UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Understanding Structure-Function Relationships in a Sesquiterpene Synthase family through the Examination of Natural and Unnatural Variants

Permalink
https://escholarship.org/uc/item/46t04154

Author
Sun, Helena Carol

Publication Date
2015

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Understanding Structure-Function Relationships in a Sesquiterpene Synthase family through the Examination of Natural and Unnatural Variants

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

in

Biology

by

Helena Carol Sun

Committee in charge:

Professor Joseph P. Noel, Chair
Professor Russell Doolittle, Co-Chair
Professor Joanne Chory
Professor Josh R. Kohn
Professor Mark H. Thiemens

2015
The Dissertation of Helena Carol Sun is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________

________________________________________

________________________________________

________________________________________

Co-Chair

________________________________________

Chair

University of California, San Diego

2015
DEDICATION

This dissertation is dedicated to my parents, Chih-Chung and Chou Min-Li, as well as all my friends, for their support and encouragement; to Joe, for his mentorship and care; and to my husband, Russ, who has been my constant companion through the long haul.
# TABLE OF CONTENTS

Signature Page.................................................................................................................................................. iii  
Dedication....................................................................................................................................................... iv  
Table of Contents............................................................................................................................................... v  
List of Abbreviations ....................................................................................................................................... viii  
List of Figures ................................................................................................................................................ x  
List of Tables ................................................................................................................................................ xii  
Acknowledgments............................................................................................................................................ xiii  
Vita.................................................................................................................................................................... xiv  
Abstract of the Dissertation ............................................................................................................................. xv  

Chapter 1  Background and Introduction ......................................................................................................... 1  
1.1 Terpenoids Overview ................................................................................................................................. 1  
1.2 Terpene Biosynthesis ............................................................................................................................... 2  
1.2.1 Biosynthesis of IPP and DMAPP precursors ......................................................................................... 3  
1.2.2 Condensations of IPP/DMAPP to form prenyl diphosphates ................................................................. 6  
1.2.3 Cyclization of prenyl diphosphate intermediates ................................................................................ 7  
1.2.4 Modifications ....................................................................................................................................... 8  
1.3 Terpene synthases ..................................................................................................................................... 9  
1.3.1 The sesquiterpene synthase EAS: structure & function ....................................................................... 9  
1.4 Origins of terpene diversity ...................................................................................................................... 14  
1.5 Enzyme plasticity, natural variation & enzyme evolution ........................................................................... 17  
1.6 Segue and chapter guides ......................................................................................................................... 18  
1.7 References ............................................................................................................................................... 20
Chapter 2  Cloning and Characterization of 5-epi-aristolochene synthase homologs from plants of the *Nicotiana* genus ................................................................. 25
  2.1 Abstract ........................................................................................................... 25
  2.2 Introduction .................................................................................................... 25
  2.3 Results and Discussion ................................................................................. 28
  2.4 Materials and Methods ............................................................................... 40
  2.5 References .................................................................................................... 44

Chapter 3  Characterization of a sesquiterpene synthase from *N. longiflora* reveals a residue position that controls product specificity ................................................................. 47
  3.1 Abstract ......................................................................................................... 47
  3.2 Introduction .................................................................................................. 47
  3.3 Results and Discussion ................................................................................ 49
  3.4 Materials and Methods ................................................................................ 60
  3.5 Acknowledgments ........................................................................................ 64
  3.6 References .................................................................................................... 65

Chapter 4  Identification of a product specificity switch residue in the TEAS family of enzymes ........................................................................................................... 67
  4.1 Abstract ......................................................................................................... 67
  4.2 Introduction .................................................................................................. 67
  4.3 Results and Discussion ................................................................................ 69
  4.4 Materials and Methods ................................................................................ 89
  4.5 Acknowledgments ........................................................................................ 93
  4.6 References .................................................................................................... 94

Chapter 5  Conclusions and Perspectives .................................................................. 96
5.1 Overview ................................................................................................................ 96
5.2 Implications and considerations ............................................................................. 98
5.3 References ............................................................................................................ 101
LIST OF ABBREVIATIONS

4-EE   4-epi-eremophilene
5-EA   5-epi-aristolochene
aa     amino acid
EAH    5-epi-aristolochene hydroxylase
EAS    5-epi-aristolochene synthase
DMAPP  dimethylallyl diphosphate
DXP    1-deoxy-D-xylulose 5-phosphate
FHP    farnesyl hydroxyphosphonate
FPP    farnesyl diphosphate
FPPS   farnesyl diphosphate synthase
GC-MS  gas chromatography – mass spectrometry
GPP    geranyl diphosphate
GGPP   geranylgeranyl diphosphate
HPLC   high performance liquid chromatography
HPS    *Hyoscamus muticus* premnaspirodiene (vetispirodiene) synthase
IPP    isopentenyl pyrophosphate
MEP    2-Methyl-D-erythritol-4-phosphate
MEV    mevalonic acid
NbEAS  5-epi-aristolochene synthase from *Nicotiana benthamiana*
NlgfS  5-epi-aristolochene synthase from *Nicotiana longiflora*
ORF    open reading frame
PSD    premnaspirodiene
SNP    single nucleotide polymorphism
TEAS   *Nicotiana tabacum* 5-epi-aristolochene synthase
TIS    TEAS-like synthase
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_m$</td>
<td>melting temperature</td>
</tr>
<tr>
<td>TPS</td>
<td>terpene synthase</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Biosynthesis of IPP and DMAP: the Mevalonate (MEV) and Non-Mevalonate (MEP/DOXP) pathways</td>
<td>4</td>
</tr>
<tr>
<td>1.2</td>
<td>Mevalonate and Non-Mevalonate Pathways for IPP/DMAPP synthesis</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>Overview and examples of terpene synthesis</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>Two options fortransoid cyclization by sesquiterpene cyclases</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>Structure of TEAS</td>
<td>11</td>
</tr>
<tr>
<td>1.6</td>
<td>Biosynthesis pathway of the phytoalexin capsidiol</td>
<td>15</td>
</tr>
<tr>
<td>1.7</td>
<td>Proposed domains evolution of terpene synthases</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>Parallel biosynthesis pathways of the phytoalexins capsidiol and solavetivone</td>
<td>26</td>
</tr>
<tr>
<td>2.2</td>
<td>Phylogenetic tree of <em>Nicotiana</em> genus</td>
<td>30</td>
</tr>
<tr>
<td>2.3</td>
<td>Cloning EAS genes from <em>Nicotiana</em> species</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>Sequence phylogeny tree of cloned EAS from <em>Nicotiana</em> species</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>Residue conservation of <em>Nicotiana</em> EAS homologs mapped on TEAS structure</td>
<td>35</td>
</tr>
<tr>
<td>2.6</td>
<td>Comparison of products from EAS from various <em>Nicotiana</em> species</td>
<td>37</td>
</tr>
<tr>
<td>2.7</td>
<td>Two <em>Nicotiana</em> EAS variants with distinctly novel product profiles as compared to TEAS</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>GC-MS product trace overlay of TEAS and NlgsS</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Residue identity at 460 directs enzyme function</td>
<td>51</td>
</tr>
<tr>
<td>3.3</td>
<td>Spatial position 460 relative to the active site</td>
<td>53</td>
</tr>
<tr>
<td>3.4</td>
<td>Map of biosynthetic pathways for TEAS, Nlgs and mutants</td>
<td>54</td>
</tr>
<tr>
<td>3.5</td>
<td>Effect of pH on product distribution</td>
<td>55</td>
</tr>
<tr>
<td>3.6</td>
<td>Melting temperatures of TEAS, Nlgs and their 460 mutants</td>
<td>57</td>
</tr>
<tr>
<td>3.7</td>
<td>Michaelis-Menten Kinetics of TEAS, Nlgs and their mutants at position 460</td>
<td>58</td>
</tr>
<tr>
<td>3.8</td>
<td>Residue conservation at position 460</td>
<td>59</td>
</tr>
<tr>
<td>4.1</td>
<td>GC-MS product trace overlay of TEAS and TIS</td>
<td>70</td>
</tr>
</tbody>
</table>
Figure 4.2  Single residue mutations of TEAS and TIS .......................................................... 72-73
Figure 4.3  Double mutants: non-additive effects of residues 289 and 508 on TEAS+Y404H.  75
Figure 4.4  Triple mutant of TEAS does not recapitulate catalytic activity of TIS ............... 76
Figure 4.5  Mutant residues farther out from the active site ................................................... 77
Figure 4.6  Kinetics of TEAS, TIS and mutants ................................................................... 78
Figure 4.7  Melting temperature effects of position 404 residues in TEAS and TIS ............. 80
Figure 4.8  Melting temperature effects of residues at position 508 ...................................... 81
Figure 4.9  Effects of pH on product distribution in TIS ...................................................... 83
Figure 4.10 Model of His404 acting as a general base ......................................................... 84
Figure 4.11 Reaction pathways map of EAS-like enzymes ................................................... 85
Figure 4.12 Interaction network of position 440 .................................................................. 87
Figure 4.13 Sequence conservation: alignment of TIS, TEAS and other sesquiterpene synthases ......................................................................................................................... 88
Figure 5.1  Cascading effect of product and substrate non-selectivity ................................. 99
LIST OF TABLES

Table 1.1  Terpene classifications by number of isoprenoid units............................................... 6
Table 1.2  Known residue functions in TEAS ............................................................................. 13
Table 2.1  Genomes sequenced to-date of Solanaceae plants .................................................... 29
Table 2.2  List of *Nicotiana* species and numeric code used in this study ............................ 33
Table 2.3  Melting temperatures of *Nicotiana* EAS enzyme variants ........................................ 39
ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Joseph Noel and all lab members, past and present, who have helped me throughout my thesis work. Joe’s enthusiasm for science was always an inspiration for me. Hyun Jo Koo for being his humorous self and for being my ever ready free-gan buddy; also for teaching me the secrets to achieving perfect kinetic curves amongst other things. Marianne Bowman for running a tight ship and being always ready & willing to help whenever needed. Gordon Louie for fielding all my computer-related & crystal data processing questions. Yong-Xia Guo, Ryan Philippe & Jing-Ke Weng for their advice & counsel on all things grad school.

The text of Chapter 3, in part, is currently being prepared for submission for publication of the material. Sun, Helena; Koo, Hyun Jo; Noel Joseph P. “Characterization of a sesquiterpene synthase from \textit{N. longiflora} reveals a residue position that control product specificity.” I am the first author of this material. Permission was obtained from all co-authors. All experiments were performed under the supervision of Joseph P. Noel.

The text of Chapter 4, in part, is currently being prepared for submission for publication of the material. Sun, Helena; Koo, Hyun Jo; Noel Joseph P. “Identification of a product specificity switch residue in the TEAS family of enzymes.” I am the first author of this material. Permission was obtained from all co-authors. All experiments were performed under the supervision of Joseph P. Noel.
VITA

2006 Bachelor of Arts, University of California, Berkeley
2006-2008 Research Associate, Sangamo Biosciences, Inc., Richmond, California
2008-2015 Graduate Student Researcher, University of California, San Diego
2015 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS


ABSTRACT OF THE DISSERTATION

Understanding Structure-Function Relationships in a Sesquiterpene Synthase family through the Examination of Natural and Unnatural Variants

by

Helena Carol Sun

Doctor of Philosophy in Biology

University of California, San Diego, 2015

Professor Joseph P. Noel, Chair
Professor Russell Doolittle, Co-Chair

Terpenes are one of the largest groups of plant secondary metabolites that are utilized for a multitude of purposes from defense to pollinator attraction in plants, and medicinal purposes in humans. The synthesis of the extensive array of terpenoid compounds is complex and nuanced; hence the following work attempts to elucidate one portion of the terpene story, focusing on 5-epi-aristolchene synthase (EAS). The terpene synthase EAS is a member of a simple three-enzyme sesquiterpene biosynthetic pathway that ultimately produces the hydroxylated sesquiterpene phytoalexin, capsidiol. Here we begin the investigation of the natural variation
found in EAS-like enzymes and their associated biochemical properties, across the ecologically diverse Nicotiana plant genus. The work entails an in-depth examination of functional diversification of an enzyme family critical to pathogen resistance in plants, where EAS-like genes were cloned, sequenced, and expressed recombinantly from several species of the Nicotiana genus of plants and the resultant enzymes were biochemically characterized. Characterization focused on three fundamental properties of enzymes, namely thermostability, kinetic efficiency, and product identity and diversity. From the aforementioned survey, two EAS-like enzyme variants were isolated that produced distinct product profiles when compared to the relatively well-characterized Nicotiana tabacum 5-epi-aristolochene synthase (TEAS). The in-depth study of these two EAS-like synthases illuminates several factors that influence the distribution and identity of the products that are synthesized by these enzymes.
CHAPTER 1

Background and Introduction

1.1 Terpenoids Overview

Terpenoids are one of the largest groups of naturally biosynthesized compounds, with over 60,000 members identified to date. Terpene compounds are found widely in bacteria, fungi and plants where they play vital roles in primary metabolism (e.g. regulating plant growth through gibberellins and phytohormones and sterols for maintaining the structure and function of cellular membranes) and perhaps more prominently in secondary or ‘specialized’ metabolism where they mediate the interactions of plants with their environment (e.g. defense against diseases and pests with noxious or toxic compounds, signaling and pollination with attractive scents). These specialized and diverse terpene metabolites increase the fitness of organisms in their complex environments and are a chemical arsenal for communication and survival.

While terpenoids are valuable in their natural provenance, several have proved useful as biomaterials and pharmaceuticals. Many terpenoids are bioactive, having antibiotic, antifungal or other therapeutic properties, like terpinene-4-ol the main antimicrobial compound in tea tree oil. Other well-known examples of sesquiterpenoid pharmaceuticals include the potent diterpenoid anticancer drug paclitaxel (Taxol®) and the sesquiterpenoid antimalarial drug artemisinin. Terpenes also have many commercial and industrial applications from use as flavors and fragrances, since terpenoids are the basis of many desirable taste and aroma in plants, to use as biofuel precursors. Bisabolane, the fully reduced form of the monocyclic
sesquiterpene bisabolene, was characterized as a substitute for D2 diesel fuel and the microbial production of the precursor bisabolene was demonstrated by Peralta-Yahya et al.\textsuperscript{15}. Amyris Biotechnologies, Inc. is also developing farnesane, the fully reduced form of the linear sesquiterpene farnesene, as an alternative biosynthetic diesel fuel; and they are also utilizing \textit{trans}-\(\beta\)-Farnesene as a renewable hydrocarbon building block for the production of several other diverse products such as cosmetics ingredients and lubricants. While plants are the bulk producers of isoprenoid compounds in nature, the rate of production is often not sufficient for industrial and pharmaceutical purposes. Hence the engineering of useful pathways into other faster growing organisms like bacteria, yeast, and algae to increase the production efficiency of valuable compound is an ongoing effort\textsuperscript{12}.

A challenge in understanding and manipulating terpene biosynthesis, and incidentally also its inherent worth, lies in its complex chemistries. Terpenoid compounds encompass a wealth of cyclic structures that differ from one another in not only in their functional groups but also in their basic carbon skeletons. These elaborate structures are achieved through some of the most complex chemical reactions occurring in nature, where terpene synthases catalyze intricate cyclizations and rearrangements to convert linear isoprenyl pyrophosphate precursors into the various multitude of multicyclic structures\textsuperscript{16}.

1.2 Terpenoid biosynthesis

Terpenes are synthesized from the combination of five-carbon isoprene units that are assembled and modified in thousands of ways. The original and simplest definition of a terpene, formulated by Otto Wallach and named the ‘isoprene rule,’ stated that all terpenes can be built from isoprene units. This more basic definition was later superseded by the Ruzicka’s biogenetic isoprene rule which stated that terpenes are compounds that can be derived by cyclization and/or other rearrangements of aliphatic precursors composed of isoprene units\textsuperscript{17}. This later definition
allowed for the inclusion of the modified terpenes, or isoprenoids, and is the definition that still
that stands today.

The four main steps in terpene synthesis are 1) formation of IPP/DMAPP; 2) condensation of IPP/DMAPP into the allylic prenyl diphosphate precursors GPP, FPP, and GGPP; 3) the rearrangement and/or cyclization of these allylic prenyl diphosphates by specific terpene synthases (TPS); and lastly, 4) any secondary chemical modifications such as oxidation-reduction reactions. These four steps are detailed in the following sections.

1.2.1 Biosynthesis of IPP and DMAPP precursors

The 5-carbon terpene precursor units isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) originate from combinations of acetyl-CoA, pyruvate, and glyceraldehyde-3-phosphate all derived from primary metabolism. The synthesis of IPP and DMAPP precursors can result from two different metabolic pathways in plants (Figure 1.1). One is the Mevalonate pathway (MEV), also known as the Mevalonic acid (MVA) pathway, which occurs in the cytosol (and certain subcellular structures) of all higher eukaryotes and several bacteria. The second is the non-mevalonate pathway, also known as the 2C-methyl-D-erythritol-4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway (MEP/DXP) pathway that functions in the plastids of plants, as well as apicomplexan protozoa and many bacteria.
Figure 1.1 Biosynthesis of IPP and DMAP: the Mevalonate (MEV) and Non-Mevalonate (MEP/DXP) pathways. Reprinted in compliance with copyright permissions 12.

In the mevalonate pathway acetyl-coA condenses with acetoacetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) which is then reduced to form mevalonate. Mevalonate is then phosphorylated twice in succession to make diphosphomevalonate and then decarboxylated to produce IPP, which can isomerize into DMAPP (Figure 1.2a). In contrast, the MEP/DXP pathway starts by condensing pyruvate with glyceraldehydes-3-phosphate (G3P) to form 1-deoxy-D-xylulose 5-phosphate (DXP). DXP is then reduced to form 2C-methyl-D-erythritol-4-phosphate (MEP) by a reductoisomerase. These previous two steps lend the pathway one of its names. MEP is then joined to a cytidine triphosphate (CTP) on one end and phosphorylated on the other, after which it circularizes by eliminating a cytidine monophosphate (CMP). The resulting ring intermediate is then reduced twice in succession to open the ring back up and form IPP/DMAPP (Figure 1.2b)
Figure 1.2 Mevalonate and Non-Mevalonate Pathways for IPP/DMAPP synthesis. Adapted & reprinted in compliance with copyright permissions from Journal of Bacteriology. (Takagi et al 2000) \(^7\).
1.2.2 Condensations of IPP/DMAPP to form prenyl disphosphates

Molecules of IPP and DMAPP are condensed in a head-to-tail fashion by isoprenyl diphosphate synthases (also named prenyltransferases) to produce linear isoprenoids of differing length, these being geranyl diphosphate (GPP, 10C), farnesyl diphosphate (FPP, 15C), and geranylgeranyl diphosphate (GGPP, 20C). These serve as the immediate precursors for the different families of isoprenoids. Geranyl pyrophosphate is formed by the head-to-tail condensation DMAPP and IPP. Geranyl pyrophosphate can further undergo additional condensation with one or two molecules of IPP to form farnesyl pyrophosphate and GGPP respectively. Repeated chain-elongation reactions can eventually form polyisoprenoids that contain from four to several thousand isoprene units. These different isoprenoid precursors are used to divide the terpenoids into classes based on the number of isoprene units included in the chain.

Table 1.1 Terpene classifications by number of isoprenoid units

<table>
<thead>
<tr>
<th>Isoprene Units</th>
<th>Carbons</th>
<th>Common Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemiterpenoids</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Monoterpenoids</td>
<td>2</td>
<td>10 camphor</td>
</tr>
<tr>
<td>Sesquiterpenoids</td>
<td>3</td>
<td>15 artemisin</td>
</tr>
<tr>
<td>Diterpenoids</td>
<td>4</td>
<td>20 major components of rosin &amp; turpentine</td>
</tr>
<tr>
<td>Sesterterpenoids</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Triterpenoids</td>
<td>6</td>
<td>30 Sterols, saponins</td>
</tr>
<tr>
<td>Sesquarterpenes</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>Tetraterpenoids</td>
<td>8</td>
<td>40 carotenoids</td>
</tr>
<tr>
<td>Polyterpenoids</td>
<td>&gt;8</td>
<td>&gt;40 rubber</td>
</tr>
</tbody>
</table>

The chain elongation chemistry consists of elimination of the pyrophosphate from DMAPP or the growing chain to form an allylic cation which is attacked by the double bond in
IPP, and a stereospecific removal of a proton to form the new C-C bond. Thus prenyl diphosphate precursors can often further be subdivided depending on the stereochemistry of the molecules: cis (Z) or trans (E), further increasing the diversity of terpenes that can ultimately be generated 24–26.

1.2.3 Cyclization of prenyl diphosphate intermediates

Prenyl diphosphate precursors undergo cyclization in the next step of terpene biosynthesis and are catalyzed through the action of monoterpene, sesquiterpene, diterpene & other respective cyclases which are also referred to as synthases. The general mechanism consists of the initial loss of the pyrophosphate from the prenyl diphosphate precursor leaving an activated carbocation intermediate that can be followed by shifting around of the cation resulting in further cyclizations and rearrangements to undergo hydride shifts and methyl or alkyl migrations including ring closures within the hydrophobic active site of the terpene cyclases until a loss of a proton, hydroxylation, or other capture of an external nucleophile occurs to neutralize the molecule. Through these cyclization reactions a remarkably diverse set of terpene hydrocarbon skeletons are generated. The studies described this dissertation will mainly focus on this step of terpene biosynthesis and its relevant enzymes. Section 1.3 will give further information on the enzymes associated with this step.

1.2.4 Modifications

Terpenes, which have the molecular formula \((\text{C}_5\text{H}_8)_n\), can be subsequently modified into terpenoid products through several means including oxidations via cytochrome P450 hydroxylases 27, moving or removing methyl groups, adding functional groups such as acyl-, aryl-, or sugar moieties via glycosyl transferases, and truncations. The modifications and added functional groups attached to the carbon skeletons contribute to the diverse properties of the terpenoids often contributing to the bioactivity of the molecules. Figure 1.3 shows an overview of
the biosynthesis of terpenes and terpenoid products and gives examples of some of the resulting molecules that can be formed from secondary modifications, from the simple limonene hydroxylation to (-)-\textit{Trans}-carveol to the elaborate, multi-step process to transform Taxadiene to Taxol.

![Figure 1.3 Overview and examples of terpene synthesis.](image)

**Figure 1.3** Overview and examples of terpene synthesis. IPP/DMAPP molecules are condensed in a head-to-tail to form prenyl diphosphate precursors of differing lengths. These precursors are then cyclized by terpene cyclases and then modified to form the various terpenoid products. Reprinted in compliance with copyright permissions 12.
1.3 Terpene synthases

Terpene synthases (TPS) are subdivided by enzymatic mechanism into type I metal dependent ionization and type II protonation dependent cyclization. Type I terpene synthases have a mainly α-helical structure that adopts a common α-bundle fold, called the terpenoid synthase fold. Terpene transferases and cyclases from most organisms share this terpene synthase fold. The active site in this fold contains a canonical aspartate-rich DDXX(XX)D/E motif along with a secondary and less well conserved aspartate-rich or (N/D)DXX(S,T)XXX(D/E) motif (referred to as the NSE/DTE motif for short) that coordinate metal ions [bolded residues indicate metal binding] \(^{28,29}\). The NSE/DTE motif appears as NSE in most microbial and fungal cyclases and DTE in most plant cyclases. The coordinated metal ion, usually Mg\(^{2+}\) and sometimes Mn\(^{2+}\), assist the departure of the substrate diphosphate leaving group.

In contrast, the type II protonation-initiated TPS has active sites that have a double α-barrel fold in which a DXDD motif provides the proton donor that triggers initial carbocation formation \(^{16,28}\). The initial structures from most groups of TPS have been solved since the mid 1990’s, including monoterpen, sesquiterpene, triterpene and isoprenyl diphosphate synthases \(^{30-34}\) with the exception of diterpene synthase which was more recently solved in 2010 \(^{35}\). The terpene synthases that will be focus of this dissertation are the type I sesquiterpene synthases, specifically a group homologous to 5-epi-aristolochene synthase (EAS).

1.3.1 The sesquiterpene synthase EAS: structure and function

There are over 300 monocyclic, bicyclic, and tricyclic sesquiterpenoids with diverse structures and stereochemistries \(^{36}\). The formation of all these cyclic sesquiterpenes is catalyzed by sesquiterpene synthases which ionize FPP to initiate cyclization by electrophilic attack of the resultant allyl cation and guide the subsequent cationic transformations and quenching that
determine the products ultimately produced. Many sesquiterpene cyclases cyclize in a transoid fashion to initiate a 1,10-closure to generate the germacryl cation or a 1,11-closure to generate the humulyl cation²⁵ (Figure 1.4). Cisoid cyclization also occurs in certain sesquiterpene cyclases, but these are not touched upon in this work.

Figure 1.4  Two options for transoid cyclization by sesquiterpene cyclases.
A 1,10-closure of the farnesyl cation generates the germacryl cation (red). A 1,11-closure of the farnesyl cation generates the humulyl cation (blue).

The first plant terpenoid cyclase to have its crystal structure determined and the second terpenoid cyclase to have its structure reported was 5-epi-aristolochene synthase cloned from *Nicotiana tabacum* (TEAS)³⁰. The structure of TEAS contains two domains, a catalytically active C-terminal domain with the α-helical class I terpenoid synthase fold, and also an N-terminal domain of unknown function, that has an α-helical fold similar to that of a class II terpenoid synthase, specifically glycosyl hydrolases³⁷.
After a substrate enters the active-site in TEAS, the structure undergoes significant conformational changes in order to close the active site to protect reactive intermediates from abortive side-reactions. Loops flanking the mouth of the active site and N-terminal domain acts as a cap to cover the active site by docking in grooves defined by the A–C and D–D1 loops on one side, and the J–K and H–H–α1 loops on the other side (Figure 1.5). The initial ionization of FPP is assisted by Mg$^{2+}$ ions coordinated by two metal binding motifs, the aspartate-rich sequence on helix D and the DTE motif on helix H which flank the mouth of the active site cavity. TEAS has three magnesium ions in the active site, Mg$^{2+}_A$ and Mg$^{2+}_C$ are coordinated by D301 and D305 of the aspartate-rich motif D$^{301}$DXXD$^{305}$ and Mg$^{2+}_B$ is chelated by D$^{444}$DTAT$^{448}$YEVE$^{452}$ of the DTE motif.

**Figure 1.5 Structure of TEAS.** Helices in N-terminus (blue) labeled 1-8, helices in C-terminus (orange) labeled A-K. Reprinted in compliance with copyright permissions from AAAS. (Starks et al. 2007) $^{30}$.
A series of mutation experiments and structure studies revealed functions for several other residues in TEAS\textsuperscript{25,30,38}, the results of which have been summarized in Table 1.2. A few residues to highlight are Y520 which helps form the eudesmyl cation intermediate and T402 & V516 which act upon the eremophilenyl cation intermediate, both of which are central waypoints in the map of EAS biochemistry downstream of germacryl cation (refer back to Figure 1.4).
Table 1.2 Known residue functions in TEAS. References: 1. Starks et al. (1997) \textsuperscript{30}. 2. Greenhagen et al. (2006) \textsuperscript{38}, 3. Noel et al. (2010) \textsuperscript{25}. Abbreviations used: 4-EE = 4-epi-eremophiline, PSD = premnaspirodiene

<table>
<thead>
<tr>
<th>Position</th>
<th>Residue</th>
<th>Description</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>R</td>
<td>Forms salt bridge with E312</td>
<td>3</td>
</tr>
<tr>
<td>264</td>
<td>R</td>
<td>Stabilises the additional negative charge on the departing diphosphate in the first step of the reaction. The A-C loop, which contains Arg264 and Arg266, translates inward toward the active site on FHP binding, positioning the side chain of Arg264 in close proximity to the C1 hydroxyl group of FHP. HJK thinks it may be critical for de-PP (ionizing FPP)</td>
<td>1</td>
</tr>
<tr>
<td>266</td>
<td>R</td>
<td>Arg266 hydrogen bonds with residues on both the J-K loop and the NH2-terminal segment in the FHP complex. Both Arg264 and Arg266 are conserved among many terpene cyclases.</td>
<td>1</td>
</tr>
<tr>
<td>273</td>
<td>W</td>
<td>The aromatic quadrupole of Trp273 stabilises the electrostatic centre of eudesmane carboxylation at C7. It abstracts a proton from C8, forming the final product, 5-epi-aristolochene.</td>
<td>1</td>
</tr>
<tr>
<td>301</td>
<td>D</td>
<td>Asp\textsuperscript{301} coordinates Mg\textsubscript{b} \textsuperscript{2+} in the native TEAS structure, and the side chain carboxyl of Glu\textsuperscript{279} provides a longer range interaction.</td>
<td>1</td>
</tr>
<tr>
<td>302</td>
<td>D</td>
<td>Asp\textsuperscript{302} demonstrates no direct metal coordination.</td>
<td>1</td>
</tr>
<tr>
<td>305</td>
<td>D</td>
<td>Asp305 provides an additional coordination bond in the enzyme with substrate analogs bound. Asp301 and Asp305 are part of a DDXXD sequence found in terpene cyclases and were prev assumed to constitute a binding site for the required divalent metal</td>
<td>1</td>
</tr>
<tr>
<td>312</td>
<td>E</td>
<td>Forms salt bridge with R115</td>
<td>3</td>
</tr>
<tr>
<td>379</td>
<td>E</td>
<td>Asp\textsuperscript{312} coordinates Mg\textsubscript{b} \textsuperscript{2+} in the native TEAS structure, and the side chain carboxyl of Glu\textsuperscript{279} provides a longer range interaction.</td>
<td>1</td>
</tr>
<tr>
<td>401</td>
<td>T</td>
<td>Its carbonyl group stabilises the carboxylation intermediates.</td>
<td>1</td>
</tr>
<tr>
<td>402</td>
<td>T</td>
<td>Its backbone carbonyl group stabilises the carboxylation intermediates (farnesyl cation). Along with 516, directs stereochem of H+ elimination off of eremophilanyl cation. S402 correlates with 4-EE &amp; a little with PSD</td>
<td>1</td>
</tr>
<tr>
<td>403</td>
<td>T</td>
<td>Its hydroxyl group stabilises the carboxylation intermediates.</td>
<td>1</td>
</tr>
<tr>
<td>441</td>
<td>R</td>
<td>It stabilises the additional negative charge on the departing diphosphate in the first step of the reaction.</td>
<td>1</td>
</tr>
<tr>
<td>444</td>
<td>D</td>
<td>It acts as a general acid and base, removing the hydroxyl proton from Tyr 520 and after the protonation of germacrede C6, redonating the proton back to Tyr 520. The side chains of Asp444, Thr448, Glu452, and one water molecule coordinate Mg\textsubscript{b} \textsuperscript{2+}. D444 along with D525-alkylation</td>
<td>1</td>
</tr>
<tr>
<td>448</td>
<td>T</td>
<td>The side chains of Asp\textsuperscript{444}, Thr\textsuperscript{448}, Glu\textsuperscript{452}, and one water molecule coordinate Mg\textsubscript{b} \textsuperscript{2+}.</td>
<td>1</td>
</tr>
<tr>
<td>452</td>
<td>E</td>
<td>The side chains of Asp\textsuperscript{444}, Thr\textsuperscript{448}, Glu\textsuperscript{452}, and one water molecule coordinate Mg\textsubscript{b} \textsuperscript{2+}.</td>
<td>1</td>
</tr>
<tr>
<td>516</td>
<td>V</td>
<td>Along with 402, directs stereochem of H+ elimination off of eremophilanyl cation</td>
<td>2</td>
</tr>
<tr>
<td>520</td>
<td>Y</td>
<td>It donates a proton, accepted from Asp 525, to germacrede C6 in the formation of the eudesmane carboxylation intermediate.</td>
<td>1</td>
</tr>
<tr>
<td>521</td>
<td>Y</td>
<td>As the J-K loop (residues 521 to 534) becomes ordered, it forms a lid that clamps down over the active site entrance in the presence of FHP.</td>
<td>1</td>
</tr>
<tr>
<td>534</td>
<td>D</td>
<td>This clamp places Tyr527 next to Trp273, forming an extended aromatic box deep within the active site pocket.</td>
<td>1</td>
</tr>
<tr>
<td>525</td>
<td>D</td>
<td>It abstracts a proton from C13 to promote the formation of the germacrene intermediate. It then donates the proton to Tyr 520 to allow it to act as an acid.</td>
<td>1</td>
</tr>
<tr>
<td>527</td>
<td>Y</td>
<td>The quadrupole of Tyr 527 stabilises the positive charge on C11 of the germacrene cation, created after the formation of the C1-C10 bond.</td>
<td>1</td>
</tr>
</tbody>
</table>
1.4 Origins of terpene diversity

The type I terpene synthase fold is widely conserved through plants, bacteria, fungi, animals, and protists, but despite both low base sequence and amino acid sequence similarity, motifs critical for catalysis are conserved (as detailed previously in section 1.3). Currently, there exist several theories about the origin and evolution of the TPS family; one that suggests that plant and bacterial TPS are the result of convergent evolution because of the low level of overall sequence similarity among terpene synthases, however because of the widespread nature of the terpene synthase fold, another theory suggests that all TPSs diverged from a common primordial bacterial ancestor and that the plant TPS arose through subsequent domain fusions and loses, horizontal gene transfer, and sequence diversification.\textsuperscript{37,39,40}

Overall there are 6 main structural domains (α, β, γ, δ, ε, and ζ) that are found in terpene synthases.\textsuperscript{41} The α unit corresponds to the terpene synthase fold and catalyzes metal-dependent ionization of isoprenyl diphosphates (corresponds to the C-terminal domain of mono- & sesquiterpene cyclases). The β unit corresponds to the N-terminal domain of mono- & sesquiterpene cyclases that are structurally similar to the glycosyl hydrolases and contributes to the active site of type II proton-initiated ionization reactions along with the γ unit since type II active sites are located at the interface of the β and γ domains. The β domain is also though to assist with type I reactions indirectly by acting as a cap for the active site pocket located in associated α units. The δ unit governs chain elongation in head-to-tail \textit{trans}-prenyl transferases. The ε unit underlies head-to-head prenyl transferases which ultimately make the tri- & tetraterpene precursors of sterols and carotenoids. ζ units produce head-to-tail \textit{cis}-prenyl transferases that synthesize the \textit{cis}-isoprenoid diphosphates involved in bacterial cell wall biosynthesis.

In the world of plant terpene cyclases, only α, β, and γ units are seen. It is widely thought that the ancestral plant terpene cyclases was an αβγ protein, which themselves appear to have
originated by fusion of α and βγ domain proteins \(^{42,43}\). Plant di- and triterpene synthases have αβγ units, while plant mono- and sesquiterpene synthases are αβ and bacterial and fungal sesquiterpene synthases are just α (Figure 1.7). It has been suggested that all modern plant terpenoid synthases originated from a common αβγ diterpene synthase followed by gene duplications and divergence \(^{3,44,45}\).

Bioinformatics predictions made about plant diterpene cyclase structures inferences made from assuming that diterpene synthases were progenitors of other terpene synthases with structural information \(^{46}\) have been confirmed experimentally by the solving of the taxadiene synthase structure \(^{35}\), leading to added confidence in proposals that many plant terpene cyclases derived from ancestral αβγ proteins.

**Figure 1.6 Proposed domains evolution of terpene synthases.**
Proposed history of domains α, β, and γ in the terpene synthase family. Reprinted in compliance with copyright permissions from The Royal Society of Chemistry. (Gao et al. 2012) \(^{45}\).
While synthases can be grouped by catalytic function, they can also be grouped by phylogenetically. Within the diverse plant TPS family, genes can be further subdivided into seven subfamilies based upon amino acid sequence. Monophyletic origin of plant TPS is supported by the common exon/intron structure found in the seven TPS subfamilies (Figure 1.8). Angiosperm monoterpane, sesquiterpene, and diterpene synthases involved in secondary metabolism (including TEAS and the other EAS homologs discussed in this dissertation) are classified as Class III enzymes based on intron/exon patterns and have a total of 7 exons, where template introns I, II, VII, IX, X and template exons 4-6 are lost.

**Figure 1.7 Exon/Intron map of plant TPS genes.**
The top figure represents the general intron/exon template; the lower figure represents class III synthases which include the Angiosperm monoterpane, sesquiterpene, and diterpene synthases involved in secondary metabolism. Exons are boxes labeled 1-15. Introns numbered I-XIV and are represented by black vertical bars. Introns are classified into three phase types according to Li (1997). Reprinted and adapted in compliance with copyright permissions from Genetics Society of America. (Trapp & Croteau 2001).
1.5 Enzyme plasticity, natural variation and enzyme evolution

While most short-chain prenyl diphosphate transferases from primary metabolism have a fairly high product specificity, many of the terpene cyclases of specialized metabolism can be extremely promiscuous. For example, humulene synthase from *Abies grandis* produces more than fifty sesquiterpenes, while TEAS produces 25 minor products alongside its major product. Interestingly, small number and even a single amino acid residue changes can be sufficient to substantially alter product outcomes. Typically, these ‘plasticity’ residues are positioned within or proximal to the active site pocket. For instance, similar but nuanced metal binding interactions among type I terpene synthases, important for stabilizing proper active site activity, are a nexus for divergent product outcomes, as demonstrated by the similar metal binding but significant functional differences of trichodiene, (+)-bornyl diphosphate and 5-epi-aristolochene synthases. Even residues in surrounding layers beyond the active site can also contribute significantly to product specificity, as demonstrated by TEAS and HPS, both of which contain the same residues in their active sites, but yield differing product profiles because of residue differences farther out. As such, the prediction of product profiles based on purely amino acid sequence is not currently possible. And more insight into structure-function mechanisms of the highly diverse TPS family is needed before we can truly understand terpene biosynthesis.

The capacity to form multiple products is not universal and some TPS are nearly specific in forming one product; even when considering synthases with high sequence homology, some might produce multiple products while others form only one product. A striking example of the issue is the transformation of seven product-specific sesquiterpene synthases from the promiscuous γ-humulene synthase accomplished by performing site-directed mutagenesis of 19
residues surrounding its active site. This raises the question: what structural elements control product specificity and plasticity? There are many theories on how TPSs can dictate their product outcomes, such as by restricting their substrate and reaction intermediates to a subset of conformations or by concatenating multiple steps into concerted, asynchronous steps in order to avoid side-reactions off of reactive intermediates, but the question remains unanswered.

Nonetheless, the origin of variation in product diversity from one sesquiterpene cyclase to the next is likely the result of the transition from primary metabolism or acquired as the enzymes evolved within specialized metabolism either as a feature of environmental adaptation or as a by-product of incomplete transitioning or lack of negative selection pressure. Gene duplication, polyploidization and mutational drift are thought to be important contributors to the expansion of the TPS family since essential functions of primary metabolism can be retained while other copies of the enzymes can undergo drift and selection for neofunctionalization.

1.6 Segue and chapter guides

While most of the effort to understand enzyme structure and function has concentrated within the active site, since it is the most obvious area to attempt to deconstruct enzyme catalytic mechanisms, findings alluded to earlier by Greenhagen et al. indicate that ecto-catalytic site residues can contribute significantly to product specificity by redirecting reaction pathways. These residue positions were initially identified through comparisons of phylogenetic variation of related enzymes that performed very different catalytic functions. Hence, the core inspiration behind the body of work described in this dissertation is to utilize information on functional diversification provided in nature as a source to unravel the mysteries of enzyme structure-function. Chapter 2 begins this exploration by cataloging enzyme diversity of a specific metabolic pathway through the Solanaceae family of plants while Chapter 3 and 4 follow up with in-depth
structure-function analysis of two interesting enzyme variants that were discovered during the work described in Chapter 2. The final chapter contextualizes the results with regard to the central question of how terpene synthases neofunctionalize and diversify.
1.7 References

1. Dictionary of Natural Products. at <http://dnp.chemnetbase.com>


CHAPTER 2

Cloning and Characterization of 5-epi-aristolochene synthase homologs from plants of the

*Nicotiana* genus

2.1 Abstract

Enzyme diversification and neofunctionalization were examined in this study through surveying homologs of EAS, FPP and EAH. Homologs were sequenced and cloned from several species of *Nicotiana* and from a few select species outside the genus within the *Solanaceae* plant family in order to evaluate what variations of the enzymes have appeared through the diversification of the plant lineages. Cloned enzyme variants were evaluated for functional changes both biochemically and thermodynamically revealing a greater amount of enzyme diversity than originally anticipated at the onset of the investigation.

2.2 Introduction

Plants, rooted in place, must employ special tactics to interact and communicate with biotic and abiotic factors in their surrounding environments. Instead of fleeing, plants often produce chemical compounds to deter herbivory, pests and pathogens \(^1-^5\) and in so doing, increase their fitness for survival. Of these natural compounds, terpenoids are the largest and most diverse class. The accretion of novel metabolic activities is hence an important contributor to evolutionary change, but is not well understood. To partly address the question of how plants amassed such an extensive arsenal of natural chemicals, we focus on a specific plant
sesquiterpene biosynthetic pathway that has undergone extensive divergence in a large family of geographically dispersed plants within the Solanaceae plant family.

Many members of the Solanaceae family produce sesquiterpenoid phytoalexins in response to cell-wall damage or challenge by various fungal elicitors. In one species of the family, *Nicotiana tabacum* (common tobacco) produces the antifungal substance capsidiol through a three enzyme pathway. In another species, *Hyoscyamus muticus* (Egyptian henbane) contains a homologous pathway that produces the antifungal solavetivone instead (Figure 2.1). Both compounds are phytoalexins that afford protection against different plant pathogens. Since the base pathway has spread widely throughout the entire Solanaceae plant family, it provides an ideal opportunity to study the emergence of new enzyme functionalities and metabolic evolution in context of what nature has already attempted.

![Figure 2.1 Parallel biosynthesis pathways of the phytoalexins capsidiol and solavetivone.](image)

In one pathway found in tobacco, TEAS turns FPP into 5-EA which then is turned by EAH into the phytoalexin capsidiol. In the other pathway found in henbane HPS turns FPP into premnaspirodiene and then is oxidized into solavetivone by HPO.
Interestingly, the enzymes in the second step of the pathway, tobacco 5-epi-aristolochene synthase (TEAS) and the cognate synthase henbane prennaspirodiene synthase (HPS), have a relatively high 78% amino acid identity, yet they cyclize the same substrate, farnesyl diphosphate (FPP) into two distinct products, 5-epi-aristolochene (5-EA) and prennaspirodiene (PSD), respectively. Currently, predicting product profiles based on TPS amino acid sequence is not possible. Even with relatively high sequence homology, there is no simple correlation of protein sequence with the structure of the cyclized terpene product so the biochemical function of an unknown terpene synthase must therefore be assessed by direct experimentation, at least perhaps until the finer nuances of enzyme structure-function are elucidated.

In a study by O’Maille et al., the interconversion of a minimal set of 9 residue positions surrounding the active site of TEAS and HPS was sufficient to completely transform their catalytic activities to that of the other enzyme. Furthermore this study examined the functional changes associated with theoretical intermediary enzymes, when a library of 512 ($2^9$) mutant enzymes representing every possible TEAS-HPS ‘ancestral’ sequence intermediate was constructed in the TEAS enzyme scaffold and assayed for product distribution. Interestingly, in this in vitro experiment a continuous spectrum of the products, in terms of product ratios, was observed, leading to the question of what sequence space has nature actually sampled.

Studying the natural variation that occurs in EAS-like enzymes as they diverged throughout several related plant species allows an in-depth examination of the functional diversification of an enzyme family critical to plant adaptation in the face of varying pathogens. This study also provides a starting point to understand how mutational variation affects fundamental properties associated with biological catalysis. Hence, EAS homologs were cloned, sequenced, expressed as enzymes and then biochemically characterized from several species of *Nicotiana*. Characterization of these EAS-like enzymes focused on thermostability, kinetic properties, and product identity and diversity.
2.3 Results and Discussion

In 2009, at the beginning of this study, very little Solanaceae sequence information was available and it was not until a few years into the project that Solanaceae genomes started to get sequenced (Table 2.1). In fact, the only other full-length EAS gene sequences available at the time, in addition to TEAS, was from *N. attenuata*, which is very closely related to *N. tabacum*. With the paucity of genetic data available for the species in Nicotiana, it was necessary to survey the extant plant enzyme diversity to estimate how much EAS sequence diversity exists within that group of related plants. Therefore the first objective was to obtain another homolog of EAS from a more distantly related species in order to be able to design ‘universal’ amplification primers against regions outside the coding sequence. Since Nicotiana is believed to have originated in South America, due to the South American-centric distribution pattern of the species, the Australian species *N. benthamiana* appeared to be a good starting candidate to survey because Australian species would most likely have had the most time and opportunity to diverge (Figure 2.2); and it is also a fairly well-characterized plant and readily available in the lab.
Table 2.1 Genomes sequenced to-date of Solanaceae plants.


<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome size</th>
<th># of genes predicted</th>
<th>Year of completion</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Solanum lycopersicum</em> (tomato) cv. Heinz 1706</td>
<td>~900 Mbp</td>
<td>34,727</td>
<td>2011</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td>2</td>
</tr>
<tr>
<td><em>Solanum pimpinellifolium</em> (Currant Tomato)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Solanum tuberosum</em> (potato)</td>
<td>856 Mbp</td>
<td>39,031</td>
<td>2011</td>
<td>3</td>
</tr>
<tr>
<td><em>Solanum commersonii</em> (commerson's nightshade)</td>
<td>840 Mbp</td>
<td>36,662</td>
<td>2015</td>
<td>4</td>
</tr>
<tr>
<td><em>Solanum melongena L.</em> (eggplant)</td>
<td>833.1 Mb</td>
<td>84,446</td>
<td>2014</td>
<td>5</td>
</tr>
<tr>
<td><em>Nicotiana benthamiana</em></td>
<td>~3 Gbp</td>
<td></td>
<td>2012</td>
<td>6</td>
</tr>
<tr>
<td><em>Nicotiana sylvestris</em></td>
<td>2.59 Gbp</td>
<td></td>
<td>2013</td>
<td>7</td>
</tr>
<tr>
<td><em>Nicotiana tomentosiformis</em></td>
<td>2.22 Gbp</td>
<td></td>
<td>2013</td>
<td>7</td>
</tr>
<tr>
<td><em>Nicotiana tabacum</em></td>
<td>4.41 Gb</td>
<td>&gt;90,000</td>
<td>2014</td>
<td>8</td>
</tr>
<tr>
<td><em>Capsicum annuum</em> (Pepper)</td>
<td>~3.48 Gbp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) cv. CM334 (b) cv. Zunla-1</td>
<td></td>
<td>(a) 34,903</td>
<td>(a) 2014</td>
<td>(a) 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) 53,336</td>
<td>(b) 2014</td>
<td>(b) 10</td>
</tr>
<tr>
<td><em>Capsicum annuum</em> var. glabriusculum* (Chiltepin)</td>
<td>~3.48 Gbp</td>
<td></td>
<td>2014</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 2.2 Phylogenetic tree of *Nicotiana* genus.
Highlighted species were included in this study; highlights are color coded to indicate the species origin according to figure key. Figure adapted and reprinted with permission of Elsevier Limited from Clarkson *et al.* (2004) 20.

The genomic sequence of *N. benthamiana* EAS (NbEAS) was obtained through DNA walking, a simple method to uncover unknown genomic DNA sequences adjacent to a known sequence such as a cDNA 21, using primer sequences designed against TEAS. Sequence comparison of NbEAS and TEAS revealed protein sequences with approximately 95% identity.
Unfortunately, DNA sequences in the upstream and downstream UTRs were not conserved enough to design universal cloning primers for *Nicotiana* outside the coding region. However, the 5’ start and 3’ end coding regions were fairly conserved, so degenerate primers were designed against an alignment of available *Solanaceae* EAS sequences to target regions of about 17-21 bases just within the EAS open reading frame.

Since the amplification primers used to clone the EAS homologs had to be designed against portions in the coding region, this obscured the true identity of 6-7 amino acid residues on each terminus. 5’ and 3’ RACE was implemented to find to try and find the true end sequences. The RACE data however revealed a formidable complexity in the genomic architecture of these allopolyploid plants, fraught with multiple homeologous genes, in varied loci, and possibly several pseudogenes. This genomic complexity complicated the process of cataloging all the enzyme variants present in each species. In response, high-throughput amplicon targeted sequencing of FPPS, EAS and EAH was attempted in hopes of efficiently sorting out all the homeologous genes from each species in an efficient manner as well as quickly obtaining homology sequences of all three enzymes of the capsidiol biosynthetic pathway. The high-throughput amplicon targeted resequencing of the pathway genes and cDNA was carried out on the Ion Torrent platform. The sequenced reads were however unable to be resolved into separate alleles since the distances between SNPs was on average longer than the read lengths; and also the frequency of sequencing errors interfered with unequivocal calling of the SNPs.

In the end, EAS genomic homologs/orthologs were isolated using the original primers to amplify EAS from 29 out of 32 *Nicotiana* species using conventional PCR techniques (Figure 2.3 & Table 2.2). Ultimately, EAS cDNA sequences were cloned from a subset of the *Nicotiana* species, consisting only of those that were expressed constitutively, since attempts at inducing expression using methyl jasmonate were not effective, yielding a total of 15 different protein variants originating from 9 species (Figure 2.4). The EAS variant cDNA sequences were aligned
to reveal several residue positions of varying conservation with most of the sequence divergence occurring in the non-catalytic N-terminal (β) domain (Figure 2.5). Considered as a set, there is a 78% amino acid sequence conservation among the 15 EAS protein sequences (123 variable residues out of 548); compared in pairs, the conservation averaged about 95-98%. When these variable residue positions are compared to those in the set of ‘M9’ positions originally examined by Greenhagen et al.¹, there are 2 residue positions (436 & 516) that appear in both the Nicotiana set and the M9 set. This result was promising, since changes to the M9 residue positions (the ‘M9’ positions are 9 residues that can interconvert EAS and EAH function) serve as compelling circumstantial evidence of EAS variant neofunctionalization.

Figure 2.3  Cloning EAS genes from Nicotiana species
Nicotiana species are number coded as in Table 2.2.
### Table 2.2 List of *Nicotiana* species and numeric code used in this study.

Species origin information reference ²³.

<table>
<thead>
<tr>
<th>ID#</th>
<th><em>Nicotiana</em> species</th>
<th>Chromosomes</th>
<th>Ploidy</th>
<th>GRIN acc.#</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>N01</td>
<td>acaulis</td>
<td>24</td>
<td>2</td>
<td>PI 555468</td>
<td>South America - southern Argentina</td>
</tr>
<tr>
<td>N02</td>
<td>acuminata</td>
<td>24</td>
<td>2</td>
<td>PI 555469</td>
<td>South America - central Chile</td>
</tr>
<tr>
<td>N03</td>
<td>africana</td>
<td></td>
<td></td>
<td>PI 444472</td>
<td>Africa - central Nambia</td>
</tr>
<tr>
<td>N04</td>
<td>alata</td>
<td>18</td>
<td>2</td>
<td>PI 42334</td>
<td>South America - Uruguay, southern tip of Brazil</td>
</tr>
<tr>
<td>N05</td>
<td>attenuata</td>
<td>24</td>
<td>2</td>
<td>PI 555474</td>
<td>North America - western US</td>
</tr>
<tr>
<td>N06</td>
<td>benoidesii</td>
<td>24</td>
<td>2</td>
<td>PI 555477</td>
<td>South America - a spot in Peru</td>
</tr>
<tr>
<td>N07</td>
<td>benthamiana</td>
<td>38</td>
<td>4</td>
<td>chory lab</td>
<td>Australia - Northwest</td>
</tr>
<tr>
<td>N08</td>
<td>developandi</td>
<td>48</td>
<td>4</td>
<td>PI 555492</td>
<td>North America - southern California &amp; Baja California</td>
</tr>
<tr>
<td>N09</td>
<td>cordifolia</td>
<td></td>
<td></td>
<td>PI 555493</td>
<td>South America - islands off Chile</td>
</tr>
<tr>
<td>N10</td>
<td>forgetiana</td>
<td>18</td>
<td>2</td>
<td>PI 555501</td>
<td>South America - southern tip of Brazil</td>
</tr>
<tr>
<td>N11</td>
<td>glauca</td>
<td>24</td>
<td>2</td>
<td>PI 555686</td>
<td>South America - Bolivia, Argentina</td>
</tr>
<tr>
<td>N12</td>
<td>gossei</td>
<td>36</td>
<td>4</td>
<td>PI 230953</td>
<td>Australia - center</td>
</tr>
<tr>
<td>N13</td>
<td>knightiana</td>
<td>24</td>
<td>2</td>
<td>PI 555527</td>
<td>South America - southern Peru</td>
</tr>
<tr>
<td>N14</td>
<td>longiflora</td>
<td>20</td>
<td>2</td>
<td>PI 555534</td>
<td>South America - Uruguay, Paraguay, northern Argentina</td>
</tr>
<tr>
<td>N15</td>
<td>miersi</td>
<td>24</td>
<td>2</td>
<td>PI 555537</td>
<td>South America - central Chile</td>
</tr>
<tr>
<td>N16</td>
<td>noctiflora</td>
<td>24</td>
<td>2</td>
<td>PI 417918</td>
<td>South America - northwestern Argentina</td>
</tr>
<tr>
<td>N17</td>
<td>obtusifolia</td>
<td>24</td>
<td>2</td>
<td>PI 555543</td>
<td>North America - southwestern US</td>
</tr>
<tr>
<td>N18</td>
<td>accidentalis</td>
<td>42</td>
<td>4</td>
<td>PI 555541</td>
<td>Australia - big patch</td>
</tr>
<tr>
<td>N19</td>
<td>otophora</td>
<td>24</td>
<td>2</td>
<td>PI 235553</td>
<td>South America - central Bolivia and South into Argentina a bit</td>
</tr>
<tr>
<td>N20</td>
<td>paniculata</td>
<td>24</td>
<td>2</td>
<td>PI 241769</td>
<td>South America - western Peru</td>
</tr>
<tr>
<td>N21</td>
<td>plumaginifolia</td>
<td>20</td>
<td>2</td>
<td>PI 302476</td>
<td>South America - small patches across mid-latitude line</td>
</tr>
<tr>
<td>N22</td>
<td>quadrivalvis</td>
<td>48</td>
<td>4</td>
<td>PI 555481</td>
<td>North America - western US</td>
</tr>
<tr>
<td>N23</td>
<td>repando</td>
<td>48</td>
<td>4</td>
<td>PI 555552</td>
<td>North America - western gulf coast, Alabama</td>
</tr>
<tr>
<td>N24</td>
<td>rustica</td>
<td>48</td>
<td>4</td>
<td>PI 555692</td>
<td>South America - Peru, extending into Ecuador and Bolivia</td>
</tr>
<tr>
<td>N25</td>
<td>solanifolia</td>
<td>24</td>
<td>2</td>
<td>PI 555558</td>
<td>South America - Northern Chile</td>
</tr>
<tr>
<td>N26</td>
<td>suaveolens</td>
<td>32</td>
<td>4</td>
<td>PI 230960</td>
<td>Australia - southeast tip</td>
</tr>
<tr>
<td>N27</td>
<td>sylvestris</td>
<td>24</td>
<td>2</td>
<td>PI 555569</td>
<td>South America - a strip crossing Bolivian and Argentinian border</td>
</tr>
<tr>
<td>N32</td>
<td>tabacum (TC311) hicks broadleaf</td>
<td>48</td>
<td>4</td>
<td>PI 552397</td>
<td>originally South America - region crossing Bolivian and Argentinian border</td>
</tr>
<tr>
<td>N33</td>
<td>tabacum (TC590) KY14</td>
<td>48</td>
<td>4</td>
<td>TC590</td>
<td>originally South America - region crossing Bolivian and Argentinian border</td>
</tr>
<tr>
<td>N34</td>
<td>tomentosa</td>
<td>24</td>
<td>2</td>
<td>PI 574525</td>
<td>South America - strip crossing Peruvian and Bolivian border</td>
</tr>
<tr>
<td>N35</td>
<td>tomentosiformis</td>
<td>24</td>
<td>2</td>
<td>PI 555572</td>
<td>South America - spot in central Bolivia</td>
</tr>
<tr>
<td>N36</td>
<td>velutina</td>
<td></td>
<td></td>
<td>PI 244630</td>
<td>Australia - central</td>
</tr>
<tr>
<td>N37</td>
<td>wigandioides</td>
<td>24</td>
<td>2</td>
<td>PI 302471</td>
<td>South America - central Bolivia</td>
</tr>
<tr>
<td>N38</td>
<td>tomentosiformis</td>
<td>24</td>
<td>2</td>
<td>PI 555572</td>
<td>South America - spot in central Bolivia</td>
</tr>
</tbody>
</table>
Figure 2.4 Sequence phylogeny tree of cloned EAS from *Nicotiana* species

Sequence phylogeny tree generated using unrooted, neighbor-joining method in MEGA version 4 (Tamara, Dudley, Nei, and Kumar 2007). Amino acid sequence alignment only displays non-consensus positions; non-consensus residues are highlighted in yellow.
In order to examine presence of neofunctionalization in the EAS natural variants, the variants were expressed in bacteria, after which they were tested for enzymatic function in in vitro reactions with FPP as the input substrate. These reactions were quantified by GC-MS to reveal that natural variation did indeed alter product profiles (Figure 2.6). Some variants were
rendered effectively dead making little to no product, other variants retained TEAS-like activities producing mainly 5-epi-aristolochene, and a select few gained completely new functionalities making major products out of products that were minor products or undetectable in TEAS. Two of the most altered product profiles came from EASs cloned from *N. longiflora* and *N. tomentosa*, producing significantly more germacrene A (3-fold increase) and more α- and β-selinenes (12-fold increase collectively) respectively at the cost of 5-epi-aristolochene production (dropping 64% and 92%) when compared to TEAS (Figure 2.7). The *N. longiflora* and *N. tomentosa* EAS variants have less total carbon output as compared with TEAS, making 89% and 64% less total products respectively, so there is a loss in throughput that comes with the change in enzyme functionalities.
Figure 2.6: Comparison of products from EAS from various *Nicotiana* species

Cloned EAS from *Nicotiana* were assayed *in vitro* and the products of the enzymes were run on GC-MS for identification and quantification. Some enzymes variants lost function, some retained TEAS-like activity, and others gained all new functionalities.
Figure 2.7  Two *Nicotiana* EAS variants with distinctly novel product profiles as compared to TEAS. An EAS variant from N. longiflora (N14-3E) and a variant from N. tomentosa (N35-1E) have very distinct product profiles compared to TEAS. The N. longiflora EAS produces germacrene A as its major product (47% of total carbon output). The N. tomentosa variant produces increased amounts of several products, most prominently the α- and β-selinens which represent 50% of the total carbon output of the enzyme. Percentages in dark blue indicate total carbon output of enzyme as compared to TEAS. Percentages next to each product peak represent the product percentage as total carbon output of the enzyme; colors coded per key.

While measuring product identity and diversity is *de facto*, measuring shifts in melting temperature ($T_m$) also provides an indication of the effects of differing residues on the thermostability of the enzymes which is again critical to their function in natural settings. So, the thermostabilities of all 15 EAS variants were assayed, revealing that these variants did show a significant change in melting temperature which infers a change in stability. The melting temperatures ranged from a high of 45°C (like TEAS) to as low as 38°C; and several lacked a distinct melting temperature (Table 2.3), which may say something about enzyme dynamics, such
as a disordered or a multiple conformation state. There does appear to be a correlation with the nature of the thermal unfolding curve and the enzyme activity. For example, *N. forgetiana* (N10), *N. gossei* (N12), *N. longiflora* (N14), and *N. gossei* (N12) all have a distinct T_m and produce a main major product, while *N. repanda* (N23) has a T_m range and does not make a distinct major product. Also, EAS from *N. velutina* (N36) appears to be unfolded considering the melting curve was flat and no products were seen in the *in vitro* assay.

**Table 2.3 Melting temperatures of *Nicotiana* EAS enzyme variants.**
The melting temperatures ranged from a high of 45°C (like TEAS) to as low as 38°C; and several lacked a distinct melting temperature which may indicate a disordered or multi-conformational state.

<table>
<thead>
<tr>
<th>EAS Variant</th>
<th>approx. T_m</th>
<th>Species</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>N10-2K</td>
<td>45</td>
<td><em>N. forgetiana</em></td>
<td>South America</td>
</tr>
<tr>
<td>N12-1-6K</td>
<td>45</td>
<td><em>N. gossei</em></td>
<td>Australia</td>
</tr>
<tr>
<td>N12-1E</td>
<td>45</td>
<td><em>N. gossei</em></td>
<td>Australia</td>
</tr>
<tr>
<td>N14-1K</td>
<td>45</td>
<td><em>N. longiflora</em></td>
<td>South America</td>
</tr>
<tr>
<td>N14-2K</td>
<td>45</td>
<td><em>N. longiflora</em></td>
<td>South America</td>
</tr>
<tr>
<td>N14-3E</td>
<td>45</td>
<td><em>N. longiflora</em></td>
<td>South America</td>
</tr>
<tr>
<td>N17-1K</td>
<td>40</td>
<td><em>N. obtusifolia</em></td>
<td>North America</td>
</tr>
<tr>
<td>N23-1K</td>
<td>39-44</td>
<td><em>N. repanda</em></td>
<td>North America</td>
</tr>
<tr>
<td>N23-1-6E</td>
<td>37-43</td>
<td><em>N. repanda</em></td>
<td>North America</td>
</tr>
<tr>
<td>N23-2E</td>
<td>38</td>
<td><em>N. repanda</em></td>
<td>North America</td>
</tr>
<tr>
<td>N23-4K</td>
<td>disordered</td>
<td><em>N. repanda</em></td>
<td>North America</td>
</tr>
<tr>
<td>N27-1K</td>
<td>NA</td>
<td><em>N. sylvestris</em></td>
<td>South America</td>
</tr>
<tr>
<td>N32-1K</td>
<td>disordered</td>
<td><em>N. tabacum</em> cv hicks broadleaf</td>
<td>South America</td>
</tr>
<tr>
<td>N32-1-6E</td>
<td>disordered</td>
<td><em>N. tabacum</em> cv hicks broadleaf</td>
<td>South America</td>
</tr>
<tr>
<td>N35-1-3K</td>
<td>38-42</td>
<td><em>N. tomentosa</em></td>
<td>South America</td>
</tr>
<tr>
<td>N35-1E</td>
<td>38-42</td>
<td><em>N. tomentosa</em></td>
<td>South America</td>
</tr>
<tr>
<td>N36-1K</td>
<td>NA</td>
<td><em>N. velutina</em></td>
<td>Australia</td>
</tr>
<tr>
<td>N36-1-4E</td>
<td>NA</td>
<td><em>N. velutina</em></td>
<td>Australia</td>
</tr>
<tr>
<td>TEAS</td>
<td>40</td>
<td><em>N. tabacum</em></td>
<td>South America</td>
</tr>
</tbody>
</table>
Ultimately, this survey into the diversification of a single specialized metabolic pathway as it has spread throughout a plant family has revealed that there is much diversity even within a single species, let alone genus or family. Most species that were surveyed had at least 2-3 and many had even more homologs of EAS. In the case of *N. longiflora*, one of these EAS variants produced a very distinct product profile from its other EAS homologs. There is still much to explored territory in terms of enzyme diversification and neofunctionalization. In the course of cloning EAS homologs from *Nicotiana*, partial EAS sequences from species in other clades of *Solanaceae* were also cloned (*Cestrum, Hyoscamus, and Lycium*). Although have not yet been characterized, the protein sequence alignment shows that there is much variation to explore. For example when compared to TEAS, an EAS from *Cestrum tomentosum* has a 69.8% identity (88.5% similar) in a 523 residue overlap and an EAS from *Hyoscamus gyofferi* has a 75.2% identity (89.7% similar) in 524 residue overlap. The identity is, not unexpectedly, perhaps more divergent when compared against the *Nicotiana* EASs which had individual identities to TEAS ranging in the 95-98% range. The other question that remains beyond the scope of diversity that exists, is how these variations are able to affect enzyme function and potentially contribute to molecular and organismal evolution. This question is addressed further in Chapters 3 and 4 of this work with in-depth characterization and mutational studies on the two enzyme variants cloned from *N. longiflora* and *N. tomentosa* that produced novel product profiles.

### 2.4 Materials and Methods

*Plant Tissue & elicitation.* Plants tissue was obtained from plants grown from seeds sourced from the USDA Germplasm Resources Information Network (GRIN) under full spectrum lights (see Table 2.2 for accession IDs). Elicitation by wounding & addition of methyl jasmonate: crushed leaf tissue with glass rod to create a damage circle about 0.5cm in diameter and pipetted
on 10ul of a 20mg/L solution of jasmonic acid diluted in 40% ethanol. Bag elicitation was conducted using 2ul methyl jasmonate into 50ul ethanol soaked into a cotton ball which was placed together with a plant in a sealed plastic bag for 48 hrs. Tissues were ground under liquid nitrogen and DNA/RNA extractions were performed with DNeasy & RNeasy Plant Mini kits (Qiagen).

Obtaining EAS sequences. Degenerate primers 5’-ATGGCCTCAGCAGCAGTWG-3’ and 5’-TCAAATTYYGATGGAGTCCAC -3’ were designed from a consensus alignment of EAS-like genes from *N. benthamiana* (SGN-E1189801, SGN-E1189802, CK286192, CK287078, CK288631 & CK287079), *N. attenuata* (AF484124), *N. plumbagnifolia* (FM244695), *N. tabacum* (Q40577) from the Sol Genomics Network and GenBank. The Genome Walker kit (Clonetech) was used to obtain EAS gene sequence from *N. benthamiana*. cDNA sequences were made using SuperScript III kit (Life Technologies).

High-throughput Targeted Resequencing. FPPS, EAS, and EAH were amplified from genomic DNA and cDNA from *Nicotiana* species. Bands of interest were gel purified before library prep where sequences were barcoded by species of origin. Library prep and sequencing were performed according to Ion Torrent protocols. Sequences were assembled and analyzed using CLC Genomics Workbench (CLC bio LLC.)

Protein Expression and Purification. Enzyme cDNA were cloned into the Escherichia coli expression vector pHis8-3. Constructs were transformed into *E. coli* BL21(λDE3). Transformed *E. coli* was grown at 37°C in Terrific broth containing 50μg/ml kanamycin until an A600 value of 1.0 was reached. After induction with 0.1mM isopropyl 1-thio-β-galactopyranoside, the cultures were grown for O/N at 18°C. Cells were harvested by centrifugation, resuspended in lysis buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 20mM
imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol, 1% [v/v] Tween 20). After sonication and centrifugation, the supernatant was passed over a Ni²⁺-NTA column and washed with 10 bed volumes of lysis buffer and 10 bed volumes of wash buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 20mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol), and then the His-tagged protein was eluted with elution buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 250mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol). His-tag was cut off in an overnight incubation with thrombin at 4°C. Proteins were buffer exchanged and concentrated with a final buffer (12.5mM Tris-HCl [pH 8.0], 50 mM NaCl, 2 mM DTT) by centrifugation using 50kDa weight cut-off columns (Millipore Amicon Ultracel).

**Enzyme Products Assay.** Assays for recombinant sesquiterpene synthase activity were performed in 500ul reaction volumes in assay buffer (50mM Bis-Tris Propane, pH 8.0; 1mM DTT; 5mM MgCl₂) with 1.0μM enzyme and 100μM FPP (Isoprenoid.com). Reactions were overlaid with 500ul hexane with p-chlorotoluene as a loading standard in 2-ml screw-top glass vials and were allowed to react overnight at room temperature. The reaction products were extracted by vigorous vortexing for 3 minutes followed by centrifugation at 3200xg for 30 min. Products were analyzed by using a Hewlett–Packard 6890 gas chromatograph (GC) coupled to a 5973 mass selective detector (MSD) outfitted with a 7683 series injector and autosampler and equipped with an HP-5MS capillary column (0.25 mm i.d. × 30 m with 0.25 μm film) (Agilent Technologies). For these experiments, needle sampling depth was set to 6 mm, placing the needle in the middle of the organic layer (near the 700-μl level in the 2-ml glass vial). The GC was operated at a He flow rate of 2 ml/min, and the MSD was operated at 70 eV. Splitless injections (3 μl) were made with an injector temperature of 250 °C. The GC was programmed with an initial oven temperature of 50 °C (5-min hold), which was then increased 10 °C min⁻¹ up to 180 °C (4-min hold), followed by a 100 °C min⁻¹ ramp until 250 °C (1-min hold). A solvent delay of 5.5
min was allowed prior to the acquisition of MS data. Product peaks were quantified by integration of peak areas using Enhanced Chemstation (version E.02.02, Agilent Technologies) and expressed as percentage of the total products. Of note, germacrene A was detected as the thermally rearranged product β-elemene. Product identification was performed using MassFinder 4 software (D. H. Hochmuth, D. Joulain, W. A. König, www.massfinder.com).

*Thermal Unfolding Assay.* Experiment was performed with 8μM of enzyme in a buffer containing 12.5mM Tris [pH 8.0], 50mM NaCl, 2mM DTT. Enzymes were incubated in buffer for 15 minutes at room temperature before adding the fluorophor SYPRO Orange (Invitrogen, Carlsbad, CA) at 10x concentration. Reactions were performed in 10ul volumes in 384-well microplates. Reactions were heated from 20 to 80°C at a rate of 0.06°C/s in a thermal cycler with fluorescent detection (Roche LightCycler 480), and the fluorescence intensity was recorded at 0.1°C intervals. Fluorescence intensities were plotted as a function of temperature, and the midpoint of the unfolding transition was taken as the melting temperature.
References


CHAPTER 3

Characterization of a sesquiterpene synthase from *N. longiflora* reveals a residue position that controls product specificity

3.1 Abstract

While mining for sesquiterpene synthase homologs of the well-characterized *Nicotiana tabacum* 5-epi-aristolochene synthase (TEAS) in closely related plant species, we discovered a new highly similar sesquiterpene synthase from *Nicotiana longiflora* (NlgfS) that makes germacrene A as its major product instead of 5-epi-aristolochene. Mutation and functional characterization of TEAS and NlgfS enzymes revealed that a single residue switch at position 460, located relatively distal to the active site, determines the product specificity of these two sesquiterpene synthases. This discovery highlights that distal residues can have a dramatic affect on active site chemistry and may provide further insight into rational enzyme engineering.

3.2 Introduction

Sesquiterpenes are a diverse class of terpenes of which there are over 300 known monocyclic, bicyclic, and tricyclic products with widely varied stereochemistries. Sesquiterpenes and their biochemically modified sesquiterpenoid products are found naturally in plants and insects, as semiochemicals that help enhance the organism’s fitness, e.g. phytoalexins or volatile attractants. The chemical foundation for generating sesquiterpene scaffold & stereochemical diversity is the formation and stabilization of carbocationic intermediates.
mediated by the sesquiterpene synthases. Sesquiterpene synthases are mainly α-helical, divalent metal-dependent enzymes (generally Mg$^{2+}$) that catalyze the diphosphate cleavage, cyclization, and rearrangements of the 15-carbon isoprenoid precursor, farnesyl diphosphate (FPP) to form the wide array of sesquiterpene products. These processes can be quite complex and presently confounds attempts at rational engineering of terpene synthases.

5-epi-aristolochene synthase from common tobacco (Nicotiana tabacum), acronym TEAS, is a well-characterized sesquiterpene synthase, with extensive functional studies and crystallographic structure data. TEAS cyclizes FPP into 5-epi-aristolochene which is further hydroxylated into the sesquiterpenoid phytoalexin capsidiol and is inducible by the elicitor, methyl jasmonate, in elicitor-treated tobacco tissues. In a previous studies of TEAS, nine residue positions in and around the active site were identified to direct product selectivity by comparison to a phylogenetically related but functionally distinct sesquiterpene synthase from Hyoscamus muticus. Furthermore, these nine positions were reciprocally interconverted in all combinations and revealed that TEAS was tremendously pliable in terms of the breadth of product profile space attainable with just these mutation combinations. This pliability allows nature to explore its chemical arsenal by inexpensively substituting a single (or limited number) of residues and shaking up the makeup of its chemical production.

The objective of this study was to find novel homologs of TEAS with altered enzyme activity from within the Nicotiana genus and from these homologs, learn more about what controls catalytic specificity. To that end, we cloned an EAS homolog from Nicotiana longiflora, the Longflower Tobacco, a species of tobacco native to South America. This species is one of 75 in the Nicotiana genus and belongs in the section Alatae. N. longiflora has a haploid number n=10 (cf. N. tabacum n=24). Functional characterization of the homolog showed that it produced germacrene A at the expense of 5-epi-aristolochene. Mutagenesis experiments revealed that a single residue position was responsible for the change in product selectivity.
3.3 Results and Discussion

An EAS homolog was cloned from *N. longiflora* tissue using degenerate primers designed against an alignment EAS genes from several *Nicotiana* species. Sequence comparison of this *N. longiflora* synthase (NlgfS) and TEAS revealed that the two genes are highly similar with a nucleotide sequence match of 1554 out of 1647 bases, a 94.4% identity (GenBank Accession No. XXXXXX). The coded proteins have a similarly high identity with only 35 amino acid differences out of a total length of 548 residues (93.6% identity). NlgfS has conserved all currently known functional sites such as the metal-binding motifs DDxxD and NSE/DTE characteristic of TEAS and the sesquiterpene synthase family at large.4

To identify the activity of this putative sesquiterpene synthase, we expressed and purified it from *E. coli* and performed an in vitro enzyme assay with Farnesyl Pyrophosphate (FPP) as the substrate. The products were extracted in hexane and analyzed by Gas Chromatography-Mass Spectrometry (GC-MS). The results revealed that the product profile of NlgfS is markedly different from that of TEAS, making germacrene A as its major product instead of 5-epi-aristolochene (Figure 3.1). Note that during the high temperature conditions of GC-MS analysis germacrene A undergoes Cope rearrangement to turn into β-elemene.18 For NlgfS and TEAS respectively, germacrene A represents 47% and 11% of the products and 5-epi-aristolochene represents 28% and 73% of the products made by these two enzymes.
Figure 3.1 GC-MS product trace overlay of TEAS and NlgfS.

GC-MS product trace overlay of TEAS and NlgfS (reacted with FPP in pH 8.0) shows that NlgfS makes germacrene A as its major product (1) cis-β-elemene (2) β-elemene, the Cope rearrangement of germacrene A (3) selina-4,11-diene (4) 5-epi-aristolochene (5) 4-epi-eremophilene (6) α-selinene (7) unresolved peaks: premnaspirodiene & unknown.
Figure 3.2 Residue identity at 460 directs enzyme function.
Switching in Ile or Thr at 460 is sufficient to interconvert enzyme function. a) Mutating position 460 in NlgfS from Ile to Thr converts the function to match that of TEAS. b) Mutating position 460 in TEAS from Thr to Ile is sufficient to phenocopy NlgfS chemistry.
In order to determine which residues caused the changes in catalytic activity we concentrated on positions located near the active site. Of the 35 residue differences between the two enzymes, most are located in the non-catalytic N-terminal domain. The nearest position is 460, which is isoleucine in TEAS and threonine in NlgfS caused by a single cytosine to thymine base change. We mutated NlgfS and TEAS to interchange the residue identities at position 460. Surprisingly, this single change was sufficient to completely interconvert the catalytic activities of TEAS and NlgfS (Figure 3.2). This result is especially interesting because the distance of position 460 from the active site surface is relatively distal, approximately 9Å (Figure 3.3), and is across the active site from most of the known residues that stabilize carbocation intermediates and activate neutral intermediates that precede germacrene A and 5-epi-aristolochene formation.

Y527 stabilizes the germacrone cation \( ^{10} \) and Y520, D444, D525 protonate germacrene A to form the eudesmyl cation stabilized by W273 \( ^{12} \) and proceeds after a few more intermediates to form 5-epi-aristolochene (Figure 3.4). Additionally, position 460 is also relatively free of interactions with the rest of the protein, with the nearest distance to another side chain (D445) of about 6Å. Hence, the mechanism behind the switch for germacrene A production is not readily apparent.
Figure 3.3 Spatial position 460 relative to the active site.
Position 460 is located relatively far from the active site surface. The following residues colored green have a known function: Y527* stabilizes the germacreryl cation. Y520, D444, D525* protonate germacrene A which forms the eudesmyl cation stabilized by W273. Residue D445 (colored orange) has not been functionally characterized. Atomic coordinates from PDB ID 3M01 (*Residues not shown, located on the labile portion of the J-K loop not modeled)
The first indication of the mechanism, suggested by an observation made while testing different buffer conditions for running the in vitro assays, was that the proportion of products made is dependent on the pH of the reaction solution. Performing NlgfS and TEAS T460I reactions in higher pH environments increased the production of germacrene A and decreased the production of 5-epi-aristolochene (Figure 3.5). The pH of the reaction environment influences the
proportion of end products as a basic environment can prevent germacrene A from picking up a proton and forming the eudesmyl cation necessary for 5-epi-aristolochene formation (Figure 3.4).

It has been suggested that this effect may be the result of the higher nucleophilicity of water at basic pH versus at neutral or acidic pH \(^{19}\). However, we made a second following observation that could suggest an alternative explanation.

\[ \text{Figure 3.5 Effect of pH on product distribution} \]

Germacrene A production increases with pH: NlgfS (shown) and TEAS T460I (not shown) reacted in solutions with higher pH results in a product profile skewed towards germacrene A (2) production at the expense of 5-epi-aristolochene (4) production
We measured thermostability of the enzymes and position 460 mutants using a fluorescent dye-based thermal unfolding assay and observed that melting temperature (T_m) was dependent on the residue identity at 460. TEAS has a T_m of 42°C but TEAS T460I has a 5°C lower T_m. The converse also holds true, NlgfS has a T_m of 35°C but NlgfS I460T has a 5°C higher T_m (Figure 3.6). Therefore, an isoleucine at position 460 in either background appears to correspond to a lower T_m that indicates a higher flexibility or disorder in the enzyme. This loss of stability could stem from the loss of hydrogen bonding between T460 and D445 through a water molecule bridge when the polar Thr 460 is substituted with the non-polar Ile. Notably, D445 is adjacent to a residue, D444 mentioned above to be involved with forming the eudesmyl cation. Wymore et al postulates that the role D444 plays in creating the eudesmyl cation is more structural than due to a transfer of protons because N444 performs comparably, nor is D444 likely to function as a proton acceptor because of its electron distribution. A destabilization of the helix where this residue resides could interfere with its structural function and hence enzyme chemistry. Thus, from the thermostability and the earlier pH observations, we propose that the identity of 460 strongly contributes to the overall conformation and stability of the protein, which influences the activity of the active site to reactivate the germacrene A intermediate into the downstream eudesmyl cation, and hence causes a higher accumulation of the early termination product germacrene A at the expense of the downstream product 5-epi-aristolochene.
Figure 3.6 Melting temperatures of TEAS, NlgfS and their 460 mutants.
The melting temperature of the enzymes is higher with threonine at position 460 and lower when 460 is an isoleucine regardless of enzyme background.

Kinetics measurements were taken with FPP as the substrate to further characterize the enzymes (Figure 3.7). TEAS has a calculated catalytic activity \((k_{\text{cat}}/K_M)\) value of 3.6 \(\text{uM}^{-1}\text{min}^{-1}\), but with the mutation to 460I the value drops to 1.63 \(\text{uM}^{-1}\text{min}^{-1}\), taking a slight hit in catalytic efficiency. NlgfS has a \(k_{\text{cat}}/K_M\) value of 0.51 \(\text{uM}^{-1}\text{min}^{-1}\) and its 460 Thr mutant has a value of 0.37 \(\text{uM}^{-1}\text{min}^{-1}\). From these numbers, it appears that the efficiency of the enzyme is best when the identity of the residue is T460.
The isoleucine at position 460 in NlgsS is a rather unique in the immediate sesquiterpene synthase family of enzymes, but represents a roughly estimated 15-20% of the alleles in *N. longiflora* from combined data from clonal sequencing and targeted high-throughput sequencing.
High-throughput sequencing with 28x coverage revealed that proportion of bases that code for isoleucine versus threonine are respectively C=22(78.5%) and T=6(21.5%) and clonal sequencing resulted in 1 T(Ile) allele and 6 C(Thr) alleles. Contrastingly, in multi-alignments of EAS, the similar vetispiradiene synthase, and even germacrene A synthases sequences, the consensus residue at position 460 is a threonine (Figure 3.8). Additionally, two other EAS homologs (not included in this publication) cloned from N. longiflora retained the Ile 460 residue. Gene duplication and tandem duplication events can act as a safety net for functional divergence, since duplicate copies can undergo mutation and evolution without sacrificing the original activity.21–23. Although I460 is rare, NlgfS compares favorably kinetically ($K_M = 17.4 \mu M$) to other Thr 460 germacrene A synthases, such as one from Cichorium intybus that has a $K_M = 6.6 \mu M$ and one from Solidago canadensis that has a $K_M = 3.4 \mu M$.18,24 Looking at the nucleotide sequence, it is a single thymine to cytosine point mutation (TTT to TTC) that is behind the change from Thr to Ile. NlgfS appears to be a rare variation of the enzyme found so far only in N. longiflora, potentially being currently selected in or out as its hosts plant requires of its chemical arsenal.23,25

Figure 3.8  Residue conservation at position 460. Isoleucine at 460 is unique to the gene cloned from N. longiflora. Threonine has been conserved throughout other EAS genes, related vetispiradiene synthase genes (VS), and other germacrene A synthases (GAS).
In summation, we identified a novel sesquiterpene synthase from N. longiflora, NlgfS, which is highly related to TEAS phylogenetically by nucleotide sequence, yet produces a distinctly altered product profile. By employing structure-based mutagenesis, we identified that position 460 acts as a switch to alter the catalytic characteristics of the enzyme depending on whether the residue is threonine or isoleucine at that site. This switch position is especially interesting because of its distance from the active site and its relatively isolated spatial position. Through a combination of thermostability and pH-dependent activity studies, we hypothesize that the change from the polar threonine to a non-polar isoleucine residue at the position destabilizes the enzyme overall and causes the active site cavity to become more unstructured, unable to facilitate the addition of the proton necessary to reactivate the germacrene A intermediate into the eudesmyl cation, therefore resulting in the accumulation of the early termination product, germacrene A. This study highlights that distal residues can contribute heavily to an enzyme’s chemistry and further suggests considering regions beyond the active site surface when engineering an enzyme.

3.4 Materials and Methods

Cloning EAS homologs. Plants tissue was obtained from plants grown from seeds sourced from USDA Germplasm Resources Information Network (GRIN). N. longiflora TW 81 (accession # PI 555534) was obtained for this study. Degenerate primers 5’-ATGGCCTCAGCAGCAGTWG-3’ and 5’-TCAAATTTYGATGGAGTCCAC -3’ were designed from a consensus alignment of EAS-like sequences from N. benthamiana (SGN-E1189801, SGN-E1189802, CK286192, CK287078, CK288631 & CK287079), N. attenuata (AF484124), N. plumbagnifolia (FM244695), N. tabacum (Q40577) from the Sol Genomics Network and GenBank.
**Mutant Constructs.** Mutant constructs were made using a standard PCR-based site-directed mutagenesis strategy (QuikChange, Stratagene). DNA sequencing for verification was performed by Retrogen.

**Protein Expression and Purification.** Enzyme cDNA were cloned into the Escherichia coli expression vector pHis8-3. Constructs were transformed into *E. coli* BL21(λDE3). Transformed *E. coli* was grown at 37°C in Terrific broth containing 50μg/ml kanamycin until an A600 value of 1.0 was reached. After induction with 0.1mM isopropyl 1-thio-β-galactopyranoside, the cultures were grown for O/N at 18°C. Cells were harvested by centrifugation, resuspended in lysis buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 20mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol, 1% [v/v] Tween 20). After sonication and centrifugation, the supernatant was passed over a Ni²⁺-NTA column and washed with 10 bed volumes of lysis buffer and 10 bed volumes of wash buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 20mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol), and then the His-tagged protein was eluted with elution buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 250mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol). His-tag was cut off in an overnight incubation with thrombin at 4°C. Proteins were buffer exchanged and concentrated with a final buffer (12.5mM Tris-HCl [pH 8.0], 50 mM NaCl, 2 mM DTT) by centrifugation using 50kDa weight cut-off columns (Millipore Amicon Ultrace).

**Enzyme Products Assay.** Assays for recombinant sesquiterpene synthase activity was performed in 500ul reaction volumes in 1x Assay buffer (10mM Bis-Tris Propane [pH 7.0 unless otherwise indicated], 5mM MgCl₂) with 1.0μM enzyme and 25μM FPP (Isoprenoid.com). Reactions were overlaid with 500ul hexane spiked with p-chlorotoluene as a sample injection standard in 2-ml screw-top glass vials and were allowed to react overnight at room temperature. The reaction products were extracted into the hexane layer by vigorous vortexing for 3 minutes.
followed by centrifugation at 3200xg for 30 min. The organic layer was isolated and concentrated to a volume of 50-80 μl under N₂ gas. Products were analyzed by using a Hewlett-Packard 6890 gas chromatograph (GC) coupled to a 5973 mass selective detector (MSD) outfitted with a 7683 series injector and autosampler and equipped with an HP-5MS capillary column (0.25 mm i.d. × 30 m with 0.25 μm film) (Agilent Technologies). The GC was operated at a He flow rate of 2 ml/min, and the MSD was operated at 70 eV. Splitless injections (3 μl) were made with an injector temperature of 250 °C. The GC was programmed with an initial oven temperature of 50 °C (5-min hold), which was then increased 5 °C min⁻¹ up to 250 °C (5-min hold). A solvent delay of 5.5 min was allowed prior to the acquisition of MS data. Product peaks were quantified by integration of peak areas using Enhanced ChemStation (version E.02.02, Agilent Technologies) and expressed as percentage of the total products. Of note, germacrene A was detected as the thermally rearranged product β-elemene. Product identification was performed using MassFinder 4 software (D. H. Hochmuth, D. Joulain, W. A. König, www.massfinder.com) and unpublished data from the Noel lab (Koo, H.J.).

**Enzyme Kinetics Assay.** Assays were performed at room temperature (21.5 °C) with 0.002-0.100 μM final concentration of enzyme (depending on the enzyme) reacted in 1x Assay buffer with FPP at 0.78125, 1.5625, 3.125, 6.25, 12.5, 25, and 50μM. 4 time points were taken for each concentration. Products were quantified on GC-MS by peak area. Kinetic rates were calculated based on Michaelis-Menten kinetics.

**Thermal Unfolding Assay.** Experiment was performed with 2μM of enzyme in a buffer containing 100mM MOPS [pH 7.0], 13% PEG 8000, 200 mM Magnesium Acetate. Enzymes were incubated in buffer for 15 minutes at room temperature before adding the fluorophor SYPRO Orange (Invitrogen, Carlsbad, CA) at 10x concentration. Reactions were performed in 10ul volumes in 384-well microplates. Reactions were heated from 20 to 80°C at a rate of
0.06°C/s in a *thermal cycler with fluorescent* detection (Roche LightCycler 480), and the fluorescence intensity was recorded at 0.1°C intervals. Fluorescence intensities were plotted as a function of temperature, and the midpoint of the unfolding transition was taken as the melting temperature.
3.5 Acknowledgments

Chapter 3, in full, is currently being prepared for submission for publication of the material as it may appear in Phytochemistry. Sun, Helena C; Koo, Hyun Jo; Noel, Joseph, P.. Permission was obtained from the coauthors. The dissertation author was the primary experimentalist and author of this material. All experiments were performed under the supervision of Joseph P. Noel.

The work in this study was supported in part by The Rose Hills Foundation, the Cell and Molecular Genetics Training Program, and the Aileen S. Andrew Foundation.
3.6 References


CHAPTER 4

Identification of a product specificity switch residue in the TEAS family of enzymes

4.1 Abstract

Most terpene synthases, like Nicotiana tabacum 5-epi-aristolochene synthase (TEAS) for one, contain plasticity residues that engender diverse end products and allow for functional evolution is one such synthase. While mining for TEAS-like enzymes, we discovered a TEAS-like enzyme that produces a host of alternative compounds at the expense of the TEAS major product, 5-epi-aristolochene. Upon examination by mutation, a single residue position was discovered to act as a product specificity switch for TEAS depending on the residue identity, tyrosine versus histidine. From pH experiments, it is surmised that histidine functions as a base to remove protons and produce terminal products at a key cationic intermediate.

4.2 Introduction

Terpenoids constitute the largest and most diverse class of specialized metabolites with over 60,000 compounds known to date identified in plants, fungi, and bacteria. Synthesized from simple prenyl diphosphate precursors, geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), the tremendous chemodiversity of the terpenoid family is facilitated by terpene synthases that catalyze diverse and complex reactions such as regio- and stereospecific ring formations, proton eliminations to form double bonds, and stereospecific hydride, proton, methyl, and methylene migrations. Indeed, many terpene
synthases have the ability to synthesize multiple products from a single prenyl diphosphate substrate, with some producing over 50 different products\textsuperscript{5,6}. This catalytic plasticity has long been thought to underlie natural enzyme evolution\textsuperscript{7–10} and additionally can be a useful tool to study enzyme catalysis by providing insight on molecular design and enzyme engineering.

5-\textit{epi}-aristolochene synthase from \textit{Nicotiana tobacco} (TEAS) is a sesquiterpene synthase that cyclizes the substrate farnesyl pyrophosphate (FPP) into 5-\textit{epi}-aristolochene, which is then hydroxylated downstream to produce the phytoalexin capsidiol. TEAS is a relatively well-characterized enzyme with extensive structural and functional studies\textsuperscript{11–14}. Residues important for catalysis in TEAS and other sesquiterpene synthases are highlighted from highly conserved sites in terpene synthase alignments, such as the divalent metal (usually Mg\textsuperscript{2+}) binding motifs DDXXD/E and DTE/NSE, which help perform the initial ionization the substrate to form the farnesyl cation. From the farnesyl cation, TEAS has three subsequent ionic intermediates, the germacranyl, eudesmyl, and eremophilanyl cations that serve as a crossroads for branching pathways leading to various end products.

While TEAS has a core set of functional residues, it has also been demonstrated to harbor several plasticity residues. In one series of studies, a mutant library was created by interconverting residues differing between TEAS and a premnaspirodiene (vetispriadiene) synthase from \textit{Hyoscamus muticus} (HPS)\textsuperscript{12,13}. This library not only showed that catalytic activity could be completely swapped, altering substrate flux through the eudesmyl cation by switching a minimal set of nine residues, it also demonstrated that mutating subsets of the residues could produce product profiles that could potentially represent a continuous spectrum in terms of product ratios. While TEAS and the henbane synthase come from the same family of plants, comparing more recently diverged TEAS-like synthases from more closely related species within a single genus could reveal other more nuanced plasticity residues that control substrate flux through the different cationic crossroads.
In this study, while searching for novel natural variants of EAS from the *Nicotiana* genus, we cloned and characterized a TEAS-like gene that makes a product profile distinctly different from that of TEAS. This TEAS-like synthase (TIS) originally cloned from *N. tomentosa* is much more plastic in terms of its product output, producing an increased variety of products as well as an increased proportion of select minor products at the expense of 5-*epi*-aristolochene production.

### 4.3 Results and Discussion

Degenerate primers, designed from an alignment of EAS from several species of *Nicotiana*, were used to clone a TEAS-like synthase (TIS) from *N. tomentosa*. The TIS has a 94.9% identity with TEAS, differing by 28 amino acids residues out of a total protein length of 548. (TIS sequence deposited as GenBank Accession No. XXXXXX). TIS was expressed in *E. coli*, purified and then assayed for enzymatic activity *in vitro* with the substrate FPP by quantitative GC-MS (Gas Chromatography Mass-Spectrometry) to determine the products formed. This analysis revealed that TIS produces a markedly different product profile from that of TEAS, with an increased proportion of such products like α- and β- selinene and α-humulene while also abrogating 5-*epi*-aristolochene (5-EA) production; and it also produces some compounds not seen from TEAS (Figure 4.1). From these product profiles, TIS appears to have lost one or more residues responsible for 5-EA product specificity.
Figure 4.1 GC-MS product trace overlay of TEAS and TIS

GC-MS product trace overlay of TEAS and TIS (reacted with FPP in pH 8.0) shows that TIS loses ability to produce 5-epi-aristolochene in favor of alternate products like α- & β- selinenes. 

(1) β-elemene, the Cope rearrangement of germacrene A (2) (E)-caryophyllene (3) selina-5,11-diene (4) spirolepechinene (5) α-humulene (6) 4,5-di-epi-aristolochene (7) selina-4,11-diene (8) 5-epi-aristolochene (9) β-selinene (10) 4-epi-eremophilene (11) α-selinene (12) premnaspirodiene.

In order to understand the structural components behind the changes in the catalytic mechanisms of the enzyme, we focused on positions with residue identity differences closest to the active site cavity. Using this proximity criteria, residue positions 289, 404 and 508 were singled out as candidates to investigate. Position 404 is on the active site surface, while 289 and 508 are further than 2 Van der Waals radii from the active site. After further sequencing analysis, while the residue identities at 289 and 508 are found in nature, it appears as if His 404 in TIS is
either a rare SNP-induced mutation, since no other such sequence was found after 20x sequencing coverage. Another alternative is that it could be a mutation introduced from the cloning process since oxidized, deaminated cytosines are a source of cytosine to thymine transitions in reverse transcription \(^{15}\), which was used to clone the gene [TAT(Tyr) in TEAS to CAT(His) in TlS]. Thus H404 will be conservatively considered as an artificial mutation for any further discussion purposes.

An eight member (2\(^3\)) mutant library was constructed to pinpoint the functional residues by swapping residue identities between those of TEAS and TlS: Met/Ile 289, Tyr/His 404, and Pro/Arg 508 in each enzyme background and in every combination. The resulting mutants were examined for residue-dependent catalytic effects using the aforementioned \textit{in vitro} product assay. Mutating positions 289 and 508 from the TEAS to TlS residues had no effects on product profiles in either background (Figure 4.2a & b). However, switching the tyrosine and histidine residues at position 404 demonstrated pronounced product profile changes; histidine at 404 greatly diminished the enzyme’s ability to produce 5-EA in favor of the alternate products, but a tyrosine in the position completely restored 5-EA product specificity to TlS (Figure 4.2c). Hence the residue identity at position 404 is largely responsible for the differences in enzymatic activity between TlS and TEAS.
Figure 4.2 Single residue mutations of TEAS and TlS

Single residue mutations demonstrate that the residue at position 404 controls catalytic function. a) Switching Met/Ile at position 289 in either TEAS or TlS backgrounds has no noticeable effect on product profile. b) Switching Pro/Arg at position 508 in either background also has negligible effects on product profile. c) Swapping Tyr/His at position 404 in both backgrounds shows that the residue identity controls the catalytic function of the enzyme. (1) β-elemene, the Cope rearrangement of germacrene A (2) (E)-caryophyllene (3) selina-5,11-diene (4) spirolepechinene (5) α-humulene (6) 4,5-di-epi-aristolochene (7) selina-4,11-diene (8) 5-epi-aristolochene (9) β-selinene (10) 4-epi-eremophilene (11) α-selinene (12) prennaspirodiene
Although position 404 is largely responsible for determining the product profile, the double mutants revealed that TIS residue Arg508 in the presence of His404 displays apparent non-additivity by boosting specifically α- and β- selinene product formation (Figure 4.3). Also, Ile289 also acts in the presence of His404 to give a small increase to the 4-epi-eremophilene product. Lastly, the triple mutants show that swapping all three residue positions in TEAS to that of TIS does not fully replicate the TIS product profile, indicating that there are additional positions that contribute to the product distribution of TIS (Figure 4.4). Hence, we tested five more positions proximal to the active site for product profile effects by switching residue identities at T388M, F505I, T507F, I539N, & N540A (Figure 4.5). While these mutations all had subtle effects on the product profile produced, they did not reveal another residue position that has a clear contribution to determining product specificity.
Figure 4.3 Double mutants: non-additive effects of residues 289 and 508 on TEAS+Y404H
(1) β-elemene, the Cope rearrangement of germacrene A (2) (E)-caryophyllene (3) selina-5,11-
diene (4) spirolepechinene (5) α-humulene (6) 4,5-di-epi-aristolochene (7) selina-4,11-diene
(8) 5-epi-aristolochene (9) β-selinene (10) 4-epi-ermophilene (11) α-selinene (12)
premnaspirodiene
Figure 4.4 Triple mutant of TEAS does not recapitulate catalytic activity of TIS.
(1) β-elemene, the Cope rearrangement of germacrene A (2) (E)-caryophyllene (3) selina-5,11-diene (4) spirolepechinene (5) α-humulene (6) 4,5-di-epi-aristolochene (7) selina-4,11-diene (8) 5-epi-aristolochene (9) β-selinene (10) 4-epi-eremophilene (11) α-selinene (12) premnaspirodiene
Figure 4.5 Mutant residues farther out from the active site
(1) β-elemene, the Cope rearrangement of germacrene A (2) (E)-caryophyllene (3) selina-5,11-diene (4) spirolepechinene (5) α-humulene (6) 4,5-di-epi-aristolochene (7) selina-4,11-diene (8) 5-epi-aristolochene (9) β-selinene (10) 4-epi-eremophilene (11) α-selinene (12) premnaspirodiene

The kinetic properties of TlS and mutants were measured. TlS has a \( V_{\text{max}} \) of 19.64 nM/min, a \( K_M \) of 5.386 μM and an apparent \( k_{\text{cat}} \) of 0.655/min (Figure 4.6). TEAS has a best fit \( V_{\text{max}} \) of 24 nM/min, a \( K_M \) of 3.5 μM, and an apparent \( k_{\text{cat}} \) of 12.45/min. So while TlS is comparable to TEAS in terms of maximal velocity and substrate affinity, it has a significantly lower catalytic efficiency. Mutant enzymes with His404 showed similar kinetics numbers to TlS and mutants with Tyr404 had similar kinetics to that of TEAS. Hence, the residue identity at 404 also largely determines enzyme’s kinetic properties.
Figure 4.6 Kinetics of TEAS, TIS and mutants.
Enzyme backbones and residues are color coded, TEAS=blue, TIS=orange. Enzymes with His404 are less catalytically efficient than those with Y404.
Measuring shifts in melting temperature provides an indication of the effects of differing residues on the thermostability of the enzymes. The melting temperatures of the enzymes were measured using a dye that fluoresces when bound to the exposed hydrophobic portions of denatured proteins. The results showed that single residue changes at positions 289 and 508 did not seem to affect enzyme thermostability significantly. However, His404 causes a destabilization in either enzyme background with a drop in melting temperature of 4-6°C (Figure 4.7). Interestingly, the Arg508 mutation on its own is neutral or even slightly stabilizing, however when adding that mutation on top of His404 in the TEAS enzyme background, the enzyme becomes disordered (Figure 4.8).
Figure 4.7 Melting temperature effects of position 404 residues in TEAS and TIS

The melting temperature (minima) of the enzyme is lowered when position 404 is a histidine in both TIS and TEAS enzyme backgrounds.
Figure 4.8 Melting temperature effects of residues at position 508
Arg508 on its own has little or no effect on the melting temperature, but when present with
His404, it greatly disorders the enzyme. a) Raw data shows that TEAS+H404+R508 is highly
disordered. b) Derivative data shows melting temperature (minima) of the other three enzymes,
but TEAS+H404+R508 is not shown since the derivative plot would be misleading judging from
the raw data above.
The mutant enzymes also demonstrate a pH-dependent effect on products produced, which was first observed while testing different buffer conditions for the *in vitro* assays. TIS and His404 mutant enzymes assayed at a higher pH resulted in a greater proportion of selinene and germacrene A products and a decrease in selina-5,11-diene, 5-*epi*-aristolochene, and 4-*epi*-eremophilene (Figure 4.9). The observation of increased selinene production in the presence of higher pH together with the fact that histidine can act as a general base (pKa 6.0), suggests a mechanism where His404, especially under basic conditions, removes a proton from the acidic and reactive eudesmyl carbocation intermediate at carbons 2, 4, or 15 to form the products selina-4,11-diene, *α*-selinene, and *β*-selinene (Figure 4.10). The proton elimination thus truncates the reaction pathway that would have lead to the production of downstream products such as 5-*epi*-aristolochene (Figure 4.11). In contrast, tyrosine 404 appears to favor the methyl shift off of the eudesmyl cation to form the eremophilanyl cation and its respective end products.
Figure 4.9 Effects of pH on product distribution in TIS
TIS GC-MS products profile from in vitro product assay run in pH 7.0 versus pH 8.0. Enzyme run in higher pH results in an increase in germacrene A, (11) α-selinene, (9) β-selinene, and (12) premnaspirodiene peaks, as well as a decrease in (3) selina-5,11-diene, (8) 5-epi-aristolochene, and (10) 4-epi-eremophilene.
Figure 4.10 Model of His404 acting as a general base
A model of α-selinene in active site of the TIS crystal structure depicting the potential of His404 to act as a general base to abstract a proton off of the eudesmyl cation.
Figure 4.11 Reaction pathways map of EAS-like enzymes. 
For the eudesmyl cation, the blue arrow shows the methyl shift to form the eremophilenyl cation, the red arrow depicts an alkyl migration to change the ring structure.
The crystal structure obtained of TIS (2.6Å resolution) when compared with TEAS (PDB ID 3M01) shows that the overall enzyme structure is preserved, RMSD = 0.590Å. Looking closer at the active site, residue 404 is closest in space to residues Trp273, Cys440, and Leu512 (Figure 4.12). The aromatic quadrupole of Trp273 has been shown to stabilize the electrostatic center of the eudesmyl carbocation at C7 and also accepts a proton from C8 to form 5-epi-aristolochene 11. Due to the proximity, Y404H could potentially interfere with the function of Trp273. Residue 404 is also within Van der Waals contact of residues Val516 and Ser436. Val516 along with Y402, directs stereochemistry of the proton elimination off of the eremophilanyl cation 12. This could explain the increase of 4-epi-eremophilene seen with His404. It has been observed from crystal structures of TEAS with the FPP substrate analog, farnesyl hydroxyphosphponate, that residue 404 also has Van der Waals contact with methyl C14, which is also stabilized by residues Cys440 and Thr403 and that Cys 440 interacts with Thr402 and V516 16. Replacing tyrosine with histidine at position 404 could disrupt the contact network of C14 and potentially interfere with its migration from C7 to C3 on the eudesmyl cation, reducing flow towards the 5-EA product. Although subsequent crystal structures have shown the substrate analog in different orientations with respect to the active site, the orientation could be dynamic. Additionally, as described above, the interference with Trp273 could further reduce 5-EA products, which is in line with the product profiles seen from His404 enzymes. Further mutation experiments or crystallization of TIS with substrate or intermediate analogs could give further insight into the exact mechanism.
TIS and TEAS were compared for conservation against known selinene and α-humulene synthases (Figure 4.13). Position 289 is not heavily conserved, existing as the nonpolar methionine and isoleucine and also polar tyrosine. Position 404 is highly conserved as tyrosine. Position 508 is also not heavily conserved, existing as nonpolar proline and cysteine and also positively charged arginine. Incidentally, even amongst the species in *Nicotiana* we sequenced, Arg508 appears much more commonly than Pro508 found in TEAS, derived from *N. tabacum*.
Out of 16 alleles sequenced from 9 species of *Nicotiana*, 13 were Arg508. In considering other similar sesquiterpene synthases, it would be interesting to mutate cognate positions of 404 to see if it has a similar effect on product specificity as in TEAS and TIS.

**Figure 4.13 Sequence conservation: alignment of TIS, TEAS and other sesquiterpene synthases.** Position 404 is mostly conserved as tyrosine, while position 289 is a mixture of mostly isoleucine and methionine. Position 508 is mostly arginine.

In this study we have identified a residue position, position 404, lying on the active site surface of EAS-like synthases that appears to have significant control over $\text{H}^+$ elimination and
methyl migration on the eudesmyl cation depending on the identity of the residue at the position. A histidine at the position is thought to act as a general base and terminate products off of the eudesmyl cation by abstracting off protons. Residue position 508 appears to be in the same reaction control network as 404 as it acts in concert with position histidine 404 to increase $\alpha$-selinene and $\beta$-selinene off of the eudesmyl cation. A tyrosine at position 404 could potentially be involved in facilitating the methyl shift step necessary to advance to the eremophilenyl cation that gives rise to the 5-epi-aristolochene major product seen in TEAS. Understanding how single residues can contribute so drastically as to act as an on/off switch for product specificity is especially interesting, seeing as any insight into the structure and mechanisms behind the chemistries of specialized metabolism may lead to more precise enzyme engineering that would increase yields of select desirable metabolic products.

4.4 Materials and Methods

*Cloning EAS homologs.* Plants tissue was obtained from plants grown from seeds sourced from USDA Germplasm Resources Information Network (GRIN). *N. tomentosa* 58-G (accession # PI 574525) was obtained for this study. Degenerate primers 5’-ATGGCCTCAGCAGCAGTWG-3’ and 5’-TCAAATTTYGATGGAGTCCAC -3’ were designed from a consensus alignment of EAS-like sequences from *N. benthamiana* (SGN-E1189801, SGN-E1189802, CK286192, CK287078, CK288631 & CK287079), *N. attenuata* (AF484124), *N. plumbagnifolia* (FM244695), *N. tabacum* (Q40577) from the Sol Genomics Network and GenBank. Sequencing of EAS homologs was done by in part by conventional sanger sequencing and also by ion semiconductor sequencing (Ion Torrent).
**Mutant Constructs.** Mutant constructs were made using a standard PCR-based site-directed mutagenesis strategy (QuikChange, Stratagene). DNA sequencing for verification was performed by Retrogen.

**Protein Expression and Purification.** Enzyme cDNA were cloned into the Escherichia coli expression vector pHis8-3. Constructs were transformed into *E. coli* BL21(λDE3). Transformed *E. coli* was grown at 37°C in Terrific broth containing 50μg/ml kanamycin until an A600 value of 1.0 was reached. After induction with 0.1mM isopropyl 1-thio-β-galactopyranoside, the cultures were grown for O/N at 18°C. Cells were harvested by centrifugation, resuspended in lysis buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 20mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol, 1% [v/v] Tween 20). After sonication and centrifugation, the supernatant was passed over a Ni²⁺-NTA column and washed with 10 bed volumes of lysis buffer and 10 bed volumes of wash buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 20mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol), and then the His- tagged protein was eluted with elution buffer (50mM Tris-HCl [pH 8.0], 500mM NaCl, 250mM imidazole [pH 8.0], 20mM β-mercaptoethanol, 10% [v/v] glycerol). Histag was cut off in an overnight incubation with thrombin at 4°C. Proteins were buffer exchanged and concentrated with a final buffer (12.5mM Tris-HCl [pH 8.0], 50 mM NaCl, 2 mM DTT) by centrifugation using 50kDa weight cut-off columns (Millipore Amicon Ultrapure).

**Enzyme Products Assay.** Assays for recombinant sesquiterpene synthase activity was performed in 500μl reaction volumes in 1x Assay buffer (10mM Bis-Tris Propane [pH7.0 unless otherwise indicated], 5mM MgCl₂) with 1.0μM enzyme and 25μM FPP (Isoprenoid.com). Reactions were overlaid with 500μl hexane with p-chlorotoluene as a loading standard in 2-ml screw-top glass vials and were allowed to react overnight at room temperature. The reaction products were extracted into the hexane layer by vigorous vortexing for 3 minutes followed by
centrifugation at 3200xg for 30 min. The organic layer was isolated and concentrated to a volume of 50-80 μl under N₂ gas. Products were analyzed by using a Hewlett-Packard 6890 gas chromatograph (GC) coupled to a 5973 mass selective detector (MSD) outfitted with a 7683 series injector and autosampler and equipped with an HP-5MS capillary column (0.25 mm i.d. × 30 m with 0.25 μm film) (Agilent Technologies). The GC was operated at a He flow rate of 2 ml/min, and the MSD was operated at 70 eV. Splitless injections (3 μl) were made with an injector temperature of 250 °C. The GC was programmed with an initial oven temperature of 50 °C (5-min hold), which was then increased 5 °C min⁻¹ up to 250 °C (5-min hold). A solvent delay of 5.5 min was allowed prior to the acquisition of MS data. Product peaks were quantified by integration of peak areas using Enhanced ChemStation (version E.02.02, Agilent Technologies) and expressed as percentage of the total products. Of note, germacrene A was detected as the thermally rearranged product β-elemene. Product identification was performed using MassFinder 4 software (D. H. Hochmuth, D. Joulain, W. A. König, www.massfinder.com) and unpublished data from this lab (Koo, H.J.).

Enzyme Kinetics Assay. 0.002-0.100 μM final concentration of enzyme (depending on the enzyme) reacted in 1x Assay buffer with FPP at 0.78125, 1.5625, 3.125, 6.25, 12.5, 25, and 50 μM. 4 time points were taken for each concentration. Products were quantified on GCMS by peak area. Kinetic rates were calculated based on Michaelis-Menten kinetics.

Thermal Unfolding Assay. Experiment was performed with 2μM of enzyme in a buffer containing 100mM MOPS [pH 7.0], 13% PEG 8000, 200 mM Magnesium Acetate. Enzymes were incubated in buffer for 15 minutes at room temperature before adding the fluorophor SYPRO Orange (Invitrogen, Carlsbad, CA) at 10x concentration. Reactions were performed in 10ul volumes in 384-well microplates. Reactions were heated from 20 to 80°C at a rate of 0.06°C/s in a thermal cycler with fluorescent detection (Roche LightCycler 480), and the
fluorescence intensity was recorded at 0.1°C intervals. Fluorescence intensities were plotted as a function of temperature, and the midpoint of the unfolding transition was taken as the melting temperature.

**Protein Crystallization and Structure Solving.** The TIS was crystallized by the hanging drop vapor diffusion method over a 0.5-mL reservoir. The crystallization solution condition is 100mM MOPSO, pH7, 13-15% PEG 8000, 200mM Mg(OAc)₂. Initially, small, irregular crystals were grown from drops containing protein at 5 mg/mL. Larger crystals were then grown by microseeding. Crystals were frozen in soak solution also containing 20% v/v ethylene glycol as a cryoprotectant. Data collection was performed at Advanced Light Source (ALS), beamline 8.2.1. The data were processed using the Collaborative Computational Project No. 4 (CCP4) software package. The initial crystallographic structure solutions were obtained through molecular replacement analyses using the TEAS-E,E-FPP complex (PDB ID 3M01) as the search model with AutoMR in Phenix. Model adjustments were performed using COOT and rounds of refinement were conducted using Phenix. The TIS crystals are of the orthorhombic space group P2₁2₁2₁, with unit cell dimensions of a = 102.95, b = 105.12, c = 121.36 Å, α = β = γ = 90°. Resolution is 2.6 Å. Structure deposited in Protein Data Bank (PDB ID XXXX).
4.5 Acknowledgments

The text of Chapter 4, in full, is currently being prepared for submission for publication
of the material. Sun, Helena; Koo, Hyun Jo; Noel, Joseph P.. “Identification of a product
specificity switch residue in the TEAS family of enzymes”. Permission was obtained from the
coauthors. The dissertation author is the first author of this material. All experiments were
performed under the supervision of Joseph P. Noel.

The work in this study was supported in part by The Rose Hills Foundation, the Cell and
Molecular Genetics Training Program, and the Aileen S. Andrew Foundation.
4.6 References


2. Dictionary of Natural Products. at <http://dnp.chemnetbase.com>


CHAPTER 5

Conclusions and Perspectives

5.1 Overview

Part of the beauty and also a complication in understanding terpene synthase function is its great plasticity. On one hand, highly divergent terpene synthases at the sequence level can make the same products and similar product profiles; on the other hand with just single amino acid changes, the same terpene synthase can also make very different products with different specificity. Clearly some residue positions must contribute more to product specificity or identity while others perform other functions such as conferring stability. In Chapter 2, the survey of extant EAS sequences in *Nicotiana* revealed a wealth of variation that occurs within a single species of plant, let alone genus or even family of plants. Several of these EAS variants produced unique product profiles. Upon further examination of two of these EAS variants (covered in Chapters 3 and 4), it was discovered that the residue changes that contributed to these changes were in unexpected positions within the enzyme structure.

In the EAS variant from *N. longiflora* (NlgfS), residue position 460 was responsible for what appears to be an early termination step in the EAS pathway scheme to produce germacrene A instead of other downstream end products. The most puzzling fact was that position 460 is located away from the active site and does not appear to have any direct method to affect the active site cavity when examined through structural modeling. Ultimately, the mechanism behind this behavior was attributed to a change in the structural stability of the enzyme that results in an
early termination product. A different, but ultimately similar situation was seen with TIS. Although the functional change in TIS was mainly attributed to a disruption in residue position 404 on the active site surface, the mechanism behind the change in product specificity was similarly attributed to an early termination event in the EAS pathway scheme. These conclusions were both inferred from the product distributions in the context of the overall pathway progression as well as thermostability and pH experiments. Contrastingly, in a recent study by Lauchli et al.\(^1\), improved selectivity of multi-product terpene synthase was achieved through directed evolution and the mechanism was hypothesized to be a disfavoring of an additional cyclization before quenching by the mutant residues. In all cases, residue changes that affect cyclization pathways, foldamer conformations, or quenching of the intermediates play an important role in directing product specificity and identity\(^2,3\).

While the mechanisms behind the product specificity and identity are being steadily unraveled, predicting terpene synthase function from sequence alone still remains elusive; and attempts to computationally model and hypothesize enzyme function still faces many challenges\(^4\). Part of the difficulty may lie with the polyphyletic origin of terpene synthases which may obfuscate relationships between structure and function\(^5\). The partitioning of enzyme residue networks into functional groups has been an emerging, sometimes termed sectors, where cohorts of amino acids are grouped by function and co-evolve\(^6-8\). This method of evaluating groupings of residue positions by function and co-evolution may be useful in unraveling at least some part of the mystery of enzyme structure-function. It is likely, if these theories hold true, that position 460 in NlgfS and position 508 in TIS, both active-site distal residues that affect enzyme stability (and in the case of residue 460 function as well) may be part of the stability network of residues in EAS and potentially the broader sesquiterpene family of enzymes. This stability network could potentially be thought of as a specificity network, since enzyme stability affects the accessibility of reactive intermediates to early termination/quenching reactions and may affect
the support of specific foldamer conformations of the intermediates resulting in multiple products from a single enzyme.

5.2 Implications and Considerations

New function can arise in terpene synthases through multiple levels of change, from functional domains being added or lost or from multiple or even single residue changes. In this body of work, the focus was terpene product specificity and the ability to neofunctionalize. However, terpene synthases can be specific and non-specific, not only for products but also for substrates\textsuperscript{9–11}; for example, (E)-β-farnesene synthase from peppermint can turn FPP into (E)- and (Z)-β-farnesene and δ-cadinene from, but also can also use GPP to make limonene, terpinolene and myrcene\textsuperscript{12}. Substrate promiscuity contributes to the diversity of the TPS family since metabolism is built around connected pathways; so if multiple substrates are accepted by a non-specific enzyme resulting even more variety of products, then these products can then become substrates of further downstream non-specific enzymes, thus propagating product diversity onward ad infinitum (Figure 5.1).
Strangely enough, the TPS family can produce members efficient and specific enough to power primary metabolism and yet is versatile enough to accommodate for specialized biochemical needs through secondary metabolism. But, how much do shifts in product spectrum really contribute to enzyme evolution when the principle behind evolution is selection on the organismal level? It has been described that changes in morphology are due mostly to changes in gene expression/regulation, while changes in physiology are mainly due to gene mutations/splicing\textsuperscript{13,14}. If you consider enzyme function as physiology and the regulation and expression of these enzymes (along with the any inherent toxicity and metabolic burden) as morphology, then the story of terpene synthase evolution only gets more complex. While there is ample support for the positive selection of essential metabolism, specialized metabolism seems to merely avoid negative selection since promiscuous enzymes do not contribute enough throughput to be selected against\textsuperscript{15}. How then, do the fledgling enzymes of specialized metabolism make the leap towards positive selection; perhaps just simple accumulation? In any case, \textit{in vivo} studies

**Figure 5.1 Cascading effect of product and substrate non-selectivity**

If enzymes are non-selective for both products and substrates, the effect on product diversity can be multiplied throughout the length of a biosynthetic pathway. Circular and ovoid molecules are of the same start from a common precursor; square and diamond molecules share a common precursor.
will be necessary to prove the connection between enzyme promiscuity and any eventual enzyme
divergence as a result of selection.

Beyond simple scientific curiosity, one goal driving the understanding of enzyme
structure and function is the wish to harness enzyme function for human purposes from
generating pharmaceutical drugs to enhanced agricultural crops, from flavor and scent
compounds to biofuels and other renewable chemicals of interest. The outcome of metabolic
engineering has not always been predictable because product ratios vary depending on enzyme
specificity, catalytic efficiency, and the availability of appropriate precursors from upstream
steps. Hence, a more thorough understanding of what factors contribute to enzyme function,
especially in terms of residue identity and position within the enzyme structure, is critical work
that must continue to be explored.
5.3 References


