Biosynthetic Engineering for the Assembly of Better Drugs

Vikramaditya Ganapati Yadav
B. A. Sc. (Hons. Co-op.) Chemical Engineering, University of Waterloo (2007)

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

May 2013

Massachusetts Institute of Technology © 2013
All rights reserved

Signature of the author

Department of Chemical Engineering
28 May 2013

Certified by

Gregory N. Stephanopoulos
W. H. Dow Professor of Chemical Engineering & Biotechnology
Thesis Advisor

Accepted by

Patrick S. Doyle
Professor of Chemical Engineering
Chairman, Committee for Graduate Students
The declining prospects of innovator pharmaceutical companies have been attributed to their inability to discover leads that bind to novel protein domains. All of the 21,000-odd drug products that have ever been approved by the US FDA bind to just 130 of the over 10,000 functional protein domains that exist in the human body. The high degree of drug redundancy, in turn, stems from the industry’s persistence with combinatorial chemistry to synthesize drug candidates. Not only is the chemical space being screened for biological activity usually one that has been previously interrogated, but the products of combinatorial chemistry too are simply not "drug-like." So dire are the odds that launching a single synthetic compound necessitates screening as many as 100,000 molecules from a combinatorial library.

In comparison, as many as 20 molecules of the 7,000-odd polyketide natural products that have been screened thus far have made it to the market. Nevertheless, drug companies remain underwhelmed as synthetic inaccessibility has heretofore stymied drug prospection in natural product space. However, recent applications of genetic approaches to decipher the biosynthetic code that guides the assembly of several natural product families have revealed that living systems methodically and repetitively combine carbohydrate-derived building blocks in metabolic unit reactions to generate unprecedented molecular diversity. By clustering combinations, permutations
and mutations of these biosynthetic genes into operons to be expressed as heterologous metabolic pathways in tractable microbial hosts, one can readily generate pharmacophores that are more likely to bind to novel targets. Ordered combinations of known genes generate pharmacophores that are natural products in toto, whereas permutations and mutations produce molecules that possess a scaffold that is an unnatural isomer of a natural product. The atom economy of microbial synthesis of pharmaceutical intermediates and active ingredients also far exceeds those of conventional schemes.

We recently demonstrated the feasibility of this new paradigm for drug discovery by integrating two of the four metabolic unit reactions involved in taxane assembly into the metabolism of E. coli to produce an unsubstituted, 20-carbon drug scaffold at 1 g/L titers. The taxane family of molecules includes, among others, the blockbuster anti-cancer drug, taxol. We then proceeded to express the first of many cytochrome P450 monooxygenases that oxidize the taxane scaffold in the third metabolic unit reaction. However, plant cytochrome P450s such as those involved in taxane biosynthesis are notoriously promiscuous and non-selective. A likely candidate that meets the synthetic requirements must first be selected, following which it must be customized to enhance its substrate specificity, product selectivity and activity. The use of structure-driven approaches is preferred herein as the structure-activity insights gained from redesigning one active site could potentially be extended to re-engineering downstream homologues. However, conventional methodologies such as X-ray crystallography are fairly lengthy and laborious. Ergo, we combined phylogeny-guided mutagenesis with a model-driven assessment of the active site of taxadiene-5α-hydroxylase, the cytochrome P450 that catalyzes proto-oxidation of taxane scaffold, to investigate its
catalytic mechanism. The reactive topography of the active site was then charted to relate the impact of structural features to activity and selectivity, opening up the future possibilities to specifically tune the reactivity of cytochrome P450s to catalyze the regio- and stereospecific oxidation of the taxane scaffold for prospection of targeted regions of natural product chemical space.

Thesis advisor: Gregory N. Stephanopoulos

W. H. Dow Professor of Chemical Engineering & Biotechnology
To my parents, Gautam & Manali

कर्मण्येव वाधिकारस्ते मा फले भु कदाचन ।
मां कर्मफलहेतुर्भू: मा ते संज्ञेऽस्तवकर्मणि ॥
ACKNOWLEDGEMENTS

Isaac Newton remarked “If I have seen further, it is by standing on the shoulders of giants”. I have been privileged to learn the art and science of research from the very best. Thank you, Professor Stephanopoulos. You have been a phenomenal guru, a fantastic mentor and a great well-wisher. To Professor Prather, my academic mother – your humour, energy, advice and support have been instrumental in the completion of this document.

I still remember the day when I received my high school graduation certificate. It’s funny how people develop selective amnesia when all of their friends have performed better than them on a test, the state examinations no less! I guess all those years of trying my level best to skip homework had finally taken their toll. All I remember is that sinking feeling when a school teacher who was a friend of the family quipped, “What happened? Your parents might not be too pleased” when I told him my final grades. Nonchalant until that precise moment, I was suddenly overcome with the fear of having deeply disappointed my parents. After all, in India, a son’s state examination grades are as much a status symbol as the car one drives or the salary one draws. Mummy and Daddy – thank you for the advice you gave me at that moment when I was looking for the biggest hole in the ground to hide myself. You have been my best friends, my teachers, my role-models and my guides for all my life. Were it not for your love and support, I would not have made it this far.
Daddy – thank you for being the best friend and confidante a lad could ask for. You have always been my beacon, guiding me through all waters. Your poise has been a source of strength. Your wit and humour have always dispelled the dark clouds away. You are my Charioteer. But I must confess – I have kept aside a bigger slice of the cake for Mummy. You and I both know why. Mummy – you are my Architect. You gave me the strongest of foundations in mathematics. You helped me with my science projects in school. You made learning history and geography stimulating. You kept me company at 3 am on practically every examination night. Even in university, you were my alarm clock, making sure I woke up on time for classes, examinations and theses defences. This degree belongs more to you than to me.

Gautam – thank you for your camaraderie and love. Your counsel has been a source of great strength. Your mischief has been a source of great enjoyment and entertainment. You have always brought calm over me when I was beset by tempests. I thank Lord Ganapati for blessing me with the privilege of having you by my side. You are my General. Manali – thank you for your limitless love and patience. Your passion energizes me. Your support is my castle. Your serenity is infectious and has helped me sail through the turbulence. Like Mummy, you too sat by my side as I worked through the night to complete my thesis. I am truly blessed. But you do owe me some cake. You have robbed me of many ‘bachelorship is bliss’ jokes.

Lastly, to all my too-many-to-list-by-name friends and well-wishers at MIT – thank you for the hours of fun and learning. You have made this journey a memorable one.
# TABLE OF CONTENTS

**Preface** 19

1 **The Target-Rich, Lead-Poor Imbalance** 23
   1.1 Drug discovery in the pre-Hatch-Waxman era 23
   1.2 Deficiencies in the old paradigm 24
   1.3 The industrialization of drug discovery 24
   1.4 New paradigm, same old problems 27
   1.5 The inadequacies of combinatorial chemistry 28
   1.6 Effect of the needle in a haystack paradigm on the industry's business model 29

2 **Natural Products & the Pharmaceutical Industry** 31
   2.1 The historical use of natural products as therapies 31
   2.2 What makes natural products good drugs? 32
   2.3 Why is industry interest in natural products declining? 34
   2.4 Alternatives for probing natural product space 37

3 **Biosynthetic Engineering for Terpenoid Synthesis** 39
   3.1 The terpenoid superfamily 39
   3.2 Modularity of terpenoid biosynthesis 40
   3.3 Metabolic & enzyme engineering for targeted & diversity-oriented biosynthesis 44
   3.4 Summary of recent progress 46
   3.5 Possible challenges moving forward 47

4 **Mapping the Taxane Biosynthetic Pathway Using Cheminformatics** 51
   4.1 Current understanding of the taxane pathway 52
   4.2 Description of the cartography methodology 54
   4.3 Constraints for ensuring accuracy of the cartographic methodology 58
   4.4 Computational assessment of the reactivity landscape of the early taxane pathway 59
   4.4.1 Homology modeling 59
   4.4.2 Substrate docking 61
   4.5 Results 62
   4.5.1 Pathway map 62
   4.5.2 The reactive landscape 63
   4.5.2.1 Comments about computational assessment of the P450 reaction mechanism 63
   4.5.2.2 Using the diffusion of oxygen into the active site as a constraint 67
   4.5.2.3 Method validation 68
   4.5.2.4 Results of the reactivity landscape investigations 70
4. 6. Implications of the computational investigations on metabolic engineering

5 Re-engineering Taxadien-5α-hydroxylase
5. 1. Background about P450 monoxygenases
5. 2. Materials & methods
   5. 2. 1. Strains & plasmids
   5. 2. 2. Culturing conditions
   5. 2. 3. GC-MS analysis of hydroxylated taxanes
   5. 2. 4. Protein engineering for optimal expression of taxadiene 5α-hydroxylase in E. coli
   5. 2.5. Site-directed mutagenesis
5. 3. Results
   5. 3. 1. Taxadiene 5α-hydroxylase is a promiscuous enzyme
   5. 3. 2. Temporal profile of taxadiene proto-oxidation
   5. 3. 3. The utility of traditional mutagenic practices to improve P450 reactivity
      5. 3. 3. 1. Phylogeny-guided mutagenesis of taxadiene 5α-hydroxylase
      5. 3. 3. 2. Mutations to taxadiene 5α-hydroxylase based on structural inferences
   5. 3. 4. Computational assessment of the taxadiene 5α-hydroxylase catalytic cycle
   5. 3. 5. Computational assessment of taxadiene 5α-hydroxylase mutants
5. 4. Conclusions & perspectives
   5. 4. 1. Comparing the methodologies
   5. 4. 2. The possibility of multiple substrates in a single active site
   5. 4. 3. Directed evolution versus in silico modeling
   5. 4. 4. Implications for metabolic engineering

6 Technoeconomic Analysis of Microbial Paclitaxel
6. 1. Background & assumptions
6. 2. Comparing the economics of the two competing approaches
6. 3. Summary of the economic assessment
6. 4. Providing unprecedented access to inaccessible chemical space

7 Putting It All Together: A Vision for Drug Discovery
   7. 1. What is biosynthomics?
   7. 2. The informed design of superior pharmacophores
   7. 3. Cheminformatics: An integral part of pharmacophore design
   7. 4. Mining the natural world
   7. 5. Continuous manufacturing as a tool for discovery
   7. 6. Metabolic engineering The ultimate demonstration of continuous manufacturing
   7. 7. Final words

Chapter References
Appendix A: List of taxanes analyzed for mapping the taxane biosynthetic pathway 141
Appendix B: Taxus cDNA library 177
Appendix C: Deduced taxane metabolic pathway 193
Appendix D: Sample PDB files for heme moiety of P450 H1 & docking code 197
Appendix E: NMR raw data 207
LIST OF FIGURES

1. 1. The target-rich, lead-poor imbalance 28

2. 1. The disparity between natural products and synthetic molecules 34
2. 2. Decline in the approval of natural product-based pharmaceuticals 35

3. 1. Modularity of terpenoid biosynthesis 43
3. 2. Diversity-oriented biosynthesis 45
3. 3. Modularization and multivariate analysis 49
3. 4. The Jones-Firn Hypothesis 50

4. 1. Overview of the pattern-finding algorithm 57
4. 2. Summary of the taxane biosynthetic pathway 62
4. 3. The P450 catalytic cycle 65
4. 4. The 'snapshot' methodology 66
4. 5. Substrate docking 67
4. 6. Diffusion of oxygen as a constraint 68
4. 7. Method validation via assessing docking of Vitamin D3 into active site of CYP105A1 69
4. 8. Taxane biosynthetic reactivity landscape 71
4. 9. TOS of paclitaxel 73

5. 1. Map of plasmid expressing the P450 system 78
5. 2. Overview of the strain and enzyme engineering protocols used in the study 80
5. 3. Promiscuity & poor selectivity of taxadiene 5a-hydroxylase 82
5. 4. Characterized structures of some of the products 83
5. 5. Daily variation in eicosanoid production titers 85
5. 6. Assessment of the daily variation in the rate of change in the concentrations of the 16 eicosanoids 86
5. 7. Speculative reaction scheme 87
5. 8. Targets for phylogeny-guided mutagenesis 88
5. 9. Activity and selectivity comparisons between the mutant and parental P450 89
5. 10. Differentiating the two types of selectivity improvements 90
5. 11. Grouping the selectivity improvements 91
5. 12. The lysine-aspartate dyad 92
5. 13. Influence of mutations to the dyad and amino acids at the periphery of the active site 93
5. 14. Improving reaction selectivity is a non-trivial challenge 94
5. 15. The oxygen rebound mechanism 95
5. 16. Overview of the competing proton abstractions from taxadiene 96
5. 17. Hypothesized reaction mechanism 97
5. 18. Comparing mutation efficacy using the computational methodology 99
5. 19. A generalized algorithm for in silico enzyme engineering 101

6. 1. Semi-synthesis of paclitaxel from plant-derived 10-deacetylbaccatin III (10-DAB) 107
6. 2. Total synthesis of taxadiene 114
6. 3. Biosynthesis of taxadiene 115

7. 1. Fragment-based pharmacophore discovery 120
7. 2. Overview of the biosynthetics paradigm 121
7. 3. Two scaffolds, one binding site 123
7. 4. Metabolic engineering, the ultimate demonstration of reaction intensification 127
LIST OF TABLES

3. 1. Titers & yields for conversion of glycerol to taxadiene by the engineered *E. coli* strain 46

4. 1. Truncations & estimated pocket volumes of the 16 P450 monooxygenases investigated 64

6. 1. Raw material costs for reaction scheme in Figure 6.1 107
6. 2. Raw material costs for reaction scheme in Figure 4.9. 110
6. 3. Total manufacturing costs for plant-based and microbial-based manufacturing schemes 112
Every commentary that one reads about Big Pharma these days paints the same, grim picture\(^1\). So bleak is the outlook in these reports that it seems implausible that innovator pharmaceutical companies were once amongst the most profitable and envied corporations in America\(^2\). Even addressing the Pfizers and Mercks of this world as ‘innovator’ drug companies seems misplaced nowadays considering the extent to which they outsource drug discovery and clinical trials to contract research organizations (CROs)\(^3\)-\(^5\). Yet, such was the hegemony of these companies over the markets that, in the 1960s, a prominent senator even advocated passing reforms to curtail the drug-makers’ burgeoning profits\(^6\)-\(^7\).

So how did Big Pharma land itself in this current predicament?

Prior to the 1930s, the United States Food & Drug Administration (FDA) was nothing more than a titular regulatory agency whose powers extended to merely inspecting drugs for possible adulterations and seizing adulterated products\(^7\)-\(^8\). In fact, so lenient was regulation during this era that drugs could not only be marketed without testing their safety, but they could also be directly purchased by consumers without a prescription from a doctor or physician. This perilous situation
continued uninterrupted for many years, until finally in 1937, it all came to a head when several individuals succumbed to chemical poisoning after consuming a liquid formulation of sulphanilamide that had been prepared using diethylene glycol as the solvent. That diethylene glycol's toxic effects on humans were fully known and well understood at the time aptly sums up the FDA's alarming lack of regulatory oversight.

Duly pressed, the government enacted the Federal Food, Drug, and Cosmetic Act (FFDCA) a year later, and, it then became mandatory for drugs to be tested for their safety by the FDA prior to market release. Other than this change, however, the provisions of the new act did not really expand the powers of the FDA. Clauses pertaining to the effectiveness of marketed drugs were still absent from the FFDCA, and, the FDA remained restricted to functioning as a passive regulatory body that could only object to the sale of unsafe drugs (as opposed to approving its release into the market). Moreover, the issue of consumers being able to purchase medication without a doctor's prescription also remained unaddressed.

It wasn't until the passage of the Humphrey-Durham Amendment in 1951 that it finally became mandatory for drugs to be sold only by prescription. While some older drugs such as aspirin were exempted from its provisions, the Humphrey-Durham Amendment to the FFDCA successfully managed to stratify the drug demand-supply chain, thereby making it considerably easier for drug manufacturers to market their products. In this new drug market, pharmaceutical companies were only required to sell their products to consumers through 'agents' — the doctors, and, instead of
manipulating the preferences of the average consumer, drug manufacturers would only have to focus their marketing efforts on swaying the opinions of doctors. In due course, the practice of ‘detailing’ – aggressively promoting a drug-maker’s products to a doctor, mainly via personal visits to the clinic by the company’s sales representatives – would take root.

The ballooning profits of the pharmaceutical companies in the decade that followed would once again catch the attention of the federal government. This time around, the influential senator from Tennessee, Estes Kefauver, was one among those who saw a profound flaw in the dichotomous structure of pharmaceutical sales and even advocated implementing reforms to curb the rapidly growing profits of drug companies. While this demand failed to materialize, the horrific thalidomide disaster that coincided with these debates eventually led to the passage of the Kefauver-Harris Amendment to the FFDCA that greatly expanded the role of the FDA to actively approving the release of a drug into the market only following validation of its safety and efficacy through extensive clinical testing. In a way, Senator Kefauver did get his wish – by mandating that a drug be extensively tested for safety and efficacy in lengthy clinical trials, the FDA was prolonging the time it took for the drug to enter the market, thus reducing its effective duration of patent protection and curtailing the drug-maker’s profits. Patent protection and profit maximization would now become buzzwords in drug discovery, development and marketing.

The very mention of patents is a veiled reference to the threat that innovator pharmaceutical companies constantly face from generic or copycat drug manufacturers – a threat that had now
escalated with the passage of the Kefauver-Harris Amendment. However, drugs manufactured by
generic drug producers were still required to undergo extensive testing prior to marketing and this
requirement often placed generic manufacturers – who tended to be much smaller than their
innovative counterparts – at a significant disadvantage. Unsurprisingly, by the 1980s, riding the
wave of their considerable competitive advantage, innovator drug producers continued to account
for roughly 85% of the American drug market⁶.

Suspecting a lack of competitiveness in the pharmaceutical industry, the government would
intervene yet again – this time, by passing the Hatch-Waxman Act in 1984. Formulated with the
specific aim of remedying the competitive disparity in the drug markets, the Hatch-Waxman Act
greatly simplified the regulatory requirements for the manufacture and marketing of off-patent drugs
by generic manufacturers, and, by 1996 – now amidst fiercer competition than before – the share of
innovator pharmaceutical companies in the American pharmaceutical market had eroded to 50%.

To counter what had now become a clear and present danger from generic drug-makers, innovator
pharmaceutical companies would adopt two paradigms that would eventually become the defining
traits of the pharmaceutical industry as we know it today – direct-to-consumer (DTC) advertising
and the needle in a haystack model of drug discovery and development. While one strategy
emphasized the ‘industrialization’ of drug discovery to reduce R&D timelines and lengthen the
duration of market exclusivity granted by patent protection, the other would, in due course, earn
innovator drug companies roughly $4.20 in sales for every dollar it invested in advertising⁷.
1.1. Drug discovery in the pre-Hatch-Waxman era

Prior to implementation of the Hatch-Waxman Act, drug discovery was a team-based exercise and each team would usually focus on a specific therapeutic class. Biochemical or cell-based assays were generally unavailable at the time and ‘lead molecules’ were selected from a library of molecules synthesized by a team of chemists based on their ability to elicit desired effects in whole animal models. More often than not, lead molecules were natural products or their synthetic analogues. While team sizes would vary from project to project, each chemist in a team was expected to contribute between 50-150 molecules annually. Each molecule in the library would be uniquely designed and synthesized individually in gram quantities and at high purities. Once a lead had been identified, ‘lead optimization’ followed suit. Herein, analogues would be synthesized in order to explore structure-activity relationships (SARs) to possibly modify the molecule in order to improve its pharmacological and drug metabolism and pharmacokinetic (DMPK) properties. Once a lead had been sufficiently improved, the drug discovery team would then prepare pre-formulations for preclinical testing.
1. 2. Deficiencies in the old paradigm

The candidates that emerged from such a discovery pipeline, in general, fared quite well in clinical testing. The probability of clinical approval was sufficiently high, and, once a drug candidate gained clinical approval, it assured a company sizable revenues for several years. In fact, in most cases, these earnings comfortably exceeded the company’s original R&D investments. Yet, despite these successes, many in the industry were concerned with what they viewed to be glaring deficiencies in this paradigm of drug discovery. Then, lead generation, identification and optimization were viewed as inseparable functions of the same drug discovery team, implying that drug discovery was sequential and slow\(^1\). Also, the ‘black box’-like nature of the \textit{in vivo} animal models meant that exclusive reliance on such models for evaluating DMPK properties such as absorption, distribution, metabolism and excretion (ADME) increased the risk of selecting a drug candidate that might work in the animal it was tested in, but not in humans\(^9\). Escalation in competition from generic drug-makers that followed the passage of the Hatch-Waxman Act merely accentuated these concerns.

1. 3. The industrialization of drug discovery

At about the same time as the passage of the Hatch-Waxman Act, the use of biochemical and cell-based assays and computational tools in drug discovery was rapidly gaining popularity. \textit{In vitro} assays greatly aided medicinal chemists to understand the molecular mechanisms of action of the drugs that they synthesized, and, as a result, allowed them to design superior drugs against molecularly well-
defined targets. A clear understanding of a drug’s molecular mechanism of action was also beneficial from a regulatory perspective. Cognizant of these benefits, innovator drug companies responded by mandating \textit{in vitro} and \textit{in silico} assessments of the drug’s pharmacological and DMPK properties prior to preclinical testing, and, soon, the use of biochemical and cell-based assays as well as \textit{in silico} tools for drug discovery became pervasive across the industry\textsuperscript{10}.

As part of the new paradigm, molecules that gave a positive result in an \textit{in vitro} screening assay were referred to as ‘primary hits’, and if the activity of a primary hit and its degree of specificity for the target could be repeatedly confirmed, it was designated as a ‘confirmed hit’\textsuperscript{10}. When the activity of a confirmed hit was corroborated in more elaborate assays such as organ bath assays, its selectivity towards a range of closely-related targets had been determined, and its chemical structure had been unambiguously proven by re-synthesis, it was then called a ‘validated hit’. If the \textit{in vivo} pharmacological properties and pharmacokinetic profile of the validated hit were deemed to be acceptable and it met a set of business criteria, the molecule would then become a lead compound, which, following numerous rounds of lead optimization, would later become a ‘drug candidate’. In comparison, in the pre-Hatch-Waxman era, a lead compound rarely underwent standalone pharmacokinetic profiling since these characteristics were (presumably) satisfactorily addressed by the whole animal models\textsuperscript{9}.

Despite their obvious benefits, exhaustively assessing potential leads \textit{in silico}\textsuperscript{11} and \textit{in vitro} prior to \textit{in vivo} pharmacology and DMPK studies significantly prolonged discovery timelines, diminishing
R&D productivity. Faced with ever-increasing competition from generic drug-makers, it soon dawned that the throughput of chemical synthesis had to rise in order to compensate for the increased testing molecules now underwent prior to in vivo testing. The scale and throughput of drug discovery had to rise by several orders of magnitude and this necessitated greatly accelerating lead generation and optimization by multiplexing and automating chemical synthesis and screening.

Accordingly, the next wave of technological adoption across the industry duly focused on massively parallelizing chemical synthesis and screening via miniaturization of reaction systems and assays, all under robotic control. In keeping with this new vision, combinatorial chemistry soon became central to drug discovery\textsuperscript{12-14}. Combinatorial chemistry generates molecular diversity by multiplicatively combining sets of precursors to yield a vast collection of discrete compounds which, depending on their functional groups, could be further reacted combinatorially to produce a larger library of molecules. This library can then be rapidly screened and pruned to yield smaller, more focused libraries whose members can then be selectively manipulated to improve their physicochemical and/or pharmacological properties\textsuperscript{14-16}.

Information technology, too, became pervasive throughout the drug discovery pipeline – right from improving the quality of lead generation through use of virtual screens\textsuperscript{17,18}, ‘bar-coding’ molecules\textsuperscript{14} \textit{viz.} storing data about a molecule’s synthesis and its performance in an assay, to retrieving and utilizing information about a molecule in real-time. Aided by sterling developments in robotics and
instrumentation\textsuperscript{19}, the omics revolution\textsuperscript{20}, as well as new developments such as the lab-on-a-chip\textsuperscript{21, 22} technology, the speed, quality and information density of assays also improved by leaps and bounds.

\textbf{1. 4. New paradigm, same old problems}

For all the exponential improvements in workflow efficiency, the ensemble of available reactions in the synthetic chemistry toolbox remained unchanged. If anything, the number of reactions in the toolbox may have reduced on account of the kinetic incompatibility of several reactions with the timescales imposed by automation. Unsurprisingly, in the three decades following passage of the Hatch-Waxman Act, R&D expenditures by innovator drug companies have ballooned six-fold\textsuperscript{23}. Yet, new drug approvals by the FDA have remained relatively unchanged\textsuperscript{24}, and candidate attrition rates and development times have risen markedly during this period\textsuperscript{2, 25, 26}. In fact, all of the 21,000-odd drug products that have ever been approved by the FDA act on no more than 324 unique molecular targets\textsuperscript{27} and, to make matters worse, this statistic has not changed appreciably following the transition to multiplexed, automated chemical synthesis and screening. Even if one were to generously assume that proteins sharing >35\% identity exhibit related pharmacology, the entire pharmacopoeia is estimated to perturb only 1,048 genes – a mere 3.5\% of the entire human genome or 130 unique functional protein domains from the more than 10,000 folds and 16,000 protein families. In fact, over 50\% of all drugs target just 4 gene families, and, of the new molecular entities that were approved in the past three decades, only 6\% targeted a previously undrugged protein domain (\textbf{Figure 1.1}). Drug discovery in its present multiplexed, automated form, it appears, is
highly redundant, and, the industry’s declining prospects rest solely on the inability of innovator drug companies to discover leads that bind to novel protein domains – and not throughput, as was initially deduced.

![Image](image.png)

**Figure 1.1. The target-rich, lead-poor imbalance:** Of the roughly 10,000 unique protein domains in the human body, only as many as 130 (or 1.3%) have been successfully drugged over the course of the past century. Implementation of the multiplexed, automated paradigm for drug discovery, which is now in its fourth decade, did not alter the synthetic capabilities of the industry. Instead, it merely made the industry more efficient at what it had been doing for close to a century. As a consequence, while there is unanimity in which targets the industry should be pursuing to reverse its declining fortunes, the absence of revolutionary new synthetic techniques implies that only 6% of all NMEs approved during the last three decades targeted domains from the undrugged 98.7%.

### 1.5. The inadequacies of combinatorial chemistry

The high degree of drug redundancy in small-molecule drug discovery – which some experts have labeled as a ‘target-rich, lead-poor’ imbalance\(^2\) – stems from the industry’s use of combinatorial
chemistry for generating compound libraries. In the absence of innovation in chemical syntheses, the
chemical space being screened for biological activity is usually one that has been previously
interrogated\(^{29}\) and, as such, is unlikely to yield hits bearing steric and electronic features capable of
binding novel protein domains. Even on the off chance that combinatorial chemistry outputs a lead
that binds to a new target, it is highly probable that the candidate will exhibit clinically relevant
polypharmacology\(^{30}\).

Furthermore, not only are leads generated by combinatorial chemistry promiscuous, but since the
precursors used in their syntheses are predominantly aromatic, they bind to their intended targets
quite weakly. Aromaticity introduces planarity into the molecules, which imparts them with
rotational flexibility. This often has detrimental entropic consequences for strong and specific
binding\(^ {31}\). That combinatorial chemistry yields only a single NME for every 100,000 molecules that
are screened points to how stacked the odds are against it.

1. 6. Effect of the needle in a haystack paradigm on the industry’s business model

The low probability that a drug candidate had of gaining clinical approval, allied with the exorbitant
costs\(^ {32}\) associated with installing and maintaining the technological base required for automated,
high-throughput lead generation, made it imperative for an approved drug to achieve exceptionally
high sales in order to recoup its significant development costs\(^ {33},\,34\). In fact, so enormous were these
costs that only a third of all drugs approved in the first half of the 1990s were actually profitable\(^ {35}\).
As a consequence, innovator drug companies began targeting only those drugs that were capable of becoming therapeutic standards for chronic diseases that affected a large section of the population, or, in other terms, drugs that stood a good chance of possibly earning millions to even billions of dollars each year – the so-called ‘blockbuster’ drugs.

The shift in companies’ business models is best illustrated by the noticeable rise in market prevalence of blockbuster drugs during the past few years. In 1995, for example, only 17 drugs were blockbusters\(^5\), whereas in 1999 and 2000, those figures rose to 35 and 44 respectively. The rise in market concentration of blockbuster drugs coincided with a similar surge in Big Pharma’s profits, and, by 2001, despite constituting only 20% of the global marketplace, blockbuster drugs accounted for nearly 45% of total sales – or 70% of the industry’s profits\(^10\).

The innovator companies’ decision to pursue quantity over quality drastically altered their technical capabilities. Their chemistry know-how – the very knowledge that formed the basis of their existence and success – more or less stagnated. Perhaps what hurt the most was that the pursuit of quantity forced these companies to forego natural products – a rich source of therapeutic molecules that contributed roughly half of the 877 small-molecule NMEs approved by the FDA between 1981 and 2002.
Chapter 2

NATURAL PRODUCTS & THE PHARMACEUTICAL INDUSTRY

2.1. The historical use of natural products as therapies

Natural products, particularly those from plants, have been used to treat human ailments since antiquity\(^{36}\). In fact, the plant-based traditional medicine systems of India, China and Egypt date as far back as 4000 years\(^{37}\). So economical and effective are these remedies that it comes as no surprise that plant-based medicine forms the bedrock of healthcare in much of the developing world\(^{37}\). In developed nations too, plant products continue to be extensively used as food additives and nutritional supplements. This widespread popularity of plant products for their perceived medicinal properties ultimately prompted the scientific community to elucidate their pharmacological mechanisms and, in due course, several secondary metabolites exhibiting potent therapeutic properties would be characterized and isolated.

The latter half of the 20\(^{\text{th}}\) century witnessed several additions to the collection of bioactive plant compounds, prominently vinblastine, an anti-cancer agent isolated from the Madagascar periwinkle, paclitaxel, an anti-cancer compound produced by the Pacific yew, and digitalis, a compound isolated
from the purple foxglove and used to treat heart conditions\textsuperscript{38}; and several of these compounds would eventually become blockbuster molecules. These developments inevitably attracted the interest of several large pharmaceutical corporations and by the 1970s, bioprospecting – the process of screening, isolating and identifying bioactive secondary metabolites – became the dominant paradigm for drug discovery\textsuperscript{56}.

2.2. What makes natural products good drugs?

Most small molecule drugs elicit their therapeutic effect by selectively inhibiting or augmenting enzyme activity. In fact, the more selective the interaction, the more efficacious the drug is. It follows that ‘drug-like’ molecules are those that can selectively interact with specific biological targets, possess physical properties that facilitate transport by biological fluids across biological membranes, and achieve the right balance between retention time and binding equilibrium.

As natural products play vital roles in reproduction, development, defence and symbiosis – especially in sessile plants\textsuperscript{39} – the strength and selectivity of their interactions with their designated targets largely determines the evolutionary fate of their hosts. As a result, it is believed that the architecture and physicochemical and physiological properties of these molecules have been stringently selected by evolutionary pressures to bear diverse ring systems and chiral centres, exhibit a high degree of oxygenation, and contain a large number of solvated hydrogen-bond donors and acceptors\textsuperscript{40}, which,
in turn, allows these molecules to precisely localize, selectively bind and be efficiently transported within biological systems\textsuperscript{41}.

The limitations of combinatorial chemistry, which yields a single NME for every 100,000 molecules that are screened, are in stark contrast to the fecundity of natural product chemical space\textsuperscript{42}. For instance, 20 or so of the 7,000-odd polyketides to have ever been screened for bioactivity have made it to the market, translating to an approval rate of 1 in 350. In fact, whereas Nature has provided the lion's share of the leads that bind to previously undrugged protein domains, \textit{de novo} combinatorial chemistry has yielded just one drug in the past three decades that is linked with a novel target (Section 1.4)\textsuperscript{43}.

The disparity\textsuperscript{44} between the drug-likeness\textsuperscript{45} of natural products and compounds generated via combinatorial chemistry (Figure 2.1), in turn, can be attributed to the ensemble of reactions that is used to generate the latter. A good number of chemical transformations such as enantioselective synthesis and oxidation, to name a couple, that are kinetically or technologically incompatible with the timescales imposed by high-throughput synthesis and screening have been largely done away with, severely curtailing access to some of the very structural features that are believed to make natural products more drug-like.

Nevertheless, drug prospectors aren't migrating to the natural product space with any particular urgency (Figure 2.2).
2.3. Why is industry interest in natural products declining?

Bioprospecting poses several scientific and non-scientific challenges that make it unappealing for pharmaceutical companies in the current business climate. On account of the great strength and selectivity with which natural products bind to their designated targets, these molecules are almost never produced in copious quantities by their hosts. For example, six fully developed yew trees (Taxus spp.), which take about 200 years to grow, cumulatively produce no more than a single dose
of the blockbuster anticancer drug, paclitaxel\(^{46}\). Additionally, as natural products are generally synthesized via highly modular secondary metabolic pathways, most pharmacoactive natural products have several inactive analogues. Paclitaxel, for instance, is just one of hundreds of closely-related taxanes that are synthesized by the yew species. As a consequence, isolation and structural characterization of natural products continues to be encumbered, slow and not necessarily unambiguous. Natural product extracts aren't nearly as 'screen friendly' as synthetic compounds.

![Bar chart showing the number of NMEs approved by the FDA from 1992 to 2007.](chart.png)

**Figure 2.2. Decline in the approval of natural product-based pharmaceuticals:** The total number of natural product and natural product-derived NMEs approved by the US FDA has been in steady decline since the late 1990s. The number of approvals in 2012 was not any greater than 2007 levels. However, if one excludes the number of repurposed NMEs, namely, molecules that were once approved for activity against a particular target but are now approved for activity against a newer target, the statistics are much worse.

On the manufacturing front, total synthesis of natural products, for all its exquisiteness\(^{47}\), remains a cottage industry, and, synthetic routes for large-scale production of a majority of these molecules
either do not exist or are commercially infeasible. A synthon is defined as a structural unit, or building block, within a molecule that is the product of a specific synthetic unit operation\textsuperscript{48}, and, the combinatorial conjugation of synthons generates vast populations of diverse molecular structures, which are then screened to identify hits\textsuperscript{49,50}. Despite their best efforts, synthetic chemists have been unable to effectively de-construct natural products into synthons, stymying natural product prospection as well as lead identification and optimization. In fact, since process development does not normally commence until the latter half of Phase II clinical trials\textsuperscript{51}, bioprospecting frequently yielded hits that had to be disregarded as late as the latter stages of Phase II clinical trials owing to insurmountable difficulties associated with their synthesis or procurement – all this after considerable amounts had already been spent on their development. While commercial considerations and material sourcing difficulties accounted for only 5\% of candidate attritions from clinical trials conducted in 1991, this number had risen to 28\% in 2001\textsuperscript{52}. Process development continues to remain a major stumbling block in the manufacture of natural products and only those products that were extractable in generous quantities from relatively easy-to-culture organisms have ever realized commercialization.

The supply of source material from which natural products are extracted is also often unreliable and prone to seasonal and environmental variations. In fact, it is not uncommon for promising leads to be forsaken since their native hosts are critically endangered or large-scale production of the drug could be detrimental to the biodiversity of the host’s natural habitat. Take the case of medicinal
plants, a source that has yielded about a fifth of all clinically approved natural products. According to one estimate, roughly a quarter of medicinal plant species now face the threat of extinction.

2.4. Alternatives for probing natural product space

Recent applications of genetic approaches to decipher the biosynthetic code that guides the assembly of polyketides have revealed that living systems methodically and repetitively combine building blocks produced by a handful of enzyme-catalyzed unit reactions to generate unprecedented molecular diversity. On the back of these developments, and aided by sterling advances in microbial genetics and industrial biotechnology, the recent explosion in the volume of gene and protein data, the development of more precise techniques for studying cellular metabolism and genetic regulation and the exponentially declining costs of genome sequencing and oligonucleotide synthesis, microbial metabolic engineering is poised to become the fulcrum of new technologies for probing natural product space. Already, significant strides in developing novel microbial strains for the production of artemisinic acid, a precursor to the antimalarial drug, artemisinin; the nutraceutical, lycopene; the production of several antibiotics; levopimaradiene, a gateway precursor to the ginkgolide family of pharmaceuticals and health supplements; and taxadiene, the first dedicated metabolite in the biosynthesis of taxol, a blockbuster anticancer drug. In principle, all that is required is to transplant the genes of the biosynthetic pathway yielding the target product into the host, eliminate unnecessary or physiologically less significant native reactions, and lastly, re-tool the cellular regulatory networks via rational or random approaches to ensure optimal production. This strategy
offers numerous advantages over *de novo* chemical synthesis and plant cell cultures in the production of plant secondary metabolites as it is readily scalable, consumes lower quantities of resources, is benign to the environment and uses inexpensive sugar-based carbon inputs. A synergism of bioprocess optimization and metabolic engineering promises high titre production of these therapeutic molecules and offers a commercially viable alternative to the production of bioactive plant-derived chemicals.
3. 1. The terpenoid superfamily

Natural products are estimated to total about 170,000 molecules, of which, a vast majority are plant secondary metabolites\(^1\), and, hard as it is to believe, this immense chemical diversity is the product of just 3 metabolic pathways. Nearly 60% of all natural products are believed to synthesized by the terpenoid (or isoprenoid) pathway, 30% are contributed by the related polyphenol, phenylpropanoid and polyketide pathways, and about 10% originate from the alkaloid pathway.

Terpenoids are interesting from several perspectives – not only are they the most abundant and ancient natural product family, but several terpenoids have been characterized as anti-cancer, anti-microbial, anti-allergenic, anti-inflammatory, and immunomodulatory drugs\(^1,\,2\). Some such as menthol, a counterirritant harvested in large quantities from the peppermint herb, *Mentha piperita*; artemisinin, an anti-malarial compound produced by the plant, *Artemisia annua*; abietic acid, an acidic resin extracted from conifers and used as a feedstock; and paclitaxel, a blockbuster anti-cancer drug isolated from the bark of the Pacific yew, *Taxus brevifolia*\(^2-4\) have already been commercialized by pharmaceutical companies.
3. 2. Modularity of terpenoid biosynthesis

Despite being a potential treasure chest of blockbuster molecules, interest from metabolic engineers in terpenoid biosynthesis, especially for pharmaceutical production, has been modest in comparison to the polyketide family, a substantially smaller class of molecules. This is explained, in part, by the fact that a great number of bioactive polyketides are products of multimodular megasynthases called type I polyketide synthases (PKSs), which made their study and manipulation less encumbered than the distributed terpenoid pathways. However, as our understanding of terpenoid biosynthesis has greatly matured over the past two decades, it is now apparent that re-engineering terpenoid pathways to meet a discovery program's synthetic requirements is arguably easier than modulating multienzyme PKSs that shuttle metabolic intermediates from one active site to the other.

In addition to modifying the reaction mechanism and substrate specificity of the pathway enzymes, PKS re-engineering also incurs the additional task of altering the interactions between its constituent modules as minor perturbations to the contact interfaces of the domains could detrimentally lower the activity of the entire enzyme complex. The enzymes that participate in terpenoid assembly, in contrast, act individually.

Terpenoid construction (Figure 3.1) commences with the head-to-tail condensation reaction between two C₅ molecules, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate.
(DMAPP). The latter is the smallest member of a homologous series of metabolites known as the allylic pyrophosphates. The pyrophosphate moiety of the allylic pyrophosphates is a strong leaving group and the transition state π-conjugated carbocation is very stable. This makes allylic pyrophosphates strong alkylating agents. IPP combines with DMAPP to generate a 10-carbon allylic pyrophosphate, which itself then alkylates IPP in situ to produce a 15-carbon species. Successive repetitions of the same head-to-tail condensation in the same active site generate longer allylic pyrophosphates in C₅ increments. The number of repetitions is pre-programmed and enzyme-specific. Incidentally, IPP is synthesized by either the mevalonate or non-mevalonate pathways. The latter commences with the condensation of glyceraldehyde-3-phosphate with pyruvate to yield 2-methyl-(D)-erythritol-4-phosphate (MEP). In contrast, the mevalonate pathway commences with the co-condensation of acetyl-CoA to produce acetoacetyl-CoA. The non-mevalonate pathway, with a few exceptions, is specific to prokaryotes and utilizes carbon and energy sources more efficiently than the eukaryotic-based mevalonate pathway.

The head-to-tail condensation that elongates the terpenoid chain is re-implemented in the second stage in a different configuration to convert the linear allylic pyrophosphate to a dephosphorylated, polycyclic molecule. Cyclization proceeds via an intra-chain Markovnikov addition reaction between any one of the several alkenyl units in the polymer chain and the allylic carbocation that is formed at the tail end of the molecule following the departure of the pyrophosphate group. Since the alkenyl units are quite close to one another, repeating at 5-carbon intervals within the chain,
addition to any one alkene initiates a cascade of proton abstractions and bond migrations that culminates in the formation of a predominantly saturated polycycle.

The catalogue of terpenoid structures synthesized by an organism can be expanded by evolving enzymes that favor coupling between two allylic pyrophosphates over electrophilic alkylation of IPP, thereby yielding branched chains$^{66,70}$. Varying the number of elongation events that occur during the first stage or altering the configuration of the electrophilic attack during the second stage also adds to terpenoid diversity$^{71}$. Crucially — and, perhaps unique to terpenoid biosynthesis — addition of the allylic carbocation to the carbon-carbon double bond in IPP can be cis/trans. While organisms generally evolve enzymes that selectively form one geometric isomer over the other, combinatorial biocatalysis offers the potential to literally exponentiate the number of synthesizable structures by introducing geometric isomerism each time the enzymatic reaction occurs$^{70}$.

In the third stage, the polycycle is oxidized by a cytochrome P450 monooxygenase. The P450s constitute a large family of highly conserved, heme-thiolate enzymes that are present in all forms of life$^{72}$. They usually act in concert with a flavoprotein redox partner — a cytochrome P450 reductase — that shuttles electrons and protons derived from NADPH to the oxygenase's active site, whereat molecular oxygen is protonated and heterolytically cleaved to form water and a reactive iron-oxygen species that then inserts the second oxygen atom into the substrate$^{73}$. Oxidation of the hydrocarbon scaffold activates it for further functionalization in the fourth stage of terpenoid synthesis.
Chain elongation via head-to-tail condensations:

\[
\text{OPP} + \text{OPP} \rightarrow \text{OPP} + \text{OPP}
\]

General formula for head-to-tail condensation:

\[
\text{OPP} + \left(\text{OPP} \right)^n \rightarrow \text{OPP} + \left(\text{OPP} \right)^n
\]

Scaffold formation via intra-chain condensation:

\[
\text{OPP} + \text{GGPP} \rightarrow \text{Taxadiene}
\]

Scaffold activation by oxidation:

\[
\text{Taxadiene-5a-ol} + \text{O}_2 + \text{NADPH} \rightarrow \text{Taxadiene-5a-ol} + \text{NADP}^+
\]

Core diversification:

\[
\text{Taxadiene-5a-ethanoate} + \text{CoA-SH} \rightarrow \text{Taxadien-5a-ethanoate} + \text{CoA-SH}
\]

Figure 3.1. Modularity of terpenoid biosynthesis: De novo total synthesis of the C_{20} taxane scaffold involves in excess of 20 steps. It biosynthetic assembly and structural core diversification, in comparison, is considerably more straightforward.

The nucleophilic oxygen atom of the hydroxyl group is either directly attacked by strong electrophiles or forms a more nucleophilic alkoxide species that then attacks weaker electrophiles. Acylations are the most commonly observed products during this stage of terpenoid assembly. The hydroxylated scaffolds can also undergo hydroxyl hydrogen substitution reactions such as
esterifications, oxidations, halogenations, hydrosulphonations, glycosylations and even protein conjugations *in vitro*. Being oxidized, such scaffolds also allow the ready addition and replacement of bioisosteres for faster and more direct lead optimization.

### 3.3. Metabolic & enzyme engineering for targeted & diversity-oriented biosynthesis

Ordered combinations of known genes into biosynthetic operons to be expressed as heterologous metabolic pathways in tractable microbial hosts – a strategy that is referred to as biosynthetic engineering – generates pharmacophores that are natural products *in toto*, whereas permutations and mutations produce molecules that possess a scaffold that is an unnatural isomer of a natural product. Conversely, clustering combinations, permutations and mutations of the biosynthetic genes could generate libraries of natural product-like molecules (Figure 3.2).

Once a combination of enzymatic reactions that can or could produce the molecule of interest has been identified, individual active sites must then be re-engineered in order to enhance substrate specificity, product selectivity and activity. This is especially true for targeted synthesis using enzymes from secondary metabolic pathways that are known to be promiscuous and poorly selective. The mutagenized enzymes must then be combined into a pathway and enzyme expression must be suitably toggled in order to adjust their concentrations. Discriminatory expression equalizes enzyme turnovers and elevates flux through the pathway.
Metagenomics will add to our repository of natural product biosynthetic pathways. Using a microbial chassis for expressing the enzyme combinations, permutations and mutations.

Cataloguing the reactive proclivities of the enzymes, including the sub-types of the reactions they catalyze.

Generating molecular diversity using combinatorial biocatalysis and chemistry.

Using a microbial chassis for expressing the enzyme combinations, permutations and mutations.

High-throughput screening: Anti-infective discovery.

Figure 3.2. Diversity-oriented biosynthesis: Akin to the concept of 'combinatorial biosynthesis' that had been espoused by advocates of polyketide-based pharmaceuticals, controlled expression of combinations, permutations and mutations of terpenoid biosynthetic genes, when perfected, can potentially generate focused libraries of natural product-like molecules.

The engineered pathway must later be integrated into a retrofitted microbial metabolic network. Interventions to microbial metabolism can occur at one or many of the following levels: (1) enhancing the rate of substrate uptake, (2) reducing the flux to competing native pathways, (3) enhancing the flux of precursors and cofactors to the engineered pathway, (4) exporting the product to the extracellular medium in order to shift equilibrium towards product formation. The choice of
the microbial host is guided by its propensity for growth, carbohydrate substrate preference, potential toxicity of products and by-products, permissible metabolic load, and adherence to desirable bioprocessing requirements.

3.4. Summary of recent progress

Our research group recently demonstrated the feasibility of biosynthetic engineering to successfully synthesize taxadiene, the first intermediate in the paclitaxel biosynthetic pathway that bears the unique taxane scaffold, using glycerol as the primary carbon feedstock at unprecedented titers of 0.30 and 1.00 g/L in 2 mL fed-batch and 1 L batch fermentations\(^9\). For the 1 L fermentations, the optical density of the *E. coli* cultures reached a peak of \(-40\) after 5 days. Assuming an equivalency between an optical density of 1.0 and a dry biomass titer of 0.3 g/L, the overall yield of taxadiene on a dry cell weight basis is estimated to be 0.0925 g/g.

<table>
<thead>
<tr>
<th>Scale &amp; configuration</th>
<th>Highest taxadiene titer (g/L)</th>
<th>Yield (g/g glycerol)</th>
<th>Actual yield as a % of theoretical maximum yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 mL fed-batch</td>
<td>0.30</td>
<td>0.02</td>
<td>5.88</td>
</tr>
<tr>
<td>1 L batch</td>
<td>1.00</td>
<td>0.03</td>
<td>8.82</td>
</tr>
</tbody>
</table>
This work marks an important conceptual advance as 2 of the 4 biosynthetic modules of terpenoid synthesis – namely, chain elongation and cyclization – appear to function quite robustly in a simple and easily manipulable host such as *E. coli*.

### 3.5. Possible challenges moving forward

The next logical goal is demonstration of microbial-based stereo- and regiospecific scaffold oxidation for the production of activated molecules for structural core diversification, either by enzymatic or chemical approaches, to corroborate the claims of biosynthetic engineering for drug discovery. To this end, we have targeted the oxidation of taxadiene, which is presumed to first undergo hydroxylation at the 5α-position. Unfortunately, the exact sequence of substitutions following 5α-hydroxylation of the taxane scaffold is unknown and deciphering the precise order of the biotransformations has been the subject of intense research over the past two decades. Yet, despite their best efforts, researchers have simply been unable to map the taxane metabolic pathway. In fact, a large majority of the biosynthetic pathways that produce potentially valuable terpenoids are uncharacterized.

Moreover, even if a pathway has been completely catalogued, its regulatory mechanisms are often poorly understood. This implies that construction and optimization of expression of the pathway, especially balancing the catabolic, primary metabolic and secondary metabolic fluxes to ensure
equitable distribution of resources between growth, maintenance and product formation\textsuperscript{76} are fairly long-drawn. For instance, strain engineering and optimization for the production of taxadiene at the yield and productivity stated earlier (which, as shall be demonstrated in Chapter 6, ensure cost-competitiveness of taxol production using microbial metabolic engineering) consumed an equivalent number of cloning hours as work conducted to demonstrate production of amorphadiene, another unsubstituted terpenoid scaffold, in \textit{E. coli}. Here, the methodology of assessing and eliminating regulatory and pathway bottlenecks by re-defining the metabolic network as a collection of distinct modules whose constitution and overall concentrations can be efficiently adjusted using a variety of transcriptional, post-transcriptional and translational mechanisms could maximize pathway turnover with respect to resource expenditure. The practice of re-casting metabolic networks as a collection of interacting modules makes the analysis of a very complicated system considerably more tractable and allows easy assessment of the performance of module configurations using multivariate statistics (Figure 3.3).

Additionally, as evidenced from the gibberellin biosynthetic pathway\textsuperscript{77}, being components of the adaptive immune system in plants\textsuperscript{78-80}, secondary metabolic pathways comprise several promiscuous and non-selective enzymes that maximize the chemical diversity produced within the pathway, which, in turn, allows the plant to respond to a greater number of environmental stimuli without expending additional energy or increasing the information content of its genome (Figure 3.4). Although useful from the perspective of diversity-oriented synthesis, this hypothesis – called the
Jones-Firn Selection Hypothesis – presents some interesting challenges for the targeted production of molecules using metabolic engineering.

Figure 3.3. Modularization and multivariate analysis: The flux through a linear pathway is the reaction rate of the pathway's slowest step. This implies that, at steady state, the observed turnovers of several enzymes of the pathway are below their maximum values. By grouping enzymes with similar turnovers into a subset, or module, and later equalizing the turnovers of the different subsets by adjusting variables such as module constitution, promoter strength, plasmid copy numbers, RBS strength, among others, one can assess pathway turnover using multivariate statistical operations such as principal component analysis. This approach has already found extensive use in analysis of molecular events in signal transduction pathways.

Once a likely combination of enzymatic reactions that could produce the molecule of interest has been identified, they will necessitate customization or active site re-engineering to enhance substrate
specificity, product selectivity and activity to minimize flux dissipation away from the molecule of interest. Consequently, enzyme engineering becomes another critical endeavor in the goal to produce specific natural products in microbial hosts.

Figure 3.4. The Jones-Firn Hypothesis: According to Clive Jones and Richard Firn, if every secondary metabolite can be thought of as being a unique response to an environmental stimulus, it can be argued that the enzymes of secondary metabolic pathways may have evolved to become promiscuous. Substrate promiscuity or poor product selectivity maximises the structural and functional diversity of secondary metabolites produced by the host, which, in turn, allows it to respond to a greater diversity of environmental stimuli, thereby improving its evolutionary fitness. In the two hypothetical scenarios depicted here, the plant in B can respond to the same number of environmental stimuli by incurring a roughly 30% lower metabolic load compared to the plant in A.
Paclitaxel is manufactured via a semi-synthetic process that commences with extraction of its precursor, 10-deacetylbaccatin III (10-DAB), from plant cell cultures of Taxus spp. As metabolomics investigations have revealed that paclitaxel is one of several hundred closely related taxanes produced by the Taxus cell cultures, a significant proportion of the bulk price of paclitaxel, which, in 2012 was estimated to be roughly $190,000 per kg, is attributed to separation costs$2. Development of a microbial-route that can produce 10-DAB, paclitaxel or another precursor at high titers and very few by-products could drastically lower the price of the drug, thereby increasing access to more patients.

However, as discussed in Section 3.5, the paucity of information regarding taxane and several other terpenoid biosynthetic pathways has greatly restricted access to these valuable molecules through microbial metabolic engineering. For the case of the taxane pathway, all that is known is that it commences with the cyclization of the linear molecule, geranylgeranyl pyrophosphate (GGPP) to produce taxadiene$3. It has been presumed that taxadiene is then substituted at the 5α-position with
a hydroxyl group. The exact sequence of substitutions following 5α-hydroxylation of the taxane scaffold is unknown and deciphering the precise order of the biotransformations has been the subject of intense research over the past two decades. Yet, despite their best efforts, researchers have simply been unable to map the taxane metabolic pathway.

Demonstrating the successful expression of the third chemical module in terpenoid biosynthesis – namely, scaffold oxidation – in *E. coli* necessitates metabolic cartography to determine the identities of the genes for their eventual expression in the microbe. To this end, we developed a generalizable methodology to map the biosynthetic pathway by analyzing chemical substitution patterns in naturally occurring taxane metabolites as well as investigating taxane biochemistry via molecular modeling and docking studies. Identifying the sequential order of metabolic reactions in the taxane biosynthetic allows one to study the biochemistry of the pathway by correctly pairing enzymes with their substrates, providing valuable insights for protein engineering to improve enzyme activity and selectivity. This new intellectual framework for elucidating biosynthetic routes and the biochemical mechanisms of their constituent reactions could open up opportunities for investigating several more terpenoid pathways.

4.1. Current understanding of the taxane pathway

While the exact sequence of reactions in the pathway has been hitherto elusive, significant progress has nonetheless been made in the study of taxane metabolism. Among these, metabolomics
investigations on *T. cuspidata* cell cultures have been able to successfully isolate and characterize 128 taxane metabolites (Appendix A), and several putative taxane biosynthetic genes have also been sequenced (Appendix B). However, the absence of a blueprint or map of taxane metabolism has prevented the identification of metabolite-enzyme pairs, which in turn has hampered efforts to study the pathway's biochemical and regulatory mechanisms in greater detail.

Graph theory and atom mapping have been previously utilized to discover novel or non-standard metabolic pathways. Herein, genome-scale models and information about atom conservation within known pathways are utilized to design novel metabolic pathways for the production of desired molecules from defined precursors can be designed and tested. Our situation is a direct contrast to these methodologies. Additionally, since secondary metabolic pathways proceed using a structural core formation-diversification synthesis plan, atom mapping is unlikely to shed any light on pathway topology.

Some recent studies have attempted to decipher the sequence of reactions in the taxane biosynthetic pathway by systematically evaluating the reactivity of individual enzymes towards different substrates *in vitro*. For example, if an enzyme that is presented with taxadiene-5α-ol, taxadiene-10β-ol and taxadiene-5α,10β-diol as substrates yields taxadiene-10β,2α-diol as the major product, it is annotated as taxadiene-10β-ol 2α-hydroxylase. The non-reacting or partially reacting substrates are then presented to another enzyme to determine its identity and this evaluation continues until all enzyme-substrate pairs have been identified.
All that is known about the taxane metabolic scheme thus far, including annotations of genes of the pathway, has been achieved using the aforementioned methodology. However, such a method for pairing enzymes to their corresponding substrates suffers from a glaring shortcoming that assumes that taxol is the major end-product of a linear pathway. This might not be the case. As suggested by the Jones-Firn Screening Hypothesis (Section 3.5), enzymes of the taxane pathway may very well be promiscuous and non-selective, implying that the pathway could be highly branched.

4. 2. Description of the cartography methodology

The methodology utilizes the chemical structures of intermediates and products, and the putative sequences of the pathway’s genes (which is used to derive homology models of the pathway’s enzymes) as inputs and then outputs the map or sequence of reactions in the metabolic pathway. The analysis is unique in its integration of data across three strata of the metabolic hierarchy, namely metabolites, genes, and proteins and makes novel use of tools and techniques that are otherwise commonly applied to the study of chemical or biological problems90. Moreover, the algorithmic architecture of the analytical procedure ensures that it is easily programmable for use as a software package for the analysis of poorly characterized metabolic pathways.

The analysis commences by cataloguing the chemical structures of as many naturally-derived taxanes and then recasting this pictorial information into a mathematically interpretable format. Examples of
mathematically interpretable formats include the SMILES\textsuperscript{91} (simplified molecular input line entry specification) nomenclature that is frequently used to catalogue molecules in chemical databases, or the more generic, binary vector representation wherein the length, column numbers and elements of the vector represent the total number of carbon atoms in the molecule, their assigned IUPAC designations, and the type of substitution they bear, respectively. It is, however, essential that a single format be used for all the molecules to preserve consistency and allow easy analysis of metabolite structural data. Also, using mathematically interpretable formats makes programming the analysis into a software package more amenable. Structure cataloguing revealed that all naturally-derived taxanes bear a subset of the following substitutions:

i. Carbon 1 is exclusively \(\beta\)-substituted with either a hydroxyl group or an ester group.

ii. Carbon 2 is exclusively \(\alpha\)-substituted with either a hydroxyl group or an ester group.

iii. Carbon 3, which is a tertiary carbon, is never substituted. Substitution at this position is sterically hindered.

iv. A double bond is mostly observed between carbons 4 and 20. In some cases, carbon 4 is part of a 4-member ring known as the oxetane ring wherein it is directly bonded to carbon 20 and carbon 5, which, in turn, are connected to each other via an ether bond. Carbon 4 is also \(\alpha\)-substituted with an ester group when it is part of the oxetane ring. In rare cases, carbon 4
and 20 form an epoxide ring, while in even rarer cases, carbon 20 is hydroxylated by hydration of the carbon 4-carbon 20 double bond.

v. Carbon 5 is generally α-substituted with either a hydroxyl group or an ester group. In some cases, it is also a part of the oxetane ring.

vi. Carbons 6 and 8 are never substituted.

vii. Carbon 7 is exclusively β-substituted with either a hydroxyl group or an ester group.

viii. Carbon 9 is exclusively α-substituted with either a hydroxyl group or an ester group. Carbon 9 is also ketonated in many taxanes.

ix. Carbon 10 is exclusively β-substituted with either a hydroxyl group or an ester group.

x. Carbons 11 and 12 are never substituted and are connected to each other by a double bond.

xi. Carbon 13 is exclusively α-substituted with either a hydroxyl group or an ester group. Carbon 13 is also ketonated in many taxanes.

xii. Carbon 14 is exclusively β-substituted with either a hydroxyl group or an ester group.
xiii. Carbon 15 is very rarely connected to carbon 13 via an ether bond. Otherwise, it is unsubstituted.

xiv. Carbons 16, 17, 18 and 19 are never substituted.

Next, a Bayesian probability-based algorithm is used to infer relationships between the taxane metabolites. The conceptual foundations of this approach, called 'substitution pattern-finding', are similar to those of applying Bayesian probability to establish phylogenetic relationships between organisms in evolutionary biology (Figure 4.1).  

![Diagram of metabolite structures](image)

\[ n_1 = \text{Number of metabolites with hydroxylation at } 5\alpha \text{ position} \]

\[ n_2 = \text{Number of metabolites with hydroxylation at } 7\beta \text{ position} \]

\[ n = \text{Number of metabolites with hydroxylations at both, } 5\alpha \text{ and } 7\beta \text{ positions} \]

If \( n_1 > n_2 \) and \( n_2 = n \), hydroxylation at \( 5\alpha \) precedes that at \( 7\beta \) position

**Figure 4.1. Overview of the pattern-finding algorithm**
4. 3. Constraints for ensuring accuracy of the cartographic methodology

Consider a hypothetical sequence of reactions that commences with the conversion of taxadiene to taxadiene-5α-ol. Taxadiene-5α-ol is then converted to taxadiene-5α,10β-diol, which then produces taxadiene-5α-acetate-10β-ol following esterification of the 5α-hydroxyl group. If one was only provided with the chemical structures of these metabolites without any information regarding the sequence of the reactions that connect them, evaluating the substitution patterns of the metabolites can yield the putative sequence of reactions. However, merely relying on substitution occurrence frequencies could lead to confounding conclusions. Accordingly, chemical feasibility constraints have to be imposed on the pattern finding exercise for deducing meaningful and non-confounding relationships between the taxane metabolites. These constraints include inviolable rules in organic chemistry as well as information about putative reaction mechanisms. One example of an inviolable rule in organic chemistry is the impossible occurrence of ester formation prior to hydroxylation. Also, understanding the mechanism of a reaction can provide insights into aspects such as transient species stabilization via steric or electronic modifications in the molecule.

Amongst the metabolites listed above, it is evident that taxadiene is the least substituted molecule and hence must precede all the other species in the hypothetical pathway. Similarly, since the 5α-acetylation must succeed the 5α-hydroxylation, it can be concluded that taxadiene-5α-acetate-10β-ol can be produced by one of three possible routes: (1) taxadiene-5α-ol to taxadiene-5α-acetate to taxadiene-5α-acetate-10β-ol, (2) taxadiene-5α-ol to taxadiene-5α,10β-diol to taxadiene-5α-acetate-
10β-ol, or (3) taxadiene-10β-ol to taxadiene-5α,10β-diol to taxadiene-5α-acetate-10β-ol. However, the separate occurrence of taxadiene-5α-ol and taxadiene-5α,10β-diol suggests that the likeliest pathway from taxadiene to taxadiene-5α-acetate-10β-ol is via path 2.

The map of taxane metabolism as outputted by the constraints-based substitution pattern finding is then validated in the second stage of the theoretical framework – computational enzyme catalysis. Herein, bioinformatics and computational chemistry are utilized to study enzyme catalysis. To eliminate bias, we ignore enzyme annotations as they appear in GenBank.

4.4. Computational assessment of the reactivity landscape of the early taxane pathway

4.4.1. Homology modeling

Since crystallographic structures of none of the taxane biosynthetic enzymes have been hitherto characterized, homology or comparative modeling has to be utilized in order to obtain precise representations of the enzymes and their active sites. Briefly, homology modeling of an enzyme, or target, commences with template selection, and a wide variety of open-source fold-recognition or threading software can be utilized for this purpose. These programs score and rank templates based on the similarity of their structural folds to those predicted for the target enzyme. While the top-ranked templates as suggested by the fold-recognition or threading programs generally produce fairly accurate homology models, it is recommended that they be manually checked for similarity with the
target enzyme by comparing predictions for physicochemical parameters such as their secondary structures, solvent accessibility, membrane-binding domains, ordered-disordered regions as well as phenomena such as distance bonding. In the rare event that a suitable template is unavailable, one has to resort to de novo enzyme modeling. Fortunately, the likelihood of such a scenario arising for enzymes of plant secondary metabolic pathways is quite slim.

After suitable templates have been selected, their sequences are carefully aligned with that of the target using the compatibility between the aforementioned physicochemical parameters as a constraint. Open-source software is also available for this purpose. Incidentally, most fold-recognition software also produce alignments between the top-ranked templates and the targets, but these are quite crude and have to be substantially refined if they are to be used for model building.

The target-template alignment is then utilized to build homology models. Many software packages, each differing in their optimization criteria, are freely available for this task, and the quality of models that they output is fairly comparable for most families of proteins. Once the homology models have been built, they are checked for structural correctness using gauges such as a Ramchandran plot and further, if necessary.

The *Taxus* cDNA library from which the sequences were derived contains 16 P450 and 10 acyltransferase (9 acetyl- and 1 benzoyltransferase) sequences. Homology models of the taxane biosynthetic enzymes were constructed using MODELLER. Target sequences that exceeded
their templates in length were suitably truncated in order to ensure that the homology modeling program is able to capture the folds of the templates. As the cytochrome P450 monooxygenases contain heme prosthetic groups, their active sites were first located using the pocket searching algorithm, POCASSA⁹⁷, following which the heme groups were manually docked and their geometries optimized using the BFGS energy minimization algorithm. The structural database, DALI⁹⁸, was also searched for crystallographically solved homologues to validate the output of the pocket searching algorithm.

4.4.2. Substrate docking

Next, computational models of an enzyme's putative substrate and product, as suggested by the consensus map, as well as a broad selection of other taxane molecules were docked into the active site of the enzyme model using an open-source software package, AUTODOCK, to predict the preferred orientation of the ligand as well as its binding affinity (or energy of interaction) with the active site. It was hypothesized that docking a large collection of molecules into the active site of the enzyme could potentially offer several insights into its promiscuity towards substrates. The docking exercise is repeated for all enzymes, and upon its completion, one can finally compare the energies of interaction for several enzyme-metabolite pairs, thereby confirming or rejecting those suggested by the consensus map.
4.5. Results

4.5.1. Pathway map

Extending the pair-wise comparisons to encompass all substituted positions produced a putative map of taxane metabolism, whose early stages almost exclusively comprise of P450-mediated oxidations (pathway summary in Figure 4.2, entire pathway is been presented in Appendix C).

![Pathway map](image)

Figure 4.2. Summary of the taxane biosynthetic pathway: Interestingly, studies on the production of paclitaxel by plant cell cultures of *T. cuspidata* have revealed that paclitaxel, on a weight basis, accounts for about a fifth of all taxanes produced. Accounting for variations in molecular weights, the map outputted by the substitution pattern finding exercise appears to mirror actual production of paclitaxel.

While substitutions at most carbon atoms on the taxane scaffold can be definitively ordered, the temporal relation between some functionalizations such as hydroxylations at the 2α, 9α and 13α positions cannot be deduced. Also, substitution patterns suggest that oxetane ring formation is among the last modifications to occur on the scaffold.
The suggestion that the early pathway almost exclusively consists of P450-mediated reactions is not all that surprising as this biosynthetic strategy has also been observed in other natural product biosynthetic pathways. As P450 monooxygenases catalyze C-H bond functionalization instead of acylation of an O-H bond that is mediated by acyltransferases, the former are more prone to promiscuity and non-selectivity, increasing chemical diversity within the pathway. Single amino acid mutations within the active site of P450 monooxygenases could make these reaction sites ‘plastic’, which, according to the Jones-Firn Selection Hypothesis, is a more robust and reliable way to generate diversity compared to the significant alterations that might be required to re-orient a polyoxidized molecule in order to drive acylation at one hydroxyl group over another. Accordingly, the reactive landscape was subsequently investigated by assuming that the hydroxylations at the 5α- and 10β-positions precede the others and assuming all possible permutations of the 2α-, 9α- and 13α-hydroxylations.

4.5.2. The reactive landscape

4.5.2.1. Comments about computational assessment of the P450 reaction mechanism

Homology models of all 16 P450 monooxygenases in the *Taxus* cDNA library were constructed using MODELLER. The enzymes were randomly annotated as H1 through H16 in order to eliminate bias. The reductase domain was excluded from this analysis. The A-chains of the P450 enzymes, 2CIB, 3EQM, 3GW9, 3DBG, 3LD6 and 2VE3 were used as templates. The homology
models were then augmented with the heme prosthetic group. The thiolate linkage between the heme iron and cysteiny1 sulfur was first created, following which, geometry of the heme moiety was optimized using the BFGS energy minimization algorithm. The number of iterations in the BFGS algorithm was left uncapped in order to ensure convergence to the optimal geometry. Accuracy of the optimized geometry was verified by comparing the bond length of the thiolate linkage in the homology model to those that are typically observed in solved crystal structures of other P450s. The average bond length of the thiolate linkage for the 16 homology models was 2.56 Å, which compares quite favorably with the average thiolate bond length of 2.47 Å for the 6 templates.

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Truncation for modeling (# of amino acids)</th>
<th>Untruncated length (# of amino acids)</th>
<th>Pocket volume (Å³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>10</td>
<td>484</td>
<td>256</td>
</tr>
<tr>
<td>H2</td>
<td>40</td>
<td>484</td>
<td>422</td>
</tr>
<tr>
<td>H3</td>
<td>50</td>
<td>501</td>
<td>601</td>
</tr>
<tr>
<td>H4</td>
<td>50</td>
<td>500</td>
<td>748</td>
</tr>
<tr>
<td>H5</td>
<td>40</td>
<td>492</td>
<td>395</td>
</tr>
<tr>
<td>H6</td>
<td>50</td>
<td>500</td>
<td>387</td>
</tr>
<tr>
<td>H7</td>
<td>40</td>
<td>483</td>
<td>400</td>
</tr>
<tr>
<td>H8</td>
<td>50</td>
<td>503</td>
<td>239</td>
</tr>
<tr>
<td>H9</td>
<td>45</td>
<td>497</td>
<td>610</td>
</tr>
<tr>
<td>H10</td>
<td>50</td>
<td>507</td>
<td>559</td>
</tr>
<tr>
<td>H11</td>
<td>50</td>
<td>509</td>
<td>286</td>
</tr>
<tr>
<td>H12</td>
<td>35</td>
<td>485</td>
<td>250</td>
</tr>
<tr>
<td>H13</td>
<td>40</td>
<td>493</td>
<td>382</td>
</tr>
<tr>
<td>H14</td>
<td>45</td>
<td>498</td>
<td>250</td>
</tr>
<tr>
<td>H15</td>
<td>45</td>
<td>495</td>
<td>477</td>
</tr>
<tr>
<td>H16</td>
<td>50</td>
<td>499</td>
<td>474</td>
</tr>
</tbody>
</table>
The heme prosthetic group proceeds through 4 states during the P450 catalytic cycle (Figure 4.3) – the ground, activated, peroxyheme and oxyferyl states. As the modeling approach cannot capture nor predict the dynamics of substrate orientation within the active site, a ‘snapshot’ methodology (Figure 4.4) that assesses the substrate pose at each stage of the P450 cycle and its eventual influence on the reactive pose viz. the pose that the substrate adopts just prior to reaction with the oxyferyl group, is implemented.

The P450 catalytic cycle: Substrate binding within the P450 active site activates the heme iron. Binding of molecular oxygen follows suit, and electrons supplied by NADPH are shuttled to the active site shortly after. Transition of the heme iron from the peroxy to the oxyferyl state reductively cleaves the oxygen molecule to form water. The oxyferyl species then oxidizes the substrate to typically form a hydroxylated product.
In the snapshot methodology, owing to our use of the BFGS energy minimization algorithm to optimize the geometry of heme in the active site, AUTODOCK was unable to distinguish between the ground and activated states of heme, and peroxo- and oxyferyl heme. The energy trajectory for the oxidation of taxadiene, therefore, was estimated by comparing the binding affinity of taxadiene in the active site that includes heme iron in its activated state to its affinity to the oxyferyl heme-containing active site. Also, as AUTODOCK typically outputs a series of closely related poses from a single input, it was executed iteratively using the most energetically favorable output of one run as the input for the program's next implementation, until an unchanging set of substrate poses was obtained (Figure 4.5).
Figure 4.5. Substrate docking: As AUTODOCK outputs a series of the most energetically-favorable poses (which is significantly different from the series of likely poses), the exhaustive list of ≤ 9 poses is first culled to a smaller set of 2-3 poses by using an iterative algorithm that recasts the docking exercise as an unconstrained optimization. The ‘optimization’ loop is terminated when 3 successive iterations output a common set of poses with the same binding affinity.

Accuracy of the docking program was assessed by evaluating the closeness of fit of the computationally predicted pose of androstenedione in the active site of human aromatase cytochrome P450 (CYP19A1) with the crystal structure of their association.100

4.5.2.2. Using the diffusion of oxygen into the active site as a constraint

The diffusional trajectory of oxygen into the activated heme-containing active site greatly limits the translational and rotational degrees of freedom the now-destabilized substrate has in order that it may assume a more stable position. This constraint was utilized to discriminate between the
numerous poses outputted by AUTODOCK in order to identify the single likeliest pose of taxadiene in both the activated and oxyferyl heme-containing active sites (Figure 4.6). The likeliest path that oxygen takes to diffuse into the active site was calculated using CAVER$^{101}$.

![Figure 4.6](image)

**Figure 4.6. Diffusion of oxygen as a constraint:** The modeling methodology identifies two possible reaction trajectories for the oxidation of taxadiene by the P450 hydroxylase. The first panel for both schemes represents binding of taxadiene to the activated heme-containing active site. The diffusion path of oxygen is highlighted in red in the second panel, and the reactive conformation that leads to substrate oxidation is depicted in the third panel. Both trajectories, however, predict competition between abstractions of the same 3 protons by the oxyferyl heme iron. These are, in their order of favorability, the protons attached to 3-, 13α- and 20-positions.

This program re-casts the atom centers in the enzyme model as a Voronoi diagram, whose edges and vertices are then converted to a weighted graph. The Delaunay triangulation is later applied to the weighted graph and the likeliest path of oxygen diffusion is calculated using a modified Dijkstra's algorithm.

### 4.5.2.3 Method validation

The computational methodology for studying P450-mediated substrate oxidation was validated by predicting the reaction trajectories for the dihydroxylation of vitamin D$_3$ by CYP105A1 (Figure
4.7). Vitamin D$_3$ is first hydroxylated at either the 1$\alpha$- or 20-position, following which, the monohydroxylated product – 1$\alpha$-hydroxy or 20-hydroxy vitamin D$_3$ – undergoes a second hydroxylation to form 1$\alpha$,20-dihydroxy vitamin D$_3$. The modeling methodology’s outputs compare favorably to the crystallographically determined substrate poses in both, the activated heme and oxyferyl heme active sites of CYP105A1 (PDB codes: 2ZBZ and 3CV9).

The 5-10% difference between the binding affinities of the substrate to the activated heme and oxyferyl heme states of the P450 was also identified as an important constraint to identify likely poses from the other outputs of the docking program.

Figure 4.7. Method validation via assessing docking of Vitamin D$_3$ into active site of CYP105A1: The binding position of vitamin D$_3$ within the active site of CYP105A1 has been determined using X-ray crystallography for both, its activated and oxyferyl heme forms. Model validation proceeded by comparing its predictions of the reaction trajectories with the crystallographic structures of CYP105A1 bound to its substrate.
4.5.2.4. Results of the reactivity landscape investigations

The enzyme that has been annotated as taxadiene 5α-hydroxylase in the sequences derived from the *Taxus* cDNA library is re-annotated as H16 in the computational assessment described herein. It is evident that taxadiene 5α-hydroxylase is a very promiscuous enzyme (Figure 4.8).

Interestingly, none of the other 16 P450s exhibit 5α-hydroxylase activity. H6 is predicted to exhibit 10β-hydroxylase activity, and the computational methodology corroborates its annotation in GenBank. H13, on the other hand, appears to catalyze stereo- and regiospecific 7β-hydroxylation of taxadiene but is annotated as a 10β-hydroxylase in GenBank. This dissonance merits further investigation but has not been pursued in this study.

H9, on the other hand, is predicted to be a taxadiene-5α-ol 10β-hydroxylase, and, H4 is predicted to catalyze the 2α-hydroxylation of taxadiene-5α,10β,13α-triol. Additionally, the size of the active site (in Å³) does necessarily correlate with promiscuity. For instance, H9, which is estimated to possess the second largest active site among the 16 P450 monooxygenases does not exhibit any activity (as predicted by our computational methodology) to substrates bearing substitutions at more than a single carbon atom on the taxane scaffold.
Figure 4.8. Taxane biosynthetic reactivity landscape: All 16 P450 enzymes from the *Taxus* cDNA library were included in the analysis. The enzymes were designated as H1-H16. The reactivity of an enzyme towards a substrate is affirmed using the 5-10% difference in binding affinities between the activated and oxyferyl states of the enzymes, as was previously established for the dihydroxylation of vitamin D$_3$ by CYP105A1.

### 4.6. Implications of the computational investigations on metabolic engineering

An important consideration in pharmacophore design is its potential use as an input for conventional target-oriented synthesis (TOS)$^{102}$, an exercise that aims to access precise regions of chemical space – either a single molecule or a small assortment of molecules exhibiting minor variations on a chemical theme. Natural product biosynthetic pathways typically consist of a large
number of reactions that are controlled by fairly complex regulatory mechanisms. Manipulating the metabolism of simple microorganisms to accommodate the expression of every single enzyme of the paclitaxel biosynthetic pathway is a prohibitively long process and is perhaps infeasible owing to the physiological stress that the expression of such a large set of enzymes might induce within the host. Instead, synthesizing an advanced intermediate that acts as a gateway molecule for TOS via the expression of a vastly smaller set of enzymes takes advantage of the core competencies of both, metabolic engineering and synthetic chemistry.

Based on our computational analysis and bio- and retrosynthetic planning for the production of paclitaxel (Figure 4.9), we have identified (1,2α,5α,7β,10β)-5-acetyloxy-1,2,7,10-tetrahydroxy-tax-4,11-diene as a potential product of microbial metabolic engineering. Our analysis suggests that expressing H16, H9 and H6 in a strain expressing taxadiene should, in theory, yield taxadiene-2α,5α,10β-triol. Of these enzymes, H16 might have to be significantly re-engineered in order to minimize its promiscuity towards taxanes other than taxadiene.

H13, which catalyzes 7β-hydroxylation of taxadiene, and H4, which hydroxylates taxadiene-5α,10β,13α-triol at the 1-carbon, will also need to be sufficiently re-engineered to alter their substrate preference to taxadiene-2α,5α,10β-triol and its 7β- or 1-hydroxy derivative. Target yields and productivities to ensure cost-competitiveness of a microbial-based process manufacturing (1,2α,5α,7β,10β)-5-acetyloxy-1,2,7,10-tetrahydroxy-tax-4,11-diene have been calculated in Chapter 6.
Figure 4.9. TOS of paclitaxel: The synthesis of paclitaxel in our scheme commences from (1,2a,5a,7β,10β)-5-acetyloxy-1,2,7,10-tetrahydroxy-tax-4,11-diene (▲). The synthesis scheme comprises of 22 steps and the yield of paclitaxel is 1.9%. Key: a. CH₂Cl₂ (yield 99%); b. Et₂N, 2-acetoxyacetyl chloride, CH₂Cl₂ (65%); c. (NH₄)₂Ce(NO₃)₆, CH₃CN (80%); d. MeOH, Na₂CO₃ (82%); e. TESCl, pyridine (89%); f. 4-DMAP, Et₃N, benzoyl chloride, CH₂Cl₂ (82%); g. CDI, NaH, DMF (81%); h. TESCl, pyridine (85%); i. H₂O₂, CH₃CN, PCC (97%); j. dioxane, MeOH, KOH (99%); k. quinuclidine, NMO, tBuOH, NaBH₄, OsO₄, THF (64%); l. 4-DMAP, CH₃COCl, CH₂Cl₂ (79%); m. 4-DMAP, MsCl, CH₂Cl₂ (84%); n. K₂CO₃, MeOH (80%); o. HMPA, DIET (77%); p. 4-DMAP, (CH₃CO)₂O, CH₂Cl₂ (70%); q. PhLi, CH₃COOH, THF, cyclohexane (70%); r. tBuOK, (PhSeO)₂O, THF (90%); s. 4-DMAP, (CH₃CO)₂O, pyridine (59%); t. CH₃COONa, PCC, NaBH₄, benzene, MeOH (81%); u. NaH, THF (68%); v. MeOH, 32% HCl (90%). SciFinder SciPlanner, Reaxys, ARChem and published literature¹⁰⁻¹¹³ were referenced to conceive this original synthetic scheme.
As was presented in Chapter 4, H16, or taxadien-5α-hydroxylase, is a supposedly a highly promiscuous cytochrome P450 monooxygenase, and, as set out in that chapter, the reactivity of this enzyme would have to be altered in order to minimize flux dissipation away from the eventual product of interest, \((1,2\alpha,5\alpha,7\beta,10\beta)-5\text{-acetyloxy-1,2,7,10-tetrahydroxy-tax-4,11-diene}\). Traditional enzyme engineering methodologies such as directed evolution and site-directed mutagenesis each have their merits. However, as P450 homologs share \(-15\%\) amino acid sequence identity and \(-30\%\) sequence similarity with one another\(^{114}\), the use of structure-guided approaches is preferred since the insights gained from redesigning one active site could easily be extended to manipulating the reactive proclivities of another P450, drastically reducing strain development times for the production of secondary metabolites. This is particularly true for the semi-synthetic manufacturing scheme proposed in Figure 4.9, which bears a preponderance of P450-catalyzed reactions in the biosynthetic stage. Approaches that are more rapid and more economical than conventional techniques such as X-ray crystallography and NMR, though, are required.
In this chapter, we build on the modeling methodology prescribed in Chapter 4 to formulate a robust and generalizable methodology to assess the mechanistic basis for the promiscuity and lack of selectivity of plant P450 enzymes. The methodology populates one’s understanding of the biocatalytic mechanism using well-established experimental protocols such as phylogeny-driven and site-directed mutagenesis, and computational techniques such as homology modeling and substrate docking. The insights that are gained from this exercise can be applied to chart the reactive topography of the active site and identify the impact of individual structural features on stability, activity and selectivity, potentially opening up possibilities to specifically tune the reactivity of plant P450s. Such tunable production of oxidized polycyclic compounds using microbial metabolic engineering represents a significant step towards realizing more exhaustive prospection and sustainable production of natural products and their analogues for drug discovery and development.

5.1. Background about P450 monooxygenases

P450 monooxygenases constitute a large family of highly conserved, heme-thiolate enzymes that are present in all forms of life. They usually act in concert with a flavoprotein redox partner – a cytochrome P450 reductase – that shuttles electrons and protons derived from NADPH to the oxygenase’s active site, whereat molecular oxygen is protonated and heterolytically cleaved to form water and a reactive iron-oxygen species that then inserts the second oxygen atom into the substrate (Figure 4.3). In mammals, mitochondrial P450s play vital roles in the steroid biosynthetic pathway and also orchestrate the breakdown of drugs and xenobiotics. In plants, microsomal P450s are an
integral component of the secondary metabolic apparatus that plays vital roles in communication, defense, development and reproduction.

5.2. Materials & methods

5.2.1. Strains & plasmids

Our research group had previously engineered E. coli K12 MG1655 ΔareA ΔendA to synthesize taxadiene at titers of about 300 mg/L and 1 g/L in 2 mL batch cultures and 1 L fed-batch cultures, respectively. In this strain, additional copies of the rate-limiting steps of the non-mevalonate pathway dxx, ispD, ispF and idi were integrated into the chromosome of the aforementioned E. coli strain using the FRT system under the control of the Trc promoter, and, the taxadiene biosynthetic apparatus, which consists of ggps (GGPP synthase) and ttx (taxadiene synthase) was expressed on a 5-copy, pACYCDUET-1 vector under the control of a T7 promoter. For production of the hydroxylated taxanes, the taxadiene biosynthetic apparatus, along with the T7 promoter that controls its expression, was also integrated onto the chromosome of the E. coli strain and the operon was localized to the arabinose operon with a kanamycin antibiotic marker. This strain was then electro-transformed to express the P450 apparatus consisting of P450 monooxygenase and its redox partner, also on a 5-copy, pACYCDUET-1 vector (Figure 5.1). Plasmid construction and sub-cloning was performed using chemically competent DH5α cells (purchased from Invitrogen). XL10-Gold ultracompetent cells (purchased from Agilent) were used to select the P450 mutants.
5. 2. 2. Culturing conditions

Biosynthesis of taxadien-5α-ol by the engineered *E. coli* strain was investigated in 2 mL batch cultures. Pre-cultures propagated in 20 μg/mL chloramphenicol-supplemented LB medium were used to inoculate the 2 mL production cultures at a starting OD of 0.1. The production medium comprised of 5 g/L yeast extract, 20 μg/mL chloramphenicol, 0.66 g/L δ-aminolevulinic acid, 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 (CH₃COO)₂Zn, 0.06 g/L Fe(III) citrate, 0.0045 g/L thiamine, 1.3 g/L MgSO₄, 10 g/L glycerol and 0.024 g/L IPTG. The starting pH was set to 7.0 and 20 vol.% dodecane was added to the production medium to scavenge the nascent taxane. The cultures were agitated at 200 rpm at 22°C and production was ceased after 5 days by extracting the product into 1 mL of a 80-20 vol.%
solution of hexane and diethyl ether. The organic phase was later separated and analyzed for its taxane content using GC-MS.

5.2.3. GC-MS analysis of hydroxylated taxanes

1 mL of the batch cultures was combined with an equal volume of a 80-20 vol.% hexane-diethyl ether solution and vortexed for 30-45 minutes. The vortexed mixture was then centrifuged at 3000 rcf at room temperature. The organic layer so-formed was then analyzed in a Varian Saturn 3800 GC 2000 MS system housing an Agilent HP-5ms column (30 m length, 0.25 mm ID, 0.25 μm film thickness). Ultrapure helium was used as the carrier gas (flow rate of 1 mL/min). Each analysis lasted 30 minutes, wherein the oven temperature was maintained at 50°C for 1 minute, ramped up at a rate of 10°C/min until it reach 220°C, and then held at 220°C for 12 minutes. The injector and transfer line temperatures were maintained at 200°C and 250°C respectively.

5.2.4. Protein engineering for optimal expression of taxadiene 5α-hydroxylase in E. coli

The amino-terminal transmembrane binding regions of the P450 monooxygenase (GenBank AY289209) and its redox partner (GenBank AY571340) were eliminated in order to favor expression in E. coli. 24 and 74 amino acids were excised from the P450 and reductase, respectively. An 8-amino acid (MALLIAVF) leader sequence of a modified15 bovine microsomal steroid 17α-hydroxylase (GenBank NM_174304) was subsequently appended to the truncated P450 to
improve expression and stability. The chimeric P450 and the truncated reductase were then fused together using a 5-amino acid (GSTGS) linker\(^\text{116}\) (Figure 5.2).

Figure 5.2. Overview of the strain and enzyme engineering protocols used in the study

5.2.5. Site-directed mutagenesis

The Quick Change II Site-Directed Mutagenesis kit was used to mutagenize taxadiene 5α-hydroxylase in the P450-reductase complex. The mutants were constructed using a protocol recommended by the kit manufacturer, Agilent, and the mutagenesis primers were also designed using a software package hosted on the manufacturer’s website.

(http://www.genomics.agilent.com/primerDesignProgram.jsp).
5. 3. Results

5. 3. 1. Taxadiene 5α-hydroxylase is a promiscuous enzyme

As predicted by the modeling methodology outlined in Chapter 4, instead of detecting a single peak that corresponds to taxadien-5α-ol, 15 species are consistently identified in the gas chromatograms of the samples upon expression of taxadiene 5α-hydroxylase (Figure 5.3), lending credence to the hypothesis that P450s that catalyze transformations in the early stages of plant secondary metabolic pathways are promiscuous. In the case of taxadiene 5α-hydroxylase, it appears that the enzyme is also very poorly selective. The taxadiene (species B) content in the samples is negligible, suggesting that it is completely converted to downstream products.

Intriguingly, taxadien-5α-ol (K) is not the major product (Figure 5.4). That distinction goes to the cyclic ether, 5(12)-oxa-3(11)-cycloptaxane (OCT, I), a structural isomer of taxadien-5α-ol. Mass spectral analysis of the remaining 13 products confirms that they are eicosanoids, 7 of which are monooxygenated (species A, C, G, H, J, K, M and P), 3 are dioxygenated (L, N and O) and 3 are unsubstituted (D, E and F). Species G is a structural isomer of OCT and it is evident that D, E and F are non-taxanes. However, the low abundance and instability of these molecules has hitherto precluded elucidation of their structures. It is uncertain how many of the remaining 9 species are taxanes.
Figure 5.3. Promiscuity & poor selectivity of taxadiene 5α-hydroxylase: (A) Typical gas chromatograms of the organic extracts from fermentation broths populated by strains expressing only the taxadiene synthesis module (black) and the taxadiene-P450 modules (red) reveal the presence of as many as 16 eicosanoids (labeled A-P), all of which elute from the column after approximately 20 minutes of the analysis has been completed. The katharometric signal strengths of the species (y-axis) have been expressed in kilocounts and the retention times have been reported in minutes (x-axis). (B) Fragments with the highest m/z ratio were normalized to an intensity of 100% in the mass spectral plots of the 16 eicosanoids (spectral plots in black - non-hydroxylated C_{20}s, red – monooxygenated C_{20}s, and blue – dioxygenated C_{20}s).
Figure 5.4. Characterized structures of some of the products: OCT, a cyclic ether, is the major product and elutes a little after 21 minutes in the 30-minute GC-MS run. Taxadiene elutes after roughly 20 minutes and the less-volatile taxadien-5α-ol is detected close to 21.5 minutes. The katharometric signal strengths of the species (y-axis) have been expressed in kilocounts and the retention times have been reported in minutes (x-axis).

5.3.2. Temporal profile of taxadiene proto-oxidation

The monooxygenase catalytic cycle proceeds under tight thermodynamic control, wherein molecular oxygen is reductively cleaved to form water and the oxidized substrate. In some instances, the catalytic cycle is shunted and produces hydrogen peroxide instead of water. *In situ* substrate dihydroxylation, therefore, is neither mechanistically nor thermodynamically feasible for P450 monooxygenases such as taxadiene 5α-hydroxylase.

The re-diffusion of a nascent monooxygenated product into the P450 active site, followed by its participation in a second oxidation cycle appears to be the only likely explanation for formation of the dioxygenated eicosanoids. Moreover, not only is this interpretation in agreement with the
hypothesis regarding P450 promiscuity, but adding polymeric bead-conjugated silyl-based scavengers similar to the packing material of reversed-phase HPLC columns to the culture medium greatly reduces eicosanoid production, more so the production of the dioxygenated compounds\textsuperscript{117}.

Accordingly, daily production of the 16 eicosanoids was tracked (Figure 5.5) in order to derive qualitative temporal relationships between the species. The absence of taxadiene on all days of the production run points to high activity of the P450. Of the remaining 15 species, only G, H, OCT, taxadien-5\(\alpha\)-ol, M and N are detected after 1 day of culturing. The remaining eicosanoids accumulate only after 2 days. These trends possibly suggest that G, H, OCT, taxadien-5\(\alpha\)-ol and M are oxidized derivatives of taxadiene. The concentrations of D and C exhibit similar trends to that of H. Species A, C, P and J, therefore, appear to be oxidized derivatives of D, while C could be the source of L and O.

Interestingly, assessing the rates of change in the concentration of all species clusters the species into 3 groups - one comprising of G, OCT, J, taxadien-5\(\alpha\)-ol, L, P and O; another which includes A, E, F, M and N; and a third group consisting of C, D and H (Figure 5.5.). The rate of change in the concentrations of a species is simply the sum of the rate at which it is formed, the rate at which it is consumed by a downstream metabolic reaction, the rate at which it inhibits host growth or the activity of the enzyme that synthesizes it, and its volatilization rate. The similarity in the variation in rates of change in the concentrations between species from one group but the difference in the variation in rates between species from different groups provides a qualitative snapshot of their
propensities to participate in another P450 catalytic cycle, as well as their volatilities and inhibitory effects.

![Daily production titers for all eicosanoids](image)

**Figure 5.5. Daily variation in eicosanoid production titers:** Concentrations of all 16 eicosanoids (in mg/L) were measured on each day of the 5-day runs. The concentrations were calculated from the katharometric signal strengths using a calibration curve for taxadiene. Error bars represent standard deviations (n = 3). Of the 16 species, taxadiene (species B) is undetectable on all 5 days and only G, H, J, K, M and N can be detected after 24 hours of the run have been completed.
Figure 5.6. Assessment of the daily variation in the rate of change in the concentrations of the 16 eicosanoids:

Daily variations in the forward derivative of the concentrations of the species with respect to time (in mM/day) have been plotted for 15 eicosanoid species – A, C-P. Trends suggest that the compounds fall into three groups. G, OCT (I), J, taxadien-5α-ol (K), L, P and O comprise one group. The second group includes A, E, F, M and N. C, D and H form a third group. Species from two groups do not appear to react further, but that members of the first group are more volatile and/or exert a greater inhibitory effect than those of the latter. The variation in the rate of the change of C and H suggests that they are likeliest sources of the dioxygenated molecules.

The variation in their rates of change suggests that species from the first two groups do not appear to react further, but that members of the first group are more volatile and/or exert a greater inhibitory effect than those of the latter. The variation in the rate of the change of D suggests that it reacts with the P450 to yield several oxygenated non-taxanes, whereas C and H appears to be the likeliest sources of the 3 dioxygenated molecules (Figure 5.7). The formation of D, E and F is particularly intriguing. We speculate that these molecules are formed in response to the concerted effect of inhibition by the oxygenated taxanes and a high flux through the prenyl pyrophosphate biosynthetic pathway.
5. 3. 3. The utility of traditional mutagenic practices to improve P450 reactivity

5. 3. 3. 1. Phylogeny-guided mutagenesis of taxadiene 5α-hydroxylase

As mentioned in Section 4.1, the *Taxus* cDNA library from which the sequence of taxadiene 5α-hydroxylase was obtained contains an additional 15 P450 monooxygenases. It was reasoned that replacing amino acid residues within the active site of taxadiene 5α-hydroxylase with similarly positioned residues in its paralogs could perhaps shed light on the reactive contributions of each residue. Consequently, all amino acids located at the periphery of the active site cavity predicted from the modeling methodology were replaced using a multiple sequence alignment of all P450 sequences in the *Taxus* cDNA library that was then utilized to identify substitutions for the selected amino acids (Figure 5.8).
Figure 5.8. Targets for phylogeny-guided mutagenesis: Amino acid residues that fall within 3.5 Å of the periphery of the active site cavity of taxadiene 5α-hydroxylase are compared to its homologs in this multiple sequence alignment. Amino acids located at the periphery of the active site are labeled with a red circle. Numbers below the red circles denote the number of mutations investigated at that site. The multiple sequence alignment was created in WebLogo, a web-based service. The use of bits instead of probabilities is a feature of the program. 3.5 bits corresponds to a probability of 1, suggesting complete conservation across all the homologs.

Many phylogeny-derived mutants exhibit vastly improved activities compared to the parental P450 (Figure 5.9). The activity is calculated by summing the molarities of all oxygenated eicosanoids produced by the enzyme. Selectivity, on the other hand, is simply the ratio between the molarity of taxadien-5α-ol and the activity of the enzyme.
Figure 5.9. Activity and selectivity comparisons between the mutant and parental P450: Percent changes in the selectivities (gray) and activities (black) of the 32 mutants are compared to those of the parental enzyme. Selectivity is defined as the ratio of the concentration of taxadien-5α-ol to all hydroxylated eicosanoids. Activity is defined as the sum of the concentrations of all hydroxylated eicosanoids. Concentrations have been converted to nM. Many phylogeny-derived mutants exhibit vastly improved activities and selectivities compared to the parental P450.

Amino acid mutations that improve enzyme selectivity can be distributed into 3 categories. If one assumes that product formation depends on the regiospecificity of the proton that is abstracted by the oxyferyl species of the heme group in the P450 active site, competing proton abstractions could explain the distribution of products formed by a single catalytic cycle. Consequently, the first class amino acid mutations include those that improve product selectivity by directly affecting the probability of regiospecific proton abstraction viz. improvements in ‘reaction selectivity’. Product
selectivity can also be improved by disfavoring the association between the enzyme and its other substrates, \textit{viz.} controlling the promiscuity of the enzyme. Mutations that achieve this result fall into the second category – improvements in ‘overall selectivity’. The third group includes amino acid mutations that affect both, the probability of regiospecific proton abstraction and enzyme promiscuity.

\[
\begin{align*}
S_A & \xrightarrow{E} P_{A1} + P_{A2} + P_{A3} \\
S_B & \xrightarrow{E} P_{B1} + P_{B2}
\end{align*}
\]

Desired product

Activity = $[P_{A1}] + [P_{A2}] + [P_{A3}] + [P_{B1}] + [P_{B2}]$

Overall selectivity = $\frac{[P_{A1}]}{[P_{A1}] + [P_{A2}] + [P_{A3}] + [P_{B1}] + [P_{B2}]}$

Reaction selectivity = $\frac{[P_{A1}]}{[P_{A1}] + [P_{A2}] + [P_{A3}]}$

Figure 5.10. Differentiating the two types of selectivity improvements

Apparently, nearly all phylogeny-guided mutations investigated herein fall in the second category (Figures 5.10 and 5.11). The product distribution of T377S is particularly surprising as it appears to favor oxygenation of the non-taxane, C, to produce the dioxygenated eicosanoid, O.

5. 3. 3. 2. Mutations to taxadiene 5α-hydroxylase based on structural inferences

Of the amino acid mutations exhibiting improvements in activity and/or selectivity, K131E and
D309E were investigated in further detail. Not only are these residues highly conserved, but K131 and D309 form a dyad that extends laterally across the roof of the active site, parallel to the heme moiety below. It was hypothesized that modifications to these residues could potentially affect the orientation in which taxadiene and the other unsubstituted eicosanoids bind within the active site (Figure 5.12).

![Figure 5.11](image-url)

**Figure 5.11. Grouping the selectivity improvements**: The fold change of the concentration of a particular eicosanoid is simply the ratio of the concentration of the species produced by the mutant to that produced by the parental enzyme. Significant variations between the fold changes of products of the same catalytic cycle is indicative of direct manipulations to the catalytic cycle *viz.* improvements in reaction selectivities. D309 is the only mutation that exhibits quantifiable differences in the selectivity towards taxadien-5α-ol (K) that can be attributed to possible differences in the
probability of abstracting a specific proton (species G, I, K, which potentially are products of the same catalytic cycle, are colored red).

Accordingly, a series of mutants combining site-saturation mutagenesis at both locations of the dyad and other phylogeny-guided mutants were constructed and evaluated for their influence on the activity and selectivity of taxadiene 5α-hydroxylase (Figure 5.13). The number of mutations varied from a single-site mutation to as many as 6 simultaneous mutations. The mutation frequency seemed to have a deleterious effect on enzyme function. The production of oxygenated compounds was all but abolished for enzymes bearing ≥4 mutations in their amino acid sequence.

Figure 5.12. The lysine-aspartate dyad: The high degree of conservation of these residues, their functional groups and their relative sizes were instrumental in their selection as mutation targets.

A double-mutant combining substitutions to K131 and D309 with arginine exhibited the greatest improvement in selectivity. However, as before, the increase in selectivity is attributable to disfavoring the formation of the mono- and dioxygenated eicosanoids (Figure 5.14). It is evident that phylogeny-guided and site-directed mutageneses do not alter any proton abstraction
probabilities in the catalytic cycle that yields, among others, taxadien-5α-ol. A double-mutant combining substitutions to K131 and D309 with arginine exhibited the greatest improvement in selectivity.

Figure 5.13. Influence of mutations to the dyad and amino acids at the periphery of the active site: Selectivity is defined as the ratio of the concentration of taxadien-5α-ol to all hydroxylated eicosanoids. Activity is defined as the sum of the concentrations of all hydroxylated eicosanoids. Concentrations have been converted to nM. Percent changes in the selectivities (gray) and activities (black) of the 32 mutants are compared to those of the parental enzyme. The phylogeny-guided mutants were single-site mutants. Single- and multi-site mutations were investigated in the K131-D309 study. Increasing the number of mutations appears to have a deleterious effect on enzyme function. Enzyme constructs bearing ≥4 mutations formed very little product.
However, as before, the increase in selectivity is attributable to disfavoring the formation of the mono- and dioxygenated eicosanoids (Figure 5.14). It is evident that phylogeny-guided and site-directed mutageneses do not alter any proton abstraction probabilities in the catalytic cycle that yields, among others, taxadien-5α-ol.

Figure 5.14. Improving reaction selectivity is a non-trivial challenge: Like in Figure 5.11, the fold change in the concentration of particular species simply refers to the ratio of its concentration as produced by a given mutant to the concentration obtained by expressing the parental enzyme. None of the mutants constructed in the study thus far improve product selectivity by directly affecting the probability of regiospecific proton abstraction. Instead, selectivity improvements have been achieved by disfavoring the association between the P450 and its other substrates.

5. 3. 4. Computational assessment of the taxadiene 5α-hydroxylase catalytic cycle

The preponderance of amino acid mutations that improve product selectivity by destabilizing binding of competing substrates over mutations that alter the orientation of taxadiene within the
active site germinated computational investigations to understand the mechanistic basis for the enzyme's peculiar product distribution. The computational methodology discussed in Section 4.5.2 was utilized for this purpose.

Oxidation of the substrate typically proceeds via the oxygen rebound mechanism (Figure 5.15)\textsuperscript{18}. Proton abstraction by the oxygen atom of oxyferyl heme forms a radical intermediate and a hydroxy-iron species, the latter which then dissociates via a rebound mechanism to form the carbon-oxygen bond. The heme group simultaneously returns to its pre-reaction, activated state, and, radical stabilization, where possible, determines the abundance of the oxygenated products.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.15}
\caption{The oxygen rebound mechanism: The oxygen rebound mechanism involves the formation of a hydroxy-iron species, which then dissociates to insert oxygen into the substrate. Bond-breaking and bond-forming electron migrations are depicted in red and blue, respectively.}
\end{figure}

Analysis of the average position of the two likeliest poses of taxadiene in the oxyferyl state of taxadiene 5α-hydroxylase suggests that proton abstractions occur at as many as 3 positions on the taxane scaffold (Figure 5.16). Abstraction of any one of the three protons attached at the 20-position produces allyl-stabilized radical species at the 20- and 5-positions (Figure 5.17.), which
then yield taxadien-20-ol (4) and taxadien-5α-ol (5), respectively. Taxadien-5α-ol is more abundant than taxadien-20-ol as the 5-position secondary carbon radical is the more stable of the two allyl-stabilized radical intermediates.

In comparison, abstraction of the sole proton that is attached at the 3-position produces a tertiary carbon radical. However, as insertion of oxygen at this position appears to be sterically hindered, the lone electron shifts to the tertiary carbon at the 12-position. This transfer does not mitigate the steric hindrance to the oxygen rebound mechanism either. Instead, we speculate that the close proximity of the alkenyl moiety to the hydroxy-iron species and the tertiary carbon radical facilitates a concerted sequence of electron exchanges that results in the addition of the oxygen atom to the carbons at the 5- and 12-positions to form OCT (8). This explanation for ether formation is different from that of a previous study on the OCT biosynthesis\(^\text{19}\). However, oxidation of taxadiene in the reaction mechanism detailed therein does not proceed via the oxygen rebound mechanism and neither can it explain formation of, among others, taxadien-5α-ol and OCT in the same active site.
Figure 5.17. Hypothesized reaction mechanism: *In silico* models of the binding of taxadiene in the active site of taxadiene 5α-hydroxylase predict competition between the abstraction of protons from the (a) 20-, (b) 3- and (c) 13α-positions. Radical species formed via proton abstractions from the 20- and 13α-positions are allyl-stabilized. Insertion of oxygen to the 3- and 12-positions is sterically inhibited. Formation of OCT following proton abstraction from the 3-position is hypothesized to proceed via the constrained Markovnikov-like addition of oxygen to the 5- and 12-positions.
Lastly, proton abstraction from the carbon at the 13α-position produces taxadiene-13α-ol (12) and yet another cyclic ether species, 5(13)-oxa-3(11)-cyclotaxane (13). Preliminary NMR characterization of the oxygenated eicosanoids appears to support this hypothesis and has also identified species G to be 5(13)-oxa-3(11)-cyclotaxane\textsuperscript{120}.

5.3.5. Computational assessment of taxadiene 5α-hydroxylase mutants

We applied our computational methodology to investigate the higher incidence of phylogeny-guided and dyad mutants that improve product selectivity by destabilizing binding of competing substrates over mutants that alter the reaction trajectory of taxadiene oxidation (Figure 5.18).

Of all the mutants that were constructed previously, K131R, S302A and V374L exhibited deducible differences in the selectivity towards 5(13)-oxa-3(11)-cyclotaxane (G), H, OCT (I), taxadien-5α-ol (K) and M. Not only are the orientations of taxadiene in the active sites of the parental P450 enzyme and its 3 mutants visibly different, but the proton abstraction lengths predicted by the model correlate well with the differences observed in the relative proportion of the monooxygenated taxadiene derivatives produced by each enzyme. To a degree of approximation, the abstraction lengths and observed product distributions correlate quite well, underscoring the previous suspicion of the limitations of traditional mutagenic (or ‘structure-blind’) approaches to truly identify mutants that alter the reaction selectivity of the enzyme.
Figure 5.18. Comparing mutation efficacy using the computational methodology: The computational methodology successfully predicts variations in product selectivities of the 3 mutants that were investigated. Electron densities of the average structures of the two likeliest poses predicted by the model (represented as dot models) exhibit visible differences between the 4 enzymes.

5.4. Conclusions & perspectives

5.4.1. Comparing the methodologies

Phylogeny-guided and site-directed mutageneses are efficient methodologies to improve the throughput of the enzyme. However, the selectivity improvements that are achieved through the use of these techniques often result from destabilizing binding of competing substrates rather than
improving the selectivity of the catalytic mechanism. In the case of promiscuous and notoriously non-selective nodal enzymes such as taxadiene 5α-hydroxylase, the benefit of being able to rapidly probe enzyme reactivity using phylogeny-guided mutagenesis is soon lost over the inability of the methodology to identify desirable amino acid substitutions without having to construct an exhaustive collective of enzyme mutants. In comparison, the computational methodology we have prescribed in Chapter 4 can conclusively identify the mechanistic basis for non-selectivity. It is robust, generalizable and manifold more rapid than other structure-guided enzyme engineering approaches. Not only can the insights that it provides be applied to chart the reactive topography of the active site and identify the impact of individual structural features on stability, activity and selectivity, one can even possibly extend the methodology to prescribe mutational targets within the active site. For this to occur, though, NMR characterization of all interacting substrates and their oxidation products is required. The algorithms utilized in the methodology can then multiplicatively compare the binding affinities and orientations of all characterized substrates within the active site to guide the construction of mutants that favor only specific orientations of desired substrates (Figure 5.19).

5. 4. 2. The possibility of multiple substrates in a single active site

Formulation of the computational methodology assumes that only a single taxadiene substrate docks into the active site of taxadiene 5α-hydroxylase. While this assumption is well-grounded in literature,
questions still remain as this enzyme is somewhat of a novelty compared to some of the P450 monooxygenases studied in literature.

\[ \text{Desired product} \]

\[ \begin{align*}
S_A & \rightarrow E \rightarrow P_{A1} + P_{A2} + P_{A3} \\
S_B & \rightarrow E \rightarrow P_{B1} + P_{B2}
\end{align*} \]

Figure 5.19. A generalized algorithm for in silico enzyme engineering

The diffusion of two molar equivalents of taxadiene into the active site of the enzyme, if it were to occur, would not be constrained by space as the active site of taxadiene 5α-hydroxylase is roughly 474 Å³ (Table 4.1.). Evaluation of this hypothesis necessitates isolation and purification of taxadiene 5α-hydroxylase, followed by an in vitro assessment of the spectral shift that is induced by substrate binding. Comparing the spectral shift to P450 standards known to bind to single or multiple substrate molecules simultaneously would provide the much-needed constraint that is required to include or eliminate the possibility of multiple substrates binding simultaneously within the active site. As the computational methodology can account for the simultaneous binding of several substrates within the active site, this constraint could improve the predictions of the model.
5. 4. 3. Directed evolution versus in silico modeling

Given that most biological experimentation still proceeds with a 'black box' mentality and each assumption that one makes in the formulation of a utilizable model diminishes the closeness of fit of the model with the phenomenon it describes, there is no argument that directed evolution is the superior of the two approaches. However, the absence of a screen that can efficiently couple the selectivity of the enzyme to gene expression or platforms such as phage-display or the yeast two-component bait-prey system preclude the use of directed evolution\textsuperscript{121}. The most realistic demonstration of directed evolution of a selective taxane P450 hydroxylase would utilize phage-display to assess the reactivity of the enzyme in vitro (and possibly high-throughput) using taxane-binding antibodies. Paclitaxel-binding antibodies are presently available commercially and generating antibodies capable of binding to the hydroxylated products of taxadiene 5α-hydroxylase is not a far-fetched goal.

5. 4. 4. Implications for metabolic engineering

Consider a metabolic pathway comprising of 3 enzymes – A, B and C. Each enzyme catalyzes a substitution at a unique position on a chemical scaffold. If only enzymes A and C of the pathway are expressed in the microbial host to produce a unique, di-substituted product, it is vital that enzyme C be suitably re-engineered in order to ensure that it exhibits the same specificity and activity towards the mono-substituted substrate as it does towards the di-substituted substrate in the unabridged
pathway, failing which, conversion and throughput of the pathway will be low. If the enzymes are promiscuous, as is the case for the early taxane pathway, many more copies of a suitably re-engineered enzyme A would have to be expressed relative to the re-engineered enzyme C to minimize flux dissipation away from the pathway of interest. One can also maximize conversion to the product of interest by temporally inducing the expression of the respective enzymes *viz.* induce expression of enzyme A to produce the preferred intermediate and only induce the expression of C after sufficient time has been allowed for enzyme A to act. The permutations are numerous and serve to underscore the mutual dependence of enzyme engineering and metabolic engineering for the heterologous production of secondary plant metabolites and molecules that are variations of a new chemical theme.
One of the principal criticisms leveled at microbial metabolic engineering is the seeming inability of simple microbial hosts such as *E. coli* to support the elaborate enzymatic machinery that is required to synthesize molecules as complex as paclitaxel. However, as demonstrated in Chapter 5, *E. coli* is more than able to accommodate the expression of plant secondary metabolic enzymes. An *E. coli* strain expressing a mutant of the cytochrome P450 monooxygenase-reductase complex described earlier produced nearly 0.45 g/L of mono- and dioxygenated taxanes in 2 mL batch cultures, which, as shall be demonstrated herein, could be sufficient to ensure cost-competitive production of paclitaxel using the microorganism.

Instead, the biggest hurdle in realizing the production of (1,2α,5α,7β,10β)-5-acetyloxy-1,2,7,10-tetrahydroxy-tax-4,11-diene, which, in Chapter 4, was identified as a promising ‘gateway’ molecule for synthesis of paclitaxel and its taxane analogues, is the notorious promiscuity and poor selectivity of the pathway enzymes. Remedying these traits to minimize flux dissipation away from the molecule of interest is arguably the biggest challenge facing microbial metabolic engineering and protein engineering.
Presently, we do not possess a methodology that permits directed evolution of plant secondary metabolic enzymes and rational techniques for enzyme engineering such as the methodology presented in Chapter 5 are far from maturity. Nevertheless, using a technoeconomic comparison between the existing plant-based semi-synthetic process and the one that has been proposed in Figure 4.9, we identify target titers and productivities for a strain expression an optimal combination of P450 monooxygenases and acyltransferases. The titers calculated herein serve as benchmarks for future protein and metabolic engineering activities.

6.1. Background & assumptions

Paclitaxel is currently produced via a 4-step synthetic scheme from 10-DAB that has been extracted from plant cell cultures of *T. baccata* (Figure 6.1). The 10-DAB yield is roughly 160 mg/g DCW\textsuperscript{122} and the paclitaxel yield from 10-DAB is 44.7\%\textsuperscript{111}, which translates to an overall semi-synthetic yield (assuming a plant water content of 80 wt.\%) of 2.25 wt.\%. This figure represents a >200-fold increase over the paclitaxel content in the bark needles of *T. baccata*, underscoring the importance of semi-synthesis in the production of therapeutic natural products that are otherwise too miniscule to be directly extracted from their native hosts.

In comparison, production of (1,2\(\alpha\),5\(\alpha\),7\(\beta\),10\(\beta\))-5-acetyloxy-1,2,7,10-tetrahydroxy-tax-4,11-diene (for simplicity, hereinafter this molecule will be abbreviated as ATHTAX) by the previously described taxadiene-producing *E. coli* strain necessitates the expression of 6 additional enzymes (5
cytochrome P450 monooxygenase-reductase complexes and a single acetyltransferase), all of whose identities are already established. Significantly, the use of chemically-inducible promoters for the expression of the taxane biosynthetic apparatus couples production with growth, thereby enabling continuous chemical synthesis through continuous cultivation.

Figure 6.1. Semi-synthesis of paclitaxel from plant-derived 10-deacetylbaccatin III (10-DAB): The yield of paclitaxel from 10-DAB is 44.7%. Key: a. CH$_2$Cl$_2$ (yield 99%); b. Et$_3$N, 2-acetoxyacetyl chloride, CH$_2$Cl$_2$ (65%); c. (NH$_4$)$_2$Ce(NO$_3$)$_6$, CH$_3$CN (80%); d. MeOH, Na$_2$CO$_3$ (82%); e. TESCl, pyridine (89%); f. 4-DMAP, Et$_3$N, benzoyl chloride, CH$_2$Cl$_2$ (82%); g. TESCl, pyridine (85%); h. CH$_3$COCl, pyridine (85%); i. NaH, THF (68%); j. 32% HCl, MeOH (90%)
The global paclitaxel supply chain is populated by little over a dozen API producers who supply the bulk drug to numerous formulators such as BMS or Mylan that then produce finished doses of the drug for distribution to various consumer channels. The API manufacturers either produce 10-DAB internally or source the material from external providers and typically produce 50-100 kg of paclitaxel annually. Accordingly, in the economic assessment that follows, the annual production volume for the manufacturing facility is also assumed to be 100 kg. As bulk chemical prices for many of the reagents, catalysts and solvents that are required by the two semi-synthetic schemes are not readily available, SciFinder Scholar was queried to identify the lowest price for all raw materials (Tables 6.1 & 6.2). The price of ATHTAX was estimated by comparing the economics of plant and microbial cultivation. The optical density of the taxadiene producing E. coli cultures reaches a peak of ~40 in 5-day, 1 L fed-batch cultures. Assuming an equivalency between an optical density of 1.0 and a dry biomass titer of 0.3 g/L, the overall yield of taxadiene on a dry cell weight basis is estimated to be 0.0925 g/g. A 30% reduction in pathway flux was factored for every additional enzyme that is expressed within the pathway, which translates to a dry basis yield of 0.016 g/g for ATHTAX. The doubling time of the engineered E. coli strains is estimated to be about 12 hours, which is nearly 12-fold lower than the doubling time of T. baccata suspension cultures.

6.2. Comparing the economics of the two competing approaches

Despite having a dry basis yield that is 10 times lower than that of 10-DAB, microbial production of ATHTAX is nearly 275-fold more productive than the plant cultures that produce 10-DAB.
Another important distinction between the two cultures is that the *E. coli* strains utilize glycerol as the principal carbon source whereas the plant cells are cultivated using a combination of sucrose, glucose and fructose as carbohydrate feedstocks. Although consumption of glycerol by the *E. coli* cultures is estimated to be roughly 4 times greater than the mono- and disaccharide demand by the suspension cultures of *T. baccata*, the media formulation costs for plant cell cultivation are greater owing to the higher unit price of the sugars as well as higher nutrition supplementation costs. Consequently, the productivity of the microbial cultures is conservatively estimated to be 300-fold greater than that of the plant suspension cultures, which amounts to each kg of ATHTAX costing about $36,600. The raw material costs for synthesizing paclitaxel from 10-DAB are, therefore, a little over 6% greater than the competing semi-synthetic process that commences from ATHTAX.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Amount</th>
<th>Unit cost</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzaldehyde</td>
<td>66 kg</td>
<td>$12/kg</td>
<td>$818</td>
</tr>
<tr>
<td>Anisidine</td>
<td>77 kg</td>
<td>$32/kg</td>
<td>$2423</td>
</tr>
<tr>
<td>Dichloromethane</td>
<td>10548 L</td>
<td>$8/L</td>
<td>$84553</td>
</tr>
<tr>
<td>Triethylamine</td>
<td>265 kg</td>
<td>$15/kg</td>
<td>$3975</td>
</tr>
<tr>
<td>2-Acetoxyacetylchloride</td>
<td>126 kg</td>
<td>$1950/kg</td>
<td>$244725</td>
</tr>
<tr>
<td>Cerium ammonium nitrate</td>
<td>657 kg</td>
<td>$95/kg</td>
<td>$62429</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>3560 L</td>
<td>$16/L</td>
<td>$56960</td>
</tr>
<tr>
<td>Methanol</td>
<td>13455 L</td>
<td>$7/L</td>
<td>$89747</td>
</tr>
<tr>
<td>Sodium carbonate</td>
<td>3 kg</td>
<td>$2/kg</td>
<td>$7</td>
</tr>
<tr>
<td>Triethylsilylchloride</td>
<td>832 kg</td>
<td>$440/kg</td>
<td>$366282</td>
</tr>
<tr>
<td>Pyridine</td>
<td>19489 L</td>
<td>$30/L</td>
<td>$578835</td>
</tr>
<tr>
<td>4-Dimethylaminopyridine</td>
<td>3 kg</td>
<td>$88/kg</td>
<td>$251</td>
</tr>
<tr>
<td>Benzoylchloride</td>
<td>44 kg</td>
<td>$13/kg</td>
<td>$578</td>
</tr>
<tr>
<td>10-Deacetylbaccatin III</td>
<td>143 kg</td>
<td>$10.98 x 10^6/kg</td>
<td>$1.56 x 10^9</td>
</tr>
<tr>
<td>Acetyl chloride</td>
<td>87 kg</td>
<td>$47/kg</td>
<td>$4064</td>
</tr>
<tr>
<td>Sodium hydride</td>
<td>13 kg</td>
<td>$101/kg</td>
<td>$1266</td>
</tr>
<tr>
<td>Tetrahydrofuran</td>
<td>2683 L</td>
<td>$10/kg</td>
<td>$27903</td>
</tr>
<tr>
<td>32% Hydrochloric acid</td>
<td>50 L</td>
<td>$1/L</td>
<td>$73</td>
</tr>
</tbody>
</table>

**Total raw material costs = $15.7 x 10^9/kg**
The capital costs of the two processes can be related using the six-tenths power law as follows\textsuperscript{126}:

\[
\frac{C_P}{C_M} = \left(\frac{V_P}{V_M}\right)^{0.6}
\]  

(6.1)

\(C_P\) and \(C_M\) are the capital costs for the plant (P) and microbial (M) processes, respectively. \(V_P\) and \(V_M\) are the total material flows associated with the two processes.

### Table 6.2. Raw material costs for reaction scheme in Figure 4.9.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Amount</th>
<th>Unit cost</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzaldehyde</td>
<td>66 kg</td>
<td>$12/kg</td>
<td>$815</td>
</tr>
<tr>
<td>Anisidine</td>
<td>76 kg</td>
<td>$2/kg</td>
<td>$524</td>
</tr>
<tr>
<td>Dichloromethane</td>
<td>1020451 L</td>
<td>$8/L</td>
<td>$8179934</td>
</tr>
<tr>
<td>Triethylamine</td>
<td>264 kg</td>
<td>$15/kg</td>
<td>$3956</td>
</tr>
<tr>
<td>2-Acetoxyacetylchloride</td>
<td>125 kg</td>
<td>$1950/kg</td>
<td>$244335</td>
</tr>
<tr>
<td>Cerium ammonium nitrate</td>
<td>656 kg</td>
<td>$95/kg</td>
<td>$62341</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>37500 L</td>
<td>$16/L</td>
<td>$600000</td>
</tr>
<tr>
<td>Methanol</td>
<td>338329 L</td>
<td>$7/L</td>
<td>$2256656</td>
</tr>
<tr>
<td>Sodium carbonate</td>
<td>3 kg</td>
<td>$2/kg</td>
<td>$6</td>
</tr>
<tr>
<td>Triethylsilylchloride</td>
<td>15023 kg</td>
<td>$440/kg</td>
<td>$6610208</td>
</tr>
<tr>
<td>Pyridine</td>
<td>398154 L</td>
<td>$30/L</td>
<td>$11825163</td>
</tr>
<tr>
<td>4-Dimethylaminopyridine</td>
<td>3065 kg</td>
<td>$88/kg</td>
<td>$269746</td>
</tr>
<tr>
<td>Benzoylchloride</td>
<td>44 kg</td>
<td>$13/kg</td>
<td>$579</td>
</tr>
<tr>
<td>(1,2α,5α,7β,10β)-5-acetoxyl-1,2,7,10-tetrahydroxy-tax-4,11-diene</td>
<td>2431 kg</td>
<td>$36,660/kg</td>
<td>$89,12 \times 10^6</td>
</tr>
<tr>
<td>Carbonyldiimidazole</td>
<td>25064 kg</td>
<td>$75/kg</td>
<td>$1879800</td>
</tr>
<tr>
<td>Sodium hydride</td>
<td>249 kg</td>
<td>$101/kg</td>
<td>$25124</td>
</tr>
<tr>
<td>Dimethylformamide</td>
<td>205438 L</td>
<td>$5/L</td>
<td>$1022054</td>
</tr>
<tr>
<td>Periodic acid</td>
<td>1012 kg</td>
<td>$238/kg</td>
<td>$240737</td>
</tr>
<tr>
<td>Pyridinium chlorochromate</td>
<td>4287 kg</td>
<td>$250/kg</td>
<td>$1071650</td>
</tr>
<tr>
<td>Dioxane</td>
<td>257258 L</td>
<td>$28/L</td>
<td>$7095176</td>
</tr>
<tr>
<td>Potassium hydroxide</td>
<td>1873 kg</td>
<td>$6/kg</td>
<td>$11236</td>
</tr>
<tr>
<td>Quinuclidine</td>
<td>1057 kg</td>
<td>$64000/kg</td>
<td>$67621000</td>
</tr>
<tr>
<td>N-Methylmorpholine-N-oxide</td>
<td>1675 kg</td>
<td>$75/kg</td>
<td>$125633</td>
</tr>
<tr>
<td>t-Butyl alcohol</td>
<td>2222783 L</td>
<td>$8/L</td>
<td>$17782262</td>
</tr>
<tr>
<td>Sodium borohydride</td>
<td>12249 kg</td>
<td>$276/kg</td>
<td>$3380818</td>
</tr>
<tr>
<td>Osmium tetroxide</td>
<td>606 kg</td>
<td>$65000/kg</td>
<td>$3938000</td>
</tr>
<tr>
<td>Tetrahydrofuran</td>
<td>3879647 L</td>
<td>$10/L</td>
<td>$40348329</td>
</tr>
<tr>
<td>Acetyl chloride</td>
<td>423 kg</td>
<td>$47/kg</td>
<td>$19682</td>
</tr>
</tbody>
</table>
The E-factor for the semi-synthesis of paclitaxel from 10-DAB – defined as the ratio of the mass of the product to the mass of all raw materials (both in kg) – is approximately 200-fold greater than the microbial semi-synthetic process, implying that the ratio of $V_M$ to $V_P$ is 200. As a consequence, the capital costs for the microbial process are 24 times greater. Additionally, capital costs for processes handling material flows in the range of 50,000 kg/yr are roughly $40 million. Capital depreciation (D) over the lifetime of the manufacturing facility can be related by:

$$D = \frac{C}{S \cdot N}$$

(3.2)

where S is the annual output of the manufacturing plant (100 kg) and N is the lifetime of the project (10 years). Accordingly, capital depreciation costs for the botanical process too are 24-fold lower. Operation of the 3 GMP-compliant steps that are common to both processes is assumed to be supervised by 1 personnel each, and 2 personnel oversee operation of the fermenter. As both
manufacturing schemes implement chemistries that are close to being or already have been adapted to the flow configuration, the personnel required to monitor the remaining synthetic steps are assumed to scale according to the six-tenths power law. Consequently, the botanical and microbial processes are assumed to be operated by 10 and 14 personnel, respectively. Calculation of other fixed costs follows rules of thumb that are commonly applied by process engineers. Additionally, waste and treatment costs (W) are often a fifth of the raw material costs for small-to-medium volume chemical processes, and, bioprocessing costs are not itemized as these have already been factored into the material costs for 10-DAB and ATHTAX.

<table>
<thead>
<tr>
<th>Itemized costs</th>
<th>Plant</th>
<th>Microbial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of operations</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>Identities of GMP-compliant steps</td>
<td>3 (f, i, j)</td>
<td>3 (f, u, v)</td>
</tr>
<tr>
<td>Number of personnel (n)</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Total capital investment (C)</td>
<td>$40 \times 10^6</td>
<td>$960 \times 10^6</td>
</tr>
<tr>
<td>Raw material costs (R)</td>
<td>$15.7 \times 10^6/\text{kg}</td>
<td>$14.7 \times 10^6/\text{kg}</td>
</tr>
<tr>
<td>Project lifetime</td>
<td>10 years</td>
<td>10 years</td>
</tr>
<tr>
<td>Depreciation costs (D)</td>
<td>$40,000/\text{kg}</td>
<td>$960,000/\text{kg}</td>
</tr>
<tr>
<td>Waste &amp; treatment costs (W)</td>
<td>$3050/\text{kg}</td>
<td>$2754300/\text{kg}</td>
</tr>
<tr>
<td>Operating labor costs (L_O = 300,000-n)</td>
<td>$3 \times 10^6/\text{yr}</td>
<td>$4.2 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Non-operating labor costs (L_N = 0.6-L)</td>
<td>$1.8 \times 10^6/\text{yr}</td>
<td>$2.52 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Administration/overhead (O = 0.9-L)</td>
<td>$2.7 \times 10^6/\text{yr}</td>
<td>$3.78 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Supplies (P = 0.3-L)</td>
<td>$0.9 \times 10^6/\text{yr}</td>
<td>$1.26 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Maintenance (M = 0.02-C)</td>
<td>$0.8 \times 10^6/\text{yr}</td>
<td>$19.2 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Utilities (U = 0.01-C)</td>
<td>$0.4 \times 10^6/\text{yr}</td>
<td>$9.6 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Miscellaneous (M_5 = 0.01-C)</td>
<td>$0.4 \times 10^6/\text{yr}</td>
<td>$9.6 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Annual fixed costs (A = L_O + L_N + O + P + M + U + M_5)</td>
<td>$10 \times 10^6/\text{yr}</td>
<td>$50.16 \times 10^6/\text{yr}</td>
</tr>
<tr>
<td>Fixed costs (F = A / S)</td>
<td>$1 \times 10^6/\text{kg}</td>
<td>$5 \times 10^6/\text{kg}</td>
</tr>
<tr>
<td>Total manufacturing costs (= R + D + W + F)</td>
<td>$16.74 \times 10^6/\text{kg}</td>
<td>$23.41 \times 10^6/\text{kg}</td>
</tr>
</tbody>
</table>
6. 3. Summary of the economic assessment

The 40% difference between the unit price of paclitaxel produced using both schemes (Table 6.3) is simply an artifact of the high cost of benzeneseleninic acid anhydride, which individually accounts for 60% of the raw material cost for microbial semi-synthetic process. In fact, a 50% reduction in the price of this forbiddingly expensive reagent makes microbially semi-synthesized paclitaxel only 10% more expensive than 10-DAB-derived paclitaxel, whereas a >60% reduction makes the microbial-based process more economical than the plant-based one. Nevertheless, it appears that producing \( (1,2\alpha,5\alpha,7\beta,10\beta)-5\text{-acetyloxy}-1,2,7,10\text{-tetrahydroxy-tax-4,11-diene} \) at 0.0032 g/g DCW per day is not only technically feasible given the tools currently at our disposal, but is also cost-competitive with the semi-synthetic plant-based process.

6. 4. Providing unprecedented access to inaccessible chemical space

Besides competing quite favorably with plant-based systems, microbial metabolic engineering offers another significant benefit, best illustrated by the example of taxadiene. Though not pharmacoactive itself, taxadiene is of interest to medicinal chemists as it is the first intermediate in the paclitaxel biosynthetic pathway that bears the unique, \( C_{20} \) taxane scaffold. Unfortunately, like many therapeutic natural products, the taxadiene content in its native host is prohibitively low. As 9 of
paclitaxel's 11 stereocenters are located on its scaffold, total synthesis of taxadiene involves as many as 18 transformations and has an overall yield of 0.21% (Figure 6.2).

![Figure 6.2. Total synthesis of taxadiene: The scheme involves 18 steps and has an overall yield of 0.21%. Key: a. [1] Me₂tBuSiCl, DMF, [2] H₂, Pd/C (yield 62%); b. [1] m-CPBA, CH₂Cl₂, [2] NaOMe, MeOH (77%); c. Me₂tBuSiCl, imidazole, Et₂O (70%); d. [1] Me₂tBuSiCl, DMF, [2] DIBAH, toluene (71%); e. [1] H₂C=CMgBr, THF, [2] Dess-Martin periodinane, CH₂Cl₂ (65%); f. Me₂C(SeMe)₂, n-BuLi, THF (100%); g. [1] Pl₃, Et₃N, CH₂Cl₂, [2] HF, MeCN, THF (64%); h. PhCH(OOMe)₂, PPTS, CHCl₃ (88%); i. LiAlH₄, AlCl₃, Et₂O, CH₂Cl₂ (100%); j. Dess-Martin periodinane, CH₂Cl₂ (48%); k. CH₂CHMgBr, THF (100%); l. BF₃OEt₃, toluene (28%); m. LiAlH₄, THF (100%); n. PhOCSCl, NaN(SiMe)₂, THF (100%); o. Bu₃SnH, Vazo, toluene (53%); p. [1] Na/NH₃, THF, [2] Dess-Martin periodinane, CH₂Cl₂ (56%); q. MeMgBr, CeCl₃, THF (100%); r. MeO₂CNSO₂NEt₃, toluene (60%)](image)

The biosynthesis of taxadiene (Figure 6.3), however, fares much better in comparison. As was presented earlier, taxadiene assembly in *E. coli* commences with the head-to-tail condensation of isopentenyl pyrophosphate with dimethylallyl pyrophosphate (DMAPP, 21) to produce geranyl pyrophosphate (GPP, 22), which then condenses with IPP *in situ* to form farnesyl pyrophosphate (FPP, 23). Yet another *in situ* condensation reaction follows suit and IPP combines with FPP to

\[ \begin{align*}
\text{IPP} \xrightarrow{\text{a}} & \text{DMAPP} \\
\text{IPP} + \text{DMAPP} \xrightarrow{\text{b}} & \text{GPP}
\end{align*} \]
\[ \begin{align*}
\text{IPP} + \text{GPP} \xrightarrow{\text{b}} & \text{FPP}
\end{align*} \]
\[ \begin{align*}
\text{IPP} + \text{FPP} \xrightarrow{\text{b}} & \text{GGPP}
\end{align*} \]
\[ \begin{align*}
\text{GGPP} \xrightarrow{\text{c}} & \text{Taxadiene (19)}
\end{align*} \]

**Figure 6.3. Biosynthesis of taxadiene:** The pathway commences with the isomerization of IPP (20) by an isomerase to form DMAPP (21) (step a). Condensation between IPP with DMAPP forms GPP (22), which then re-condenses with IPP to form FPP (23) *in situ*. FPP further condenses with IPP to form GGPP (24). All condensation reactions occur in the active site of a single enzyme, GGPP synthase (step b). GGPP is then cyclized by taxadiene synthase to form taxadiene (19) (step c).

The sequence of 3 consecutive condensation reactions involving IPP is catalyzed by the enzyme, GGPP synthase, whereas DMAPP is an isomerization product of IPP formed by the enzyme, IPP isomerase. Taxadiene formation is catalyzed by taxadiene synthase. IPP is synthesized by the non-mevalonate pathway, which commences with the condensation of glyceraldehyde-3-phosphate with
pyruvate to yield 2-methyl-(D)-erythritol-4-phosphate (MEP). Our demonstrations of taxadiene synthesis with *E. coli* consuming glycerol as the primary carbon source achieved yields of 0.68% and 0.85% in 2 mL batch cultures and 1 L fed-batch fermentations, respectively\(^5\). The biosynthetic atom economy is also several orders of magnitude higher than that of total synthesis. Significantly, considering that the theoretical yield of taxadiene on glycerol is 11.625% (for strains utilizing glucose as the primary carbon feedstock, the theoretical yield is 20%) and that the batch and fed-batch fermentations were yet to be optimized and the metabolism of *E. coli* could be further manipulated to improve the flux to taxadiene, the improvement in overall yield of microbial-aided synthesis over total synthesis could be even more marked. Our group has already successfully elevated the yield of taxadiene at the 2 mL scale by as much as 50% by expressing a downstream cytochrome P450 monooxygenase-reductase complex that oxidizes taxadiene to form taxadien-5α-ol\(^\text{128}\).
Chapter 7
PUTTING IT ALL TOGETHER: A VISION FOR DRUG DISCOVERY

In Chapter 1, we elaborated how deficiencies in combinatorial chemistry were contributing to the pharmaceutical industry’s ‘target-rich, lead-poor’ imbalance. The decision to pursue quantity over quality, how it contributed to the decline of the industry’s synthetic capabilities and how these lamentable decisions led to companies foregoing natural product space as a source of new drugs were discussed at length in Chapter 2. In the three chapters that followed, a case was made for microbial metabolic engineering as a viable platform for the discovery of natural products and some of the technical challenges stymying the field were addressed therein. In this chapter, we shall elaborate a unifying vision that could translate unprecedented numbers of small-molecule drugs to the bedside. Called ‘biosynthomics’, our vision for re-inventing drug discovery is driven by microbial metabolic engineering and recent developments in analytics, genome sequencing, flow chemistry, informatics, oligonucleotide synthesis and retrosynthetic analysis, as well as exponential improvements in the raw computing power that is now available for drug research. Many of its individual elements are already in place and if metabolic and enzyme engineering are able to address some of the challenges that were discussed in Chapter 5, biosynthomics could become a broad paradigm for generating focused libraries of natural product-like molecules.
7. 1. What is biosynthomics?

Biosynthomics comprises four principal domains – design, synthesis, exploration and integration. Design encompasses the selection of electronic features and their steric optimization onto a rigid molecular framework in order to ensure optimal binding to a specific biological target. The conceptual underpinning of this exercise is rooted in fragment-based pharmacophore modeling\textsuperscript{129-131} and all steps are completed \textit{in silico}. The rigidity of the molecular framework, or scaffold, translates to a minimization of the substituents’ rotational entropy\textsuperscript{132}. The pharmacophores so-designed are then synthesized in a microbial chassis that has been carefully constructed via metabolic engineering\textsuperscript{133-135}. This activity constitutes the second domain of biosynthomics – synthesis.

Although secondary metabolic enzyme homologues usually combine metabolic building blocks via a common reaction mechanism, they often differ from one another in their preferences for substrates. Even homologues combining the same building blocks frequently yield products that are stereo- and regiochemically unique. The third domain of biosynthomics – exploration – expands the biosynthetic ensemble for metabolic engineering through a systematic search of Nature’s metabolic landscape\textsuperscript{53, 136, 137}. Another important consideration in pharmacophore design is its potential use as an input for conventional target-oriented synthesis (TOS)\textsuperscript{102}, an exercise that aims to access precise regions of chemical space – either a single molecule or a small assortment of molecules exhibiting minor variations on a chemical theme. Accordingly, the fourth principal activity of biosynthomics –
integration – is a modern, biosynthetic take on retrosynthesis and targets the synthesis of an advanced intermediate that can act as a gateway molecule for TOS via the expression of a vastly smaller set of enzymes takes advantage of the core competencies of both metabolic engineering and synthetic chemistry.

7.2. The informed design of superior pharmacophores

Lead selection by forward chemical genetics generally involves high-throughput screening of a combinatorial library comprising of \( >10^6 \) compounds, and, compounds that are typically selected for lead optimization exhibit \( K_d \) values of \( <1 \mu M \). Fragment-based pharmacophore design (Figure 7.1.), in comparison, commences by screening only a few thousand compounds for \( K_d \) values that fall in high \( \mu M \) to mM range\(^{138} \). High-throughput 2-D isotope-edited NMR is commonly utilized to screen the compounds\(^{139} \), and, the interactions of their molecular appendages – or fragments – with residues within the target’s binding pocket are then characterized using X-ray crystallography. Several fragments are then incorporated within a single molecular entity.

The amalgamation of multiple fragments is rationalized by the additive effect that their individual interactions have on the binding affinity of the chimeric molecule, and, the likelihood of generating a hit from a compound library synthesized using fragment-based approaches is \( \leq 3 \) orders of magnitude higher than it is for high-throughput screening of combinatorial libraries.
Fragment-based pharmacophore discovery: Inhibition of β-secretase has been shown to be an effective strategy for treating Alzheimer’s disease. Species 1 was identified by 1-D NMR screening of a fragment library. Its binding affinity was determined by surface plasmon resonance (SPR). X-ray crystallography was then utilized to unearth insights regarding the structural features of 1 that interact with the binding site of β-secretase, which led to the synthesis of species 2. The use of pharmacophore and structure-based drug design approaches was then implemented to derive species 3.

The design methodology we envision builds upon fragment-based pharmacophore discovery and commences with comparative modeling of the putative target. However, as only a third of all protein sequences can be accurately modeled in silico, high-throughput X-ray crystallography is used either to determine the structures of targets comprising of unknown or poorly characterized folds, or, in conjunction with NMR, electron microscopy and X-ray scattering, supply restraints that can aid model generation.

The target itself is discovered and validated using omics-based or clinical approaches, or, more recently, by metabolic control analysis. Omics technologies, in particular, rapidly generate vast
amounts of relevant data that can be compiled into systems-wide models for simulations to prioritize targets, accelerate hypothesis generation and testing in disease models, and design clinical trials\textsuperscript{20}.

Figure 7.2. Overview of the biosynthomics paradigm: The proposed design methodology builds upon fragment-based drug discovery. The application of cheminformatics features prominently, and, casting the steric and electronic effects of the fragments as a three-dimensional free energy relationship for informed screening of a scaffold database by minimizing the cost function is a unique addition to the approach. Panel A highlights the workflow for the entire methodology. Panel B describes how omics-derived targets are analyzed to identify optimally binding fragments (also Panel C) whose three-dimensional positioning within the active site guide scaffold design (also Panel D).

Once the topographies of the target’s binding pockets have been charted, fragments that might interact with residues in the binding pockets are screened virtually using approaches such as similarity searching\textsuperscript{42}. The catalogue of \textit{in silico} fragment ‘hits’ is subsequently refined using high-
throughput 2-D NMR, and interactions between the fragments and residues in the binding pocket are validated using high-throughput X-ray crystallography.

7.3. Cheminformatics: An integral part of pharmacophore design

Though an improvement over compounds generated using combinatorial chemistry, synthetic molecules that incorporate multiple fragments still possess rotational degrees of freedom that have detrimental entropic consequences for binding. Instead, pharmacophores consisting of fragments whose coordinates within the binding site have been greatly restricted by steric effects induced by stereo- and regiospecifically linking them to privileged, natural product-like scaffolds are more likely to interact strongly and selectively with the target. Accordingly, the next phase of pharmacophore design involves mining natural product-like scaffolds that meet the spatial constraints of the interacting fragments. The steric and electronic effects of the fragments are cast as a three-dimensional free energy relationship \(^{150, 151}\), following which, a scaffold database \(^{152}\) is screened for candidates that minimize the cost function.

A database similar to the National Cancer Institute’s ChemBank \(^{153}\) would have to be created for natural products. Estimates suggest that such a database could contain in excess of 170,000 entries \(^{61}\). Software that allows virtual screening of the structural database and analysis of the scaffolds will also be needed \(^{154}\). As the pharmacophores are intended to be gateway molecules for TOS, integration of computer-aided synthetic planning \(^{155, 156}\) into the screening software is desired. Additionally, the
software package would also include National Center for Biotechnology Information (NCBI) database search and analyses capabilities that append the cheminformatics output with information about the biosynthesis of the molecule. For instance, paclitaxel and epothilone B, a 16-membered macrolide produced by the myxobacteria, *Sorangium cellulosum*\(^\text{157}\), arrest the cancerous cell cycle in its metaphase by inhibiting tubulin depolymerization. Both molecules also share a common binding pocket, albeit they interact quite differently therein\(^\text{158}\). Pharmacophore design for the discovery of more potent and safer tubulin-binding drugs according to the methodology outlined herein would, therefore, yield structures based on the scaffolds of both paclitaxel and epothilone B (Figure 7.3). However, based on an assessment of the biosynthetic information that is presently available for both molecules\(^\text{83, 159}\) and assuming that the TOS schemes do not exhibit significant differences in their coverage of local chemical space, one is more likely to select the pharmacophore based on the epothilone scaffold. Gateway molecules, thus, can be identified through synthetic and biosynthetic analyses involving candidates that have been objectively screened from a natural product database.

![Figure 7.3. Two scaffolds, one binding site: Structural elements of paclitaxel and epothilone B that share a common interaction site within the binding pocket have been highlighted with the same color.](image_url)
It is evident that cheminformatics is central to the success of the biosynthomics drug discovery platform. Individual software elements already exist but a single program in the mold of Accelrys, Inc.’s Discovery Studio, that accesses databases of the kind described previously and supports rigorous scientific analysis of the retrieved data as discussed does not exist. The development of such tools could not be more urgent. Once available, they could transform drug discovery.

7. 4. Mining the natural world

Information about biosynthetic gene sequences encoding individual assembly steps and structural characterization of the intermediates and products that various enzyme combinations produce is crucial to the success of the biosynthomics platform. To this end, high-throughput whole genome sequencing, assembly and bioinformatics analysis will greatly expand the biosynthetic ensemble that is available for metabolic engineering. The vast majority of recently uncovered bacterial terpenoids were discovered through genome mining. Warp Drive Bio, a biotechnology start-up in Cambridge, MA has adopted a similar approach for discovering new drug candidates by leveraging its competencies in whole genome sequencing, assembly, mining and gene synthesis to produce natural products for Sanofi Aventis’ drug discovery programs.

Additionally, NMR-based comparative metabolomics should aid rapid molecular characterization, and, the development of in silico tools and methodologies for correlating the sequence and domain
organization of PKSs will accelerate annotation of more orphan PKS\textsuperscript{163-169}. Predictive capabilities for characterizing novel terpenoid biosynthetic enzymes would also have to expanded\textsuperscript{166, 170.}

7.5. Continuous manufacturing as a tool for discovery

It has been proposed that continuous manufacturing could potentially benefit drug discovery by allowing investigations of reaction conditions, scale-up from gram-scales to ton-scales to meet material requirements for clinical trials, evaluation of supply chain constraints, and rapid transfer of the technology to production\textsuperscript{171}. Click chemistry\textsuperscript{15}, DNA-templated organic synthesis\textsuperscript{172} and a very recent, multidimensional experimentation technique to evaluate reactions catalyzed by transition metal complexes\textsuperscript{173} offer improved and high-throughput prospection of chemical space. However, none of these techniques can consistently generate molecular architectures capable of rivaling the complexity of natural products. Among other platforms, automated synthesizers have lately gained acceptance as robust tools for natural product syntheses\textsuperscript{184}. However, as these units predominantly utilize batch chemistries to synthesize target molecules, their integration with continuous manufacturing faces many technical challenges.

Microreactors, on the other hand, appear to be much more promising. Linearly-linked flow reactors have been successfully used for the multistep syntheses of the cytotoxic alkaloid, (±)-oxomaritidine\textsuperscript{174}, as well as a small library of piperazines targeting a novel chemokine receptor linked to several inflammations and allergies\textsuperscript{175}. In the case of (±)-oxomaritidine, synthesis was achieved
using packed columns containing immobilized reagents, catalysts and scavengers. The \([5+2]\)
cycloaddition of quinone monoketals and styrene derivatives to produce the bicyclo[3.2.1]octanoid
scaffolds of neolignan natural products has also been demonstrated in flow. Yet, these successes are
but a few examples of the microreactor-facilitated traversal of natural product space\(^{176}\). Adapting
batch chemistries to the flow format is fairly protracted and technically challenging, thereby
rendering the design of chemical processes involving six to eight synthetic steps – as is often the case
for small-molecule drugs – impractical for drug discovery.

7. 6. Metabolic engineering _ The ultimate demonstration of continuous manufacturing

Microbial synthesis of natural products intensifies and replaces an otherwise complicated synthetic
scheme comprising of scaffold assembly and multiple functional group transformations with a single
unit operation wherein a carbohydrate feedstock is converted to the product of interest under
considerably milder operating conditions by a microbial host (Figure 7.4). Moreover, it is unlikely
that the microbial-aided synthesis of APIs and pharmaceutical intermediates will face many of the
challenges that flow synthetic schemes encounter during work-up. As the routine use of perfusion
bioreactors by biopharmaceutical companies to manufacture protein products that rapidly degrade in
the culture broth would suggest, continuous biomanufacturing and cultivation are well-established
industrial unit operations\(^{177, 178}\). This versatility makes it an important value addition to drug
discovery. Also, whereas isolation of the API and its synthetic intermediates typically incur
considerable material and energy costs in traditional, batch-based manufacturing, microbial
production of an API entirely does away with the need to isolate synthetic intermediates, greatly shrinking the footprint of the manufacturing facility. For an industry that typically generates between 25 to 100 kg of waste for every kg of product it manufactures, adoption of metabolic engineering for chemical synthesis promises quantum leaps in atom and energy economies\textsuperscript{179}. The API itself can be extracted from the culture broth through use of continuous countercurrent chromatography\textsuperscript{178}, followed by purification and crystallization to produce the bulk API.

Figure 7.4. Metabolic engineering, the ultimate demonstration of reaction intensification: E and G denote a biosynthetic enzyme and gene, respectively. R represents the reactant, S denotes the solvent and W represents the waste stream emerging from the unit operation.
7. 7. Final words

Any technology is only as good as its weakest link and much remains to be done before biosynthonics crystallizes into a viable drug discovery platform. However, many of its elements are already in place. Future advances and the convergence of analytical biochemistry, bioinformatics, computation, flow chemistry and metabolic & protein engineering will drive this revolution in drug discovery. One sincerely hopes this is the case. After all, the price of the status quo for discovering and developing a drug has already exceeded a billion dollars\textsuperscript{180}. 
Chapter References


94. http://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/multalignviewer/frame

95. Schoendorf, A., Rithner, C.D., Williams, R.M. & Croteau, R. Molecular cloning of a cytochrome P450 taxane 10β-hydroxylase cDNA from *Taxus* and functional expression in


Appendix A: List of taxanes analyzed for mapping the taxane biosynthetic pathway
<table>
<thead>
<tr>
<th>Structure: Markush</th>
<th>Substance Identification: Reaxys</th>
<th>Chemical Name</th>
<th>Molecular Formula</th>
<th>Molecular Weight</th>
<th>Linear Structure Formula</th>
<th>Type of Substance</th>
<th>Compositional Comp. Name</th>
<th>InChIKey</th>
<th>Field Availability</th>
<th>Number of Reactions</th>
<th>Number of Reference(s)</th>
<th>Isolation from Natural Product</th>
<th>Comment</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>4290290 (View in Reaxys)</td>
<td>(2αR, 4βα, 4αβ, 6βα, 9αβ, 12αβ, 10αβ, 12βα, 12αβ)</td>
<td>[2αα,4ββ,4αβα,6αβ,9αβ]</td>
<td>C47H51N</td>
<td>570.37</td>
<td>heterocyclic</td>
<td>RCIQNCONZ</td>
<td>SJXQF-MXXOVDVA-N</td>
<td>853.92</td>
<td>leaves of <em>Tausus cuspidata</em> collected in China</td>
<td></td>
<td></td>
<td></td>
<td>Journal: Dong, Hui; Luo, Lina; Zhou, Shan; Bi, Pengyu; Li, Kaif; Zhao, Jiping; Chemistry of Natural Compounds; vol. 41; 4 (2007); p. 478 - 490; Khimija Prirodnykh Soedinenii; 4 (2007); p. 392 - 393; View in Reaxys</td>
<td></td>
</tr>
<tr>
<td>4290290 (View in Reaxys)</td>
<td>(2αR, 4βα, 4αβ, 6βα, 9αβ, 12αβ, 10αβ, 12βα, 12αβ)</td>
<td>[2αβ,4βα,4αβα,6αβ,9αβ]</td>
<td>C47H51N</td>
<td>570.37</td>
<td>heterocyclic</td>
<td>RCIQNCONZ</td>
<td>SJXQF-MXXOVDVA-N</td>
<td>853.92</td>
<td>dark brown callus culture of <em>Tausus cuspidata</em> Sieb. et Zucc. (Taxaceae) Japanese yew, collected in Sendai</td>
<td></td>
<td></td>
<td></td>
<td>Journal: Bai, Jiao; To, Nobuhide; Sakai, Junichi; Kitakata, Mutsumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujun; Hirose, Katsutomi; Tomita, Akihito; Tsuno, Takashi; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4 (2005) p. 497 - 501; View in Reaxys</td>
<td></td>
</tr>
<tr>
<td>4290290 (View in Reaxys)</td>
<td>(2αR, 4βα, 4αβ, 6βα, 9αβ, 12αβ, 10αβ, 12βα, 12αβ)</td>
<td>[2αα,4ββ,4αβα,6αβ,9αβ]</td>
<td>C47H51N</td>
<td>570.37</td>
<td>heterocyclic</td>
<td>RCIQNCONZ</td>
<td>SJXQF-MXXOVDVA-N</td>
<td>853.92</td>
<td>callus culture induced from young stems of <em>Tausus cuspidata</em> Sieb. et Zucc., collected in Sendai, Japan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4290290 (View in Reaxys)</td>
<td>(2αR, 4βα, 4αβ, 6βα, 9αβ, 12αβ, 10αβ, 12βα, 12αβ)</td>
<td>[2αβ,4βα,4αβα,6αβ,9αβ]</td>
<td>C47H51N</td>
<td>570.37</td>
<td>heterocyclic</td>
<td>RCIQNCONZ</td>
<td>SJXQF-MXXOVDVA-N</td>
<td>853.92</td>
<td>(a) washing a raw material (RM) with water; (b) extracting with an organic solvent a wet RM; (c) contacting the wet RM</td>
<td></td>
<td></td>
<td></td>
<td>Patents: Chaiichem Pharmaceuticals International; US6795939: B1; (2004); View in Reaxys</td>
<td></td>
</tr>
<tr>
<td>4290290 (View in Reaxys)</td>
<td>(2αR, 4βα, 4αβ, 6βα, 9αβ, 12αβ, 10αβ, 12βα, 12αβ)</td>
<td>[2αβ,4βα,4αβα,6αβ,9αβ]</td>
<td>C47H51N</td>
<td>570.37</td>
<td>heterocyclic</td>
<td>RCIQNCONZ</td>
<td>SJXQF-MXXOVDVA-N</td>
<td>853.92</td>
<td>suspension cell cultures of <em>Tausus cuspidata</em> (P034F) and <em>Tausus x media</em> cv. Hicki</td>
<td></td>
<td></td>
<td></td>
<td>Journal: Ketcho, Raymond E. B.; Rhinner, Christopher D.; Ghi, Dayou; Kim, Yo Sun; Williams, Robert M.; Creneau, Rodney B.; Phytochemistry (Elsevier); vol. 62; 6 (2003); p. 901 - 916; View in Reaxys</td>
<td></td>
</tr>
<tr>
<td>4290290 (View in Reaxys)</td>
<td>(2αR, 4βα, 4αβ, 6βα, 9αβ, 12αβ, 10αβ, 12βα, 12αβ)</td>
<td>[2αβ,4βα,4αβα,6αβ,9αβ]</td>
<td>C47H51N</td>
<td>570.37</td>
<td>heterocyclic</td>
<td>RCIQNCONZ</td>
<td>SJXQF-MXXOVDVA-N</td>
<td>853.92</td>
<td>noted cultures of the <em>Tausus cuspidata</em> Sieb. et Zucc.</td>
<td></td>
<td></td>
<td></td>
<td>Journal: Shi, Qiu-Wei; Petetko, Tracy L.; Sautel, Francois; Mamar, Oran; Zamir, Lolita O.; Canadian Journal of Chemistry; vol. 81; 1 (2003); p. 64 - 74; View in Reaxys</td>
<td></td>
</tr>
</tbody>
</table>
420260 View in Reaxys

[Image of a chemical structure]

420260 View in Reaxys

[Image of a chemical structure]

420260 View in Reaxys

[Image of a chemical structure]

420260 View in Reaxys

[Image of a chemical structure]

420260 View in Reaxys

[Image of a chemical structure]

420260 View in Reaxys

[Image of a chemical structure]

1444248 View in Reaxys

[Image of a chemical structure]

Copyright © 2010 Elsevier Properties SA. All rights reserved. Authorized use only. Reaxys® is a trademark owned and protected by Elsevier Properties SA and used under license.
<p>| C4H4H9N O13 | 10-deacetyl taxol(7a-OH); 7-epi-10-deacetyltaxol; 10-deacetyl-paclitaxel; 10-deacetyltaxol; 10-deacetyl taxol; 10-deacetoxytaxol | 78432-77-6; 78454-10-6; 123564-13-4; 133577-34-1; 135501-18-7 | 811.883 | TYLVQKN NHU;XIP-MHFAHFC; SA-N |
| C4H4H9NO13 | heterocyclic | 932.103 | leaves of Taxus cuspidata collected in China |
| C4H5H5O14 | 3-N-(trans-2-methyl-2-butenoyl)-3-N-debenzoylpaclitaxel; cephalomannine; Cephalomannine; caphalomannine; taxol B; caphalomannine | 71610-00-9 | 831.914 | DBXFAP;JCZ ABTR-WBYYIYXIS A-N |
| C4H5H5O14 | heterocyclic | 8032151 | cactus culture induced from young stems of Taxus cuspidata Sieb. et Zucc.; collected in Sandy, Japan stimulated with methyl | 135501-18-7 |
| C4H5H5O14 | 3'-N-(trans-2-methyl-2-butenoyl)-3'-N-debenzoylpaclitaxel; cephalomannine; Cephalomannine; caphalomannine; taxol B; cephalomannine | 71610-00-9 | 831.914 | DBXFAP;JCZ ABTR-WBYYIYXIS A-N |
| C4H5H5O14 | heterocyclic | 8032151 | suspension cell cultures of Taxus cuspidata (PIDAF) and Taxus x media cv. Hickey | 135501-18-7 |
| C4H5H5O14 | 3'-N-(trans-2-methyl-2-butenoyl)-3'-N-debenzoylpaclitaxel; cephalomannine; Cephalomannine; caphalomannine; taxol B; cephalomannine | 71610-00-9 | 831.914 | DBXFAP;JCZ ABTR-WBYYIYXIS A-N |
| C4H5H5O14 | heterocyclic | 8032151 | needles of the Japanese yew Taxus cuspidata Sieb et Zucc, collected from Yagishiri Island, Hokkaido, Japan | 135501-18-7 |
| C4H5H5O14 | 3'-N-(trans-2-methyl-2-butenoyl)-3'-N-debenzoylpaclitaxel; cephalomannine; Cephalomannine; caphalomannine; taxol B; cephalomannine | 71610-00-9 | 831.914 | DBXFAP;JCZ ABTR-WBYYIYXIS A-N |
| C4H5H5O14 | heterocyclic | 8032151 | stems of Japanese yew Taxus cuspidata Sieb. et Zucc. | 135501-18-7 |</p>
<table>
<thead>
<tr>
<th>Citation</th>
<th>Formula</th>
<th>RMPUDM-RZHPVQDS AN</th>
<th>Identification</th>
<th>Physical Data</th>
<th>Spectra</th>
<th>Biological Activity</th>
<th>Bioactivity/Toxicity</th>
<th>Metabolites</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal: Oshige, Masahiko; Takeouchi, Mik, Kato, Yasutaro; Kamisuki, S. T.; Takahashi, Toshiyuki; Kuroda, Kazufumi; Sato, Noriaki; et al.; Bloorganic Medicinal Chemistry; vol. 12; 10; (2004); p. 2597 - 2602; View in Reaxys</td>
<td>C3SH42O9</td>
<td>3835-52-7</td>
<td>606.713</td>
<td>C3SH42O9</td>
<td>isocyclic</td>
<td>BAYHEZUZ RMPUDM-RZHPVQDS AN</td>
<td>16 spectra</td>
<td>11 biological activity</td>
<td>Japanese yew (Taxus cuspidata)</td>
</tr>
<tr>
<td>Journal: Shi, Qing-Wen; Petzke, Tracy L.; Sauriol, Francoise; Mamer, Orval; Zamir, Lolita O.; Canadian Journal of Chemistry; vol. 81; 1; (2003); p. 64 - 74; View in Reaxys</td>
<td>C3SH42O9</td>
<td>3835-52-7</td>
<td>606.713</td>
<td>C3SH42O9</td>
<td>isocyclic</td>
<td>BAYHEZUZ RMPUDM-RZHPVQDS AN</td>
<td>8 spectra</td>
<td>4 biological activity</td>
<td>needles of the Japanese yew (Taxus cuspidata)</td>
</tr>
<tr>
<td>Journal: Kobayashi, Jun'ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</td>
<td>C3SH42O9</td>
<td>3835-52-7</td>
<td>606.713</td>
<td>C3SH42O9</td>
<td>isocyclic</td>
<td>BAYHEZUZ RMPUDM-RZHPVQDS AN</td>
<td>7 spectra</td>
<td>6 biological activity</td>
<td>stems of the Japanese yew (Taxus cuspidata)</td>
</tr>
<tr>
<td>Journal: Kobayashi, Jun'ichi; Ogiwara, Aya; Hosoya, Hirokazu; Shigemori, Hideyuki; Yoshida, Naotani; et al.; Tetrahedron; vol. 50; 25; (1994); p. 7401 - 7416; View in Reaxys</td>
<td>C3SH42O9</td>
<td>3835-52-7</td>
<td>606.713</td>
<td>C3SH42O9</td>
<td>isocyclic</td>
<td>BAYHEZUZ RMPUDM-RZHPVQDS AN</td>
<td>6 spectra</td>
<td>5 biological activity</td>
<td>stems of the Japanese yew (Taxus cuspidata)</td>
</tr>
</tbody>
</table>
| Journal: Kondo, Tokuhashi; Yakagaku Zasshi; vol. 45; (1955); p. 861,874; disth. Ref. Nr. 524, S. 2; Chem.Abstr.; (1926); p. 767; View in Reaxys | C3SH42O9 | 3835-52-7           | 606.713        | C3SH42O9     | isocyclic | BAYHEZUZ RMPUDM-RZHPVQDS AN | 5 spectra | 4 biological activity | Isolierung aus dem Biettem von Taxus cuspidata |...
<table>
<thead>
<tr>
<th>Entry</th>
<th>View in Reaxys</th>
<th>Structure</th>
<th>Formula</th>
<th>MW</th>
<th>CAS</th>
<th>IUPAC Name</th>
<th>Full Name</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>7059440</td>
<td>View in Reaxys</td>
<td><img src="image1.png" alt="Structure Image" /></td>
<td>$C_{28}H_{40}O_8$</td>
<td>504.621</td>
<td>C28H40O8</td>
<td>isocyclic</td>
<td>KFFHSFCO KCGBBW- VCPDXWRA SA-N</td>
<td>dark brown callus culture of Taxus cuspidata Sieb. et Zucc. (Taxaceae) Japanese yew, collected in Sendai, Japan</td>
<td>Journal; Bai, Jiao; Ito, Nobuhida; Sakai, Junchi; Kikabatake, Mutumi; Fujii, Hajime; Bae, Liming; Dai, Jungui; Zhang, Shujun; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takeki; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005): p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>7059440</td>
<td>View in Reaxys</td>
<td><img src="image2.png" alt="Structure Image" /></td>
<td>$C_{28}H_{40}O_8$</td>
<td>504.621</td>
<td>C28H40O8</td>
<td>isocyclic</td>
<td>KFFHSFCO KCGBBW- VCPDXWRA SA-N</td>
<td>callus culture induced from young stems of Taxus cuspidata Sieb. et Zucc., collected in Sendai, Japan</td>
<td>Journal; Bai, Jiao; Kikabatake, Mutumi; Toyotomi, Kazuki; Fu, Lue; Zhang, Shujun; Dai, Jungui; Sakai, Junchi; Hirose, Katutoshi; Yamori, Takaao; Tomida, Akhiro; Tsuruo, Takeki; Ando, Masayoshi; Journal of Natural Products; vol. 67; 1; (2004): p. 58 - 63; View in Reaxys</td>
</tr>
<tr>
<td>6837340</td>
<td>View in Reaxys</td>
<td><img src="image3.png" alt="Structure Image" /></td>
<td>$C_{37}H_{46}O_10$</td>
<td>650.766</td>
<td>C37H46O10</td>
<td>isocyclic</td>
<td>MUTXBNFQ DJTPL-PXORYUNG SA-N</td>
<td>stems of Japanese yew Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal; Kobayashi, Jun’ichi; Shigemori, Hideyuki; Hataocyclois; vol. 47; 2; (1998): p. 1111 - 1134; View in Reaxys</td>
</tr>
<tr>
<td>6837340</td>
<td>View in Reaxys</td>
<td><img src="image4.png" alt="Structure Image" /></td>
<td>$C_{37}H_{46}O_10$</td>
<td>650.766</td>
<td>C37H46O10</td>
<td>isocyclic</td>
<td>MUTXBNFQ DJTPL-PXORYUNG SA-N</td>
<td>stems of the Japanese yew Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal; Kobayashi, Jun’ichi; Ogihara, Ayo; Horozawa, Hirokazu; Shigemori, Hideyuki; Yoshida, Naoto; et al.; Tetrahedron; vol. 50; 25; (1994): p. 7401 - 7416; View in Reaxys</td>
</tr>
<tr>
<td>3150690</td>
<td>View in Reaxys</td>
<td><img src="image5.png" alt="Structure Image" /></td>
<td>$C_{28}H_{40}O_8$</td>
<td>504.621</td>
<td>C28H40O8</td>
<td>isocyclic</td>
<td>SKJSIVQEP KBFJ-HUWLPS B-S A-N</td>
<td>dark brown callus culture of Taxus cuspidata Sieb et Zucc. (Taxaceae) Japanese yew, collected in Sendai</td>
<td>Journal; Bai, Jiao; Ito, Nobuhida; Sakai, Junchi; Kikabatake, Mutumi; Fujii, Hajime; Bae, Liming; Dai, Jungui; Zhang, Shujun; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takeki; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005): p. 497 - 501; View in Reaxys</td>
</tr>
</tbody>
</table>

Copyright © 2010 Elsevier Properties SA. All rights reserved. Authorized use only. Reaxys® is a trademark owned and protected by Elsevier Properties SA and used under license.
<table>
<thead>
<tr>
<th>Journal/Miyazaki et al.; Chemical and Pharmaceutical Bulletin; vol. 16; (1968); p. 545 - 548;</th>
<th>Taxus cuspidata; dark brown callus culture of Taxus cuspidata; Siib. et Zucc. (Taxaceae) Japanese yew, collected in Sendai, Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal/Sugiyama, Takeyoshi; Oritani, Takayuki; Oritani, Takashi; Bioscience, Biotechnology, and Biochemistry; vol. 56; 10; (1994); p. 1923 - 1924</td>
<td>View in Reaxys</td>
</tr>
<tr>
<td>Journal/Bai, Jiao; Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujian; Hirose, Katsuhiko; Tomida, Akihito; Tsuruo, Takashi; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005); p. 497 - 501</td>
<td>View in Reaxys</td>
</tr>
<tr>
<td>Journal/Bai, Jiao; Kitabatake, Mutumi; Toyozumi, Kazuki; Fu, Liwei; Zhang, Shujian; Dai, Jungui; Sakai, Junichi; Hirose, Katsuhiko; Yamori, Takak, Tomida, Akihito; Tsuruo, Takashi, Ando, Masayoshi; Journal of Natural Products; vol. 67; 1; (2004); p. 58 - 63</td>
<td>View in Reaxys</td>
</tr>
<tr>
<td>Journal/Kobayashi, Jun'ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134</td>
<td>View in Reaxys</td>
</tr>
<tr>
<td>Journal/Murakami, Ryo; Shi, Qing-wen; Oritani, Takayuki; Phytochemistry (Elsevier); vol. 52; 8; (1999); p. 1577 - 1580; View in Reaxys</td>
<td>View in Reaxys</td>
</tr>
</tbody>
</table>

**Notes:**
- **Journal:** Journal of Natural Products; vol. 67; 1; (2004); p. 58 - 63; View in Reaxys
- **Journal:** Journal of Natural Products; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys
- **Journal:** Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys
- **Journal:** Phytochemistry (Elsevier); vol. 52; 8; (1999); p. 1577 - 1580; View in Reaxys
<table>
<thead>
<tr>
<th>Identity(9)</th>
<th>Physical Data(5)</th>
<th>Spectra(8)</th>
<th>Bioactivity/Ecotox(44)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2a,5a,10-triacetoxy-14β-(2'-methyl)-butysyloxy-4(20),11-taxadiene;</strong></td>
<td><strong>C31H46O5</strong></td>
<td><strong>546.701</strong></td>
<td><strong>Identification(9) USNOOZXS AIFYCE-PHcal</strong></td>
</tr>
<tr>
<td><strong>2a,5a,10-triacetoxytaxa-4(20),11(12)-dien-14p-yl 2-methylbutyrate;</strong></td>
<td><strong>C31H46O8</strong></td>
<td><strong>546.701</strong></td>
<td><strong>View in methylbutyrate; C31H46O</strong></td>
</tr>
<tr>
<td><strong>4633976 1p-hydroxybaccatin I;</strong></td>
<td><strong>C32H44O14</strong></td>
<td><strong>30244-37-2</strong></td>
<td><strong>578.657</strong></td>
</tr>
<tr>
<td><strong>8173928 10-deacetyl-7-epi-taxol;</strong></td>
<td><strong>C45H49N1</strong></td>
<td><strong>811.883</strong></td>
<td><strong>View in 10-deacetyl-7-epi-taxol; C45H49N1</strong></td>
</tr>
<tr>
<td><strong>6032639 5o-hydroxy-2a,7β,9α,10p,13α-pentaacetoxy-4(20),11-taxadiene;</strong></td>
<td><strong>C30H42O11</strong></td>
<td><strong>84652-33-5</strong></td>
<td><strong>View in 5o-hydroxy-2a,7β,9α,10p,13α-pentaacetoxy-4(20),11-taxadiene;</strong></td>
</tr>
</tbody>
</table>

Copyright © 2010 Elsevier Properties SA. All rights reserved. Authorized use only. Reaxys® is a trademark owned and protected by Elsevier Properties SA and used under license.
5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J

5a-decinnamoyltaxinine J; 2-decinnamoyltaxinine J; decinnamoyltaxinine J; decinnamoyltaxanine J; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxy-4(20),11-taxadiene; 5a-hydroxy-2a,7p,9a,10p,13a-pentaacetoxytaxa-4(20),11-diene; 5a-decannamoyltaxinine J; 2-decannamoyltaxinine J; decannamoyl-taxinine J; decannamoyl-taxantine J
Journal: Kobayashi, Jun'ichi; Ogiwara, Aya; Hosoyama, Hirokazu; Shigemori, Hideyuki; Yoshida, Naotoshi; et al.; Tetrahedron; vol. 50; 25; (1994): p. 7401 - 7416; View in Reaxys

Journal: Kobayashi, Jun'ichi; Ogiwara, Aya; Hosoyama, Hirokazu; Shigemori, Hideyuki; Yoshida, Naotoshi; et al.; Tetrahedron; vol. 50; 25; (1994): p. 1111 - 1134; View in Reaxys

Journal: Shi, Qing-Wen; Petzke, Tracy L.; Sauriol, Francoise; Mamer, Orval; Zamir, Lolita; Canadian Journal of Chemistry; vol. 81; 1; (2003): p. 64 - 74; View in Reaxys

Journal: Bai, Jiao; Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujun; Hirose, Katutoshi; Tomida, Akihiro; Tsuruo, Takashi; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005): p. 497 - 501; View in Reaxys

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093

5843093
<table>
<thead>
<tr>
<th>CAS Number</th>
<th>Molecular Formula</th>
<th>Molecular Weight</th>
<th>Bioactivity/Ecotox</th>
<th>Journal, Year, Page Numbers</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>7558016</td>
<td>C31H46O9</td>
<td>562.701</td>
<td></td>
<td>Journal; Bai, Jiao; Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujun; Hirose, Katutoshi; Tomida, Akhiro; Tsucao, Takashi; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
<td>Cell culture induced from young stems of Taxus cuspidata Sieb. et Zucc., collected in Senda, Japan.</td>
</tr>
<tr>
<td>6553639</td>
<td>C30H42O10</td>
<td>552.657</td>
<td></td>
<td>Journal; Bai, Jiao; Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujun; Hirose, Katutoshi; Yamori, Takao; Tomida, Akhiro; Tsuruo, Takeshi; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
<td>Cell culture induced from young stems of Taxus cuspidata Sieb. et Zucc., collected in Senda, Japan.</td>
</tr>
<tr>
<td>6553639</td>
<td>C30H42O10</td>
<td>552.657</td>
<td></td>
<td>Journal; Kobayashi, Jun’ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</td>
<td>Cell culture induced from Taxus cuspidata, collected in Senda, Japan (stimulated with methyl)</td>
</tr>
<tr>
<td>6553639</td>
<td>C30H42O10</td>
<td>552.657</td>
<td></td>
<td>Journal; Shi, Qing-Wen; Petzke, Tracy L.; Saurio, Francoise; Mamer, Orval; Zamir, Lolita O.; Canadian Journal of Chemistry; vol. 81; 1; (2003); p. 64 - 74; View in Reaxys</td>
<td>Cell culture induced from Taxus cuspidata, collected in Senda, Japan (stimulated with methyl)</td>
</tr>
<tr>
<td>Structure</td>
<td>IUPAC Name</td>
<td>CAS Number</td>
<td>Molecular Weight</td>
<td>Bioactivity</td>
<td>Source</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>------------</td>
<td>------------------</td>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td><img src="image1" alt="Structure" /></td>
<td>5-hydroxytaxinine B; taxuspine F</td>
<td>C28H38O10</td>
<td>534.604</td>
<td>isoyclic</td>
<td>1.0-dehydroxybaccatin VI; 1-deoxybaccatin VI; baccatin VI</td>
</tr>
<tr>
<td><img src="image2" alt="Structure" /></td>
<td>18-dehydroxybaccatin VI; 1-dehydroxybaccatin VI; 1-deoxybaccatin VI; baccatin VI</td>
<td>C28H38O110</td>
<td>563.732</td>
<td>heterocyclic</td>
<td>1.0-dehydroxybaccatin VI; 1-deoxybaccatin VI; baccatin VI</td>
</tr>
<tr>
<td><img src="image3" alt="Structure" /></td>
<td>5a-cinnamoyloxy-9a,10β,13α-triacetoxytaxa-4(20),11-diene; 5a-cinnamoyloxy-9a,10β,13α-triacetoxytaxa-4(20),11-diene</td>
<td>C28H38O110</td>
<td>563.732</td>
<td>heterocyclic</td>
<td>1.0-dehydroxybaccatin VI; 1-deoxybaccatin VI; baccatin VI</td>
</tr>
</tbody>
</table>

*View in Reaxys*
<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Structure</th>
<th>CAS Number</th>
<th>Molecular Weight</th>
<th>Biological Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a,5a-dihydroxy-7β,9α,10β,13α-tetraacetoxy-4(20),11-taxadiene</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>C28H40O14</td>
<td>652.693</td>
<td>heterocyclic</td>
</tr>
<tr>
<td>2a,5a-dihydroxy-7β,9α,10β,13α-tetraacetoxy-4(20),11-taxadiene</td>
<td><img src="image2.png" alt="Structure" /></td>
<td>C28H40O14</td>
<td>652.693</td>
<td>heterocyclic</td>
</tr>
<tr>
<td>baccatin IV</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>C32H44O13</td>
<td>572.653</td>
<td>heterocyclic</td>
</tr>
<tr>
<td>9-dehydro-13-acetylbaccatin</td>
<td><img src="image4.png" alt="Structure" /></td>
<td>C31H40O10</td>
<td>538.638</td>
<td>isocyclic</td>
</tr>
<tr>
<td>10-O-Acetyl-5-O-cinnamoylbaccatin</td>
<td><img src="image5.png" alt="Structure" /></td>
<td>C31H38O8</td>
<td>538.638</td>
<td>isocyclic</td>
</tr>
</tbody>
</table>

*Biological Activity:*
- **Rooted cuttings of the Taxus cuspidata**
- **Stems of Japanese yew Taxus cuspidata**
- **Suspension cell cultures of Taxus cuspidata**
- **Taxus x media cv. Hicksii**
- **Taxus cuspidata (barks, branches and needles)**
- **Taxus cuspidata**
- **Taxus x media cv. Hicksii**
- **Washing a raw material (RM) with water**
- **Extracting with an organic solvent a wet RM**
- **Contacting the wet RM**
<table>
<thead>
<tr>
<th>ID</th>
<th>Molecular Structure</th>
<th>Formula</th>
<th>Molecular Weight</th>
<th>Identification</th>
<th>Spectra</th>
<th>Bioactivity/Ecotox</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>7556128</td>
<td><img src="image1.png" alt="Molecular Structure" /></td>
<td>C29H42O8</td>
<td>518.648</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image2.png" alt="Molecular Structure" /></td>
<td>C29H42O8</td>
<td>518.648</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7556128</td>
<td><img src="image3.png" alt="Molecular Structure" /></td>
<td>C29H42O8</td>
<td>518.648</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7685130</td>
<td><img src="image4.png" alt="Molecular Structure" /></td>
<td>C37H46O14</td>
<td>714.764</td>
<td>heterocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7685130</td>
<td><img src="image5.png" alt="Molecular Structure" /></td>
<td>C37H46O14</td>
<td>714.764</td>
<td>heterocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7685130</td>
<td><img src="image6.png" alt="Molecular Structure" /></td>
<td>C37H46O14</td>
<td>714.764</td>
<td>heterocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Journal:***
- Bai, Jiao; Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jinguji; Zhang, Shujun; Hirose, Katutoshi; Tomita, Akhiro; Tsurow, Takashi; Ando, Masayoshi; Journal of Natural Products; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys
- Bai, Jiao; Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jinguji; Zhang, Shujun; Hirose, Katutoshi; Tomita, Akhiro; Tsurow, Takashi; Ando, Masayoshi; Journal of Natural Products; vol. 67; 1; (2004); p. 58 - 63; View in Reaxys
- Ketchum, Raymond E. B.; Richter, Christopher D.; Giv, Deyou; Kim, You Sun; Williams, Robert M.; Croteau, Rodney B.; Phytochemistry (Elsevier); vol. 62; 6; (2003); p. 901 - 910; View in Reaxys
- Shi, Qing-Wen; Petzke, Tracy L.; Saunol, Francoise; Mamer, Orval; Zamir, Lolita 0.; Canadian Journal of Chemistry; vol. 81; 1; (2003); p. 64 - 74; View in Reaxys

---

Copyright © 2010 Elsevier Properties SA. All rights reserved. Reaxys is a trademark owned and protected by Elsevier Properties SA and used under license.
<table>
<thead>
<tr>
<th>Id</th>
<th>Structure</th>
<th>Formula</th>
<th>MW</th>
<th>CAS Number</th>
<th>Bioactivity/Phytochemicals</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>8024907</td>
<td><img src="image1" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Shi, Qing-Wen; Petzke, Tracy L.; Sauriol, Francoise; Mamer, Orvat, Z., Lottia O.; Canadian Journal of Chemistry; vol. 81: 1; (2003); p. 64 - 74; View in Reaxys</td>
</tr>
<tr>
<td>8024907</td>
<td><img src="image2" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujin; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takashi; Ando, Masayoshi; Heterocycles; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>8024907</td>
<td><img src="image3" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujin; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takashi; Ando, Masayoshi; Heterocycles; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>8024907</td>
<td><img src="image4" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujin; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takashi; Ando, Masayoshi; Heterocycles; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>8024907</td>
<td><img src="image5" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujin; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takashi; Ando, Masayoshi; Heterocycles; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>8024907</td>
<td><img src="image6" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujin; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takashi; Ando, Masayoshi; Heterocycles; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>8024907</td>
<td><img src="image7" alt="Structure" /></td>
<td>C33H40O8</td>
<td>564.676</td>
<td>8024907</td>
<td>Isocyclic, rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
<td>Journal: Ito, Nobuhide; Sakai, Junichi; Kitabatake, Mutumi; Fujisawa, Hajime; Bai, Liming; Dai, Jungui; Zhang, Shujin; Hirose, Katutoshi; Tomida, Akhiro; Tsuruo, Takashi; Ando, Masayoshi; Heterocycles; vol. 68; 4; (2005); p. 497 - 501; View in Reaxys</td>
</tr>
<tr>
<td>CAS Number</td>
<td>Name</td>
<td>Molecular Formula</td>
<td>MW</td>
<td>Spectra</td>
<td>Bioactivity/Ecotox</td>
<td>Bioavailability</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------</td>
<td>-------------------</td>
<td>----</td>
<td>---------</td>
<td>--------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>9449731</td>
<td>2a,5a,10p-triacetoxy-taxa-4(20),11-diene; 2a,5b,10p-triacetoxy-taxa-4(20),11-diene; 14-deacetoxytaxusyunnamine C</td>
<td>C26H36O6</td>
<td>562.657</td>
<td>isocyclic</td>
<td>(5) Identification</td>
<td>(3) Physical Data</td>
</tr>
<tr>
<td>7840555</td>
<td>5a-hydroxy-7b,9b,10b-triacetoxy-taxa-4(20),11(12)-taxadien-13-one; taxuspinamane K</td>
<td>C26H36O6</td>
<td>476.567</td>
<td>isocyclic</td>
<td>(6) Identification</td>
<td>(1) Physical Data</td>
</tr>
<tr>
<td>8028189</td>
<td>taxinine E</td>
<td>C37H46O10</td>
<td>650.706</td>
<td>isocyclic</td>
<td>(5) Identification</td>
<td>(1) Physical Data</td>
</tr>
<tr>
<td>8028189</td>
<td>taxinine E</td>
<td>C37H46O10</td>
<td>650.706</td>
<td>isocyclic</td>
<td>(5) Identification</td>
<td>(1) Physical Data</td>
</tr>
</tbody>
</table>

**Journal:** Kobayashi, Jun'ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2 (1998); p. 1111 - 1134. View in Reaxys.
<table>
<thead>
<tr>
<th>ID</th>
<th>View in Reaxys</th>
<th>Name</th>
<th>C37H49N</th>
<th>651.797</th>
<th>C37H49NO9 isocyclic</th>
<th>Identification(5)</th>
<th>Bioactivity/Ecotox(6)</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>6837816</td>
<td>View in Reaxys</td>
<td>taxine II</td>
<td>C37H49NO9</td>
<td>651.797</td>
<td>C37H49NO9 isocyclic</td>
<td>Identification(5)</td>
<td>Bioactivity/Ecotox(6)</td>
<td>2 stems of the Japanese yew Taxus cuspidata Sieb. et Zucc.</td>
</tr>
<tr>
<td>7150289</td>
<td>View in Reaxys</td>
<td>20a,5b,5b-trihydroxy-13b,13d-diacetyl-taxa-4(20),11-diene</td>
<td>C24H36O7</td>
<td>436.546</td>
<td>C24H36O7 isocyclic</td>
<td>Identification(5)</td>
<td>Bioactivity/Ecotox(6)</td>
<td>2 rooted cuttings of the Taxus cuspidata Sieb. et Zucc.</td>
</tr>
<tr>
<td>7236395</td>
<td>View in Reaxys</td>
<td>7β-acetoxytaxinine: taxuspine F</td>
<td>C26H38O10</td>
<td>534.604</td>
<td>C26H38O10 isocyclic</td>
<td>Identification(5)</td>
<td>Bioactivity/Ecotox(6)</td>
<td>the Japanese yew Taxus cuspidata</td>
</tr>
<tr>
<td>7242009</td>
<td>View in Reaxys</td>
<td>taxinine B</td>
<td>C37H44O11</td>
<td>664.75</td>
<td>C37H44O11 isocyclic</td>
<td>Identification(4)</td>
<td>Bioactivity/Ecotox(1)</td>
<td>2 the branches of Taxus cuspidata Sieb et Zucc.</td>
</tr>
<tr>
<td>7788786</td>
<td>View in Reaxys</td>
<td>taxuspine Z</td>
<td>C37H51N09</td>
<td>653.613</td>
<td>C37H51N09 isocyclic</td>
<td>Identification(3)</td>
<td>Bioactivity/Ecotox(1)</td>
<td>2 stems of the Japanese yew Taxus cuspidata Sieb. et Zucc.</td>
</tr>
<tr>
<td>7844183</td>
<td>View in Reaxys</td>
<td>taxuspinane D</td>
<td>C31H38O11</td>
<td>566.636</td>
<td>C31H38O11 heterocyclic</td>
<td>Identification(3)</td>
<td>Bioactivity/Ecotox(1)</td>
<td>2 stems of Taxus cuspidata Sieb. et Zucc. var. nana Rehder</td>
</tr>
<tr>
<td>Identification</td>
<td>Formula</td>
<td>MW</td>
<td>Physical Data</td>
<td>Spectra</td>
<td>Bioactivity/Ecotox</td>
<td>Source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>------</td>
<td>---------------</td>
<td>---------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7685822</td>
<td>C29H36O10</td>
<td>544.599</td>
<td>heterocyclic</td>
<td></td>
<td></td>
<td>Taxus cuspidata Siib et al. Zucc var nan Rehder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7665663</td>
<td>C37H46O11</td>
<td>666.706</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td>Taxus cuspidata Siib et al. Zucc var nan Rehder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8013086</td>
<td>C22H32O6</td>
<td>392.492</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td>Japanese Taxus cuspidata Siib et al. Zucc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8023286</td>
<td>C31H38O8</td>
<td>538.638</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td>Japanese Taxus cuspidata Siib et al. Zucc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>PubChem ID</td>
<td>PubChem Inchi</td>
<td>Molecular Formula</td>
<td>Molecular Weight</td>
<td>Molecular Description</td>
<td>Identification</td>
<td>Physical Data</td>
<td>Bioactivity/Ecotox</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Taxol</td>
<td>8024272</td>
<td>C28H40O10</td>
<td>536.62</td>
<td>isocyclic</td>
<td>FYBMOVYE QAHKMA- W0KHEYH SA-N</td>
<td>(3) Identification</td>
<td>(1) Physical Data</td>
<td>(6) Bioactivity/Ecotox</td>
</tr>
<tr>
<td>Baccatin VI; Taxus Yojiro; Tsuruo, Takashi; Bloorganic &amp; Medicinal Chemistry Letters; vol. 8; 1998; p. 1555 - 1558; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baccatin III; C31H40O11</td>
<td>470415</td>
<td>C31H40O11</td>
<td>586.652</td>
<td>heterocyclic</td>
<td>YCNPMNG GXXWHNHFJLMLRSIS A-N</td>
<td>(3) Identification</td>
<td>(3) Spectra</td>
<td>(2) Bioactivity/Ecotox</td>
</tr>
<tr>
<td>Baccatin III; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxine II; C37H49NO9</td>
<td>687737</td>
<td>C37H49NO9</td>
<td>651.797</td>
<td>isocyclic</td>
<td>UJAPNTWL SVMUJFH4NKTFTC SA-N</td>
<td>(4) Identification</td>
<td>(9) Spectra</td>
<td>(6) Bioactivity/Ecotox</td>
</tr>
<tr>
<td>Taxine II; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuspidin; C47H53NO1</td>
<td>408528</td>
<td>C47H53NO1</td>
<td>608.729</td>
<td>isocyclic</td>
<td>SOSLCGUO OOGMENHNCERTHO SA-N</td>
<td>(4) Identification</td>
<td>(2) Physical Data</td>
<td>(9) Spectra</td>
</tr>
<tr>
<td>Cuspidin; Journal of Natural Products; vol. 61; 4; (1998); p. 474 - 479; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2010 Elsevier Properties SA. All rights reserved. Authorized use only. Reaxys® is a trademark owned and protected by Elsevier Properties SA and used under license.
<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Formula</th>
<th>MW</th>
<th>Bioactivity/Ecotox</th>
<th>Identification/Physical Data</th>
<th>Spectra</th>
<th>View in Reaxys</th>
</tr>
</thead>
<tbody>
<tr>
<td>6181117</td>
<td>baccatin III 13-cinnamate; taxuspinanane J</td>
<td>C40H44O12</td>
<td>716.782</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8246897</td>
<td>taxezopidine J</td>
<td>C35H42O10</td>
<td>622.712</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8246897</td>
<td>taxezopidine J</td>
<td>C35H42O10</td>
<td>622.712</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9434037</td>
<td>5a-acetoxytaxa-4(20),11(12)-diene</td>
<td>C22H34O2</td>
<td>330.511</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9888672</td>
<td>2-deacetyl-4(20),11-deoxytaxine B; 4(20),11-deoxytaxine A</td>
<td>C33H42O7</td>
<td>550.692</td>
<td>isocyclic</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: View in Reaxys indicates that the compound can be viewed in Reaxys database. Copyright © 2010 Elsevier Properties SA. All rights reserved. Authorized use only.
<table>
<thead>
<tr>
<th>Structure</th>
<th>C26H360</th>
<th>492.566</th>
<th>C26H3609</th>
<th>iso cyclic</th>
<th>HMVND0FA GCGZSA-SZMXVPE SA-N</th>
<th>Identification(3) Bioactivity/Ecotox(3)</th>
<th>1</th>
<th>stems of Japanese yew Taxus cuspidata Siib. et Zucc.</th>
<th>Journal; Kobayashi, Jun'ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>C31H400</td>
<td>572.653</td>
<td>C31H40010</td>
<td>heterocyclic</td>
<td>BKASSTPD ANLBH-Q3HYBMO SA-N</td>
<td>Identification(3) Bioactivity/Ecotox(1)</td>
<td>1</td>
<td>stems of Japanese yew Taxus cuspidata</td>
<td>Journal; Kobayashi, Jun'ichi; Hosoyama, Hirokazu; Wang, Xiaoxia; Shigemori, Hideyuki; Sudo, Yojio; Tsuruo, Takashi; Bioorganic &amp; Medicinal Chemistry Letters; vol. 8; 812; (1998); p. 1555 - 1558; View in Reaxys</td>
</tr>
<tr>
<td>Structure</td>
<td>C37H51N09</td>
<td>653.813</td>
<td>C37H51NO9</td>
<td>iso cyclic</td>
<td>CXXBHEBNB EFREKT-W02FEXK SA-N</td>
<td>Identification(3) Bioactivity/Ecotox(3)</td>
<td>1</td>
<td>stems of Japanese yew Taxus cuspidata Siib. et Zucc.</td>
<td>Journal; Kobayashi, Jun'ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</td>
</tr>
<tr>
<td>Structure</td>
<td>C37H46011</td>
<td>666.766</td>
<td>C37H46011</td>
<td>iso cyclic</td>
<td>N0MGXQG PVHIH0V- LUAYAUSYS A-N</td>
<td>Identification(3) Bioactivity/Ecotox(1)</td>
<td>1</td>
<td>stems of Japanese yew Taxus cuspidata</td>
<td>Journal; Kobayashi, Jun'ichi; Hosoyama, Hirokazu; Wang, Xiaoxia; Shigemori, Hideyuki; Sudo, Yojio; Tsuruo, Takashi; Bioorganic &amp; Medicinal Chemistry Letters; vol. 8; 812; (1998); p. 1555 - 1558; View in Reaxys</td>
</tr>
<tr>
<td>Structure</td>
<td>C37H460110</td>
<td>667.797</td>
<td>C37H460110</td>
<td>iso cyclic</td>
<td>GODKREDP NCAKTL- GLACRIUJS A-N</td>
<td>Identification(3) Bioactivity/Ecotox(3)</td>
<td>1</td>
<td>stems of Japanese yew Taxus cuspidata Siib. et Zucc.</td>
<td>Journal; Kobayashi, Jun'ichi; Shigemori, Hideyuki; Heterocycles; vol. 47; 2; (1998); p. 1111 - 1134; View in Reaxys</td>
</tr>
<tr>
<td>Structure</td>
<td>C47H61N01</td>
<td>848</td>
<td>C47H61NO1</td>
<td>heterocyclic</td>
<td>CXIWJTIMC KSBSL-Y3RHRHNIJ SA-N</td>
<td>Identification(3) Bioactivity/Ecotox(1)</td>
<td>1</td>
<td>stems of Japanese yew Taxus cuspidata</td>
<td>Journal; Kobayashi, Jun'ichi; Hosoyama, Hirokazu; Wang, Xiaoxia; Shigemori, Hideyuki; Sudo, Yojio; Tsuruo, Takashi; Bioorganic &amp; Medicinal Chemistry Letters; vol. 8; 812; (1998); p. 1555 - 1558; View in Reaxys</td>
</tr>
<tr>
<td>Compound</td>
<td>Molecular Formula</td>
<td>Molecular Weight</td>
<td>Homology</td>
<td>Identification</td>
<td>Bioactivity/Ecotox</td>
<td>Source</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------------</td>
<td>------------------</td>
<td>----------</td>
<td>----------------</td>
<td>-------------------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-deacetyl taxinine</td>
<td>C33H40O8</td>
<td>554.676</td>
<td>heterocyclic</td>
<td>Identification(4)</td>
<td>Physical Data(1) Spectra(4)</td>
<td>Journal: Sh., Qing-Wen; Oritani, Takayuki; Zucker, Ryo; Oritani, Takayuki; Phytochemistry (Elsevier); vol. 63; 8; (1999); p. 924 - 929; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a,7b,10b,13e-pentacacetoxoy-4p,20-epoxy-taxa-11-en-1p-ol</td>
<td>C39H48O14</td>
<td>740.802</td>
<td>heterocyclic</td>
<td>Identification(4)</td>
<td>Physical Data(1) Spectra(4)</td>
<td>Journal: Sh., Qing-Wen; Oritani, Takayuki; Zucker, Ryo; Oritani, Takayuki; Phytochemistry (Elsevier); vol. 63; 8; (1999); p. 924 - 929; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2'-hydroxytaxinine</td>
<td>C35H42O10</td>
<td>622.712</td>
<td>isocyclic</td>
<td>Identification(3)</td>
<td></td>
<td>Journal: Murakami, Ryo; Shi, Qing-wen; Oritani, Takayuki; Phytochemistry (Elsevier); vol. 52; 8; (1999); p. 1577 - 1580; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-cinnamoyltaxicin-triacetate</td>
<td>C37H44O11</td>
<td>664.75</td>
<td>isocyclic</td>
<td>Identification(3)</td>
<td></td>
<td>Journal: Murakami, Ryo; Shi, Qing-wen; Oritani, Takayuki; Phytochemistry (Elsevier); vol. 52; 8; (1999); p. 1577 - 1580; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a,9a,10b-triacetoxy-5p-cinnamoyltaxa-4(20),11-dien-13a-ol</td>
<td>C35H44O9</td>
<td>608.729</td>
<td>isocyclic</td>
<td>Identification(4)</td>
<td>Physical Data(1) Spectra(6)</td>
<td>Journal: Sh., Qing-wen; Oritani, Takayuki; Zhao, Ding; Murakami, Ryo; Oritani, Takayuki; Planta Medica; vol. 66; 3; (2000); p. 294 - 299; View in Reaxys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8604709</td>
<td>View in Reaxys</td>
<td>2a,7b,9a,10b-tetraacetoxy-6b-[2R,(3S)-N-dimethyl-3-phenylisoserioxy]-taxa-4(20),11-dien-13-one</td>
<td>C39H51NO1 AOCT-</td>
<td>Identification(4) Physical Data(1) Spectra(7)</td>
<td>1</td>
<td>seeds of Taxus cuspidata Sie et Zucc, collected in Toyama Prefecture, Japan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8667237</td>
<td>View in Reaxys</td>
<td>5a-cinnamoyl-2a,9a,10b-tetraacetoxy-taxa-4(20),11-dien-13-one</td>
<td>C39H51N</td>
<td>Identification(4)</td>
<td></td>
<td></td>
<td>needles of Japanese Yew, Taxus cuspidata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8668050</td>
<td>View in Reaxys</td>
<td>5a-cinnamoyl-2a,7b,9a,10b-tetraacetoxy-taxa-4(20),11-dien-13-one</td>
<td>C37H44O12</td>
<td>Identification(4) Physical Data(1) Spectra(6)</td>
<td>1</td>
<td>needles of Japanese Yew, Taxus cuspidata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9457045</td>
<td>View in Reaxys</td>
<td>2a,7b,9a,10b,13a-pentaacetoxy-taxa-4(20),11-dien-13-one</td>
<td>C37H46O10</td>
<td>Identification(4)</td>
<td></td>
<td></td>
<td>suspension cell cultures of Taxus cuspidata (P33A) and Taxus x media cv. Hick's</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9461829</td>
<td>View in Reaxys</td>
<td>8-dihydro-13-acetylbaccatine III</td>
<td>C37H46O12</td>
<td>Identification(4)</td>
<td></td>
<td></td>
<td>suspension cell cultures of Taxus cuspidata (P33A) and Taxus x media cv. Hick's</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9452532</td>
<td>View in Reaxys</td>
<td>5a,7b,9a,10b,13a-pentaacetoxy-2a-benzoyloxytaxa-4(20),11-dien-13-one</td>
<td>C37H46O12</td>
<td>Identification(4)</td>
<td></td>
<td></td>
<td>suspension cell cultures of Taxus cuspidata (P33A) and Taxus x media cv. Hick's</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2010 Elsevier Properties SA. All rights reserved. Reaxys® is a trademark owned and protected by Elsevier Properties SA and used under license.
<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Name</th>
<th>Structure</th>
<th>Molecular Formula</th>
<th>Molecular Weight</th>
<th>Physical Data</th>
<th>Bioactivity</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>11195631</td>
<td>2a,14p-diacetoxy-10β-ethoxystaxa-11,4(20)-dien-5α-ol</td>
<td>![Structure Image]</td>
<td>C26H40O6</td>
<td>448.6</td>
<td>Identification(4)</td>
<td>Spectra(4)</td>
<td>Heartwood of Taxus cuspidata collected in Toyama Prefecture, Japan</td>
</tr>
<tr>
<td>11195632</td>
<td>2a,14p-diacetoxy-10β-methoxystaxa-11,4(20)-dien-5α-ol</td>
<td>![Structure Image]</td>
<td>C25H38O6</td>
<td>434.573</td>
<td>Identification(4)</td>
<td>Spectra(4)</td>
<td>Heartwood of Taxus cuspidata collected in Toyama Prefecture, Japan</td>
</tr>
<tr>
<td>11195633</td>
<td>2a-anthoxy-10α-ethoxystaxa-11,4(20)-dien-5α,14β-diol</td>
<td>![Structure Image]</td>
<td>C24H38O5</td>
<td>406.563</td>
<td>Identification(4)</td>
<td>Spectra(4)</td>
<td>Heartwood of Taxus cuspidata collected in Toyama Prefecture, Japan</td>
</tr>
<tr>
<td>11298427</td>
<td>2o,5α,7β,9α,13α-pentahydroxy-10β-acetoxytaxa-4(20),11-diene</td>
<td>![Structure Image]</td>
<td>C22H34O7</td>
<td>410.508</td>
<td>Identification(4)</td>
<td>Spectra(4)</td>
<td>Needles of Taxus cuspidata collected in Tsutsuzigaoka Park, Sandai city, Japan</td>
</tr>
<tr>
<td>15840365</td>
<td>O-deacetyltaxine II</td>
<td>![Structure Image]</td>
<td>C35H47N O8</td>
<td>609.76</td>
<td>Identification(3)</td>
<td>Spectra(4)</td>
<td>Stems, leaves and seeds of the Japanese yew Taxus cuspidata Sieb. et Zucc. (Taxaceae) collected at Sapporo, Hokkaido</td>
</tr>
<tr>
<td>15840266</td>
<td>taxezopidine</td>
<td>![Structure Image]</td>
<td>C37H49NO10</td>
<td>667.797</td>
<td>Identification(4)</td>
<td>Spectra(4)</td>
<td>Stems, leaves and seeds of the Japanese yew Taxus cuspidata Sieb. et Zucc. (Taxaceae) collected at Sapporo, Hokkaido</td>
</tr>
</tbody>
</table>

Copyright © 2010 Elsevier Properties SA. All rights reserved. Reaxys® is a trademark owned and protected by Elsevier Properties SA and used under license.
<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>System</th>
<th>Molecular Formula</th>
<th>Molecular Weight</th>
<th>Structural Identification</th>
<th>Physical Data</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>15882421</td>
<td>Taxine NA-13</td>
<td>C37H47N O10</td>
<td>665.781</td>
<td>C37H47NO1</td>
<td>1</td>
<td>Identification(4) Physical Data(2) Spectra(3)</td>
<td>the needles and young stems of Taxus cuspidata collected in Abayama, Sendai, Japan</td>
</tr>
<tr>
<td>15381179</td>
<td>Taxezopidine O</td>
<td>C36H49N O9</td>
<td>639.786</td>
<td>C36H49NO9</td>
<td>1</td>
<td>Identification(3) Physical Data(2) Spectra(9)</td>
<td>seeds of Taxus cuspidata Sieb. et Zucc.; collected at Sapporo, Hokkaido, Japan</td>
</tr>
</tbody>
</table>
Appendix B: *Taxus* cDNA library

I. Cytochrome P450s

a. H1

MDTFIQHESSPLLTLTAVILGTLILLILSGKQYRRESSRKLPPGNMGFPLIGETIALISDTPR
KFIDDRVKKFGLVFKTSLIGHPAPVICGSSANRFLLSLNEEKLVRMSLPNAVKLGLGQDCV
MGKTGVEHGIVRTALARALGPQALQNYVAKMSSEIEHHINQKWGKDEVKLPLIRSL
VFSISTLFFGINDEHHQKRLHLLETVMGLVSIPLDLFGPTRFKLYARSLDEIMMS
VIERRRSDLRSAASSDQDLLSLVTFTKDERGNSFADKEILDNSFLHLALYHTISPLT
LIFKLLSSSPECYENIAQEQLIEILGNKKDREEISWKDLKDMKYTWQAVQETLRFPPVY
GYIREALTDIDYGTTIPKGRALCPHTTHSKEEFDEPEEFPSRFEDQGRHVAPYT
FIPFGGGLRICAGWFAKMEIILFMHDFVKTFSHFDPNEKISRDPLPPPVPVKGFSIKP
FPRSP

b. H2

MLIEMDTFVQLESSPVLTLTLILLFIFCSKQYRSSKLPPGNMGFPLIGETIALASQTP
DKFFGDRMKKKFGKVKTLQHPTIVLCGSSGNRFLLSLNEEKLVRMFPNNSSSKLLGQ
DSVLGKIGEEHRIVRTALARCLGPQALQNYVSMSSEIQRHINQKWGKGEVKMLPLIR
SLVFSIATLFFGITDEQQQERLHHLETVRTGLLCIPLDLPGTTFRKALHARSKLDEIMS
SVIEERRNDLRLGAASSQDQDLSVLLTFKDERGNPFADKEILDNSFLLLHALYDTTISPL
TLVFKLVSNPECYENIAQELOLEIRRNKKGDEISDIALKVMKTYTWQAVQETLRMCPP
VYGNFRKALTDIHYDGYTPKGWRLCSPYTHSKVEYDDPEKKFPRFSRFEEQGRDVA
PYTFIPFGGGGLRCPGREFAKMEILVFMMH芙VFKAFSSFPVDPEKISTDPSIPVPNFS
INLVPRS

c. H3

MDSFSFLKSMEAKFGQVIHRDQSSSTALLSALTAAVAIFLVLFRFKSRPSTNFPPP
FGFPFGETIQFLRALRSESPHMFFDERLKKFRGVFKTSLTGHTAVFCGPAGNRFIYS
NEHKLVQSSGPNSFVKLGVQQQIVTGTKGEEHRIFGLVLEFLGMHALQSYTPKMSSKIQ
ENINKHWKGDDEVNMLPSIRQLFSSISSLFFDINDDEQEQQLKTLLETILVGTLSVPLDI
PGSNFRKALRARSKLDEILSRLIESRRKDMRGIASTSKNLLSVLAFKDERGNPLTDTE
ILDNFSFMLHASYDATTVSPTCPUKLLSANPECYEKVVQEQQLGILGNKKDGEMCWDNL
KAMKYTWQAAQETMRLFPAPAFGSFRKVIADIHHDGYIPKGWKAMVTYNSTSRKEEYF
DEPNFKPSRFGDGYVAPYTFLPGAIGIRCPEGWEFKALMILLFIHHFVKNFSGYP
LDTKEKISGDPFPPLPKNGFPIKLFPR-}

d. H4

MDALSLVNSTVAKFNEVTQLQAASPAILSTALTAIAGIIVLVLVITSKRRSSLKLPPGKLGLPF
IGETLEFKALRSQDLRQFVEEREKGFKGRVFKTSLLGKPTVILCPAGNRVLVLSNEEKL
HVSWSAQARIILGLNSVAVKRGDDHRVLRLVALAGFLGSAGLQLYIKMSALIRNHINEK

178
WKGKDEVNVLSLVRDLVMDSAILFFNIYDKRQKQLHEILKIILASHFGIPLNIPGFLYR
KALKGSLKRKILSALLEKRKDELRSRSSLASSNQDLSSVLSSRDERGKGPSDEAVLDNC
FAMLDASYDTTTSQMELKLMLSSNPECFEKVQEQLIAENKKEEGEEITMNDKIKAMKY
TWQVLQESLRLSMSPVFGTLLRKTMDNINHDGYTIPKGWQTVVTTYSTHJKDIYFQPD
KFMPSRFEEDGHLDAYTFPVPGGGRRTCPGWEAYAKEILLFLHHFKAFSGYTPTPDP
HERICGYPVPLVPKVKGFPKLIARS

e. H5

MDTIRASFGGEVIQPEYESPLISVALAAFLGIVIFSIFSSTRESYVNLPPGNLGLPFIGETIQF
LGALQSEKPHTFDERRVKFGKVKTSLIDPTVVLCPAGNRLVLSNEDKKVQSAGP
KSFLKLFGESVAAKREESHRILRSLGRFLGPHALQNYIGKMNSEMQRHDDKWKG
KEVKEVLPLVRGLESFATSLFFNINDRQREQLHLGLDDTILVSGMTIPLNIPGTFLRKAVK
ARAKLDEILFALIENRRRELRSGLNSGNQDLSSLTFLKDEKGNPLTDKEILDNFSVMLH
ASYDTEVSPTVLILKLLASNPECYEKVVQEQGLILASKKEEGEENVWKDLKAMPYTWQAI
QEPFLRMPQQLLECEFEEASLIFSWKAIQFQKGQCLGQILVNGREEFFNFPDKFKPSHE
EGKPLDPIFIFPGAGVRICAGWEFAKAELLFLVFHFPVKNFGCIIIDNPEKISGDPFPPL
PTSGQLMKLIPRS

f. H6

MDSFNFLRGIGADFGGFIQFQSSPAVLSSLITTILGVLLLWFFLHHKNGSSVTLPPGNLG
FPFIGETIPFLRALRSETQPFTFDERRVKGFFKTRVGHPTVVLCPGEGNRFFLSNED
KLVQASLPNSSEKLIGKYSILSKRGEEHRILRAALARFLRPQALQGYVAKMSSEIQHHIK
QKWKGNEDEVKVLPIRRTLIFNIASSLFFGINDEHQEQLHELLEALVGLSLSVPLDFPGT
RFRKALDARSKLDEILSSLMESRRDDLRLGTASENQDLLSVLLTFKDERGPNLTDKEIF
DNFSFMLHASYDTTVSPTGLMLKLLFSSPDCCYEKLQVEQLGIVGNKKEGEEISWNDLK
AMKYTCKVVQESMRMLPPVFGSYRKAITIHYDGYTIPKGWNIWSPYTTHGKEEYFN
EADKFMPSRFEEGKYVAPYTFLPGAGLRVCPEGWEFAKTEILLFVHHFITTFSSYIIPDP
KDKISGDPFPPLPTNGFSMKLFTRS-

g. H7

MDTLIQIQSSPDGLSTLTAVVLVLLIFRYKHRSALKLPPGNNLGLPFIGETITFASQPP
QKFLNERGKKGFPVFKTSLIGHTPVLCGSSGNRFLLSNEEKLVRMSLPNSYMKLGLQ
DSSLGKTGQEHRIVRTALGRFLGPQELQNHVAKMSSDIQHHINQKWKGNEVKVLPLI
RNLVFSIATSLFFGINDEHQQRLEILHLLLLETVMGAVCIPLAFPGSGFRKALQARSELGIL
ISLMKIRRSDLRSAASSNQDLLSVLLTFKDERGPNLTDKEILDNFSVLLHGLYDITTISPL
TLIFKLMSSNTCEYENVQVEQLEILSHREKGEEIGWKLKSMKTTYWQLAIQETLRFMPPPV
YGNFRKALTDIHYDGYTIPKGWRVLCSPFTTHSNEEYFNEPDEFRPSRFEGQGGKNVPS
YTFIPFGGGGLRICPGWEFAKTEMLLFIHYVFKTFFSYPVPDPNEKISADPLASFPVNGFS
VKLFPRS

h. H8
MDAFNILGPAAKLNGVQLGSYTDRISSITVVAFITILLLLMLRWSQSSVKLPPGNGFG
FPLIGETLQLLRAFRSNTTQFFDERQKFGCVFKTSVLGERTVVLCPGSPGNRLVLAN
QNKVVESSWPSAFIKLIGEDSIANTNGEKHRILRAALLRLYLGPSGLQNYVVGKMREIEH
HINEKWKGDKDEVKLVDVLVRKNVFSVATALFFGVPNDEERKRIRPPSILRLKHFAOSFSIPL
DFPGTSSYRLAEALKLDKLSSLISERSEEMLRLSGLASGNDLVSVTLLTFKDEGGNPLT
DKEILDNFSGLLHASYDITTSALTLLTFKLMSISSAECYDKVVEQLRIVSNKKEGEEISLKL
DLKDMKTYTQVQETLMFPLFGSFRKTIADIQYDGYTIPKGWKVLWATYTTGRDE
YFSEPQKFRPSRFEEGGKHVAPYTFLPFEGGERTCPGYFESKTHILLFIHQVFVKFTGY
IPLDNPESISANPLPLPLANGFVKLQFRS-

i. H9

MDSFIFLRSIGTKFGQLESSPAILSLTLAPILAIILLLLFRYNHRSVKLPPGKLGFPLGETI
QLLRTLSETQPKFFDDRLKKFGPVYMTSIGHPTVVLCPGAPNKLVLSNEDKLVEME
GPKSFMKLIGEDSIARKGEDHRLRLTRLARLGLAQALQNYLGRMSSEIGHHFNKWK
GKDEVKLPLVRLGFSASTLFFDVNDGHQKQLHLLETILVGLSSVLDFPGRTYRKG
GLQARKLDEILSSLIKRRRRDLRSGIASSQQDLSSVLLTFRDEKKNSLTDQGILDNFSAMFHASYDTPVAPMALIFKLLYSNPEYHEKVQFQLEIIGNKKEGEEISWKLDSMKYTW
QAVQESLRMYPPVFIFRKAITDIHYDGYTIPKGWRLCSPYTTLREEYFPEPEEFRP
SRFEDEGRHVTPYVFFGGLRTCPGWEFSKIEILLLVHHFVKNFSSYIPVDPNEKVL
SDPLPPLPANGFSIKLFPRS
j. H10

METKFGQLMQLFLEFLPFLTPIGLALVLLHLFRHRNRSSVKLPPGKLGFPGVIGETIQFLRAL
RSQTPQKFFDDRVQKFGVFKTSLIGNPLVVMCGPAGNRLVLSNEDKLVLQLEAPNSL
MKLMGQDSLAKRQEDHRTRALARFLGPGQALQNYMTKISSRTEHHMNEKWKGKD
EVRTLPLIRELIFSNASSLFFDINDEHQOERLHHLLEAVVVGSMSIPLDFPGTRLRKALQ
ARSKLDIEILSSLIKSKRRKDLVSGIASDDQDLLSVLLTFKDERGNPLTDKEILDNFSLLLHA
SYDDTTVSPMVLTLKLSSNPECYEKVVQEQLGIVANKRIGEESWKDLKAMKYTWQVV
QETLRMFPPFGSFRKAMVIDYDYTGYPGWMILWTTYGTHLREEYFNEPLKFRPSR
FEEDGRVTPTVFIPFGGARTCPGEFSKTEILLFIHVHFVRTFSSYLPDSNEKISADPF
PPLPAGFSIJKLASADPFPLPANGFSIKLFPRSQSN

k. H11

MDVFYPLKSTVAKFNECFPAILFIVLSAVAGIVLPLLLFLRSKRRSSVGGLPPGKLGYPFIG
ESLLFLKALRSNTVEQFOLDERVKNFGNVFKTSLIGHTPVLCGPAGNRLILANEEKLVQ
MSWPKSSMKLMGEXSITAKRGEGHMIIRSLQGFFSPGALQKYIGQWMSKTENHNEK
WKGNDOVSVMVVALVGDVFISACLFFNINENKHERERELFELIIEIAVGVLAVPVDPDGFAY
HRALQARSLNAILSLHEKREKMDLSSGLATSNQDLLSVLTFKDFRNGPCSDEEEILDN
FSGLLHGSYDDTTSAMACVFKLSSNPECYEKVVQEQLGILSNKLEGDEITWVDVSM
KTYTWQVVQETLRRLYPFSFRQAITDIHYNGYIIPKGWKLWWTYTPTHPKEMYFSEPE
KFLPSRFDAQEGKLAVAPYTLFPGGGQRSCPGWEFSKMEILLSVHFVKTFSTFTPVDP
AEIIARDSLCPLPSNFGSVKLFPFPRSYSLHTGNQVKKI

i. H12

MDALKQLEVSPSILFVTAVMAGIILFFRSHSSVKLPPGNLGFPFLVGETLQFVRSLGS
STPQQFIEERMSKFGDVFKTSIIGHPTVVLCPAGNRLVLSNENKLVQMSWPSSMMKL
IGEDCLGGKTQEGHRIVRAALTRFLGPQALQNHFAKMSGGIQRHINEKWKGKDEATVL
PLV/KDLVFSVASRFLLFFGITEEHQELQHLNLEVILVGSVPLNIGFSYHKAIQARATLA
DIMTHLEIKRRNLRAGTASENQDLSVLLFTFTDERGNSLADKEILDNFSLLHGSYDS
TNSPLTMLIKVLASHPESYEKVAQEQFGLSTKMEGEEIAWKDLKEMKYSWQV/VQETL
RMYPPIFGTFRKAIDDIHYNGTIPKGKWLWWTYSTQKEEYFKDADFQFKPSFEEEG
KHVTPYTLPFGGMRVCPGWEAKMETTLLFLHHFVKAFSGLKAIMPNKLSGKPLPP
LPVNGLPKLYSRS

m. H13

MDSFTFVTIKMGKIWQVIQEYILSLLTAILLLFFRYRNKSSSHKLPPGNLGFPFIETIQF
LRSLRSQTEFFFDERVKKFGPVFKTSIGAPTVICGAAGSRVLVSNEDKLVQMESPSS
SLKKLMGNSILYKREEEHRLRALSFLGPGQALQTYIAKMSTEIERHINEKWKGKEEV
KTLPLIRGLVQSSASSFLFDINDEPQQERLHHHLVESLVAEGSMARLDFPGTRFRKAVEA
RSKLDEALHSIKSRRDSLSSGKASSQDLSVLLFSKDERGNPLRDEEILDNFSLILHA
SYDTTTISPMVLTLKLLSSNPECYDQVQEIQFGILANKKEGEEISWKDLKAMKyTQVV
QETLRMFPPFLGSFRKAMVDMINYDGYTIPKGWIVLWTYTHVKEEYFNEPGKFRPSRF
FEHDGRHVPAYTFPLPGGLRTPGWEFSKTEILLFIHHFVKTFSYGLPDPNEKISAD
PFPPLPANGFSIKLFPFRS

n. H14

MDALYKSTVAKFNEVTQLDCSTESFSIALSAIAGILLILLLLLFRSRHSSLKLPPGKLGPFI
GESFIFLRALRSNSLEQFFDEVRKFKGFLVKTSLIGHTPVTVLCPAGNLRLSNEEKLVQ
MSWPAPQFMKLMGENSATRGRGDIHVMRSAEALAGFFGPGALQSYIKMNTEIQSHINE
KWKGKDEVPVLPLVRELVFNISAILFFEYDKEQDRHLKLEEILVGSFALPDPIDPGGF
HRALQGRAKLNKIMSLIKKKREDGCSLDQQPRRISCLFSCLSEMSTKGLPHPMEILDN
FSSLHLASYDFTTSPMALIFKLLSSNPCYQKVQEQLEILSNKEEGEEITWKDLKAMK
YTWQVAQETLRMFPPVFGTFRKAIDDIYDGTNSKRGPRLWTYTHPKDLYFNEPEK
FMPSRFDEQGKHVPAYTFPLPFGGQRSCVGWEFSKMEILLFVHHFVKTFSYTPDP
DEKISGDPLPLPSKGFSIKLFPFRP

o. H15

MDAMDLTVAKFKEFTQLQSSAILTVSGIVVIDLLRSSKRRSSLKLPKGKLGPLLIGESL
SFLWALRSNTLEQFVDKRVKKYGNVFKTSLLQPTVLCAAGNRLLSNQEKLLSR
VSDRVAKLTGDTSISVIAGDSHRRIIRAAVAGFLPGAPLKHIGEMSAHIRNHINQVWKGK
DEVNVSLARELVFAMSASLFLNINDREEEHQLHKTLETILPGYFSVPINFPGFAFKAL
EGNSKRRKHFSVLQEKRRDLVSLASRTQDLLSVLLEYEDDKGNPLTDEEVLDNISA

184
p. H16

II. Acyltransferases

a. A1

MGRFNVDMIERVIVAPCLQSPKNILHLPIDNKTRGLTNSVNASQVSVPQADPAKTI
REALSKVLVYYPPFAGRLRTENGDEVCTGEGAVFVEAMADNLSVLQDFNEYDP
SFQQLVFNLREDVNLHLLTVQVTRFTCGGFVVGTRFHHSVDGKGIGQLLKKMGE
MARGEKFPSLEPIWEN MVKPDIMYMQLQFDHFDFIHPPP LNLEKSIQASMIVSFERINYIK
RCMMEECKEFFSAFEVVVALIWLARTKSFRIPPNEYVIKIPIDMNSFDSPPLPGYGYGN
AIGNACAMDNVKDLLNGSLLYALMLIKKSKFALNENFKSRILTKPDSTLDANMKHENVG
CGDWRNRLGFYEADFGWGNANVNVSPMQQQREHELMQNYFLFLRSAKNMIDGIKILMF
MPASMVKPFSKIEEMEVTINKYVAKICNSKL

b. A2

MEKTDLHVNLIEKVVMGSPPLPKTTLQSSIDNLPGVRGSIFNALLIYNASPSPTMISA
DPAPKIREALAKILVYYPPFAGRLRETEGNDLEVCTGEMAMFLEAMADNELSVLVGDF
DDSNPSFQQLLFSLPLDTNFKLSSLVQVTRFTCGGFVVGVSFHHSVCDGRGAQF
LKGLAEMARGEVKLSLEPIWENRELVKLDKYLQFFHFHFNEFLRAPSIVEKIVQTYFIIDFET
INYIKQSVMEECKEFCSSFEVASAMTWIARTRAFQIESEYVKILFGMDMRNSFNPPLP
SGYGNISIGTACAVDNVQDLSGSSLRAIMIJKSKVSLNDFNKSRAVVKPSKEDVNMN
HENVAFADWSRGLGFDEVDFGWGAIVSVSPVQQQSAQMNYFLFLKPSKNKPDGI
KILMFPLSKMKSFKIEMEAMMKKYVAKV

c. A3

MAGSTEFVVRSLERMVAPSQPSKPAFLQLSTLNLPGVRENFNTLLVYNASDRVSV
DPAPKIRVQALSKLVLVYSPFAGRLRKKENGDLVECTGEGALFVEAMADTDLSVLGDL
DDYSPSLQLLFCLPPDITEDIHPLVVQVTRFTCGGFVVGVSFCHGIDCDLGAGQFLI
AMGEMARGEIKPSSEPIWKRELLKPEDPLYRFQYYHFQLICPPSTFGKIVQGSLVITSET
INCIKQCLREESKEFCSAFEVVSALAWIARTRALQIPHSENVKLFAMDRKLFNPPLSK
GYYGFIVGTVCAMDNKVSLSGSLLRVVRRIIIKAKVSLNEHTSTIVTPRSGSDESINYE
NIVGFGRRLRGLGFDEVDGFVGVHADNVSLVQHLKDVSVQVQYFLIFIRPPKNNPGIKI
LSFMPPSIVKSFKFEMETMTNKYVTKP
d. A4
MEKSGSADLHVNIIERVVVAPCQPTPKTIQLQSSIDKMGGFGANVLLVFQASHGVSDP
AKTIREALSKTLVFYFPFAGRLRKKEDGDIEVECIEQGALFVEAMADNDLSVRDLDEY
NPLFRQLQSSLSDLTDYKDLHLMTQVPFTPCTCGFVMGTSHQHSCIDGNGLGQFFKS
MAEIVRGEVKPSIEPIWNRELVKPEDYIHQLQLYVSEFIRPPLVEKVQGTSVLFHISEFIKHI
KRCIMEESKSFSSFEIVTAMVWLARTRAFQIPHEDVTLAMDARRSFDPPIPKGYY
GNVIGTTEAKDNVHNLLSGLHALTVIKSMSSFYENMTSRLVNPSTLDLSMKYENV
VSLSDWRLGHNEVDFGWGNAINVSTLQQWENEVAIMPTFFTLQTPKNIPEGIKILMF
MPPSREKTFIEIEVEAMIRKYLTKVSHEL

c. A5
MKKTGSAEFHVNMIEFMVRPCLPSPKTLPLSAIDNMARAFSNVLLVYAHANMDRVSA
DPAKVIREALSVKLVYYPFAGRLRNKENGGELEVECTQGVLFLEAMADSDLVLTDL
DNYNPSFQQLIFSLPQDITEDLHLIIVQVRFTCGFVVGANVYGSACDAKGFGQFL
QSMAEMARGEVKPSIEPIWNRELVKLEHCPFRMSHLQIIFAPVIEEKFVQTSLVINFEII
NHIRRRIMEERKESLSSFEIVAALVWLAKIKAFQIPHSENVKLLFAMDLRRSFNPPLPHG
YYGNAFGIACAMDNVHLLSGLLRTIMIKKSKFSLHKELNSKTVMSSSVVDVNTKFED
VVSISDWRHSIYYEVDGFWGDAMNVSTMLQQQEHEKSLPTYFSLQSTKNMPDGIKM
LMFMPPSKLKKFKIEIEAMIKKYVTKCPSKL

f. A6

MEKAGSTDFHVKKFDPVMVAPSLPSPKATVQLSVVDSTICRGIFNTLLVFNAPDNISA
DPVKIIIREALKVLVYYPFLAGRLRSKEIGELEVETCHGDGALFVEAMVEDTISVLRDLD
LNPSFQQLVFWHPIDTAIEDLHLIVQQVTRFTCGGIAVGVTLPHSVCDGRGAAQFVTAL
AEMARGEVKPSLEPIWNRNPPEDPLHQLNQ5FDSCPMPLEELGQASFVINVTDIE
YMKQCVMECECN SECVVAALVWIARTKALQPIHTENVKLFAMDLRLKLFNPPLPN
GYYGNAIGTAYAMDQ DLDNGSLLRAMIMIKKAKDLKDNYSRVTVTNPYSLDVNKK
SDNILALSDWRRLGFYEADFGWGGGPLNVSSLQRLENGLPMFSTFLYLLPAKNKSDGIK
LLSCMPPTTLKSFKIVMEAMIEKYVSKV

g. A7

MEKGNASDVPHELHVQICERVVMVKPCVPSPSNLVLQLSAVDRLPGMKFATFSAVLVYN
ASSHSIFANPAQIIRQALSKVLQYYPFAGRIRQKENEELEVETCHGELFVEALVNDNL
SVLRDLDONAQNYEQLLPSLNPIQVQDLHPILQVRFTCGGFVVGFGHGCADARG
GTQFQLQGLADMARGETKPLVEPWNRELKPEDLMHLQFHKFGLIRQPLKLDIECQAS
FTINSEIINYIKQVIEECNEIFSAFEVVVALTWIARTKAFQIPHNENVMMLFMDARKYF
NPPLPKGYGYGNAIGTSCVIENVQDLLNGSLRAVMITKSKIKPLIENLRSRIVANQSGVD
EEIKHENVVGFDWRRLGFHEVDFGSGDAVNSIPOQQRLEDQMLARMNKLFLRPYKD
MPNGIKILMFMDPSSRVLKFKDEMEAMIKYMPKA

h. A8

MEKLHVDIIEVKVAPCLPSKEIQLSLLDNLRCYVSFLVYVRSTVSANPAKTIREA
LSKVLVYYSPFAGRLRNKENGDLEVECGSVEAVAFVEMADNELSVLQLDEYCTSLK
QLIFTVMDTKIEDLHLSVQTVTCTGGGFCVGSFYHTICDGKGLGQFLQGMSEISKGA
FKPLSLPVPNREMVKPEHLMLQFNFEFVPHPKFKKIVKASIELFETINCFKQCM
EECKENSTFEIAALIWLAKTKSFQIDPSENKLMFAVDMMRTSDFPPPKGYYGNIVGI
AGAIDNKELSSGLRLIIQIKTIKSLKDNFISRRLMKPSTLDVNMKHESETVLLGDWRN
LGYYEADCGCGNLNSVIPMDQQIHESPVQSRFMLLLRRSKMNMQNGIKLMSMPESMA
KPFKSEMKFTIKKYVTGACFSEL

i. A9

MEKAGSSTEFHVKISDPVMPCPPSKTILQLSAVDNYPAVRGLNLDDLVLVNASNTIS
ADPATVIREALSKVLVYYFPFAQRMRNKGDELEVDCGTGEALFVEAMADDNLSVLG
GFYHDNPAGKLLYLDSPLTDPIHDLHPLVQTVTCTGGGFCVGGLSLEDHSICDGRGAGQ
FLKALAEMARGAEKPSLEPIWNREIIPDLRLQFYHESMRPPPIVEEIVQASIIVNSE
TISNIQYMEECECSSFAFEVVAALAWLARTRAFFIPHTENVKLLFAVDTRRSDFPPPLP
KGYYGNAAGNACAMDNVQDLNLGSLRAVMIKSSKVSLNENIRAKTVMRPSAIDVNM
KHESTVGLSDLRHLGFNEVDFGWGDALNASLVQHGVIQQNYFLFLQPSKNMNGGIKIA
MFMPQSKVPKPFKIEMEALISKYATKV

j. B (benzoyltransferase)

MGRTNVDIEMVRAPCLQSPKNILHLPIDNKTGRGTLTNSVYNASNRVSVSADPAKT
RELSKVLVYPPFAGRLRTNEDGDEVECGEVGVEAMADNDLSVLQDFNEYDP
SFQQLVFNLREDVNPEDHLITVQVFTRCTCGGFVVGTRFHHSVSDGKGGQLLLKGMGE
MARGEFKPSLEPIWNREMVRKPDIMYLQFDHFDFIHPLNLEKSIQASMVISFERINYIK
RCMMEECKEFFSAFEVVALIWHARSKGPSFRIPPNEYVKIIIFPIDMRNSFDSPLPKGYGN
AIGNLACAMDNVZLDLLNGSSLYALMLIKKKSKFALLENFSDKRLTKPSTLDANMKHENVVG
CGDWRNLFYEDDFGWGNAVNVSMPQQQREHELAMQNYFLFLRSAKNSIDIKILMF
MPASMVKPKFIEMEVTINKYVAKICNSKL

III. Cytochrome P450 reductase

a. CPR1

MQANSNTVEGASQGKLLDISRLDHIFALLLNNGKGGDLGAMTGSALILTENSQNLMLT
TALAVLVAFCVFFVFWRGSDTQPVPRTPLKEEDEEEDDSAKKKTIFGFTQTG
TAEGFAKALAAEAKARYEKAVFKVLDNYAADDQYEKELKKEKLAFMLATYGDGE
PTDHAARFYKWFLEKREPWLDSLTYGVFLGNRQYEHHFNKVAKADVLEIQGAK
RLVPVGLGDDDQCIEDDFTAWREQWPELDQLLRDEDEPSTSATPYTAIPEYRVEIY
DSVVSVEETHALKQNGQAVYDIHHPCRSNVAVRRELHTPLSDRSCIHEFDISDTGLI
YETGDHVGVHTENSIEVTVEAAKLGLGYQLDITFSVHGKDGDGTPLGGSSLPPFPGPC
TLRTALARYADLLLNPPRKAAFLALAAHASDPAEAERLKFLSSPAGKDEYSQWVTASQR
SLLEIMAEFPSAKPPLGVFFAIAAPRLOPRYYSISISSSPRFAPSRIHVTCAVLGYSPTGR
HKGVCSNWMKNSLPSEETHDCSWAPVFVRQNSFKLPADESTTPIVMVGPGETGPAPFR
GFLQERALKQAGEKLGPAVLFFGCRNMDYIYEDELGVEKFILNTLIVAFSREGA
TKEYVQHJKLMKASDTWSLIAQGGYLYVCGDAKGMARDVHLTHLVQEQESVDSSK
AEFLVKKLQMDGRLDIW
Appendix C  Deduced taxane metabolic pathway
Appendix D: Sample PDB files for heme moiety of P450 H1 & docking code

Whereas MODELER successfully generates PDB files for the polypeptide sequence of the enzymes, the heme moiety has to docked, energy-optimized and then linked to the polypeptide chain manually. The algorithm utilized in Chapter 4 is wholly novel and atomic coordinates of the heme group have been provided herein as a reference.

a. Ground state

| HETATM 3795  | CBB HEM 600 | -13.705 | -7.698 | 59.813 | 1.00  | 56.34 | C  |
| HETATM 3796  | CBC HEM 600 | -21.741 | -4.086 | 60.876 | 1.00  | 54.16 | C  |
| HETATM 3797  | FE HEM 600  | -17.839 | -7.090 | 64.655 | 1.00  | 55.01 | C+3|
| HETATM 3798  | CAB HEM 600 | -13.943 | -8.656 | 60.714 | 1.00  | 54.29 | C  |
| HETATM 3799  | CAC HEM 600 | -20.430 | -4.080 | 60.621 | 1.00  | 53.50 | C  |
| HETATM 3800  | C2B HEM 600 | -14.057 | -8.795 | 63.248 | 1.00  | 53.56 | C  |
| HETATM 3801  | C2C HEM 600 | -18.476 | -5.703 | 60.617 | 1.00  | 53.67 | C  |
| HETATM 3802  | C2A HEM 600 | -16.583 | -7.878 | 68.595 | 1.00  | 53.10 | C  |
| HETATM 3803  | C3D HEM 600 | -20.340 | -4.930 | 67.231 | 1.00  | 54.61 | C  |
| HETATM 3804  | C2D HEM 600 | -20.826 | -4.472 | 66.012 | 1.00  | 54.14 | C  |
| HETATM 3805  | C3A HEM 600 | -15.499 | -8.438 | 67.935 | 1.00  | 53.10 | C  |
| HETATM 3806  | C3B HEM 600 | -14.554 | -8.364 | 62.030 | 1.00  | 54.03 | C  |
| HETATM 3807  | C3C HEM 600 | -19.490 | -5.021 | 61.270 | 1.00  | 53.94 | C  |
| HETATM 3808  | CMA HEM 600 | -14.555 | -9.479 | 68.479 | 1.00  | 52.53 | C  |
| HETATM 3809  | CMD HEM 600 | -22.035 | -3.604 | 65.798 | 1.00  | 53.79 | C  |
| HETATM 3810  | CMB HEM 600 | -12.835 | -9.641 | 63.472 | 1.00  | 52.87 | C  |
| HETATM 3811  | CMC HEM 600 | -18.147 | -5.623 | 59.153 | 1.00  | 52.76 | C  |
| HETATM 3812  | CHA HEM 600 | -18.447 | -6.389 | 67.900 | 1.00  | 53.35 | C  |
| HETATM 3813  | CHB HEM 600 | -14.755 | -8.521 | 65.595 | 1.00  | 53.12 | C  |
| HETATM 3814  | CHC HEM 600 | -16.564 | -7.081 | 61.342 | 1.00  | 53.29 | C  |
| HETATM 3815  | CHD HEM 600 | -20.099 | -4.703 | 63.658 | 1.00  | 53.36 | C  |
| HETATM 3816  | CAA HEM 600 | -17.028 | -8.260 | 69.989 | 1.00  | 53.34 | C  |
| HETATM 3817  | CAD HEM 600 | -20.994 | -4.737 | 68.577 | 1.00  | 55.31 | C  |
| HETATM 3818  | C1A HEM 600 | -17.265 | -7.104 | 67.652 | 1.00  | 52.88 | C  |
| HETATM 3819  | C1B HEM 600 | -14.908 | -8.296 | 64.226 | 1.00  | 52.92 | C  |
| HETATM 3820  | C1C HEM 600 | -17.742 | -6.364 | 61.593 | 1.00  | 53.10 | C  |
| HETATM 3821  | C1D HEM 600 | -19.967 | -4.953 | 65.031 | 1.00  | 53.77 | C  |
| HETATM 3822  | C4A HEM 600 | -15.581 | -8.029 | 66.611 | 1.00  | 53.03 | C  |
| HETATM 3823  | C4D HEM 600 | -19.213 | -5.698 | 66.947 | 1.00  | 54.08 | C  |
| HETATM 3824  | C4B HEM 600 | -15.714 | -7.638 | 62.305 | 1.00  | 53.89 | C  |
| HETATM 3825  | C4C HEM 600 | -19.345 | -5.294 | 62.632 | 1.00  | 53.55 | C  |
| HETATM 3826  | CBA HEM 600 | -17.758 | -9.609 | 69.999 | 1.00  | 55.72 | C  |
| HETATM 3827  | CBD HEM 600 | -21.969 | -5.882 | 68.860 | 1.00  | 58.09 | C  |
| HETATM 3828  | CGA HEM 600 | -18.221 | -9.945 | 71.386 | 1.00  | 58.55 | C  |
| HETATM 3829  | CGD HEM 600 | -22.655 | -5.679 | 70.176 | 1.00  | 60.08 | C  |
| HETATM 3830  | NA HEM 600  | -16.641 | -7.190 | 66.442 | 1.00  | 52.37 | N  |
HETATM 3831 NB HEM 600  -15.919  -7.593  63.649  1.00  53.29  N
HETATM 3832 ND HEM 600  -18.991  -5.703  65.605  1.00  53.53  N
HETATM 3833 NC HEM 600  -18.293  -6.139  62.818  1.00  52.75  N
HETATM 3834 O1A HEM 600  -17.423  -9.757  72.413  1.00  60.38  O
HETATM 3835 O1D HEM 600  -22.112  -6.139  71.281  1.00  59.70  O
HETATM 3836 O2A HEM 600  -19.447  -10.375  71.585  1.00  58.24  O
HETATM 3837 O2D HEM 600  -23.781  -5.004  70.236  1.00  60.88  O
CONECT 3347 3797
CONECT 3797 3347
CONECT 3797 3830 3831 3832
CONECT 3797 3833
CONECT 3795 3798
CONECT 3796 3799
CONECT 3798 3795 3806
CONECT 3799 3796 3807
CONECT 3800 3806 3810 3819
CONECT 3801 3807 3811 3820
CONECT 3802 3805 3816 3818
CONECT 3803 3804 3817 3823
CONECT 3804 3803 3809 3821
CONECT 3805 3802 3808 3822
CONECT 3806 3798 3800 3824
CONECT 3807 3799 3801 3825
CONECT 3808 3805
CONECT 3809 3804
CONECT 3810 3800
CONECT 3811 3801
CONECT 3812 3818 3823
CONECT 3813 3819 3822
CONECT 3814 3820 3824
CONECT 3815 3821 3825
CONECT 3816 3802 3826
CONECT 3817 3803 3827
CONECT 3818 3802 3812 3830
CONECT 3819 3800 3813 3831
CONECT 3820 3801 3814 3833
CONECT 3821 3804 3815 3832
CONECT 3822 3805 3813 3830
CONECT 3823 3803 3812 3832
CONECT 3824 3806 3814 3831
CONECT 3825 3807 3815 3833
CONECT 3826 3816 3828
CONECT 3827 3817 3829
CONECT 3828 3826 3834 3836
CONECT 3829 3827 3835 3837
CONECT 3830 3797 3818 3822
CONECT 3831 3797 3819 3824
CONECT 3832 3797 3821 3823
CONECT 3833 3797 3820 3825
CONECT 3834 3828
CONECT 3835 3829
CONECT 3836 3828
CONECT 3837 3829
END
b. Activated state

<table>
<thead>
<tr>
<th>Residue</th>
<th>Charge</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>3795 CBB HEM</td>
<td>HETATM</td>
<td>-13.705 -7.698 59.813 1.00 56.34 C</td>
</tr>
<tr>
<td>3796 CBC HEM</td>
<td>HETATM</td>
<td>-21.741 -4.086 60.876 1.00 54.16 C</td>
</tr>
<tr>
<td>3797 FE HEM</td>
<td>HETATM</td>
<td>-17.464 -6.649 64.627 1.00 55.01 FE+3</td>
</tr>
<tr>
<td>3798 CAB HEM</td>
<td>HETATM</td>
<td>-13.943 -8.656 60.714 1.00 54.29 C</td>
</tr>
<tr>
<td>3799 CAC HEM</td>
<td>HETATM</td>
<td>-20.430 -4.080 60.621 1.00 53.50 C</td>
</tr>
<tr>
<td>3800 C2B HEM</td>
<td>HETATM</td>
<td>-14.057 -8.795 63.248 1.00 53.56 C</td>
</tr>
<tr>
<td>3801 C2C HEM</td>
<td>HETATM</td>
<td>-18.476 -5.703 60.617 1.00 53.67 C</td>
</tr>
<tr>
<td>3802 C2A HEM</td>
<td>HETATM</td>
<td>-16.583 -7.878 68.595 1.00 53.10 C</td>
</tr>
<tr>
<td>3803 C3D HEM</td>
<td>HETATM</td>
<td>-20.340 -4.930 67.231 1.00 54.61 C</td>
</tr>
<tr>
<td>3804 C2D HEM</td>
<td>HETATM</td>
<td>-20.826 -4.472 66.012 1.00 54.14 C</td>
</tr>
<tr>
<td>3805 C3A HEM</td>
<td>HETATM</td>
<td>-15.499 -8.438 67.935 1.00 53.10 C</td>
</tr>
<tr>
<td>3806 C3B HEM</td>
<td>HETATM</td>
<td>-14.554 -8.364 62.030 1.00 54.03 C</td>
</tr>
<tr>
<td>3807 C3C HEM</td>
<td>HETATM</td>
<td>-19.490 -5.021 61.270 1.00 53.94 C</td>
</tr>
<tr>
<td>3808 CMA HEM</td>
<td>HETATM</td>
<td>-14.555 -9.479 68.479 1.00 52.53 C</td>
</tr>
<tr>
<td>3809 CMD HEM</td>
<td>HETATM</td>
<td>-22.035 -3.604 65.798 1.00 53.79 C</td>
</tr>
<tr>
<td>3810 CMH HEM</td>
<td>HETATM</td>
<td>-12.835 -9.641 63.472 1.00 52.87 C</td>
</tr>
<tr>
<td>3811 CMC HEM</td>
<td>HETATM</td>
<td>-18.147 -5.623 59.153 1.00 52.76 C</td>
</tr>
<tr>
<td>3812 CHA HEM</td>
<td>HETATM</td>
<td>-18.447 -6.389 67.900 1.00 53.35 C</td>
</tr>
<tr>
<td>3813 CHB HEM</td>
<td>HETATM</td>
<td>-14.755 -8.521 65.595 1.00 53.12 C</td>
</tr>
<tr>
<td>3814 CHC HEM</td>
<td>HETATM</td>
<td>-16.564 -7.081 61.342 1.00 53.29 C</td>
</tr>
<tr>
<td>3815 CHD HEM</td>
<td>HETATM</td>
<td>-20.099 -4.703 63.658 1.00 53.36 C</td>
</tr>
<tr>
<td>3816 CAA HEM</td>
<td>HETATM</td>
<td>-17.028 -8.260 69.989 1.00 53.34 C</td>
</tr>
<tr>
<td>3817 CAD HEM</td>
<td>HETATM</td>
<td>-20.994 -4.737 68.577 1.00 55.31 C</td>
</tr>
<tr>
<td>3818 C1A HEM</td>
<td>HETATM</td>
<td>-17.265 -7.104 67.652 1.00 52.88 C</td>
</tr>
<tr>
<td>3819 C1B HEM</td>
<td>HETATM</td>
<td>-14.908 -8.296 64.226 1.00 52.92 C</td>
</tr>
<tr>
<td>3820 C1C HEM</td>
<td>HETATM</td>
<td>-17.742 -6.364 61.593 1.00 53.10 C</td>
</tr>
<tr>
<td>3821 C1D HEM</td>
<td>HETATM</td>
<td>-19.967 -4.953 65.031 1.00 53.77 C</td>
</tr>
<tr>
<td>3822 C4A HEM</td>
<td>HETATM</td>
<td>-15.581 -8.029 66.611 1.00 53.03 C</td>
</tr>
<tr>
<td>3823 C4D HEM</td>
<td>HETATM</td>
<td>-19.213 -5.698 66.947 1.00 54.08 C</td>
</tr>
<tr>
<td>3824 C4B HEM</td>
<td>HETATM</td>
<td>-15.714 -7.638 62.305 1.00 53.89 C</td>
</tr>
<tr>
<td>3825 C4C HEM</td>
<td>HETATM</td>
<td>-19.345 -5.294 62.632 1.00 53.55 C</td>
</tr>
<tr>
<td>3826 CBA HEM</td>
<td>HETATM</td>
<td>-17.758 -9.609 69.999 1.00 55.72 C</td>
</tr>
<tr>
<td>3827 CBD HEM</td>
<td>HETATM</td>
<td>-21.969 -5.882 68.860 1.00 58.09 C</td>
</tr>
<tr>
<td>3828 CGA HEM</td>
<td>HETATM</td>
<td>-18.221 -9.945 71.386 1.00 58.55 C</td>
</tr>
<tr>
<td>3829 CGD HEM</td>
<td>HETATM</td>
<td>-22.655 -5.679 70.176 1.00 60.08 C</td>
</tr>
<tr>
<td>3830 NA HEM</td>
<td>HETATM</td>
<td>-16.641 -7.190 66.442 1.00 52.37 N</td>
</tr>
<tr>
<td>3831 NB HEM</td>
<td>HETATM</td>
<td>-15.919 -7.593 63.649 1.00 53.29 N</td>
</tr>
<tr>
<td>3832 ND HEM</td>
<td>HETATM</td>
<td>-18.991 -5.703 65.605 1.00 53.53 N</td>
</tr>
<tr>
<td>3833 NC HEM</td>
<td>HETATM</td>
<td>-18.293 -6.139 62.818 1.00 52.75 N</td>
</tr>
<tr>
<td>3834 O1A HEM</td>
<td>HETATM</td>
<td>-17.423 -9.757 72.413 1.00 60.38 O</td>
</tr>
<tr>
<td>3835 O1D HEM</td>
<td>HETATM</td>
<td>-22.112 -6.139 71.281 1.00 59.70 O</td>
</tr>
<tr>
<td>3836 O2A HEM</td>
<td>HETATM</td>
<td>-19.447 -10.375 71.585 1.00 58.24 O</td>
</tr>
<tr>
<td>3837 O2D HEM</td>
<td>HETATM</td>
<td>-23.781 -5.004 70.236 1.00 60.88 O</td>
</tr>
</tbody>
</table>

CONECT 3347 3797
CONECT 3797 3347
CONECT 3797 3830 3831 3832
CONECT 3797 3833
CONECT 3795 3798
CONECT 3796 3799
CONECT 3798 3795 3806
CONECT 3799 3796 3807
c. Peroxyheme state

HETATM 3795 CAA HEM 600 -17.028 -8.260 69.989 1.00 53.34 C
HETATM 3796 CAB HEM 600 -13.943 -8.656 60.714 1.00 54.29 C
HETATM 3797 CAC HEM 600 -20.430 -4.080 60.621 1.00 53.50 C
HETATM 3798 CAD HEM 600 -20.994 -4.737 68.577 1.00 55.31 C
HETATM 3799 NA HEM 600 -16.641 -7.190 66.442 1.00 52.37 N
HETATM 3800 CBA HEM 600 -17.758 -9.609 69.999 1.00 53.72 C
HETATM 3801 CBB HEM 600 -13.705 -7.698 59.813 1.00 56.34 C
HETATM 3802 CBC HEM 600 -21.741 -4.086 60.876 1.00 54.16 C
HETATM 3803 CBD HEM 600 -21.969 -5.882 68.860 1.00 58.09 C
HETATM 3804 NB HEM 600 -15.919 -7.593 63.649 1.00 53.29 N
HETATM 3805 CGA HEM 600 -18.221 -9.945 71.386 1.00 58.55 C
| HETATM 3806 CGD HEM | -22.655 -5.679 70.176 | 1.00 60.08 C |
| HETATM 3807 ND HEM | -18.991 -5.703 65.605 | 1.00 53.53 N |
| HETATM 3808 CHA HEM | -18.447 -6.389 67.900 | 1.00 53.35 C |
| HETATM 3809 CHB HEM | -14.755 -8.521 65.595 | 1.00 53.12 C |
| HETATM 3810 CHC HEM | -16.564 -7.081 61.342 | 1.00 53.29 C |
| HETATM 3811 CHD HEM | -20.099 -4.703 63.658 | 1.00 53.36 C |
| HETATM 3812 CMA HEM | -14.555 -9.479 68.479 | 1.00 52.53 C |
| HETATM 3813 CMB HEM | -12.835 -9.641 63.472 | 1.00 52.87 C |
| HETATM 3814 CMC HEM | -18.147 -5.623 59.153 | 1.00 52.76 C |
| HETATM 3815 CMD HEM | -22.035 -3.604 65.798 | 1.00 53.79 C |
| HETATM 3816 C1A HEM | -17.265 -7.104 67.652 | 1.00 52.88 C |
| HETATM 3817 C1B HEM | -14.908 -8.296 64.226 | 1.00 52.92 C |
| HETATM 3818 C1C HEM | -17.742 -6.364 61.593 | 1.00 53.10 C |
| HETATM 3819 C1D HEM | -19.967 -4.953 65.031 | 1.00 53.77 C |
| HETATM 3820 O1 HEM | -16.413 -4.844 64.539 | 1.00 30.00 O |
| HETATM 3821 O1A HEM | -17.423 -9.757 72.413 | 1.00 60.38 O |
| HETATM 3822 O1D HEM | -22.112 -6.139 71.281 | 1.00 59.70 O |
| HETATM 3823 C2A HEM | -16.583 -7.878 68.595 | 1.00 53.10 C |
| HETATM 3824 C2B HEM | -14.057 -8.795 63.248 | 1.00 53.56 C |
| HETATM 3825 C2C HEM | -18.476 -5.703 60.617 | 1.00 53.67 C |
| HETATM 3826 C2D HEM | -20.826 -4.472 66.012 | 1.00 54.14 C |
| HETATM 3827 O2 HEM | -16.659 -4.287 63.318 | 1.00 30.00 O |
| HETATM 3828 O2A HEM | -19.447 -10.375 71.585 | 1.00 58.24 O |
| HETATM 3829 O2D HEM | -23.781 -5.004 70.236 | 1.00 60.88 O |
| HETATM 3830 C3A HEM | -15.499 -8.438 67.935 | 1.00 53.10 C |
| HETATM 3831 C3B HEM | -14.554 -8.364 62.030 | 1.00 54.03 C |
| HETATM 3832 C3C HEM | -19.490 -5.021 61.270 | 1.00 53.94 C |
| HETATM 3833 C3D HEM | -20.340 -4.930 67.231 | 1.00 54.61 C |
| HETATM 3834 C4A HEM | -15.581 -8.029 66.611 | 1.00 53.03 C |
| HETATM 3835 C4B HEM | -15.714 -7.638 62.305 | 1.00 53.89 C |
| HETATM 3836 C4C HEM | -19.345 -5.294 62.632 | 1.00 53.55 C |
| HETATM 3837 C4D HEM | -19.213 -5.698 66.947 | 1.00 54.08 C |
| HETATM 3838 NC HEM | -18.293 -6.139 62.818 | 1.00 52.75 N |
| HETATM 3839 FE HEM | -17.464 -6.649 64.627 | 1.00 55.01 FE+3 |
| HETATM 3840 H1 HEM | -15.917 -3.654 63.002 | 1.00 53.51 H |

CONNECT 3827 3840
CONNECT 3840 3827
CONNECT 3347 3839
CONNECT 3795 3800 3823
CONNECT 3796 3801 3831
CONNECT 3797 3802 3832
CONNECT 3798 3803 3833
CONNECT 3799 3816
CONNECT 3799 3834
CONNECT 3799 3839
CONNECT 3800 3795 3805
CONNECT 3801 3796
CONNECT 3802 3797
CONNECT 3803 3798 3806
CONNECT 3804 3817
CONNECT 3804 3835
CONNECT 3804 3839
CONNECT 3805 3800
CONNECT 3805 3821
CONNECT 3805 3828

201
CONECT 3806 3803
CONECT 3806 3822
CONECT 3806 3829
CONECT 3807 3819
CONECT 3807 3837
CONECT 3807 3839
CONECT 3808 3816
CONECT 3808 3837
CONECT 3809 3817
CONECT 3809 3834
CONECT 3810 3818
CONECT 3810 3835
CONECT 3811 3819
CONECT 3811 3836
CONECT 3812 3830
CONECT 3813 3824
CONECT 3814 3825
CONECT 3815 3826
CONECT 3816 3799
CONECT 3816 3808
CONECT 3816 3823
CONECT 3817 3804
CONECT 3817 3809
CONECT 3817 3824
CONECT 3818 3810
CONECT 3818 3825
CONECT 3818 3836
CONECT 3819 3807
CONECT 3819 3811
CONECT 3819 3826
CONECT 3820 3827 3839
CONECT 3821 3805
CONECT 3822 3806
CONECT 3823 3795
CONECT 3823 3816
CONECT 3823 3830
CONECT 3824 3813
CONECT 3824 3817
CONECT 3824 3831
CONECT 3825 3814
CONECT 3825 3818
CONECT 3825 3832
CONECT 3826 3815
CONECT 3826 3819
CONECT 3826 3833
CONECT 3827 3820
CONECT 3828 3805
CONECT 3829 3806
CONECT 3830 3812
CONECT 3830 3823
CONECT 3830 3834
CONECT 3831 3796
CONECT 3831 3824
CONECT 3831 3835
CONECT 3832 3797
d. Oxyferyl state

HETATM 3795  CAA HEM  600  -17.028  -8.260  69.989  1.00  53.34  C
HETATM 3796  CAB HEM  600  -13.943  -8.656  60.714  1.00  54.29  C
HETATM 3797  CAC HEM  600  -20.430  -4.080  60.621  1.00  53.50  C
HETATM 3798  CAD HEM  600  -20.994  -4.737  68.577  1.00  55.31  C
HETATM 3799  NA HEM  600  -16.641  -7.190  66.442  1.00  52.37  N
HETATM 3800  CBA HEM  600  -17.758  -9.609  69.999  1.00  55.72  C
HETATM 3801  CBB HEM  600  -13.705  -7.698  59.813  1.00  56.34  C
HETATM 3802  CBC HEM  600  -21.741  -4.086  60.876  1.00  54.16  C
HETATM 3803  CBD HEM  600  -21.969  -5.882  68.860  1.00  58.09  C
HETATM 3804  NB HEM  600  -15.919  -7.593  63.649  1.00  53.29  N
HETATM 3805  CGA HEM  600  -18.221  -9.945  71.386  1.00  58.55  C
HETATM 3806  CGD HEM  600  -22.655  -5.679  70.176  1.00  60.08  C
HETATM 3807  ND HEM  600  -18.991  -5.703  65.605  1.00  53.53  N
HETATM 3808  CHA HEM  600  -18.447  -6.389  67.900  1.00  53.35  C
HETATM 3809  CHB HEM  600  -14.755  -8.521  65.595  1.00  53.12  C
HETATM 3810  CHC HEM  600  -16.564  -7.081  61.342  1.00  53.29  C
HETATM 3811  CHD HEM  600  -20.099  -4.703  63.658  1.00  53.36  C
HETATM 3812  CMA HEM  600  -14.555  -9.479  68.479  1.00  52.53  C
HETATM 3813  CMB HEM  600  -12.835  -9.641  63.472  1.00  52.87  C
HETATM 3814  CMC HEM  600  -18.147  -5.623  59.153  1.00  52.76  C
HETATM 3815  CMD HEM  600  -22.035  -3.604  65.798  1.00  53.79  C
HETATM 3816  C1A HEM  600  -17.265  -7.104  67.652  1.00  52.88  C
HETATM 3817  C1B HEM  600  -14.908  -8.296  64.226  1.00  52.92  C
HETATM 3818  C1C HEM  600  -17.742  -6.364  61.593  1.00  53.10  C
HETATM 3819  C1D HEM  600  -19.967  -4.953  65.031  1.00  53.77  C
HETATM 3820  O1 HEM  600  -16.644  -5.148  64.446  1.00  30.00  O
<table>
<thead>
<tr>
<th>HETATM</th>
<th>3821 O1A HEM</th>
<th>600</th>
<th>-17.423</th>
<th>-9.757</th>
<th>72.413</th>
<th>1.00</th>
<th>60.38</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>HETATM</td>
<td>3822 O1D HEM</td>
<td>600</td>
<td>-22.112</td>
<td>-6.139</td>
<td>71.281</td>
<td>1.00</td>
<td>59.70</td>
<td>O</td>
</tr>
<tr>
<td>HETATM</td>
<td>3823 C2A HEM</td>
<td>600</td>
<td>-16.583</td>
<td>-7.878</td>
<td>68.595</td>
<td>1.00</td>
<td>53.10</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3824 C2B HEM</td>
<td>600</td>
<td>-14.057</td>
<td>-8.795</td>
<td>63.248</td>
<td>1.00</td>
<td>53.56</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3825 C2C HEM</td>
<td>600</td>
<td>-18.476</td>
<td>-5.703</td>
<td>60.617</td>
<td>1.00</td>
<td>53.67</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3826 C2D HEM</td>
<td>600</td>
<td>-20.826</td>
<td>-4.472</td>
<td>66.012</td>
<td>1.00</td>
<td>54.14</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3827 O2A HEM</td>
<td>600</td>
<td>-19.447</td>
<td>-10.375</td>
<td>71.585</td>
<td>1.00</td>
<td>58.24</td>
<td>O</td>
</tr>
<tr>
<td>HETATM</td>
<td>3828 O2D HEM</td>
<td>600</td>
<td>-23.781</td>
<td>-5.004</td>
<td>70.236</td>
<td>1.00</td>
<td>60.88</td>
<td>O</td>
</tr>
<tr>
<td>HETATM</td>
<td>3829 C3A HEM</td>
<td>600</td>
<td>-15.499</td>
<td>-8.438</td>
<td>67.935</td>
<td>1.00</td>
<td>53.10</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3830 C3B HEM</td>
<td>600</td>
<td>-14.554</td>
<td>-8.364</td>
<td>62.030</td>
<td>1.00</td>
<td>54.03</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3831 C3C HEM</td>
<td>600</td>
<td>-19.490</td>
<td>-5.021</td>
<td>61.270</td>
<td>1.00</td>
<td>53.94</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3832 C3D HEM</td>
<td>600</td>
<td>-20.340</td>
<td>-4.930</td>
<td>67.231</td>
<td>1.00</td>
<td>54.61</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3833 C4A HEM</td>
<td>600</td>
<td>-15.581</td>
<td>-8.029</td>
<td>66.611</td>
<td>1.00</td>
<td>53.03</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3834 C4B HEM</td>
<td>600</td>
<td>-15.714</td>
<td>-7.638</td>
<td>62.305</td>
<td>1.00</td>
<td>53.89</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3835 C4C HEM</td>
<td>600</td>
<td>-19.345</td>
<td>-5.294</td>
<td>62.632</td>
<td>1.00</td>
<td>53.55</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3836 C4D HEM</td>
<td>600</td>
<td>-19.213</td>
<td>-5.698</td>
<td>66.947</td>
<td>1.00</td>
<td>54.08</td>
<td>C</td>
</tr>
<tr>
<td>HETATM</td>
<td>3837 NC HEM</td>
<td>600</td>
<td>-18.293</td>
<td>-6.139</td>
<td>62.818</td>
<td>1.00</td>
<td>52.75</td>
<td>N</td>
</tr>
<tr>
<td>HETATM</td>
<td>3838 FE HEM</td>
<td>600</td>
<td>-17.464</td>
<td>-6.649</td>
<td>64.627</td>
<td>1.00</td>
<td>55.01</td>
<td>FE</td>
</tr>
</tbody>
</table>

**CONECT**
- 3347 3838
- 3795 3800 3823
- 3796 3801 3830
- 3797 3802 3831
- 3798 3803 3832
- 3799 3816
- 3799 3833
- 3799 3838
- 3800 3795 3805
- 3801 3796
- 3802 3797
- 3803 3798 3806
- 3804 3817
- 3804 3834
- 3804 3838
- 3805 3800
- 3805 3821
- 3805 3827
- 3806 3803
- 3806 3822
- 3806 3828
- 3807 3819
- 3807 3836
- 3807 3838
- 3808 3816
- 3808 3836
- 3809 3817
- 3809 3833
- 3810 3818
- 3810 3834
- 3811 3819
- 3811 3835
- 3812 3829
- 3813 3824
- 3814 3825
- 3815 3826
- 3816 3799

204
e. Sample code for substrate docking

```
receptor = taxadiene5ahydroxylase.pdbqt
ligand = taxadiene.pdbqt
out = output_iteration_1.pdbqt

center_x = -22.519
center_y = -19.74
center_z = -10.502

size_x = 40
size_y = 40
size_z = 40

exhaustiveness = 9
```
Appendix E: NMR raw data
Taxadien-5α-ol
Species K
Oxacyclotaxane (OCT)

Species I
Cyclic ether
Species G

Retention time (minutes)

Chemical shift (ppm)