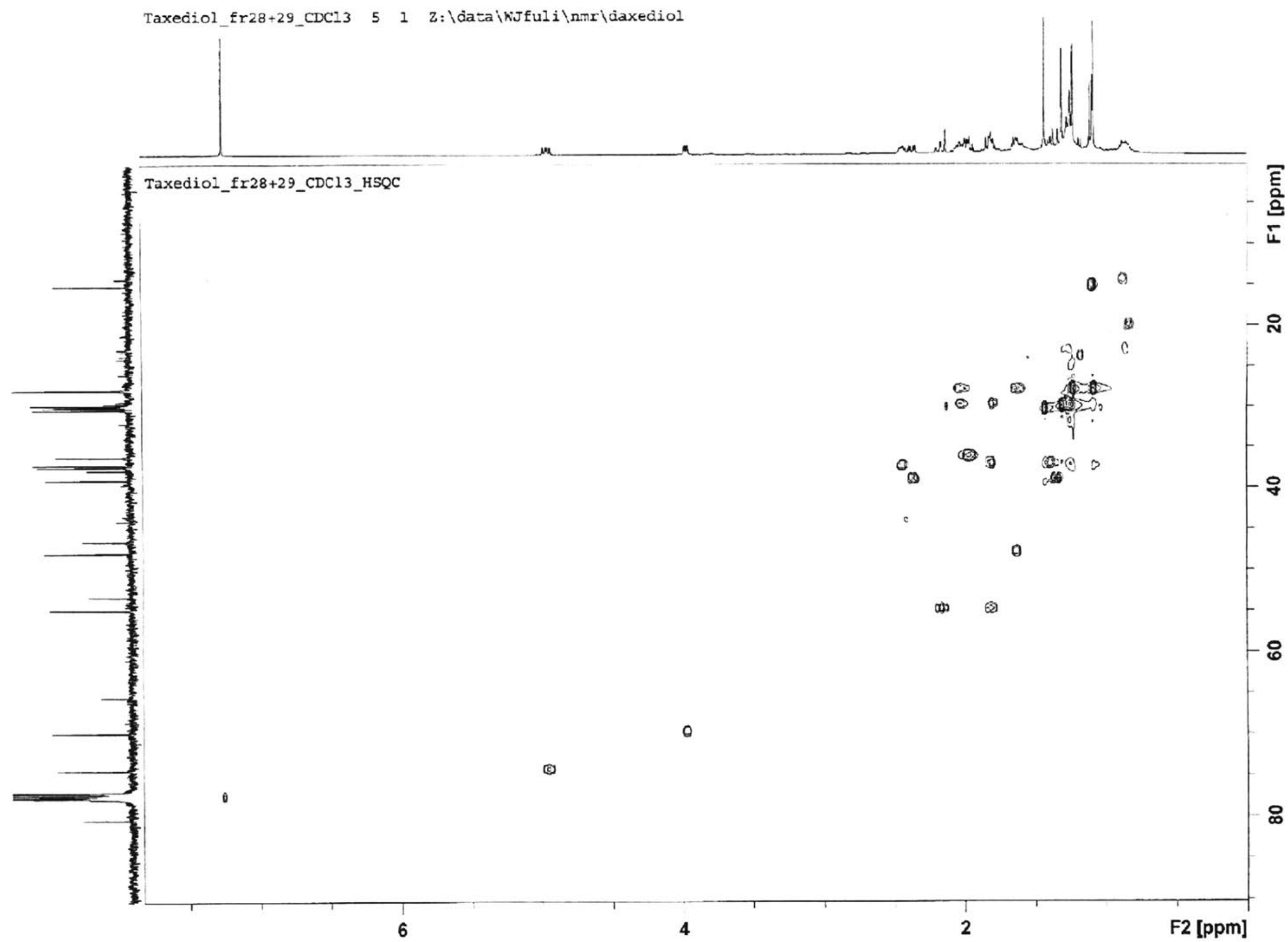

Figure B-16: HSQC NMR of 5(12)-oxa-3(11)-cyclo-taxan-10-ol