UNIVERSITY OF CALIFORNIA, SAN DIEGO

Exploring Structural and Functional Features of Enzymes Across Isoprenoid Biosynthesis: From Archaeal Isopentenyl Phosphate Kinase of Primary Metabolism to Plant Terpene Cyclases of Specialized Metabolism

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

in

Chemistry

by

Nikki Dellas

Committee in Charge:

Professor Joseph P. Noel, chair
Professor Elizabeth Komives, co-chair
Professor Michael Burkart
Professor Ronald Burton
Professor Gourisankar Ghosh

2010

The Dissertation of Nikki Dellas is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2010

DEDICATION

for Anderson

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

2F. 2-fluoro

4EE. 4-epi eremophilene

5-EA. 5-epi aristolochene

Å. Angstroms

AAK. Amino acid kinase

ADP. Adenosine diphosphate

ADS. Abietadiene synthase

AID. Average interneighbor distance

AMPPNP. Adenylyl imidodiphosphate

ATP. Adenosine triphosphate

ATPγS. Adenosine 5'-(gamma-thiotriphosphate)

C2F. Cis-2-fluoro

CCW. Counterclockwise

CDP. Copalyl diphosphate

CDS. Copalyl diphosphate synthase

CW. Clockwise

DMAPP. Dimethylallyl diphosphate

DNA. Deoxyribonucleic acid

DTT. Dithiothrietol

DXP. 1-deoxy-D-xylulose 5-phosphate

EES. Epi-eremophilene synthase

EST. Expressed sequence tag

FARM. First aspartate-rich motif

FHP. Farnesyl hydroxyphosphonate

FomA. Fosfomycin resistance A

FP. Farnesyl phosphate

FPP. Farnesyl diphosphate

FPPS. Farnesyl diphosphate synthase

GCMS. Gas chromatography mass spectrometry

GGPP. Geranylgeranyl diphospahte

GGPPS. Geranylgeranyl diphosphate synthase

GP. Geranyl phosphate

GPP. Geranyl diphosphate

GPPS. Geranyl disphosphate synthase

HPS. Hyoscyamus premnaspirodiene synthase

IP. Isopentenyl phosphate

IPK. Isopentenyl phosphate kinase

IPP. Isopentenyl diphosphate

IPPβS. Isopentenyl β-thiodiphosphate

KS. *Ent*-kaurene synthase

mg. Milligrams

Mg. Magnesium

$MgCl_2$. Magnesium chloride

ml. Milliliters

Mn. Manganese

MVA. Mevalonate

NaCl. Sodium Chloride

NADH. Nicotinamide adenine dinucleotide, reduced

NAGK. N-acetylglutamate kinase

NMR. Nuclear magnetic resonance

NPP. Nerolidyl diphosphate

PAS. Patchouli alcohol synthase

PCR. Polymerase chain reaction

PDB. Protein data bank

PEG. Polyethylene glycol

PSD. Premnaspirodiene

SARM. Second aspartate-rich motif

SCOPE. Structure-based combinatorial protein engineering

SDS-PAGE. Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SIM. Single ion mode

TEAS. Tobacco 5-epi aristolochene synthase

TIC. Total ion count

UDP. Uridine diphosphate

µl. Microliters

UMPK. Uridine monophosphate kinase

WT. Wild type

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Joseph P. Noel, for all of his guidance, support, and generosity both inside and outside of the lab throughout the past five years. His good nature, positive outlook, and enthusiasm will continue to inspire me throughout life.

I would also like to thank all of my labmates including those in both the Noel lab and the Wang lab, especially Paul O'maille, Marianne Bowman, Gordon Louie, and Jeffrey Takimoto. Paul O'maille was a postdoc in the Noel lab whom I worked closely with when I first joined the lab and continued to work with for several years thereafter. He was a great friend with a very unique and creative sense of humor. Marianne Bowman is the lab manager in the Noel lab; she was instrumental in maintaining lab functionality on a variety of levels, but was also a wonderful person to talk to for support and advice. Gordon Louie is a staff scientist in the lab who helped me with all aspects of crystallography: from data collection to crystallographic refinement. Jeffrey Takimoto is a graduate student in the Wang lab who has been an incredibly supportive and loyal friend, and was always there for me, especially during more recent struggles.

I would like to thank my friends and family for their love and support. Specifically, I would like to thank my mom for teaching me persistence and encouraging me to indulge not only my scientific side but also my creative site throughout most of my life. I would like to thank my dad for teaching me that life is not only about hard work, but also about generosity, selflessness, and having fun with the people that are close to you. I would like to thank my brother Tim for all of the hilarious memories we have together, doing the random things that we end up doing, and for being someone I could always laugh with, share music with, and be myself around. I would like to thank my sister Meg for her positive attitude, compassion for others, and love for life. She has an amazing ability to inspire happiness in those around her

and, as we have grown up, I no longer think of her as my younger sister, but instead as one of my peers. In more recent years, I have confided in her and she has become an incredibly important person in my life, despite the fact that we haven't lived in the same city for nearly ten years. I would also like to thank Uncle Tom, Aunt Lauren, and their three kids Sophie, Jonas, and Drew, who been infinitely hospitable since I moved to San Diego.

I would like to thank Court Heller for his love and support throughout the bulk of my graduate career. I would especially like to thank him for enduring the pains and celebrating the successes of graduate school with me.

I would like to thank The National for their music.

Ultimately, I would like to thank Greg Macias (a.k.a. Percy Robinson). He has made planets align.

The text of chapter 2, in full, is a reprint of the material as it appears in Nature Chemical Biology 2008, Vol. 18, pp. 3039-3042. Permission was obtained from the co-authors. I was the third author of this work. As mentioned in the manuscript, Paul O'Maille designed the study, conducted experiments, analyzed data and wrote the manuscript, Arthur Malone conducted experiments and developed small-scale protein purification, I conducted experiments, analyzed data and contributed revision to the manuscript, B Andes Hess Jr. conducted quantum mechanics calculations and contributed revisions to the manuscript, Lidia Smentek conducted quantum mechanics calculations, Iseult Sheehan conducted experiments, Bryan Greenhagen and Joseph Chappell designed the study and contributed revisions to the manuscript, Gerard Manning analyzed data, developed the biosynthetic tree and chemical distance analysis, and contributed revisions to the manuscript, and Joseph P. Noel designed the study, analyzed the data and wrote the manuscript. This research was performed under the supervision of Joseph P. Noel.

The text of chapter 3, in full, is a reprint of material as it appears in ACS Chemical Biology 2010, 5 (4), pp 377–392, with the exception of the section under supporting information entitled "computational details" which was excluded. Permission was obtained from all co-authors. I am second author of this work. Paul O'Maille wrote the manuscript, and was also involved with protein purification, GCMS data analysis, crystallization experiments, and crystallographic data processing, structure solution and refinement. I was responsible for protein purification, GCMS data analysis, crystallization experiments, crystallographic data processing, structure solution, refinement, and contributed revisions to the manuscript. Juan Faraldos was responsible for organic synthesis, NMR characterization of sesquiterpenes, and contributed revisions to the manuscript. Yuxin (Marilyn) Zhao was responsible for chemical synthesis of *cis*-FPP. B. Andes Hess Jr. and Lidia Smentek were responsible for all computational studies. The research included in the manuscript was performed under the supervision of Robert Coates and Joseph P. Noel (who also contributed revisions and helped write the manuscript).

The text of chapter 4, in part, is currently being prepared for submission for publication of the material. Dellas, Nikki; Noel, Joseph P. I am the first author of this material. All experiments were performed under the supervision of Joseph P. Noel.

The text of chapter 5, in full, is a reprint of the material as it appears in ACS Chemical Biology 2010, 5(6), pp 589-601. I am the primary author of this paper. The research was performed under the supervision of Joseph P. Noel.

The text of chapter 6, in part, has been submitted for publication of the material as it may appear in Chemical Communications, 2010, Dellas, Nikki; Manning, Gerard, Noel, Joseph P. I am the first author of this paper. Gerard Manning and Joseph P. Noel are the corresponding authors. I was responsible for all gene cloning, enzyme expression, purification,

and kinetic characterization of IPK and its homologs. Gerard Manning was responsible for the bioinformatic and phylogenetic analysis of IPK and its homologs. All experiments were performed under the supervision of Joseph P. Noel.

VITA

Education

2010 Ph.D., Chemistry
University of California, San Diego

2007 M.S., Chemistry
University of California, San Diego

2005 B.S., Chemistry
Carnegie Mellon University

Publications

1. **Dellas, N.**; Noel, J.P. A Conserved Amino Terminal Motif in Patchouli Alcohol Synthase Controls Product Distribution. *Manuscript in preparation.*

2. **Dellas, N.**; Manning, G.; Noel, J.P. Isopentenyl Phosphate Kinase Homologs Outside of Archaea Suggest a Bifurcating Mevalonate Pathway in a Diversity of Eukaryotes. *Submitted to Chem Commun.*

3. **Dellas, N**.; Noel, J. P., Mutation of archaeal isopentenyl phosphate kinase highlights mechanism and guides phosphorylation of additional isoprenoid monophosphates. *ACS Chem Biol* **2010,** *5* (6), 589-601.

4. Noel, J. P.; **Dellas, N.**; Faraldos, J. A.; Zhao, M.; Hess, B. A., Jr.; Smentek, L.; Coates, R. M.; O'Maille, P. E., Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. *ACS Chem Biol* **2010,** *5* (4), 377-392.

5. Faraldos, J. A.; O'Maille, P. E.; **Dellas, N.**; Noel, J. P.; Coates, R. M., Bisabolyl-derived sesquiterpenes from tobacco 5-epi-aristolochene synthase-catalyzed cyclization of (2Z,6E)-farnesyl diphosphate. *J Am Chem Soc* **2010,** *132* (12), 4281-9.

6. O'Maille, P. E.; Malone, A.; **Dellas, N**.; Andes Hess, B., Jr.; Smentek, L.; Sheehan, I.; Greenhagen, B. T.; Chappell, J.; Manning, G.; Noel, J. P., Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nature Chem Biol* **2008,** *4* (10), 617-623.

7. Dasgupta, R.; Hirschmann, M. M.; **Dellas, N**., The effect of bulk composition on the solidus of carbonated eclogite from partial melting experiments at 3 GPa. *Contrib. Mineral. Petrol.* **2005,** *149* (3), 288-305.

ABSTRACT OF THE DISSERTATION

Exploring Structural and Functional Features of Enzymes Across Isoprenoid Biosynthesis: From Archaeal Isopentenyl Phosphate Kinase of Primary Metabolism to Plant Terpene Cyclases of Specialized Metabolism

by

Nikki Dellas

Doctor of Philosophy in Chemistry

University of California, San Diego, 2010

Professor Joseph P. Noel, Chair

Professor Elizabeth Komives, Co-Chair

Isoprenoid biosynthesis constitutes an immensely diverse, highly branched network of pathways that spans both primary and secondary (specialized) metabolism in all organisms. The mevalonate (MVA) pathway or the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway operate in a given organism to produce the two important building blocks for all downstream isoprenoids: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In Archaea, the biosynthesis of these two vital building blocks remains unclear. The current hypothesis is that Archaea utilize an alternative mevalonate pathway that follows the canonical pathway up until the biosynthesis of phosphomevalonate. At this point, a decarboxylation event followed by a phosphorylation event produces the essential building block, IPP. The latter step is catalyzed by isopentenyl phosphate kinase (IPK). In this work, we solved the structure of IPK from *M. jannaschii* and successfully used it toward: 1) the

design of a deeper active site pocket for binding and catalysis of longer chained isoprenoid monophosphates; 2) the identification and characterization of active IPK homologs in other kingdoms of life. This work contributes towards the design of a synthetic metabolic pathway and reveals new information about the potential existence of a bifurcated mevalonate pathway in all plants and certain other eukaryotic organisms.

Farnesyl diphosphate is directly derived from the building blocks IPP and DMAPP and is an essential metabolic intermediate for a variety of downstream primary and secondary metabolic pathways including cholesterol biosynthesis and terpenoid biosynthesis, respectively. Sesquiterpene cyclases (synthases) are part of terpenoid biosynthesis and catalyze the cyclization of farnesyl diphosphate into one or more sesquiterpene products; these chemicals play important biological roles in defense and communication, especially in plants. Here, we explore a variety of mutant and wild type plant sesquiterpene cyclases in attempt to understand several concepts: 1) how these enzymes traverse a defined catalytic landscape to biosynthesize disparate products without compromising their catalytic activities; 2) the structural and functional differences associated with turnover of *cis*- and *trans*-FPP by wild type and promiscuous cyclase mutants; 3) how certain sesquiterpene synthases utilize an Arg-Pro motif within the amino terminal domain to interact with the catalytic C-terminal domain and modulate product profile complexity.

Chapter 1

Introduction

1.1. Isoprenoid biosynthetic pathways

Isoprenoid biosynthesis constitutes a complex series of branched pathways that results in the production of a variety of essential and specialized metabolites across all kingdoms of life. These essential metabolites include (but are not limited to) squalene, hopanoids, and steroids (important for membrane structure in Archaea, Bacteria, and Eukarya, respectively),[1,2] dolichols (N-linked glycosylation and membrane anchorage of sugars in eukaryotes and archaea,[3] terpenes (plant defense and communication), carotenoids (photoprotection for certain prokaryotes and plants[4], prenylquinones (mitochondrial electron transport),[5] and gibberellins (plant growth and development, for review see Hedden et al 1997).[6]

All metabolites discussed above originate from the two essential five-carbon building blocks of isoprenoid biosynthesis: isopentenyl diphosphate (IPP) and its stereoisomer, dimethylallyl diphosphate (DMAPP). One molecule of DMAPP reacts with one, two, or three molecules of IPP via a prenyltransferase (isoprenoid diphosphate synthase) to generate geranyl diphosphate (GPP), farnesyl diphosphate (FPP), or geranylgeranyl diphosphate (GGPP), respectively. These three compounds are then utilized in different ways by a variety of enzymes to biosynthesize a repertoire isoprenoid products. DMAPP can be produced either in conjunction with IPP or is made through isomerization of IPP via an IPP isomerase (IPPI).

The current hypothesis is that IPP (and DMAPP) biosynthesis occurs through one of two distinct metabolic pathways: the mevalonate (MVA) pathway or the more recently discovered 1-deoxy-D-xylulose 5-phosphate (DXP) pathway (also known as the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway).[7-9]

1.1.1. The DXP pathway

The DXP pathway consists of the following steps: 1) Condensation of glyceraldehyde 3-phosphate (G3P) and the "activated acetaldehyde" of pyruvate (Pyr) catalyzed by the enzyme DXP synthase (DXPS) to produce DXP[10]; 2) reduction of DXP to 2-C-methylerythritol-4-phosphate (MEP) by the enzyme DXP reductoisomerase (DXR); 3) coupling of MEP and cytidine triphosphate (CTP) by the enzyme 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (CMS) to generate 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME); 4) phosphorylation of CDP-ME by the enzyme 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK) to produce 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP); 5) conversion of CDP-MEP to 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEcPP) by the enzyme 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS); 6) ring-opening reduction of MEcPP to (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP) by the enzyme HMP-PP synthase (HDS)[11,12] and 7) reductive dehydration of HMB-PP to a mixture of IPP and DMAPP by the enzyme HMB-PP reductase (HDR).[11,13] These steps are detailed in Figure 1.1.

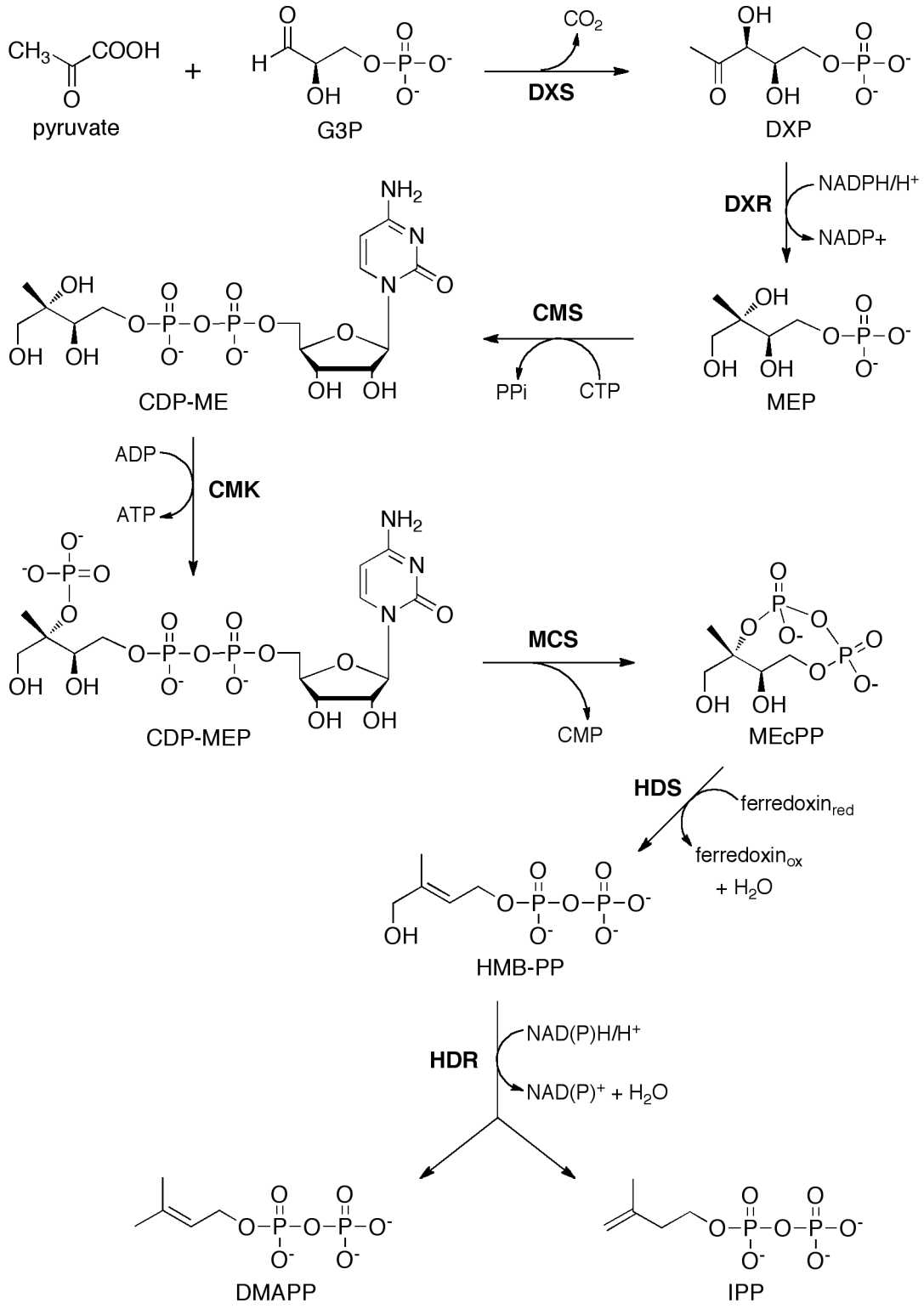


Figure 1.1. The DXP pathway

1.1.2. The MVA pathway

The MVA pathway consists of the following steps: 1) condensation of acetyl-CoA with acetoacetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) via the enzyme HMG-CoA synthase; 2) reduction of HMG-CoA to mevalonate by HMG-CoA reductase (note: this is the rate limiting step of cholesterol biosynthesis[14] targeted by statin drugs[15]); 3) phosphorylation of mevalonate to phosphomevalonate by the enzyme mevalonate kinase (MVK); 4) phosphorylation of phosphomevalonate to diphosphomevalonate (DPM) by the enzyme phosphomevalonate kinase (PMK); 5) decarboxylation of DPM to generate IPP via the enzyme DPM decarboxylase (DPM-DC) and 6) isomerization of IPP to generate DMAPP via the enzyme IPP isomerase (IPPI). These steps are detailed in Figure 1.2.

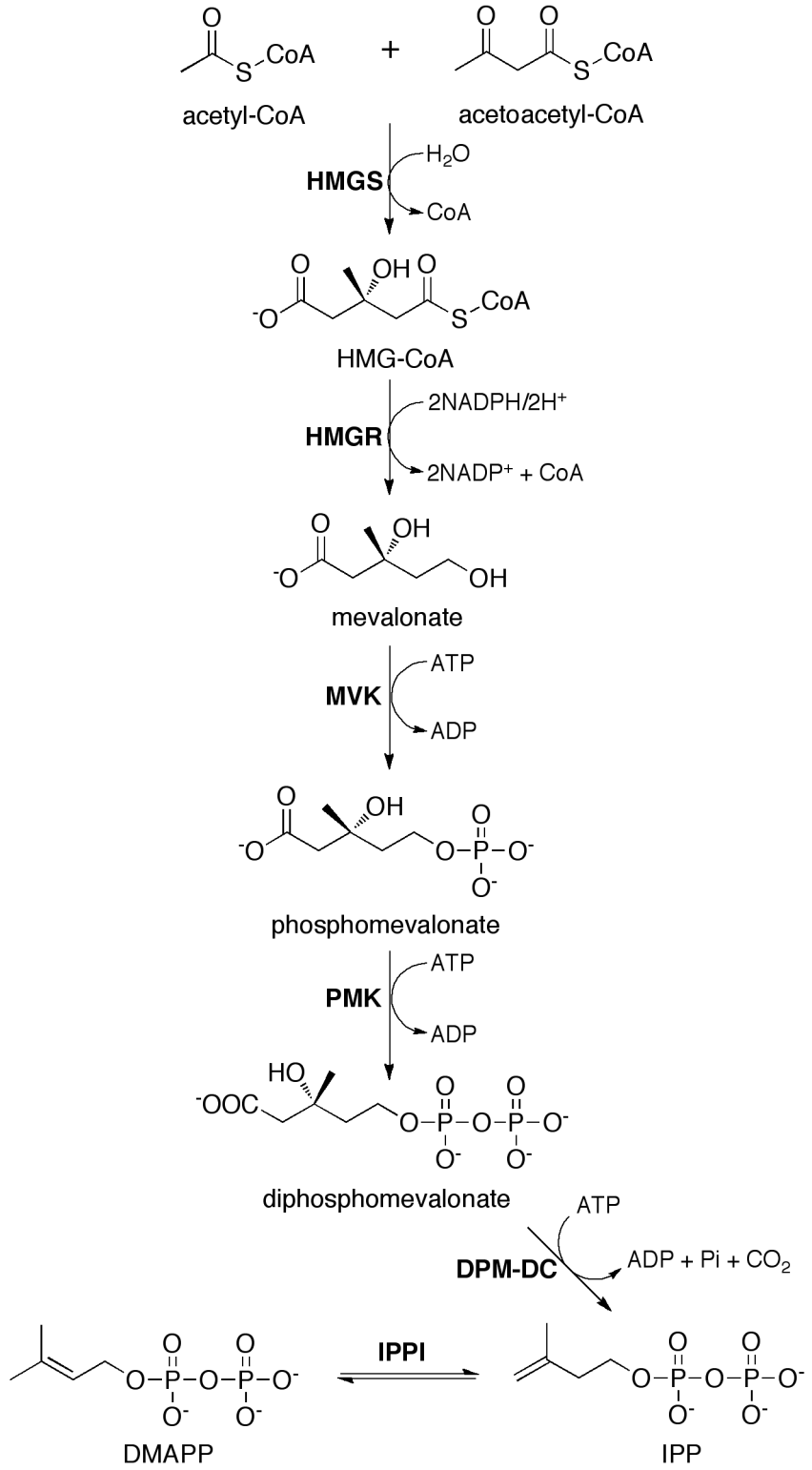


Figure 1.2. The MVA pathway

1.1.3. Isoprenoid biosynthesis across the three domains of life

Most organisms contain gene orthologs for either one or both pathways. In general, the MVA pathway is found in eukaryotes (and certain bacteria) while the DXP pathway is found in most bacteria and plastid-bearing eukaryotes. Plants contain both pathways: the DXP pathway operates in the plastids while the MVA pathway operates in the cytosol (although recent results suggest sub-cellular localization of certain MVA pathway enzymes to the ER, mitochondria, and peroxisomes).[16,17,18] Archaea contain gene orthologs for the first four listed steps of the MVA pathway, but are missing the last three steps, including those catalyzed by PMK, DPM-DC, and IPPI[19]. The current view is that Archaea use a modified mevalonate pathway to generate IPP and DMAPP[19, 20]. One proposal for evolutionary modification entails a reversal of the phosphorylation and decarboxylation events that follow PMK biosynthesis in the classical mevalonate pathway. This modification would include decarboxylation of PMK to isopentenyl phosphate (IP), followed by phosphorylation of IP to IPP, generating the same end product as in the classical MVA pathway (Figure 1.3). The recent successful isolation and characterization of an archaeal isopentenyl phosphate kinase (IPK) that can perform the latter reaction circumstantially support the proposed modified pathway.[19, 20] Chapter 5 details more recent findings with regard to this modified MVA pathway, particularly with regard to the unexpected discovery of its existence outside of the Archaeal domain of life. These findings challenge our current understanding on what was thought to be a widely accepted biosynthetic pathway.

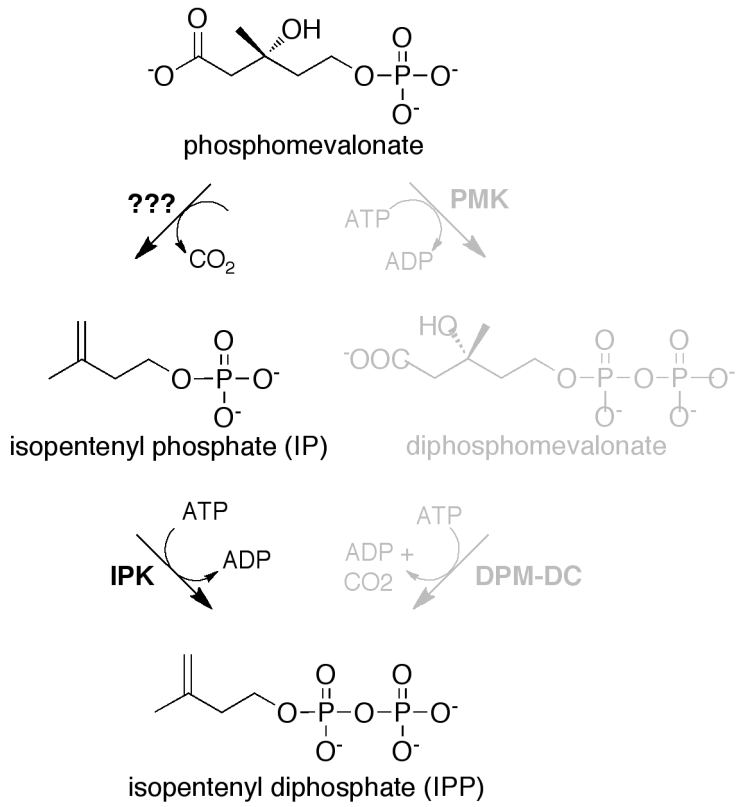


Figure 1.3. Proposed alternative mevalonate pathway in Archaea

1.2. Short-Chain Prenyl Diphosphate Synthases

GPP synthase (GPPS), FPP synthase (FPPS), and GGPP synthase (GGPPS) belong to a division of prenyltransferases known as "short-chain prenyl diphosphate synthases" that catalyze the iterative transfer of one, two, or three molecules of IPP, respectively, to either DMAPP or a growing prenyl diphosphate chain.[21] The GPPS reaction mechanism includes ionization of the pyrophosphate group of DMAPP, forming an electrophilic carbocation, electrophilic addition of the double bond of IPP to the carbocation C1 atom of DMAPP, and proton abstraction from C2 of the resulting C10 carbocation to generate GPP[22] (Figure 1.4). The mechanisms for FPPS and GGPPS proceed with one and two more iterations, respectively, to generate the appropriate C_{15} and C_{20} prenyl diphosphate products. These

reactions are an important part of primary metabolism in many organisms; for example, in eukaryotes, FPPS is vital for the downstream production of all sterols including cholesterol.

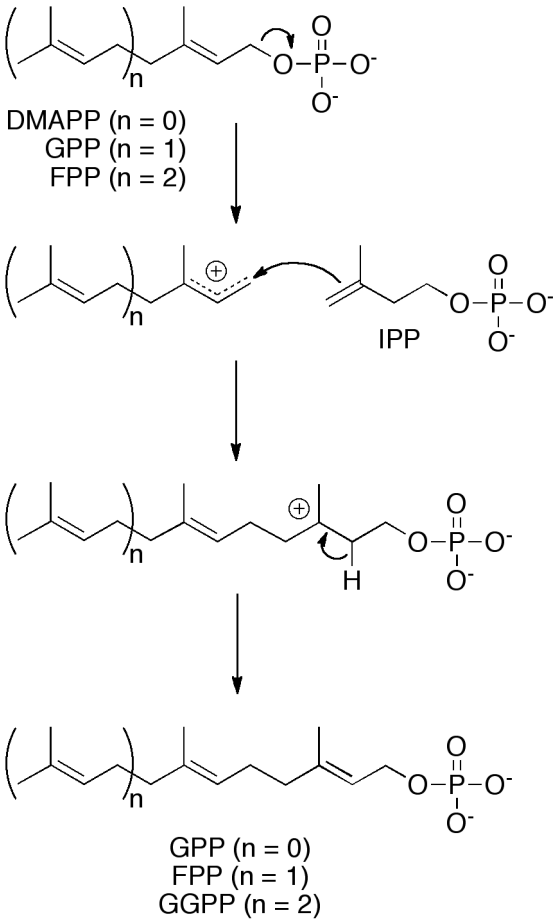


Figure 1.4. General mechanism for short-chain prenyl diphosphate synthases

In general, these three types of prenyltransferases share catalytic machinery and a conserved structural scaffold within which these reactions occur. In 1994, the crystal structure of avian FPPS was published, representing the first prenyl diphosphate synthase to be structurally characterized[23]. The structure consists of a homodimeric arrangement, where each monomer encompasses of a bundle of α-helices; ten of these helices surround the active site cavity.[21, 23] Two highly conserved aspartate-rich motifs, known as the "first aspartate-rich

motif" (FARM, represented as $DDx_{2-4}D$) and the "second aspartate-rich motif" (SARM, represented as DDXXD), lie on opposite ends of the active site.[24, 25] More recently published crystal structures of FPPS from *E. coli* complexed with DMAPP or DMAPP and IPP demonstrate conformational changes associated with different phases of the elongation reaction[26]. In the presence of DMAPP and Mg^{2+}, two Mg^{2+} ions coordinate to FARM and the diphosphate group on the allylic substrate.[23, 26] The binding of IPP triggers secondary structural changes that close the active site and squeeze out water; These changes are accompanied by the coordination of a third Mg^{2+} ion to SARM[26].

Most short-chain prenyl diphosphate synthases do not demonstrate a high degree of product promiscuity.[24, 25] There are several features that govern chain-length specificity in short-chain prenyl diphosphate synthases. One hallmark is size of the active site pocket: a larger pocket can accommodate longer-chained products[24] Another attribute is the presence or absence of amino acids at specific locations directly upstream of FARM;[23, 24, 27, 28] in some cases these residues may protrude into the active site tunnel, marking the floor of the active site and preventing further chain elongation.[23, 24] A third notable feature that modulates chain-length specificity is the presence of extra residues in FARM (DDXXXXD compared to DDXXD); this structural component results in products with shorter chain lengths.[23, 29]

1.3. Terpene synthases: function and mechanism

Terpene synthases (cyclases) encompass a family of enzymes playing critical roles in the secondary metabolism and chemical ecology of plants, bacteria, fungi and marine organisms.[30] These enzymes catalyze the cyclization of their respective isoprenyl diphosphate substrate (either GPP, FPP or GGPP) into a variety of chemically complex products that often contain a number of chiral centers. There are three distinct classes of terpene synthases:

monoterpene, sesquiterpene, and diterpene synthases. Monoterpene synthases catalyze the cyclization of GPP (a C_{10} prenyl diphosphate) into one or more monoterpene products. One example is S-linalool synthase, which produces the fragrant monoterpene S-linalool that is used to attract a moth pollinator to a specific plant species.[31] Sesquiterpene synthases catalyze the cyclization of FPP (a C_{15} prenyl diphosphate substrate) into one or more sesquiterpene products. One example is (E)-beta-caryophyllene synthase, which produces the sesquiterpene (E)-beta-caryophyllene as its major product; this molecule contributes to the airborne defense response for certain plants against herbivores.[32] Diterpene synthases catalyze the cyclization of GGPP (a C_{20} prenyl diphosphate) into one or more diterpene products. One example is taxadiene synthase, which produces the hydrocarbon core, taxadiene, of the pharmaceutically relevant anti-cancer agent known as Taxol™.[33]

Although many monoterpenes and sesquiterpenes function as signaling molecules to attract pollinators, ward off enemies, or communicate with their external environment, other sesquiterpenes (and diterpenes) can additionally be either directly or indirectly used for medicinal purposes. One popular example of a sesquiterpene synthase that produces such a precursor is amorpha-4,11-diene synthase, whose product can be derivatized to the anti-malarial drug known as artemisinin.[34]

For this reason, the search for ways to overproduce such valuable compounds is ongoing.[35] Overexpression of MVA pathway enzymes in *S. cerevisiae,*[36, 37] overexpression of the DXP pathway in *E. coli,*[38] or heterologous expression of the MVA pathway in *E. coli*[39, 40] are three common methodologies that have effectively produced significant quantities of such terpenes. However, each method has certain drawbacks. For example, heterologous expression of the MVA pathway in *E. coli* has lead to difficulties associated with metabolic flux through the pathway and with cell growth,[39] while overexpression of MVA pathway

enzymes in *S. cerevisiae* causes the non-productive accumulation of farnesol, which is usually considered an unwanted byproduct.[36] FPP-induced feedback inhibition of mevalonate kinase of the MVA pathway has also been reported.[41] Nevertheless, some of these methods have successfully produced concentrations of terpenes at over 100mg/liter of culture and continuing efforts will most likely improve this number.[37]

In general, the terpene cyclase reaction begins with Mg^{2+} or Mn^{2+} assisted ionization of the pyrophosphate group on the substrate, which is usually accompanied by electrophilic cyclization to generate a secondary or tertiary carbocation intermediate.[42] The highly reactive acyclic or cyclic carbocation intermediate can then undergo further transformations including ring closures and hydride shifts within the hydrophobic active site through other closures and migrations until proton abstraction or hydroxylation quenches this cycle by means of water or an active site side chain. This reaction, termed "ionization-dependent cyclization" takes place in the C-terminal catalytic domain of terpene cyclases.

A highly conserved aspartate-rich motif termed the "**DD**XX(**D**/**E**)" motif coordinates two of the three divalent metal cations (in the case of Mg^{2+}, these are usually denoted Mg_A^{2+} and Mg_C^{2+})[43, 44] that are responsible for lowering the activation barrier for pyrophosphate ionization and subsequent allylic carbocation stabilization; this motif is structurally and functionally conserved with the FARM motif in prenyl diphosphate synthases (residues in bold denote those involved with metal ion coordination)[45]. Another conserved motif present in all terpene cyclases that coordinates the third divalent cation (often referred to as Mg_B^{2+}) is the (**N,D**)DXX(**S,T**)XXX(**E,D**) motif, abbreviated as the NSE/DTE motif[46]. This motif is found as **N**DXX**S**XXX**E** in most fungal and bacterial terpene cyclases, and as **D**DXX**T**XXX**E** in most plant terpene cyclases[46].

1.3.1. Monoterpene synthases

Monoterpene synthases (cyclases) are a division of terpene synthases that turn over the C_{10} isoprenoid GPP using an "ionization-dependent cyclization" mechanism. Plant monoterpene cyclases usually contain a plastid localization sequence, which consists of approximately fifty additional residues flanking the amino-terminus.[47] Given that monoterpene synthases accept the shorter C10 substrate, GPP, the double bonds are not initially oriented properly to enable electrophilic cyclization of the nascent carbocation. Therefore, following initial pyrophosphate ionization, an isomerization event must occur, which generates the stable intermediate linalyl diphosphate via a two-step reaction entailing reattachment of the pyrophosphate to C3 and accompanying rotation about the C2-C3 bond[48] (Figure 1.5). Since roughly one-third of all characterized monoterpene synthases produce acyclic products,[49] this isomerization event is not always necessary; however it is a prerequisite for the generation of any cyclic monoterpene. A pair of Arg residues located directly C-terminal to the plastid localization sequence have been implicated in the isomerization mechanism. For example, in limonene synthase, truncation or mutation of the arginine pair renders the protein inactive towards geranyl diphosphate (the native substrate) however the enzyme catalyzes the reaction to completion when provided with the isomerized version of the substrate, linalyl diphosphate.[47] Nevertheless, there is debate with regard to the precise function of this motif, especially since it exists in certain sesquiterpene synthases (as either an Arg-Arg pair or an Arg-Pro pair) that do not require an isomerization event. A mutational analysis of both the Arg-Pro and Arg-Arg pairs in several sesquiterpene synthases (detailed in Chapter 6) implicates a broader role for this motif in reaction modulation.

The recent discovery of a *cis*-GPP synthase (called neryl diphosphate synthase, or NPPS) capable of producing *cis*-derived neryl diphosphate (NPP) suggests an alternative

mechanism for derivation of cyclic monoterpenes which would not involve isomerization of GPP;[50] In fact, successful characterization of the NPP-utilizing β-phellandrene synthase fully supports this hypothesis.[50] Another recent publication analogous to this in a sesquiterpene cyclase reports utilization of the cis-derivative of FPP, (Z,E)-FPP, as its substrate[51].

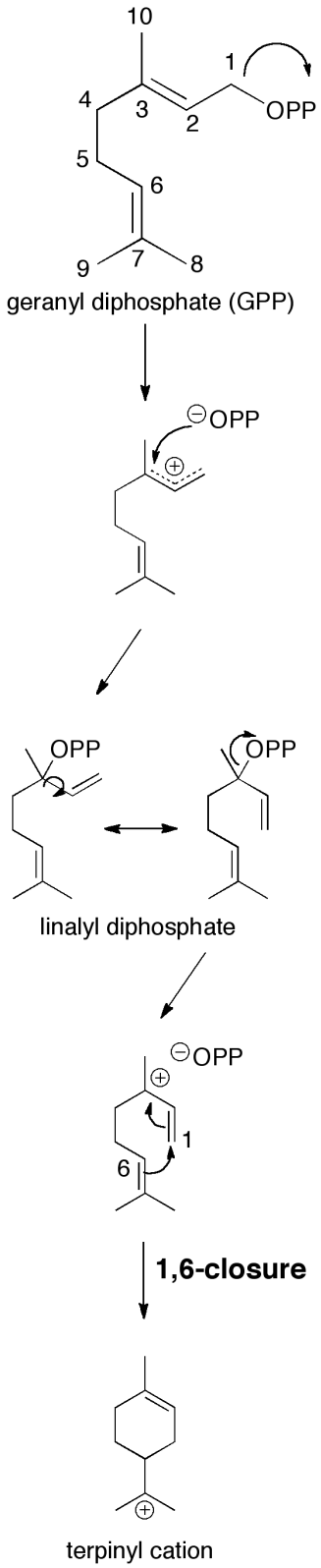


Figure 1.5. Geranyl Cation Cyclization

1.3.2 Sesquiterpene Synthases

Sesquiterpene synthases (cyclases) are a well-studied division of terpene synthases that turn over the C_{15} isoprenoid FPP using an "ionization-dependent cyclization" mechanism. Following initial pyrophosphate loss, many sesquiterpene cyclases employ the transoid cyclization pathways (termed "transoid synthases")[52] that include an initial 1,10-closure or 1,11-closure, generating the germacradienyl cation or the humulyl cation, respectively (Figure 1.5). These central carbocation intermediates are shuttled through a cascade of rearrangements within the enzyme's active site to generate a repertoire of different sesquiterpene products. Additionally, certain sesquiterpene synthases employ the cisoid cyclization pathway (termed "cisoid synthases")[52] by performing an initial isomerization event (analogous to that occurring in monoterpene synthases) to generate nerolidyl diphosphate prior to pyrophosphate re-ionization and subsequent 1,6-closure or 1,7-closure to generate the bisabolyl cation or cycloheptenyl cation, respectively[49]; one such example is amorpha-4,11-diene synthase[53] (Figure 1.6). A variety of FPP synthases can produce minor amounts of (*Z,E*)-FPP in addition to the *all-trans* major product.[54] This finding indicates that in some organisms, more than one substrate may be available to sesquiterpene cyclases. A recent paper reports the discovery of an sesquiterpene cyclase analogous to the NPP-utilizing monoterpene cyclase in that it uses the cis-derivative of FPP, (Z,E)-FPP, as its substrate. Incidentally, 5-epi-aristolochene synthase from *Nicotiana tabacum* (TEAS) can produce a small number of *cis*-derived products among a majority of trans-derived products, suggesting that some of these enzymes have the catalytic machinery necessary to perform both reactions.[55] Chapter 3 explores the structural and functional capabilities of TEAS when given either (*E,E*)- or (*Z,E*)-FPP.

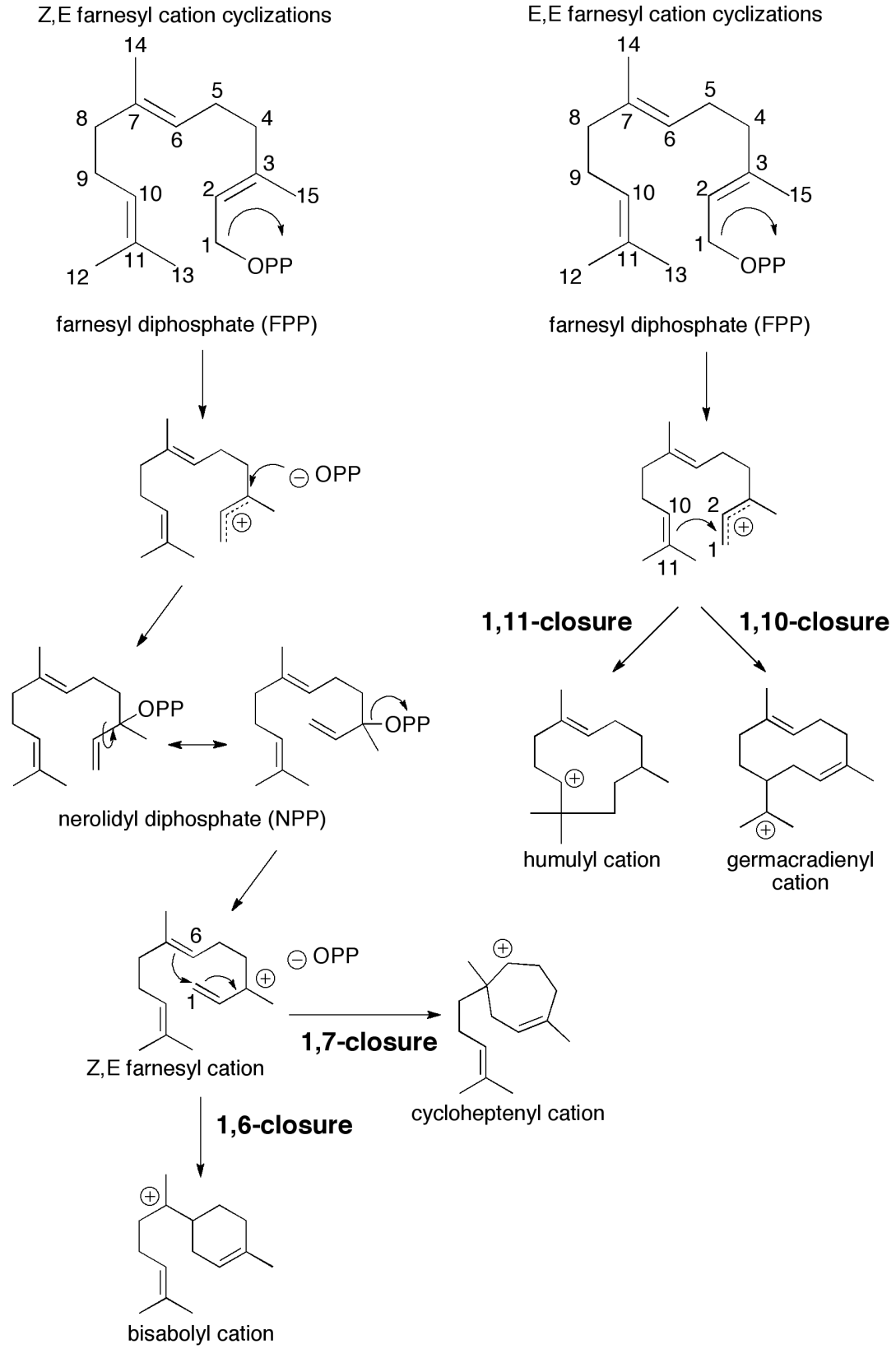


Figure 1.6. Farnesyl cation cyclization pathways

Although bacterial sesquiterpene synthases usually produce only one product, plant sesquiterpene synthases exhibit varying degrees of product diversity (catalytic promiscuity)[56]. For example, humulene synthase from *Abies grandis* produces more than fifty cisoid- and transoid-derived sesquiterpenes[57]. Such high levels of product diversity can be indicative of relaxed pyrophosphate binding within the active site.[57] Patchouli alcohol synthase from *Pogostemon cablin* synthesizes at least thirteen all-trans derived sesquiterpene products in addition to the major product (-)-patchoulol at approximately 37%.[58] By contrast, TEAS synthesizes approximately 79% 5-epi aristolochene in addition to twenty-five minor products.[52, 55] In general, variation in product diversity from one sesquiterpene cyclase to the next is most likely a reflection of both the degree of evolutionary refinement (as these enzymes transitioned from primary metabolism[59] or traversed through a landscape within specialized (secondary) metabolism[60]) and environmental adaptation (where a "chemical library" or "cocktail" of compounds from one sesquiterpene synthase possesses broader protection for a sessile organism within an ecosystem[59]). Current research involving specificity transformations is guided by such underlying themes. For example, a highly promiscuous sesquiterpene cyclase can be tuned to produce one major product, as shown by Yoshikuni *et al* (2006), where γ-humulene synthase was used as a platform to engineer seven distinct sesquiterpene synthases each with its own major product.[61] Interconversion of two highly specific plant sesquiterpene cyclases is demonstrated in work by Greenhagen *et al* (2006), where mutation of nine amino acid positions in 5-epi-aristolochene synthase and eight positions in a premnaspirodiene synthase was sufficient for interconversion of the two enzyme activities.[62] Intriguingly, interconversion of these two enzymes was accomplished through mutation of amino acids that were mostly second tier to the active site and were not directly in

contact with the farnesyl diphosphate substrate, which suggests that tuning these enzymes toward production of an alternative product is not always obvious.

1.3.3. Diterpene Synthases

Diterpene synthases (cyclases) are a division of terpene cyclases that cyclize C_{20} prenyl diphosphate substrates. Although certain diterpene cyclases (such as taxadiene synthase[63]) rely solely on the "ionization-dependent cyclization" mechanism, some require an additional step involving "proton-initiated cyclization" prior to "ionization-dependent cyclization" (Figure 1.7). For example, copalyl diphosphate, which is formed from GGPP via "proton-initiated cyclization," is the substrate for certain diterpene cyclases such as *ent*-kaurene synthase and abietadiene synthase. In higher plants and bacteria, *ent*-kaurene biosynthesis requires two separate cyclases: (-)-copalyl diphosphate synthase (CPS, formerly known as *ent*-kaurene synthase A[64]) which performs "proton-initiated cyclization" of GGPP to (-)-copalyl diphosphate ((-)-CDP), and *ent*-kaurene synthase (KS, formerly *ent*-kaurene synthase B[64]) which performs the "ionization-dependent cyclization" of (-)-CDP to ent-kaurene.[65-67] These two reactions are accomplished by one bifunctional enzyme in lower level plants such as moss[68] and in fungi[69]. Abietadiene synthase (ADS) is another bifunctional diterpene cyclase that contains two active sites: one in the N-terminal domain that performs proton-initiated cyclization to generate (+)-copalyl diphosphate and the other in the C-terminal domain that performs "ionization-dependent cyclization" to eventually generate abietadiene (FIGURE).[70, 71] The universal DDXXD motif remains conserved throughout all bifunctional and monofunctional diterpene cyclases and, as mentioned previously, is important for the ionization-dependent reaction[70]. An additional motif, the DXDD motif, is important for catalysis of the proton-initiated cyclization reaction[72]. The spatial orientations of these motifs

in the context of two common terpene cyclase folds will be discussed in the following section, which details several tertiary structural elements conserved among terpene cyclases. Bifunctional diterpene cyclases such as ADS contain a 240 amino acid N-terminal insert whose structure and function remain unknown, although there has been speculation that this "insertional element" plays some role in the proton-initiated reaction such as shielding the active site from water or premature release of a reactive carbocation intermediate into bulk solvent[72-74].

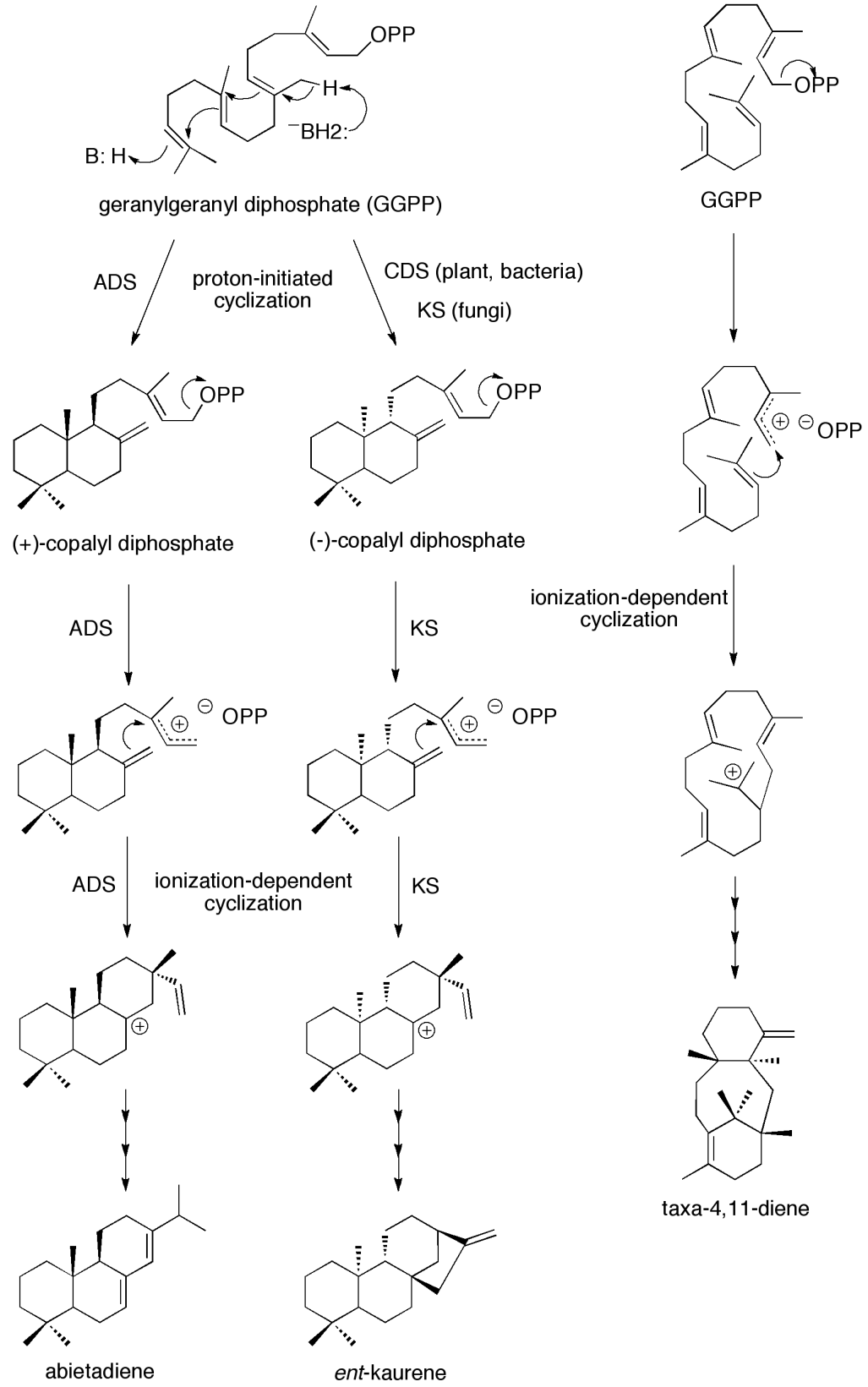


Figure 1.7. Geranylgeranyl cation cyclization pathways

1.4. Terpene synthases: structure

1.4.1. The terpene cyclase fold

The class I terpene cyclase fold and class II terpene cyclase fold are two common tertiary structural features observed among mono-, sesqui-, and diterpene cyclases. The class I terpene cyclase fold is an α-helical fold where the ionization-dependent cyclization reaction takes place[44]. The class II terpene cyclase fold is an α-barrel fold that carries out the proton-initiated cyclization reaction[44].

Monoterpene and sesquiterpene synthases share many similar structural features. All plant mono- and sesquiterpene synthases contain both an N-terminal and C-terminal domain. The N-terminal domain structurally aligns with the catalytic core of glycosyl hydrolases[75] and possesses some structural homology to the class II terpene cyclase fold;[76] however, no function has been assigned to this domain other than that it is thought to be involved with capping of the active site in the C-terminal domain.[77] The C-terminal domain in mono- and sesquiterpene cyclases and the single domain of bacterial and fungal terpene cyclases contains the class I terpene cyclase fold and accompanying DDXXD and NSE/DTE motifs necessary for the ionization-dependent cyclization. Despite sequence divergence between terpene cyclases and prenyltransferases, short-chain prenyltransferases share this class I terpene cyclase fold.[78]

Like mono- and sesquiterpene cyclases of plant origin, diterpene cyclases contain an N-terminal domain and C-terminal domain that have class II and class I terpene cyclase folds, respectively. Although the N-terminal domain of monofunctional diterpene cyclases (such as taxadiene synthase) is inactive, the N-terminal domain of bifunctional diterpene cyclases (such as ADS) contains the conserved DXDD motif and is able to perform the proton-initiated cyclization event[44]. Monofunctional diterpene cyclases have mutations in the DXDD motif

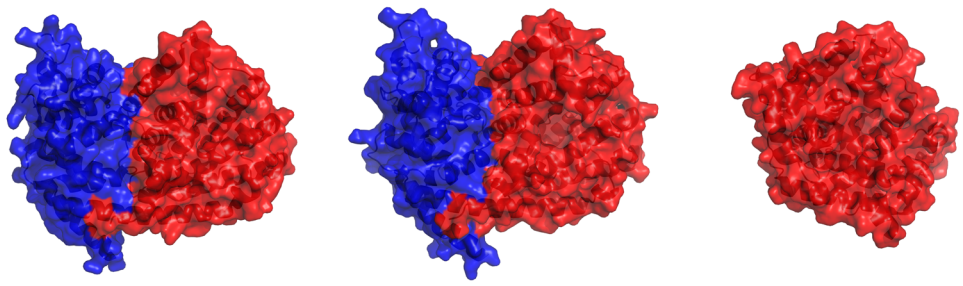
that render them incapable of performing the proton-initiated reaction.[72] The C-terminal domain of monofunctional diterpene cyclases contains the class I terpene cyclase fold, the DDXXD and NSE/DTE motifs, and performs the ionization-dependent cyclization reaction similarly to mono- and sesquiterpene cyclases. There are additional cases where the bifunctional diterpene cyclase exists as two separate enzymes, as is the case with CDS and KS (discussed in a previous section); these two enzymes are structurally and functionally homologous to the N-terminal and C-terminal domain in ADS and contain the class II and class I terpene cyclase folds, respectively.

1.4.2. Crystal structures of monoterpene and sesquiterpene synthases

To date, there are three crystal structures of monoterpene cyclases and seven crystal structures of sesquiterpene cyclases. The three monoterpene synthase crystal structures are from plants, and include (+)-bornyl diphosphate synthase from *Salvia officinalis* (sage)[77], limonene synthase from *Mentha spicata* (peppermint)[79], and 1,8-cineole synthase from *Salvia fruticosa* (Greek sage)[80]. The seven sesquiterpene synthase crystal structures include two from plants (5-epi-aristolochene synthase from *Nicotiana tabacum*[75] and δ-cadinene synthase from *Gossypium arboreum*[81]), three from fungi (trichodiene synthase from *Fusarium sporotrichioidies*[82], aristolochene synthase from *Aspergillus terreus*[83], and aristolochene synthase from *Penicillium roqueforti*[84]), and two from bacteria (pentalenene synthase from *Streptomyces sp.*[78] and epi-isozizaene synthase from *Streptomyces coelicolor*[85]).

In general, all crystal structures of terpene cyclases to date share a high degree of structural homology considering their sequence similarity is quite low, which suggests early evolutionary divergence followed by significant sequence diversification[86]. Plant monoterpene and sesquiterpene cyclases contain both an N-terminal α-barrel domain and C-terminal α-

helical domain, while bacterial and fungal sesquiterpene cyclases have one domain (corresponding to the C-terminal domain of plant terpene cyclases) (Figure 1.8). The helices comprising the C-terminal domain are named according to the same nomenclature as that used for short-chain prenyl diphosphate synthases (Figure 1.9).



plant monoterpene synthases | plant sesquiterpene synthases | bacterial/fungal sesquiterpene synthases

Figure 1.8. Global Structure of monoterpene and sesquiterpene cyclases from various kingdoms of life. N-terminal domain is colored blue, C-terminal domain is colored red.

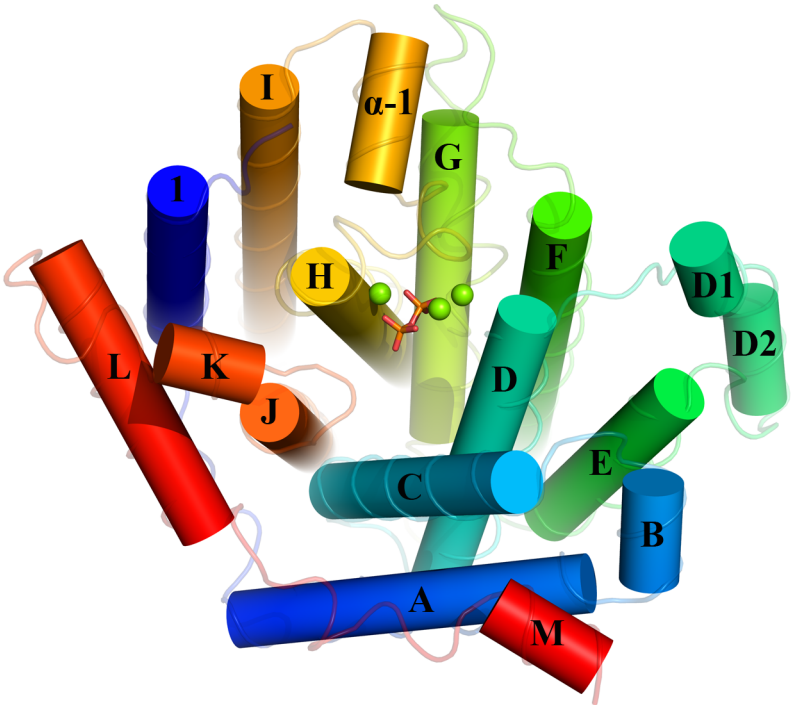


Figure 1.9. The catalytic C-terminal domain of terpene cyclases (image designed based on the crystal structure for trichodiene synthase complexed with three Mg^{2+} ions and pyrophosphate (pdb ID: 2PS5).[87]

The general terpene cyclase active site contains a hydrophobic region and a hydrophilic region: the former stabilizes the isoprenoid chain through hydrophobic interactions and the latter coordinates magnesium ions and stabilizes the pyrophosphate moiety. The two metal-binding motifs, including the DDXXD motif (located on helix D) and the NSE/DTE motif (located on helix H), are mostly conserved throughout all structures and coordinate up to three Mg^{2+} or Mn^{2+} ions. Most structures of terpene cyclases exhibit some dynamic character in one or more secondary structural elements upon substrate binding; these movements aid in exclusion of water from the active site to promote completion of the carbocation mechanistic cascade. Although all terpene cyclases described here share

considerable structural homology, there are several noteworthy structural differences between monoterpene and sesquiterpene cyclases, between sesquiterpene cyclases from different kingdoms of life, and between individual cyclases. These differences are outlined below.

1.4.3 Divalent metal ion coordination

In general, most terpene cyclases require three divalent metal ions to assist in pyrophosphate ionization and subsequent catalysis. In the first published crystal structure of a terpene cyclase, 5-epi-aristolochene synthase (TEAS), Mg_A^{2+} and Mg_B^{2+} bind in the unliganded enzyme and Mg_C^{2+} additionally binds in the presence of the substrate analog, farnesyl hydroxyphosphonate (FHP)[75]. The first two Mg^{2+} ions coordinate with octahedral geometry to the DDXXD motif and NSE/DTE motif, respectively; the third Mg^{2+} ion binds in close proximity to Mg_A^{2+} and, in the presence of FHP, also coordinates with octahedral geometry to the **D**DXX**D** motif, the phosphate moiety, and several water molecules[75] (Figure 1.10).

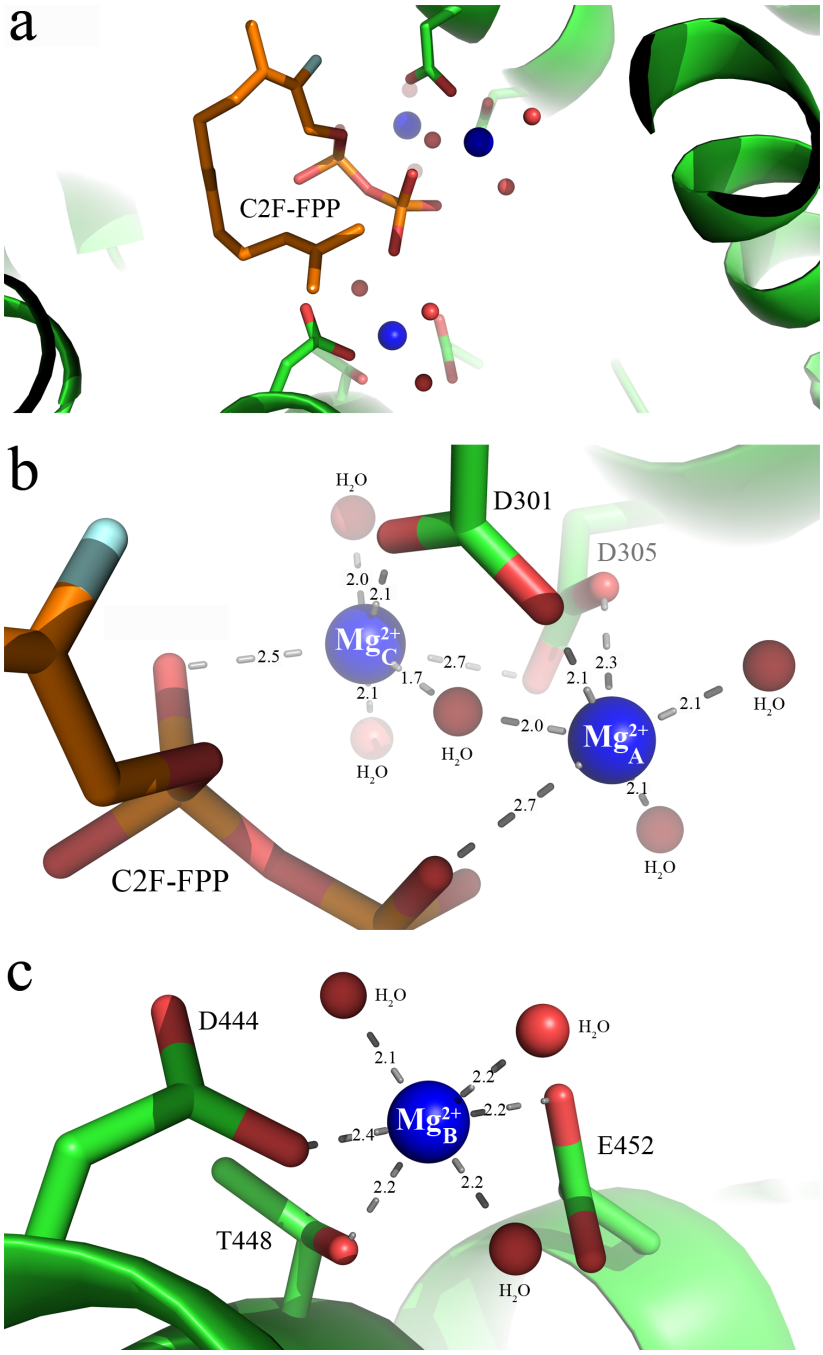


Figure 1.10. Magnesium ion coordination in the active site of 5-epi-aristolochene synthase complexed with magnesium and the fluorinated substrate analog, C2F-FPP (pdb ID: 3M00[52]). a) Overview of the active site. b) close-up view of coordination sites for Mg_A^{2+} and Mg_C^{2+}. c) close-up view of coordination sites for Mg_B^{2+}.

This example is one of many variations on what is observed for divalent metal ion coordination in the active site of a terpene cyclase. For example, in the unliganded structure of (+)-bornyl diphosphate synthase, only one magnesium ion is coordinated to the **D**DXX**D** motif and the other two are missing, whereas the substrate-analog bound structure shows three Mg^{2+} ions in locations that are consistent with what is observed for TEAS[77]. In comparing various crystal structures of aristolochene synthase from *A. terreus*, monomer D shows either one or two Mg^{2+} ions bound (either Mg_B^{2+} or Mg_B^{2+} and Mg_C^{2+}) in the presence of pyrophosphate or substrate analog; however, the other three monomers show only substrate analog with no accompanying divalent metal ion coordination[83]. Such monomeric differences are thought to represent snapshots of various phases of the terpene cyclase reaction; however, these results also highlight that Mg_B^{2+} plays a very important role in properly orienting the pyrophosphate moiety of the substrate for catalysis[83]. Fungal trichodiene synthase and bacterial epi-isozizaene synthase both coordinate three divalent magnesium ions, however in contrast to structures of plant terpene cyclases which coordinate Mg_A^{2+} and Mg_C^{2+} with the first and last aspartic acid in the DDXXD motif, these two enzymes only coordinate with the first aspartic acid[82, 85]. Additionally, the second aspartic acid of the motif plays a role in the hydrogen-bonding network between the substrate and surrounding residues, and mutation at this position causes significant loss of activity[85, 88]. Notably, bacterial and fungal terpene cyclases usually contain an NSE motif (instead of a DTE motif as seen in most plant terpene cyclases); thus, divalent metal ion coordination in terpene cyclases appears to have evolved slightly differently in plants compared to fungi and bacteria. The most interesting example of metal ion coordination is in δ-cadinene synthase. This sesquiterpene cyclase contains the conventional DDXXD motif that binds Mg_A^{2+} and Mg_C^{2+}, however it is missing the highly conserved NSE/DTE motif and instead contains another DDXXD/E motif that coordinates the

Mg_B^{2+} ion[81]. Both δ-selinene synthase and γ-humulene synthase also contain this additional DDXXD motif[57]. This second motif corresponds to SARM in short-chain prenyl diphosphate synthases.

1.4.4. Ligand-induced structural changes

In general, upon substrate (or pyrophosphate) binding, terpene cyclases close their active sites to accommodate the ligand, to exclude water, and to initiate pyrophosphate ionization.[82] Plant terpene cyclases adopt more subtle changes upon ligand binding compared to fungal and bacterial terpene cyclases[81, 82, 85] For example, superposition of apo and ligand-bound structures of 5-epi-aristolochene synthase from tobacco (TEAS) generates a root mean square deviation (rmsd) for C^{α} atoms of 0.43Å[75] while a similar superposition in fungal trichodiene synthase generates an rmsd for C^{α} atoms of 1.4Å.[82] Some plant terpene cyclase structures show ordering of the following motifs when complexed with ligand: the A-C loop, the J-K loop, part of helix H, and the amino-terminus[75, 77, 79]. Others, such as δ-cadinene synthase from cotton, do not demonstrate any such conformational changes[81]; In contrast, fungal and bacterial crystal structures show a large degree of movement in several or all of the following motifs: helices 1, D, H, J, K, L, and loops 1-A, D-D1, F-G, H-α1, J-K, and K-L.[82, 85, 89] Fungal and bacterial structures may undergo such drastic conformational changes to compensate for the fact that they lack the amino-terminal domain that the plant enzymes have to protect the active site from highly reactive water.[81] Additionally, fungal and bacterial terpene cyclases usually produce one specific product compared to those of plant origin, suggesting that they may adopt a more rigid active site contour on which to template the substrate[83].

1.4.5. Substrate analogs

Ongoing efforts are aimed towards complexation of terpene cyclases with substrate-like and reaction-like mimics and/or inhibitors to gain insight on each terpene cyclase reaction mechanism. Complexes of monoterpene cyclases with GPP, linalyl diphosphate (the isomerized version of GPP), and various carbocation mimics have been reported. In the case of (+)-bornyl diphosphate synthase, a variety of aza analogs were synthesized and complexed with the enzyme to mimic the carbocation intermediates generated throughout the reaction; these mimics were somewhat successful, although the geometry at the nitrogen in the aza analog is different than that of the planar carbocation center. A similar result is observed in epi-isozizaene synthase complexed with the benzyl triethylammonium cation (BTAC), which is meant to mimic the bisabolyl cation (the first cation formed in the mechanism) but also has different geometry than the naturally occurring carbocation intermediate. Fluorinated substrate analogs, including 2-fluoro-GPP (2F-GPP), 2-fluoro-FPP (2F-FPP), and 12,13-fluoro-FPP (difluoro-FPP, or DF-FPP) are most commonly used as substrate mimics for terpene cyclases[52, 77, 89]. These analogs are usually non-hydrolyzable due to the presence of the fluorine atom, which withdraws negative charge from the proximal carbon-carbon double bond via the inductive effect, rendering (in most cases) inability for pyrophosphate ionization and electrophilic cyclization. Aristolochene synthase is the exception since it is able to hydrolyze 2F-FPP into a stable intermediate, 2-fluorogermacrene A (but cannot complete the reaction to generate aristolochene)[89]. In some cases, the electron density for the isoprenoid tail of the substrate or substrate analog is less clearly defined, which is most likely a reflection of a more dynamic substrate. For example, the electron density for the isoprenoid moiety of nearly any given FPP analog is much more clearly defined in the active site of aristolochene synthase compared to TEAS,[75, 89, 90] which is perhaps correlated to the fact that the former produces

aristolochene exclusively[91], while the latter produces a variety of minor products in addition to 5-epi-aristolochene.[55]

1.5. Emergence of terpene synthases from primary metabolism

There are several theories on the evolution of terpene cyclases. Based on intron/exon organization in plant angiosperms and gymnosperms, one theory suggests that the ancestral terpene synthase was a diterpene synthase of primary metabolism (such as KS or CDS) that underwent gene duplication and divergent evolution to create the present day plant terpene cyclases.[64] Due to lack of sequence similarity, large differences in intron/exon organization, and large phylogenetic distances between clades, microbial and plant terpene synthases were thought to have undergone convergent evolution.[64] More recently, however, a theory that incorporates a hierarchy of levels of evolution involving triterpene synthases, bacterial diterpene synthases, and eventually plant diterpene synthases, has come into view.[92] This theory, based on structural, functional, and sequence comparisons, suggests that the first bacterial class I diterpene cyclases were created from the ancient triterpene synthase (containing the DXDD motif and performing the proton-initiated reaction), while the first bacterial class II diterpene cyclase was created from an ancestor of the class II terpene cyclase fold (containing the DDXXD motif and performing the ionization-dependent reaction). Class I and class II diterpene cyclase domains then fused together to create modern day plant diterpene cyclases, which eventually, through the loss of several exons, evolved into present day plant monoterpene and sesquiterpene synthases.[92] This theory speculates that all terpene cyclases were derived from bacterial ancestors, and that bacteria eventually transferred these genes to plants.[92]

1.6. Conclusions

Isoprenoid biosynthesis constitutes a network of biosynthetic pathways that spans primary and secondary metabolism. In primary metabolism, the MVA pathway and DXP pathway produce the two essential building blocks for biosynthesis of all downstream isoprenoids: IPP and DMAPP. The MVA and DXP pathways are not as well understood as once thought, especially in archaea where some enzymes in the mevalonate pathway have still not been identified. Work discussed in chapters 5 and 6 focuses on resolving such issues by using the crystal structure of an archaeal kinase as a starting point towards the discovery of MVA pathway alternatives both within and outside of this domain of life.

The short-chain prenyl diphosphate synthases represent a family of enzymes that bridges the primary and secondary metabolic pathways of isoprenoid biosynthesis. Using the two essential IPP and DMAPP building blocks, these enzymes synthesize GPP, FPP, and GGPP, which are then substrates for all downstream primary and secondary metabolic enzymes in this pathway, including the monoterpene, sesquiterpene, and diterpene synthases of secondary metabolism, respectively.

In the case of all terpene cyclases, the idea that such product diversity can be created from one substrate is fascinating and has been explored here in three different ways. Chapter 1 addresses how a chemical profile can change throughout the landscape of sequence space that exists between two sesquiterpene cyclases. Chapter 2 analyzes both structural and functional effects associated with both substrate and product promiscuity among wild type TEAS and a promiscuous mutant. Chapter 3 discusses how an extensive mutational analysis at an amino terminal motif in patchoulol synthase (PAS) has demonstrated its importance in maintaining a chemically complex product profile.

REFERENCES

1. Novakova, Z.; Surin, S.; Blasko, J.; Majernik, A.; Smigan, P., Membrane proteins and squalene-hydrosqualene profile in methanoarchaeon Methanothermobacter thermautotrophicus resistant to N,N'-dicyclohexylcarbodiimide. *Folia microbiologica* **2008,** *53* (3), 237-240.

2. Ourisson, G.; Rohmer, M.; Poralla, K., Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annual Review of Microbiology* **1987,** *41*, 301-333.

3. Eichler, J.; Adams, M. W., Posttranslational protein modification in Archaea. *Microbiology and molecular biology reviews : MMBR* **2005,** *69* (3), 393-425.

4. Bartley, G. E.; Scolnik, P. A., Plant carotenoids: pigments for photoprotection, visual attraction, and human health. *Plant Cell* **1995,** *7* (7), 1027-38.

5. Trumpower, B. L., New concepts on the role of ubiquinone in the mitochondrial respiratory chain. *J Bioenerg Biomembr* **1981,** *13* (1-2), 1-24.

6. Hedden, P.; Kamiya, Y., GIBBERELLIN BIOSYNTHESIS: Enzymes, Genes and Their Regulation. *Annu Rev Plant Physiol Plant Mol Biol* **1997,** *48*, 431-460.

7. Arigoni, D.; Sagner, S.; Latzel, C.; Eisenreich, W.; Bacher, A.; Zenk, M. H., Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement. *Proc Natl Acad Sci U S A* **1997,** *94* (20), 10600-5.

8. Eisenreich, W.; Schwarz, M.; Cartayrade, A.; Arigoni, D.; Zenk, M. H.; Bacher, A., The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chemistry & biology* **1998,** *5* (9), R221-33.

9. Rohmer, M., The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural product reports* **1999,** *16* (5), 565-574.

10. Sprenger, G. A.; Schorken, U.; Wiegert, T.; Grolle, S.; de Graaf, A. A.; Taylor, S. V.; Begley, T. P.; Bringer-Meyer, S.; Sahm, H., Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proc Natl Acad Sci U S A* **1997,** *94* (24), 12857-62.

11. Rohdich, F.; Zepeck, F.; Adam, P.; Hecht, S.; Kaiser, J.; Laupitz, R.; Grawert, T.; Amslinger, S.; Eisenreich, W.; Bacher, A.; Arigoni, D., The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proc Natl Acad Sci U S A* **2003,** *100* (4), 1586-91.

12. Nyland, R. L., 2nd; Xiao, Y.; Liu, P.; Freel Meyers, C. L., IspG converts an epoxide substrate analogue to (E)-4-hydroxy-3-methylbut-2-enyl diphosphate: implications for IspG catalysis in isoprenoid biosynthesis. *J Am Chem Soc* **2009,** *131* (49), 17734-5.
13. Grawert, T.; Span, I.; Eisenreich, W.; Rohdich, F.; Eppinger, J.; Bacher, A.; Groll, M., Probing the reaction mechanism of IspH protein by x-ray structure analysis. *Proc Natl Acad Sci U S A* **2010,** *107* (3), 1077-81.

14. Brown, M. S.; Dana, S. E.; Goldstein, J. L., Regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity in human fibroblasts by lipoproteins. *Proc Natl Acad Sci U S A* **1973,** *70* (7), 2162-6.

15. Furberg, C. D.; Adams, H. P., Jr.; Applegate, W. B.; Byington, R. P.; Espeland, M. A.; Hartwell, T.; Hunninghake, D. B.; Lefkowitz, D. S.; Probstfield, J.; Riley, W. A.; et al., Effect of lovastatin on early carotid atherosclerosis and cardiovascular events. Asymptomatic Carotid Artery Progression Study (ACAPS) Research Group. *Circulation* **1994,** *90* (4), 1679-87.

16. Sapir-Mir, M.; Mett, A.; Belausov, E.; Tal-Meshulam, S.; Frydman, A.; Gidoni, D.; Eyal, Y., Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. *Plant Physiol* **2008,** *148* (3), 1219-28.

17. Carrero-Lerida, J.; Perez-Moreno, G.; Castillo-Acosta, V. M.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D., Intracellular location of the early steps of the isoprenoid biosynthetic pathway in the trypanosomatids Leishmania major and Trypanosoma brucei. *Int J Parasitol* **2009,** *39* (3), 307-14.

18. Hartman, I. Z.; Liu, P.; Zehmer, J. K.; Luby-Phelps, K.; Jo, Y.; Anderson, R. G.; DeBose-Boyd, R. A., Sterol-induced dislocation of 3-hydroxy-3-methylglutaryl coenzyme A reductase from endoplasmic reticulum membranes into the cytosol through a subcellular compartment resembling lipid droplets. *J Biol Chem* **2010,** *285* (25), 19288-98.

19. Smit, A.; Mushegian, A., Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome research* **2000,** *10* (10), 1468-1484.

20. Grochowski, L. L.; Xu, H.; White, R. H., Methanocaldococcus jannaschii uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *Journal of Bacteriology* **2006,** *188* (9), 3192-3198.

21. Ogura, K.; Koyama, T., Enzymatic Aspects of Isoprenoid Chain Elongation. *Chem Rev* **1998,** *98* (4), 1263-1276.

22. Burke, C. C.; Wildung, M. R.; Croteau, R., Geranyl diphosphate synthase: cloning, expression, and characterization of this prenyltransferase as a heterodimer. *Proc Natl Acad Sci U S A* **1999,** *96* (23), 13062-7.

23. Tarshis, L. C.; Yan, M.; Poulter, C. D.; Sacchettini, J. C., Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-A resolution. *Biochemistry* **1994,** *33* (36), 10871-7.

24. Tarshis, L. C.; Proteau, P. J.; Kellogg, B. A.; Sacchettini, J. C.; Poulter, C. D., Regulation of product chain length by isoprenyl diphosphate synthases. *Proc Natl Acad Sci U S A* **1996,** *93* (26), 15018-23.

25. Wang, K.; Ohnuma, S., Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends Biochem Sci* **1999,** *24* (11), 445-51.

26. Hosfield, D. J.; Zhang, Y.; Dougan, D. R.; Broun, A.; Tari, L. W.; Swanson, R. V.; Finn, J., Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J Biol Chem* **2004,** *279* (10), 8526-9.

27. Ohnuma, S.; Narita, K.; Nakazawa, T.; Ishida, C.; Takeuchi, Y.; Ohto, C.; Nishino, T., A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. *J Biol Chem* **1996,** *271* (48), 30748-54.

28. Lee, P. C.; Petri, R.; Mijts, B. N.; Watts, K. T.; Schmidt-Dannert, C., Directed evolution of Escherichia coli farnesyl diphosphate synthase (IspA) reveals novel structural determinants of chain length specificity. *Metab Eng* **2005,** *7* (1), 18-26.

29. Ohnuma, S.; Hirooka, K.; Ohto, C.; Nishino, T., Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase. Two amino acids before the first aspartate-rich motif solely determine eukaryotic farnesyl diphosphate synthase activity. *J Biol Chem* **1997,** *272* (8), 5192-8.

30. Gershenzon, J.; Dudareva, N., The function of terpene natural products in the natural world. *Nature chemical biology* **2007,** *3* (7), 408-414.

31. Pichersky, E.; Lewinsohn, E.; Croteau, R., Purification and characterization of S-linalool synthase, an enzyme involved in the production of floral scent in Clarkia breweri. *Arch Biochem Biophys* **1995,** *316* (2), 803-7.

32. Kollner, T. G.; Held, M.; Lenk, C.; Hiltpold, I.; Turlings, T. C.; Gershenzon, J.; Degenhardt, J., A maize (E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell* **2008,** *20* (2), 482-94.

33. Hezari, M.; Lewis, N. G.; Croteau, R., Purification and characterization of taxa-4(5),11(12)-diene synthase from Pacific yew (Taxus brevifolia) that catalyzes the first committed step of taxol biosynthesis. *Archives of Biochemistry and Biophysics* **1995,** *322* (2), 437-444.

34. Bouwmeester, H. J.; Wallaart, T. E.; Janssen, M. H.; van Loo, B.; Jansen, B. J.; Posthumus, M. A.; Schmidt, C. O.; De Kraker, J. W.; Konig, W. A.; Franssen, M. C., Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis. *Phytochemistry* **1999,** *52* (5), 843-54.

35. Kirby, J.; Keasling, J. D., Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu Rev Plant Biol* **2009,** *60*, 335-55.

36. Asadollahi, M. A.; Maury, J.; Moller, K.; Nielsen, K. F.; Schalk, M.; Clark, A.; Nielsen, J., Production of plant sesquiterpenes in Saccharomyces cerevisiae: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol Bioeng* **2008,** *99* (3), 666-77.

37. Ohto, C.; Muramatsu, M.; Obata, S.; Sakuradani, E.; Shimizu, S., Overexpression of the gene encoding HMG-CoA reductase in Saccharomyces cerevisiae for production of prenyl alcohols. *Appl Microbiol Biotechnol* **2009,** *82* (5), 837-45.

38. Morrone, D.; Lowry, L.; Determan, M. K.; Hershey, D. M.; Xu, M.; Peters, R. J., Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl Microbiol Biotechnol* **2010,** *85* (6), 1893-906.

39. Martin, V. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D., Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature biotechnology* **2003,** *21* (7), 796-802.

40. Pitera, D. J.; Paddon, C. J.; Newman, J. D.; Keasling, J. D., Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metabolic engineering* **2007,** *9* (2), 193-207.

41. Fu, Z.; Voynova, N. E.; Herdendorf, T. J.; Miziorko, H. M.; Kim, J. J., Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism. *Biochemistry* **2008,** *47* (12), 3715-24.

42. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R., Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **1998,** *95*, 4126-4133.

43. Cane, D. E.; Xue, Q.; Fitzsimons, B. C., Trichodiene synthase. Probing the role of the highly conserved aspartate-rich region by site-directed mutagenesis. *Biochemistry* **1996,** *35* (38), 12369-12376.

44. Christianson, D. W., Structural biology and chemistry of the terpenoid cyclases. *Chemical reviews* **2006,** *106* (8), 3412-3442.

45. McGarvey, D. J.; Croteau, R., Terpenoid metabolism. *Plant Cell* **1995,** *7* (7), 1015-26.

46. Cane, D. E.; Kang, I., Aristolochene synthase: purification, molecular cloning, high-level expression in Escherichia coli, and characterization of the Aspergillus terreus cyclase. *Archives of Biochemistry and Biophysics* **2000,** *376* (2), 354-364.

47. Williams, D. C.; McGarvey, D. J.; Katahira, E. J.; Croteau, R., Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* **1998,** *37* (35), 12213-12220.

48. Rajaonarivony, J. I.; Gershenzon, J.; Croteau, R., Characterization and mechanism of (4S)-limonene synthase, a monoterpene cyclase from the glandular trichomes of peppermint (Mentha x piperita). *Arch Biochem Biophys* **1992,** *296* (1), 49-57.

49. Degenhardt, J.; Kollner, T. G.; Gershenzon, J., Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **2009,** *70* (15-16), 1621-37.

50. Schilmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E., Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc Natl Acad Sci U S A* **2009,** *106* (26), 10865-70.

51. Sallaud, C.; Rontein, D.; Onillon, S.; Jabes, F.; Duffe, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; Causse, M.; Tissier, A., A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato Solanum habrochaites. *Plant Cell* **2009,** *21* (1), 301-17.

52. Noel, J. P.; Dellas, N.; Faraldos, J. A.; Zhao, M.; Hess, B. A., Jr.; Smentek, L.; Coates, R. M.; O'Maille, P. E., Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. *ACS chemical biology* **2010,** *5* (4), 377-392.

53. Picaud, S.; Olofsson, L.; Brodelius, M.; Brodelius, P. E., Expression, purification, and characterization of recombinant amorpha-4,11-diene synthase from Artemisia annua L. *Arch Biochem Biophys* **2005,** *436* (2), 215-26.

54. Thulasiram, H. V.; Poulter, C. D., Farnesyl diphosphate synthase: The art of compromise between substrate selectivity and stereoselectivity. *Journal of the American Chemical Society* **2006,** *128* (49), 15819-15823.

55. O'Maille, P. E.; Chappell, J.; Noel, J. P., Biosynthetic potential of sesquiterpene synthases: Alternative products of tobacco 5-epi-aristolochene synthase. *Archives of Biochemistry and Biophysics* **2006,** *448* (1-2), 73-82.

56. Cane, D. E., How to evolve a silk purse from a sow's ear. *Nat Chem Biol* **2006,** *2* (4), 179-80.

57. Little, D. B.; Croteau, R. B., Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases delta-selinene synthase and gamma-humulene synthase. *Archives of Biochemistry and Biophysics* **2002,** *402* (1), 120-135.

58. Deguerry, F.; Pastore, L.; Wu, S.; Clark, A.; Chappell, J.; Schalk, M., The diverse sesquiterpene profile of patchouli, Pogostemon cablin, is correlated with a limited number of sesquiterpene synthases. *Archives of Biochemistry and Biophysics* **2006,** *454* (2), 123-136.

59. Yoshikuni, Y.; Keasling, J. D., Pathway engineering by designed divergent evolution. *Curr Opin Chem Biol* **2007,** *11* (2), 233-9.

60. O'Maille, P. E.; Malone, A.; Dellas, N.; Andes Hess, B., Jr.; Smentek, L.; Sheehan, I.; Greenhagen, B. T.; Chappell, J.; Manning, G.; Noel, J. P., Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nature chemical biology* **2008,** *4* (10), 617-623.

61. Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D., Designed divergent evolution of enzyme function. *Nature* **2006,** *440* (7087), 1078-1082.

62. Greenhagen, B. T.; O'Maille, P. E.; Noel, J. P.; Chappell, J., Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (26), 9826-9831.

63. Lin, X.; Hezari, M.; Koepp, A. E.; Floss, H. G.; Croteau, R., Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of taxol biosynthesis in Pacific yew. *Biochemistry* **1996,** *35* (9), 2968-77.

64. Trapp, S. C.; Croteau, R. B., Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **2001,** *158* (2), 811-832.

65. Sun, T. P.; Kamiya, Y., The Arabidopsis GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis. *Plant Cell* **1994,** *6* (10), 1509-18.

66. Morrone, D.; Chambers, J.; Lowry, L.; Kim, G.; Anterola, A.; Bender, K.; Peters, R. J., Gibberellin biosynthesis in bacteria: separate ent-copalyl diphosphate and ent-kaurene synthases in Bradyrhizobium japonicum. *FEBS Lett* **2009,** *583* (2), 475-80.

67. Saito, T.; Abe, H.; Yamane, H.; Sakurai, A.; Murofushi, N.; Takio, K.; Takahashi, N.; Kamiya, Y., Purification and Properties of ent-Kaurene Synthase B from Immature Seeds of Pumpkin. *Plant Physiol* **1995,** *109* (4), 1239-1245.

68. Hayashi, K.; Kawaide, H.; Notomi, M.; Sakigi, Y.; Matsuo, A.; Nozaki, H., Identification and functional analysis of bifunctional ent-kaurene synthase from the moss Physcomitrella patens. *FEBS Lett* **2006,** *580* (26), 6175-81.

69. Toyomasu, T.; Kawaide, H.; Ishizaki, A.; Shinoda, S.; Otsuka, M.; Mitsuhashi, W.; Sassa, T., Cloning of a full-length cDNA encoding ent-kaurene synthase from Gibberella fujikuroi: functional analysis of a bifunctional diterpene cyclase. *Bioscience, biotechnology, and biochemistry* **2000,** *64* (3), 660-664.

70. Vogel, B. S.; Wildung, M. R.; Vogel, G.; Croteau, R., Abietadiene synthase from grand fir (Abies grandis). cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis. *J Biol Chem* **1996,** *271* (38), 23262-8.

71. Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B., Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J Am Chem Soc* **2001,** *123* (37), 8974-8.

72. Peters, R. J.; Croteau, R. B., Abietadiene synthase catalysis: conserved residues involved in protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate. *Biochemistry* **2002,** *41* (6), 1836-42.

73. Xu, M.; Hillwig, M. L.; Prisic, S.; Coates, R. M.; Peters, R. J., Functional identification of rice syn-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. *Plant J* **2004,** *39* (3), 309-18.

74. Peters, R. J.; Carter, O. A.; Zhang, Y.; Matthews, B. W.; Croteau, R. B., Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. *Biochemistry* **2003,** *42* (9), 2700-7.

75. Starks, C. M.; Back, K.; Chappell, J.; Noel, J. P., Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **1997,** *277* (5333), 1815-1820.

76. Wendt, K. U.; Schulz, G. E., Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure (London, England : 1993)* **1998,** *6* (2), 127-133.

77. Whittington, D. A.; Wise, M. L.; Urbansky, M.; Coates, R. M.; Croteau, R. B.; Christianson, D. W., Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (24), 15375-15380.

78. Lesburg, C. A.; Zhai, G.; Cane, D. E.; Christianson, D. W., Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science (New York, N.Y.)* **1997,** *277* (5333), 1820-1824.

79. Hyatt, D. C.; Youn, B.; Zhao, Y.; Santhamma, B.; Coates, R. M.; Croteau, R. B.; Kang, C., Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (13), 5360-5365.

80. Kampranis, S. C.; Ioannidis, D.; Purvis, A.; Mahrez, W.; Ninga, E.; Katerelos, N. A.; Anssour, S.; Dunwell, J. M.; Degenhardt, J.; Makris, A. M.; Goodenough, P. W.; Johnson, C. B., Rational conversion of substrate and product specificity in a Salvia

monoterpene synthase: structural insights into the evolution of terpene synthase function. *The Plant Cell* **2007,** *19* (6), 1994-2005.

81. Gennadios, H. A.; Gonzalez, V.; Di Costanzo, L.; Li, A.; Yu, F.; Miller, D. J.; Allemann, R. K.; Christianson, D. W., Crystal structure of (+)-delta-cadinene synthase from Gossypium arboreum and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry* **2009,** *48* (26), 6175-83.

82. Rynkiewicz, M. J.; Cane, D. E.; Christianson, D. W., Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98* (24), 13543-13548.

83. Shishova, E. Y.; Di Costanzo, L.; Cane, D. E.; Christianson, D. W., X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* **2007,** *46* (7), 1941-1951.

84. Caruthers, J. M.; Kang, I.; Rynkiewicz, M. J.; Cane, D. E.; Christianson, D. W., Crystal structure determination of aristolochene synthase from the blue cheese mold, Penicillium roqueforti. *The Journal of biological chemistry* **2000,** *275* (33), 25533-25539.

85. Aaron, J. A.; Lin, X.; Cane, D. E.; Christianson, D. W., Structure of epi-isozizaene synthase from Streptomyces coelicolor A3(2), a platform for new terpenoid cyclization templates. *Biochemistry* **2010,** *49* (8), 1787-97.

86. Reardon, D.; Farber, G. K., The structure and evolution of alpha/beta barrel proteins. *FASEB J* **1995,** *9* (7), 497-503.

87. Vedula, L. S.; Jiang, J.; Zakharian, T.; Cane, D. E.; Christianson, D. W., Structural and mechanistic analysis of trichodiene synthase using site-directed mutagenesis: probing the catalytic function of tyrosine-295 and the asparagine-225/serine-229/glutamate-233-Mg2+B motif. *Arch Biochem Biophys* **2008,** *469* (2), 184-94.

88. Lin, X.; Cane, D. E., Biosynthesis of the sesquiterpene antibiotic albaflavenone in Streptomyces coelicolor. Mechanism and stereochemistry of the enzymatic formation of epi-isozizaene. *J Am Chem Soc* **2009,** *131* (18), 6332-3.

89. Shishova, E. Y.; Yu, F.; Miller, D. J.; Faraldos, J. A.; Zhao, Y.; Coates, R. M.; Allemann, R. K.; Cane, D. E.; Christianson, D. W., X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *J Biol Chem* **2008,** *283* (22), 15431-9.

90. Shishova, E. Y.; Di Costanzo, L.; Cane, D. E.; Christianson, D. W., X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* **2007,** *46* (7), 1941-51.

91. Felicetti, B.; Cane, D. E., Aristolochene synthase: mechanistic analysis of active site residues by site-directed mutagenesis. *J Am Chem Soc* **2004,** *126* (23), 7212-21.

92. Cao, R.; Zhang, Y.; Mann, F. M.; Huang, C.; Mukkamala, D.; Hudock, M. P.; Mead, M. E.; Prisic, S.; Wang, K.; Lin, F. Y.; Chang, T. K.; Peters, R. J.; Oldfield, E., Diterpene cyclases and the nature of the isoprene fold. *Proteins* **2010,** *78* (11), 2417-32.

Chapter 2

Quantitative Exploration of the Catalytic Landscape Separating Divergent Plant Sesquiterpene Synthases

2.1. ABSTRACT

Throughout molecular evolution, organisms create assorted chemicals in response to varying ecological niches. Catalytic landscapes underlie metabolic evolution, wherein mutational steps alter the biosynthetic properties of enzymes. We report the first systematic quantitative characterization of a catalytic landscape underlying the evolution of sesquiterpene chemical diversity. Based on our previous discovery of a set of 9 naturally occurring amino acid substitutions that functionally inter-converted orthologous sesquiterpene synthases from *Nicotiana tabacum* and *Hyoscyamus muticus*, we created a library of all possible residue combinations (29 = 512) in the N. tabacum parent. The product spectra of 418 active enzymes to reveal a rugged landscape where several minimal combinations of the 9 mutations encode convergent solutions to the inter-conversions of parental activities. Quantitative comparisons indicate context dependence for mutational effects - epistasis - in product specificity and promiscuity. These results provide a measure of the mutational accessibility of phenotypic variability among a diverging lineage of terpene synthases.

2.2. INTRODUCTION

The acquisition of innovative metabolic activities is a major force shaping evolutionary change, but one that is poorly understood. Metabolic adaptability is particularly crucial for a plant's capacity to synthesize specialized chemicals affording protection against microbial pathogens[1-3], elicitation of symbiotic relationships[4], attractive[5] and repellent[6] activities and physiological responses to local environments (as reviewed[7]). Understanding the evolution of metabolic function at the molecular level requires knowledge of the distribution of enzymatic properties through accessible mutational changes, and hence defining the catalytic landscape. Currently, there is no reported measurement of the catalytic landscapes

underlying secondary (specialized) metabolism. The physicochemical constraints relating sequence variation and metabolic output raise important fundamental questions concerning catalytic complexity and natural selection. For instance, how does a particular biosynthetic property like catalytic efficiency or substrate/product specificity vary amongst extant and probable ancestral sequences? Moreover, how likely is the emergence of new product specificities in a family of diverging biosynthetic enzymes?

To experimentally approach these questions in the current work, we relied upon i.) a model system composed of a pair of closely-related secondary metabolic enzymes, ii.) a simplified set of naturally occurring mutations which interconvert a defined catalytic property that is functionally divergent between the parental sequences, iii.) structure-based combinatorial protein engineering (SCOPE)[8] to provide a means of creating an enzyme lineage representing putative evolutionary intermediates in one parental background, and iv.) a gas chromatography-mass spectrometry (GC-MS) method[9] for measuring the catalytic properties (recording the chemical readout) of the enzyme library. Therefore, quantitative comparison of catalytic properties of enzymes across these simplified and probable lineages provides a direct measure of functional variation likely correlated with phenotypic variation in the defense response of parental species. Moreover, this comprehensive study explores the mutational accessibility of alternative biosynthetic properties over this experimentally defined region of sequence space.

Terpene synthases of secondary metabolism are a diverse enzyme family responsible for the biosynthesis of complex chemicals. Tobacco 5-epi-aristolochene synthase (TEAS) and henbane premnaspirodiene synthase (HPS) are closely related (75% amino acid identity) enzymes yet cyclize ionized farnesyl diphosphate (FPP, 1) to form 5-epi-aristolochene (5-EA, 2) and premnaspirodiene (PSD, 3), respectively. These structurally distinct terpene

hydrocarbons are precursors of antifungal phytoalexins in solanaceous plants[10, 11]. Mechanistically, TEAS and HPS share a common carbocation intermediate during an electrophilic cyclization reaction, but differ in directing either a methyl or methylene migration, respectively (Figure 1 panel a). Density functional theory calculations indicate both of these alkyl shifts to be endothermic (~3 kcal/mol), with the methylene shift's transition state of lower energy (Figure 2.1).

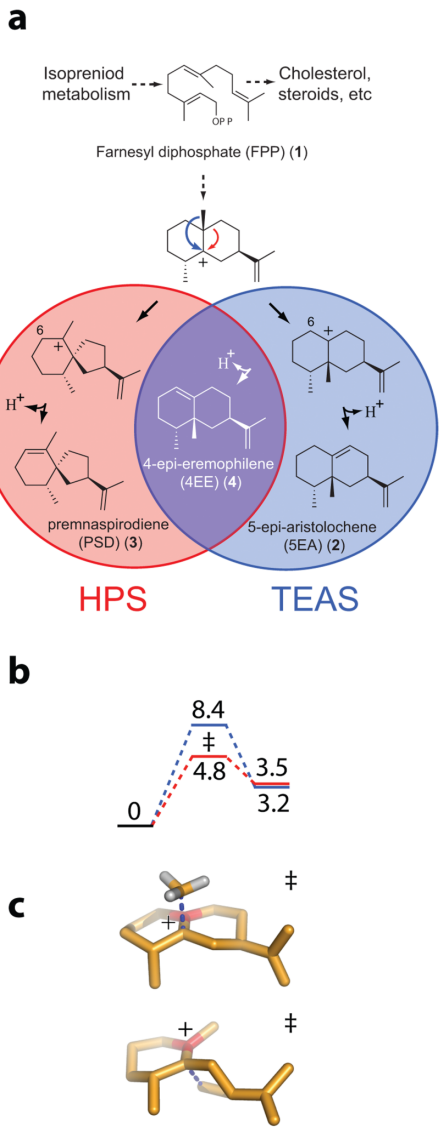


Figure 2.1. Terminal cyclization steps of TEAS and HPS terpene synthases. (a) TEAS and HPS exert differential conformational control on a common carbocation intermediate to produce 5-EA and PSD. The discovery of 4-EE biosynthetic activity supports hybridization of the final two biosynthetic steps in TEAS and HPS, involving a methyl migration shared with TEAS and a final deprotonation at C6 shared with HPS. (b) Proposed reaction coordinate for the methyl (blue) and alkyl (red) migrations extending from a common carbocation intermediate (defined as zero energy) through a transition state (‡), leading to the penultimate carbocations of their respective reaction pathways. Calculated energies are expressed in units of kcal mol-1. (c) Conformations of the methyl (top) or alkyl (bottom) migration transition states as calculated from density functional theory calculations. Carbon atoms are shown in gold and the carbocation center (+) in red. Dashed blue lines indicate newly forming bonds. Hydrogen atoms are omitted for clarity.

We previously used a structure-guided approach to identify a functionally linked subset of 9 amino acid residues from the 135 naturally occurring amino acid differences between TEAS and HPS (Figure 2.2). Mutational swaps of this nine-residue subset are sufficient to interconvert the product specificities of the encoded mutant proteins in the background of each parent enzyme[12]. Eight of these nine amino acid substitutions are achievable by single nonsynonymous nucleotide changes per codon (Figure 2.2, panel a). However, position 406 (TEAS numbering) requires a two-base change to interconvert between Tyr and Leu in TEAS and HPS, respectively, suggesting the possible intermediacy of Phe at this position in a common ancestor. Notably, only two of the nine amino acid differences are localized on the active-site surface, whereas the remainder are scattered throughout the second tier (Figure 2.2, panel d). Replacing these two active-site residues of TEAS with their HPS counterparts redirects the final deprotonation-neutralization step to produce 4-epi-eremophilene (4-EE, Compound 4), a terpene not previously identified in nature[12]. Thus, the resulting 4-epi-eremophilene synthase (EES) represents an intermediate enzyme with hybrid parental activities (Figure 2.1, panel a).

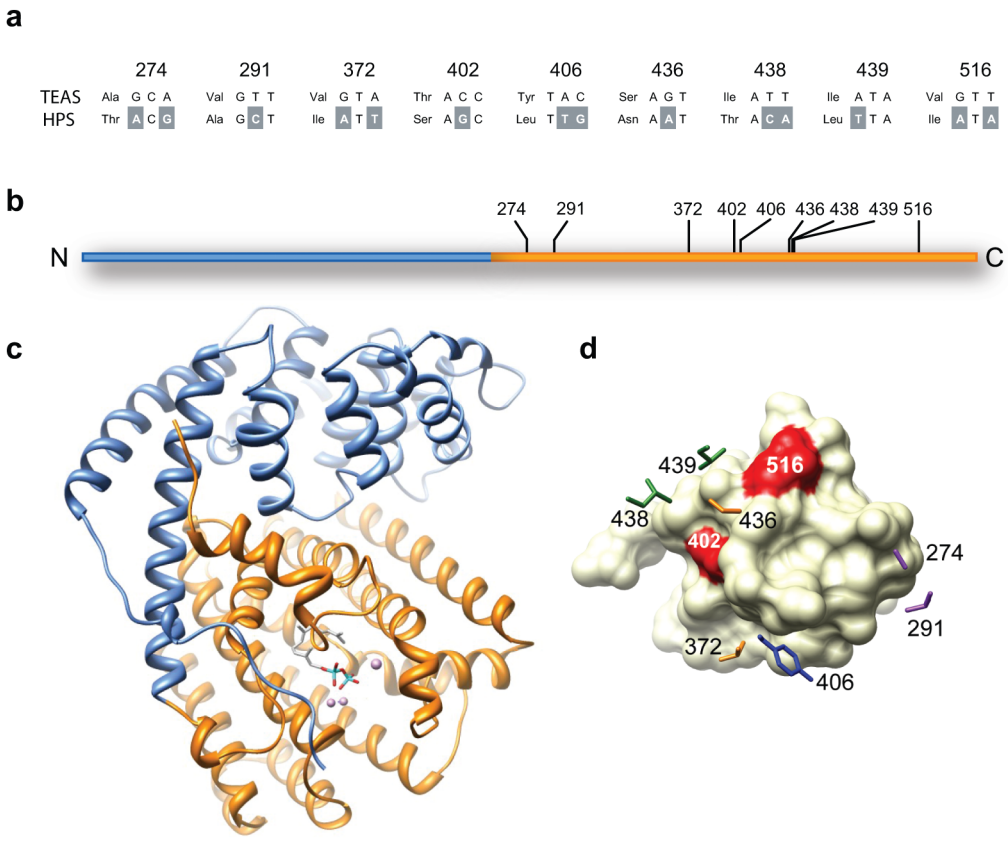


Figure 2.2. Overall structure of TEAS and location and identity of M9 residues. (a) Nucleotide and amino acid identity of substitutions between TEAS and HPS. Shading indicates nucleotide substitutions in HPS relative to the TEAS reference sequence. (b) The primary structure is composed of N-terminal (blue) and C-terminal (gold) terpenoid synthase domains. Amino acid positions of the M9 library are indicated using TEAS numbering. (c) Tertiary structure of TEAS (PDB ID 5eat) shown as ribbons, with domains colored as in a and Mg2+ and farnesyl diphosphate modeled into the active site. (d) Spatial distribution of M9 library residues. The active site is rendered as a continuous van der Waals surface (positions 402 and 516 highlighted in red), and second-tier residues (colored side chains) are arrayed behind the active site proper.

Here we investigated the natural distribution of these activities by constructing a phylogenetic tree from available TEAS-like and HPS-like sequences from related solanaceous plants (Figure 2.3 panel a, Table 2.2). Although the product spectra of terpene synthases cannot be readily predicted from traditional phylogenetic analyses[13, 14], a clear functional division was apparent between the tobacco and pepper synthases compared to their orthologs

in tomato, potato and henbane. This division was also apparent at the level of our nine-residue subset, with the exception of the *Capsicum annuum* synthase (Figure 2.3 panel b). This TEAS-like enzyme differs from both TEAS- and HPS-like groups at three positions and, most notably, contains a threonine at position 438 like HPS, suggesting that the first mutational steps in the TEAS-HPS divergence occurred at these positions. Evaluating the functional divergence of TEAS and HPS within the context of these nine amino acid substitutions provides a simplified experimental system to address the broader issue of how prevalent—and hence evolvable—these parental and alternative biosynthetic activities are throughout the intervening lineages connecting these extant enzymes. Measuring the distribution of biosynthetic activities over this sequence space defines a functionally relevant portion of the overall catalytic landscape and provides a window into the complex functional terrain underlying the evolution of these enzymes. Although variation at other positions may have contributed to the functional divergence of TEAS and HPS in meaningful ways, this focused set of functionally important residues makes it experimentally tractable to quantitatively characterize a catalytic landscape of secondary metabolism to biochemical resolution.

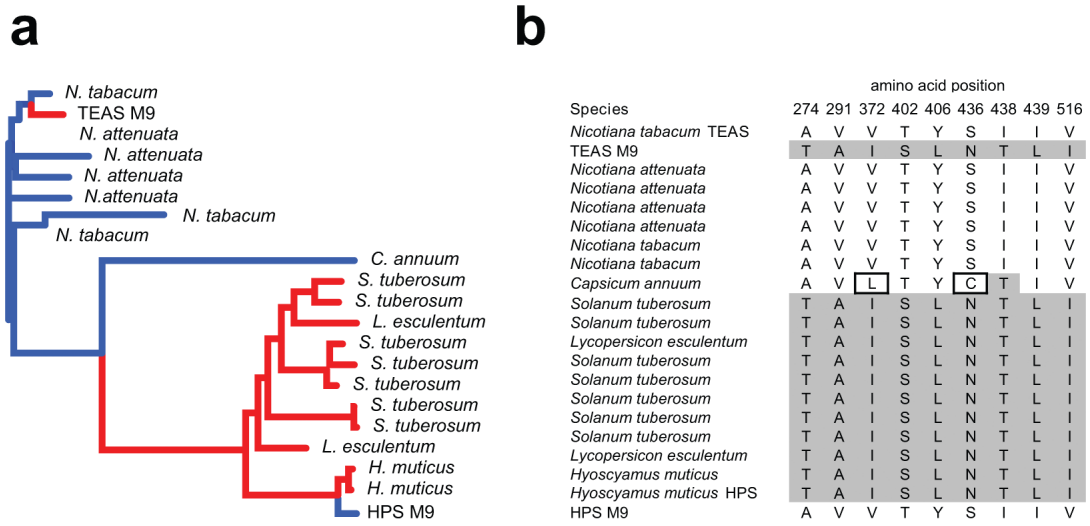


Figure 2.3. Phylogenetic distribution of solanaceous TEAS- and HPS-like terpene synthases. (a) An unrooted phylogenetic tree of 5-EA and PSD terpene synthases created from available sequences (Table 2.2). Branches are colored according to established or putative functions as TEAS-like (blue) or HPS-like (red). (b) Sequence alignment of the M9 residue positions of the sequences in a, with HPS-like residues shaded in gray. Boxes mark residues of C. annuum that differ from both TEAS and HPS. Of note, the previous taxonomic classification *Lycopersicon esculentum* is used here, consistent with the database entries for their respective proteins. However, the most recent taxonomic nomenclature has been changed to *Solanum esculentum*.

2.3. RESULTS AND DISCUSSION

2.3.1. Creation and characterization of the M9 lineage

Using SCOPE, we constructed a gene library encoding all permutations of nine amino acid substitutions in TEAS (29 = 512 combinations) previously shown to functionally interconvert TEAS and HPS[15]. The resulting library, termed the M9 lineage, represents the nodes along all possible mutational pathways (9! = 362,880) between wild-type TEAS and the nine mutant HPS-like forms (TEAS M9). The combinatorial exploration of variation at these diverging positions using SCOPE therefore captures a portion of the functionally relevant genetic variation leading to the current extant sequences. We cloned and identified mutant genes by sequencing, resulting in 432 unique combinations (Table 2.3). We then expressed

and purified individual mutants from Escherichia coli, leading to the recovery of 418 active proteins. We developed an in vitro biochemical assay for increased sample throughput and analysis of terpene synthases using GC-MS[9], which provided quantified chemical fingerprints and catalytic activities of the M9 library proteins (Tables 2.4 and 2.5).

To quantitatively assess product specificity, the catalytic property defined here as the relative proportion of products, we calculated the product percentages of each mutant from their respective total ion chromatograms. PSD, 5-EA and 4-EE were the dominant products observed across a wide spectrum of mutants, accompanied by a collection of minor products that were grouped together and treated as a single product class for simplicity. This assumption may introduce error, as the ionization efficiencies of these minor components are as yet unknown; however, their inclusion contributes to the measure of product specificity and promiscuity.

A three-dimensional scatter plot shows how the product specificities of mutants distribute throughout chemical space (Figure 2.4 panel a). The three dominant products (5-EA, 4-EE and PSD) define a two-dimensional triangular plane, and the collective minor products contribute the third dimension of the tetrahedron. Mutants with varying degrees of catalytic promiscuity radiate uniformly from a cluster of TEAS-like activities, together forming a continuum with more HPS-like mutants. By contrast, EES-like enzymes are rare, appearing as a sparsely populated subgroup. Subdividing the scatter plot into three smaller tetrahedrons of equal volume geometrically defines product specificity as >50% 5-EA, 4-EE or PSD, whereas the central volume represents promiscuous activities (Figure 2.4 panel b). The majority of mutants were promiscuous (51%), showing expanded product distributions and upregulation of other TEAS minor products, predominantly germacrene A (Compound 5)[16], a neutral intermediate along the TEAS and HPS cyclization pathways and the major product of a

closely related family of plant synthases. Kinetic analyses of select members of the library with diverse product specificities revealed that most mutants possess catalytic activities (k_{cat}) within tenfold of wild-type TEAS, indicating that most combinations of mutations altered product specificity without significantly compromising the overall catalytic rate (Table 2. 5).

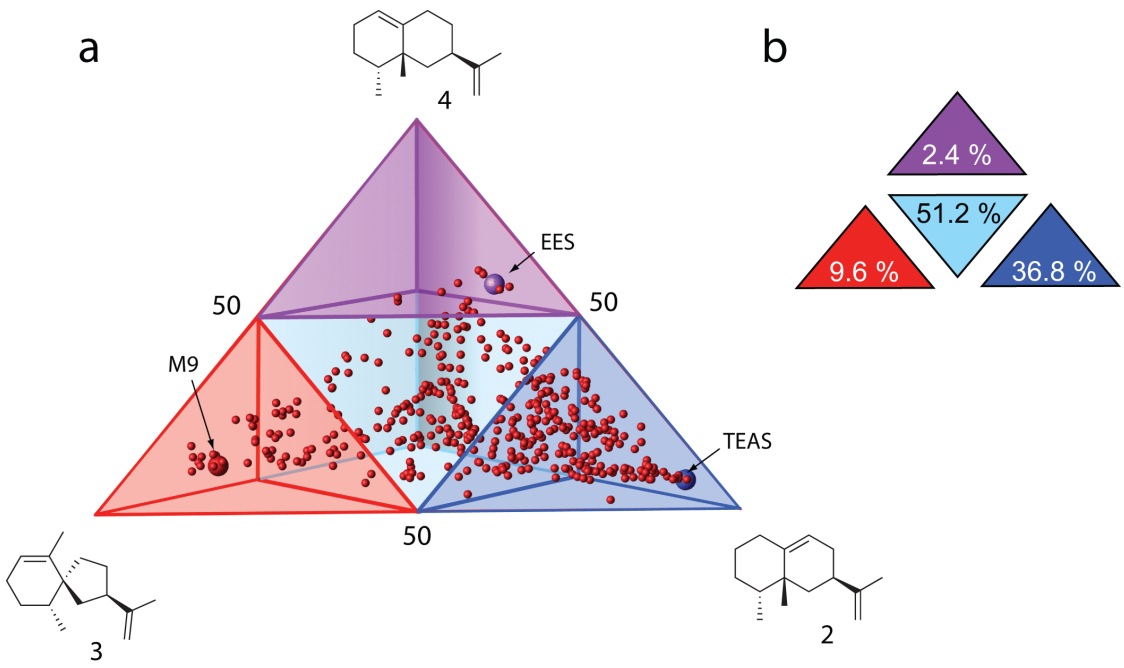


Figure 2.4. Activities of the M9 lineage. (a) A three-dimensional scatter plot of the product output (chemical space). The x, y and z axes correspond to percentages of the major products 5-EA (Compound 2), 4-EE (Compound 4) and PSD (Compound 3), respectively (Table 2.4). Each sphere represents 1 of the 418 active mutant proteins from the M9 library, with wild-type TEAS, M9 and EES highlighted as enlarged spheres. The tetrahedron encompassing the scatter plot was partitioned to represent each of the major reaction products by choosing the midpoint of each axis for subdivision into geometrically equivalent tetrahedrons. Each shaded volume (blue, 5-EA; purple, 4-EE; red, PSD) indicates product specificity of 50% or greater. Mutants in the remaining central volume (cyan) are defined as promiscuous. (b) Schematic of the scatter plot in a, summarizing the distribution of activities where the number of mutants in each quadrant is expressed as a percentage of the total number characterized.

2.3.2. Biosynthetic tree of the M9 lineage

To quantitatively examine the distribution of biosynthetic activities across the M9 library, we performed a sum of least squares pairwise comparison of chemical product

proportions. The resulting 'chemical' distance matrix was condensed to produce an unrooted neighbor-joining 'biosynthetic tree' (Figure 2.5). This tree showed several distinct clusters or clades separating TEAS- and HPS-like activities at either end. Annotating each clade with chromatograms from representative mutants highlighted the common product profiles that define its members. For example, a promiscuous clade near the tree center was marked by elevated production of germacrene A in mutant 8. Sequence analysis of the HPS-like clade revealed a clustering of mutants into distinct groups based on sequence, indicating that several convergent solutions exist along a subset of synthetic lineages (Figure 7). By comparison, members of the EES-like clade generally possessed diverse sequences but showed a strict dependence on the T402S active-site mutation for EES activity.

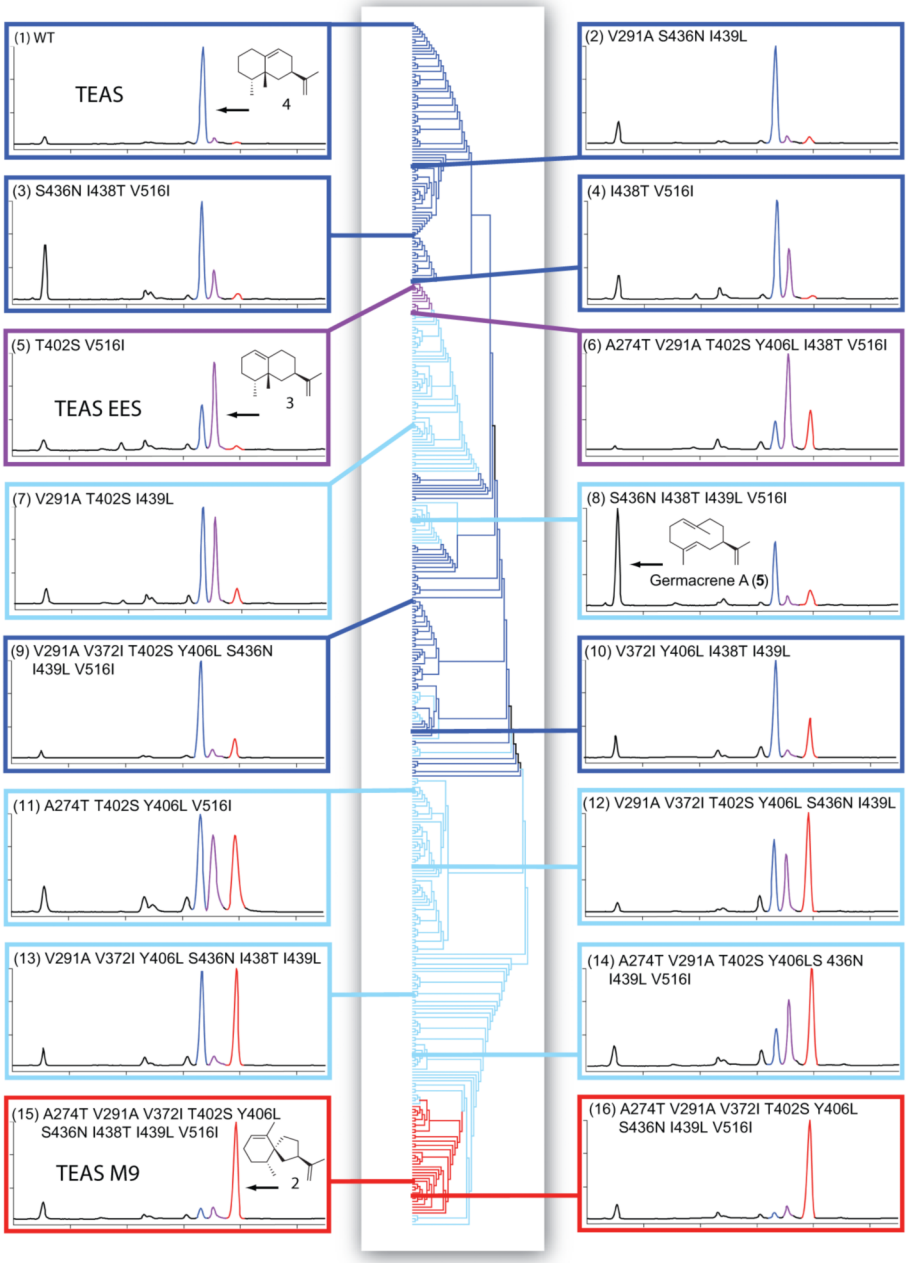


Figure 2.5. Biosynthetic tree of the M9 library. A similarity-based cluster diagram was constructed to quantitatively organize the M9 library according to terpene product spectra from the pairwise alignment of product proportions for each of the 418 active mutants (described in Methods). Clades are colored according to the major reaction product (defined in Figure 2.4 panel a), with representative chromatograms identified and numbered branching off each major clade. Product peaks in the chromatograms are colored blue for 5-EA (Compound 2), purple for 4-EE (Compound 4) and red for PSD (Compound 3).

2.3.3. Chemical distances of mutational effects

Values from the chemical distance matrix are a measure of changes in product spectra between mutants and hence provide a quantitative basis to assess the influence of each mutation on product outcome. To determine whether one of the nine positions is most crucial for controlling product specificity, we considered the effect of mutating a particular position in the background of all other possible combinations of mutations. We calculated the average chemical distance of each of the nine positions in this manner and found them to be comparable, each having a large s.d. of 50% or greater than each residue's average distance (Table 2.6). This result indicated that no single position is more important than another, suggesting that a position's influence on the control of product specificity is context dependent.

2.3.4. Quantifying mutational context

Given the importance of context, we investigated the accessibility of alternative product specificities in various mutant backgrounds. To quantitatively examine context effects, we tabulated chemical distances for a subset of 236 mutants for which there were complete data for all single mutational neighbors (permutations that differed at only one position). For example, all TEAS single mutants were characterized and represented the mutational neighbors that were a single, coding mutational step away from wild-type TEAS. However, some permutations of the 512 combinations were not identified by sequencing or did not produce recombinant protein; these were absent from the final dataset and hence omitted from this analysis.

The average interneighbor distance (AID) was calculated for each mutant; specific examples show how this index relates to chemical space (scatter plot) and sequence space

(alignment with mutational neighbors; Figure 2.6 panels a–c). For a mutant with a low AID, most mutations registered negligible to modest effects on product output, as evident from the clustering of most mutational neighbors in a small region of chemical space (Figure 2.6 panel a). By contrast, mutants with a high AID showed a broad scattering of mutational neighbors throughout chemical space (Figure 2.6 panel c), with demonstrable long jumps between highly specific TEAS-like to EES- or HPS-like activities by single mutational steps. Hence, the activities of some mutants are highly sensitive to mutational perturbations. For a promiscuous mutant with a moderate AID (Figure 2.6 panel b), nearly half of the mutational steps tightened product specificity. Considering the larger trends throughout the M9 library, the AID for the subset of 236 mutants was plotted as a simple histogram (Figure 2.6 panel d). Plotting the AID as a function of the number of mutations revealed that the distribution of averages was similar across the library (Figure 2.6 panel e).

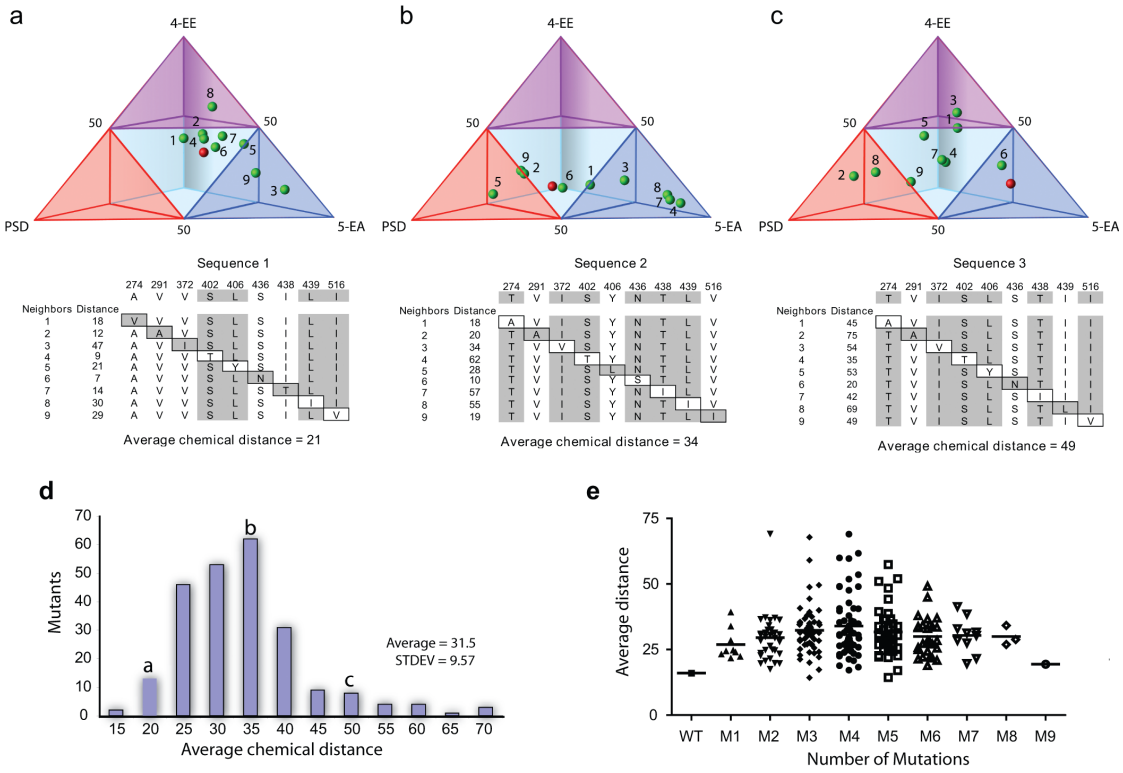


Figure 2.6. AID in chemical and sequence space. (a–c) A representative mutant (unlabeled red sphere) is shown in chemical space, along with all nine possible single-mutant neighbors (numbered green spheres) to show short (a), medium (b) and high (c) AID. Sequences of each representative mutant are referenced across the top of the three alignment tables, with mutational neighbors and distances listed below. Each mutated position is boxed, and residues of HPS origin are indicated with shading. (d) AID for a subset of 236 mutants was plotted as a simple histogram, where the shoulders and apex of the distribution are labeled 'a', 'b' and 'c' to correspond to representative mutants above. (e) The distribution of AID as a function of the number of accumulated HPS substitutions was plotted, where M1 refers to all single mutants, M2 to all double mutants, and so on.

2.3.5. Discussion

The emerging picture from these experiments is a rugged landscape in which alternative catalytic specificities are mutationally accessible, requiring as little as a single base change in the coding gene. Single mutations on average exert moderate effects, relaxing product specificity while upregulating 5-EA, 4-EE, PSD or other TEAS minor products, consistent with postulates that specific activities arise from catalytically promiscuous

ancestors[11-19]. Most mutations are additive, but rapid or punctuated changes in product output are not rare. In fact, such hot spots (AID > 50) account for 7% of the mutants analyzed thus far, indicating considerable biosynthetic potential for rapid evolutionary jumps throughout the M9 lineage. This rapid adaptability may be unique to terpene cyclases, stemming in part from the subtle energetic differences between competing cyclization pathways (Figure 2.1 panel b) that ultimately govern product specificity. By implication, TEAS-HPS predecessors had the potential for frequent shifts between PSD, 4-EE and 5-EA biosynthesis to generate rapidly changing chemical repertoires throughout evolution.

Although both additive and punctuated specificity changes have been observed in terpene cyclases[20-22], this is the first effort to quantitatively measure the frequency and distribution of these enzymatic properties over a catalytic landscape. To quantitatively describe this landscape, it was essential to use a simple Euclidean distance metric, a chemical distance that is generally applicable to mapping how any experimentally defined enzymatic property is distributed throughout sequence space. In the current work, this metric provided a means to quantitatively compare product spectra of terpene synthases, assess the effects of mutations in different backgrounds—particularly mutational neighbors—and construct a biosynthetic tree to quantitatively organize the M9 enzyme lineage by functional relatedness.

Structural and phylogenetic information has been invaluable in guiding mutagenesis experiments leading to the successful interconversion of terpene cyclase substrate[23] and product specificities[12, 23, 24]. In the absence of phylogenetically derived models, applying saturation mutagenesis to the active site of a notably promiscuous terpene cyclase, γ-humulene synthase, has made considerable engineering advances to generate specific activities[21]. By contrast, the work here explores phylogenetically relevant biosynthetic interrelationships that extend product specificity control beyond the active site. Characterizing the reciprocal M9

lineages in HPS will be of great future interest to address the contribution of alternative protein backgrounds to the product specificity landscape.

Only recently have efforts been made to characterize the underlying adaptive landscapes of molecular evolution[25-28] or to trace the evolutionary origins of the four fundamental isoprenoid-based coupling reactions[29]. Our study provides the first experimental measure of the complex functional terrain evident in secondary metabolism from the construction and biochemical characterization of intervening lineages between a pair of extant and diverging enzymes. Although it is tempting to speculate that 4-EE was the dominant product of a TEAS-HPS common ancestor on the basis of its hybrid mechanistic origins, this activity represents less than 3% of the total library. Also, greater product specificity for PSD is achievable with fewer than nine mutations. For example, an M8 (mutant 226, Table 2.4, Table 2.7) produces 81% PSD, versus 72% by M9. However, the native henbane enzyme produces 97% PSD, indicating that additional mutations beyond the nine considered here contribute to this high degree of product specificity. The facile phenotypic change from minimal sequence changes uncovered by our work suggests that it is extremely difficult to make accurate assignments of ancestral function in this pervasive class of secondary metabolic enzymes. This result has broader implications for reconstructing ancestral proteins and ascribing ancient functions; one must consider the extent of phenotypic variation among a population of putative intermediates encompassed by a 'probabilistic guess' of the most likely ancestor.

Connecting catalytic landscapes of secondary metabolism to fitness landscapes of organisms remains an enormous future challenge. Antibiotic resistance or primary metabolic functions in microbes have direct survival consequences easily measured in laboratory evolution experiments, but the fitness contributions of secondary metabolism, which are of particular relevance to speciation in complex organisms, are intrinsically more difficult to

study. This arises in part from the myriad roles of secondary metabolites in the greater chemical ecology of host organisms. In the current work, relating the in vitro results to in vivo functions involves several caveats; numerous factors including transcription, translation and solubility surely contribute to enzyme evolution, and possible effects of the cellular environment on enzymatic activity must also be considered. There is precedence, however, for the correspondence of in vitro and in vivo product profiles of terpene cyclases[24], so the observed plasticity of terpene cyclase enzymatic function from in vitro biochemical measurements is likely to approximate the activities of these enzymes in their native environment. More extensive sequencing efforts and biochemical annotation of terpene synthases from the larger family of solanaceous plants will both address the degree to which mutant combinations of the M9 lineage reflect the actual evolutionary lineages and provide valuable insights toward understanding the role of secondary metabolism in shaping the evolution of the Solanaceae.

The observation that HPS-like activity is achievable by many combinations of mutations lying outside the active site (Tables 2.7, Table 2.8) highlights the crucial importance of epistasis. This phenomenon has been documented and described in other enzyme systems[25-27] and is manifested here as a highly interdependent network of interactions in the protein that ultimately controls product specificity. These functionally crucial yet distal mutations are not surprising, given the effects of other distant mutations on protein and enzyme function[30, 31]. Modulation of isoprenoid cyclization by discrete ensembles of peripherally distributed residues is suggestive of energetic networks spread throughout the protein structure[32], which may allow greater adaptive capabilities. As recently noted, the interface of enzyme adaptability with intrinsic and induced substrate reactivity underlies the emergence of cyclic diversity in secondary metabolism[33]. Our quantitative exploration of the catalytic landscape of

the M9 lineage provides a first glimpse into the functional effects of quantum evolutionary change on specialized biosynthesis.

2.4. METHODS

2.4.1. Library construction

SCOPE mutant gene synthesis was done using published methods[15]. A library encoding 512 mutants comprising all permutations of the original TEAS M9 mutants[12] was produced, with 432 unique combinations identified by DNA sequencing (Table 2.3). The library was created as gene sets consisting of a series of discrete mixtures to reduce oversampling. Mutant TEAS genes were subcloned into pH9GW, an in-house expression vector encoding nine N-terminal histidine residues, using the Gateway system (Invitrogen) according to the manufacturer's instructions.

2.4.2. Biosynthetic tree construction

Terpene products from GC-MS analyses were quantified by integration of product peak areas and transformed into percentages, where 5-EA, 4-EE, PSD and all remaining products represented four groups adding up to 100% (Table 2.4). A distance (d) between every pair of mutants was calculated by the sum of least squares:

$$d = \mathrm{sqrt}\ [\ (w_1 - w_2)^2 + (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2]$$

where product profile 1 has coordinates w1, x1, y1 and z1, and product profile2 has coordinates w_2, x_2, y_2 and z_2. The variables w, x, y and z correspond to 5-EA, 4-EE, PSD and the sum of remaining products, respectively. A large n x n matrix was dimensionally reduced into a standard phylogenetic tree to show which mutants cluster together in space. An unrooted N-J tree was produced using MEGA 3.1 software[34].

2.4.3. Sequencing

Plasmid DNA was prepared by the Microarray Core Facility at The Salk Institute, and DNA sequencing was done by Eton Biosciences.

2.4.4. Vial assay characterization

In vitro assays of purified recombinant enzyme were conducted in duplicate according to a previously published method[9]. Products were quantified by integration of peak areas from the gas chromatography trace using Agilent ChemStation software and expressed as a percentage of the total products. Notably, germacrene A was detected as the thermally rearranged product β-elemene. Authentic standards of 5-EA, 4-EE and PSD were used for instrument calibration and absolute product quantitation for kinetic measurements (Supplementary Methods online).

2.4.5. Protein expression

pH9GW expression vectors were transformed into E. coli BL21(λDE3) and plated on LB growth media with 50 μg/ml kanamycin selection. Colonies were transferred to 1 ml liquid media (LB with kanamycin) in 96-well plates followed by 16 hrs growth with shaking at 37˚C at 275 rpm. Cultures were diluted 10-fold into 5 ml of TB growth media with kanamycin in 24-well round bottom plates covered with micro-porous tape, followed by growth with shaking at 37˚C at 275 rpm until cultures reached OD600 ≥ 1.5. Protein expression was induced by addition of IPTG to 0.1 mM followed by growth with shaking at 20˚C at 275 rpm for 5 hrs. Cells were harvested by centrifugation and cell pellets were frozen at -20˚C.

2.4.6. Purification of library proteins

Pellets from 5 ml expression cultures were re-suspended by adding 0.8 ml of lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 10% glycerol (v/v), 10 mM β-mercaptoethanol, and 1% (v/v) Tween-20, pH 8) containing 1 mg/ml lysozyme and 1 mM EDTA directly to frozen pellets followed by shaking at room temperature (25˚C) at 350 rpm for 30 minutes. Next, 10 µl of benzonase solution (850 mM $MgCl_2$ and 3.78 U/ul benzonase (Novagen) was added followed by additional shaking at 350 rpm for 15 min. Lysates were passed through a Whatman unifilter 96-well plate and collected in another Whatman plate containing 100 µl bed-volume of superflow Ni-NTA resin (QIAgen), preequilibrated with wash buffer using a vacuum manifold. Each well was washed with 2 ml lysis buffer, followed by 1.5 ml wash buffer (lysis buffer lacking Tween- 20). Resin was air dried prior to addition of 150 µl elution buffer (wash buffer containing 250 mM imidazole), incubated for 10 min, followed by centrifugation to recover eluted protein. Protein recovery (~0.25 µg per 5 mL culture) and purity (approximately 95%) were verified by SDS-PAGE analysis and UV at an absorbance value of 280nm. An equal volume of 100% glycerol was added to eluted samples followed by long term storage at -20˚C.

2.4.7. Kinetic Measurements

Enzyme kinetics was performed using the vial assay[9] under the following modified conditions. Reactions were composed on a 500 µl scale using a 3- component buffer system[35] (25 mM 2- (N-Morpholino) ethanesulfonic acid (MES), 50 mM Tris, and 25 mM 3-(Cyclohexylamino) propanesulfonic acid (CAPS)) at pH = 7.0 with 10 mM MgCl2 and a fixed substrate concentration of 300 uM farnesyl pyrophosphate (FPP). For enzyme quantitation, proteins samples were denatured in 6M guanidinium-HCl prior to measuring UV absorbance

at 280 nM. Protein concentrations were calculated using theoretical extinction coefficients[36]. Serial enzyme dilutions ranging from 1 to 300 μM were incubated with substrate at room temperature with an ethyl acetate overlay for 12 minutes prior to quenching by vortexing. Authentic standards of 5-EA (2), 4-EE (4), PSD (3) were used for instrument calibration and quantitation. The slope of the calibration curves (instrument response as a function of analyte concentration) defines ionization efficiencies of these analytes, found to be nearly identical over the linear range of detection employed in GC experiments:

Table 2.1. Ionization efficiencies of 5-EA, 4-EE, and PSD

	Compounds		
Parameters	5-EA (2)	4-EE (4)	PSD (3)
Instrument mode	TIM[a]	TIM	TIM
Slope (TIC/10^5)	19.99 ± 0.097	19.46 ± 0.109	19.25 ± 0.156
Correlation coefficient r^2	0.999	0.999	0.999
Concentration range (μM)	50-2	50-2	50-2

[a]TIM mode refers to the mass spectrometer detection settings in which all ions derived from a given compound are counted and contribute to the instrument signal in the total ion chromatogram.

Under conditions of excess substrate, the reaction follows first order kinetics (zero order with regards to substrate), $v = k_o[E]_o$, where the apparent rate constant ko is considered the turnover number $k_{cat,App}$.

2.5. SUPPORTING INFORMATION

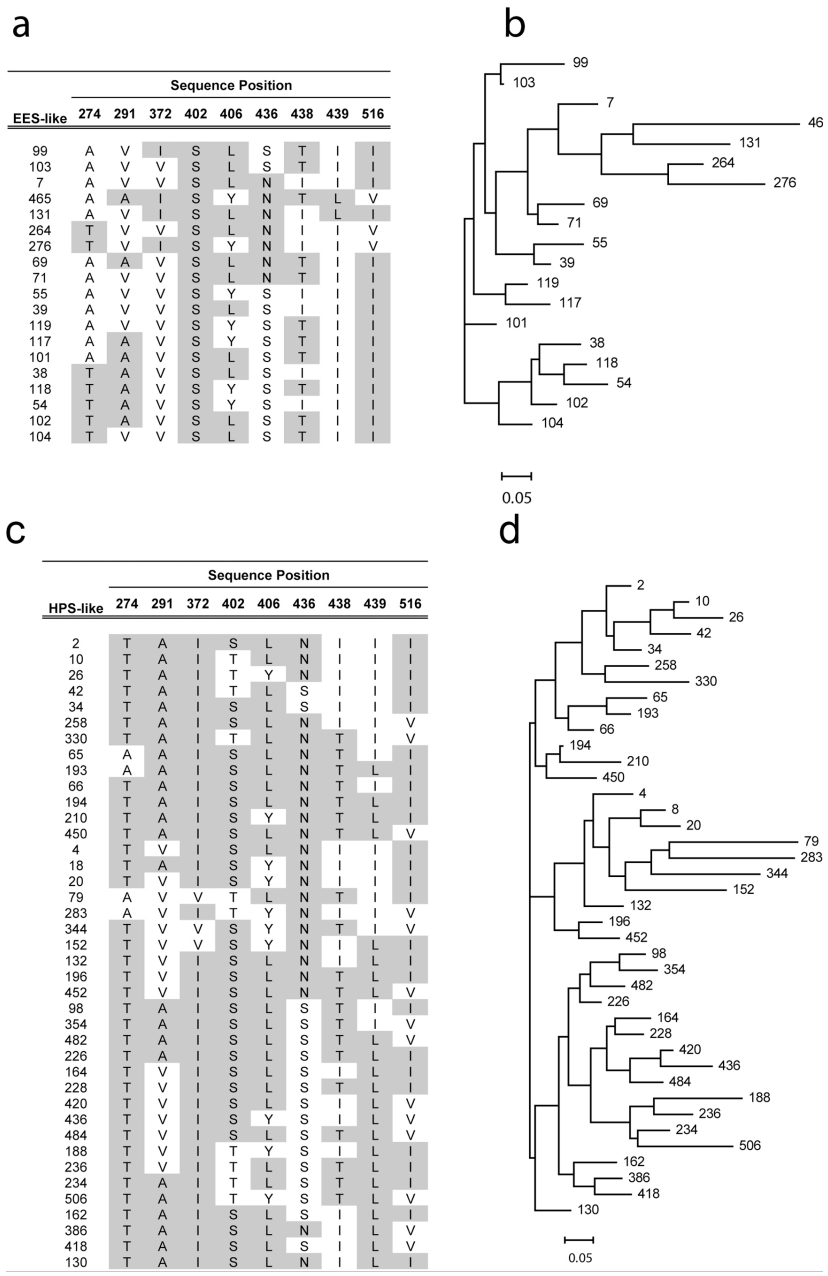


Figure 2.7. Similarity-based cluster diagram of the EES-like and HPS-like mutant clades. (a) Sequences from the M9 library encoding enzymes producing greater than 50% 4-EE (4) as their major product were compiled and aligned, where sequence 55 corresponds to the TEAS mutant EES. Positions shaded in gray signify mutations to the structurally equivalent HPS residue. (b) An unrooted N-J phylogenetic tree was constructed from the input sequences in part a using ClustalW (http://www.ebi.ac.uk/clustalw/). (c) Sequences from the M9 library encoding enzymes producing greater than 50% PSD (3) as their major product were compiled and aligned. Positions shaded in gray signify mutations to the 5 structurally equivalent HPS residue, where sequence number 194 corresponds to TEAS M9. (d) An unrooted N-J phylogenetic tree was constructed from the input sequences in part d using ClustalW (http://www.ebi.ac.uk/clustalw/).

Table 2.2. Sequences of Solanaceous putative and characterized 5-EA and PSD terpene synthases.

	amino acid position [a]										
Species	**274**	**291**	**372**	**402**	**406**	**436**	**438**	**439**	**516**	**Accession**	**Activity [d]**
Nicotiana tabacum TEAS [b]	A	V	V	T	Y	S	I	I	V	AAA19216.1	5-EAS
TEAS M9	T	A	I	S	L	N	T	L	I		PSDS
Nicotiana attenuata	A	V	V	T	Y	S	I	I	V	AAP05761.1	5-EAS
Nicotiana attenuata	A	V	V	T	Y	S	I	I	V	AAP05762.1	5-EAS
Nicotiana attenuata	A	V	V	T	Y	S	I	I	V	AAO85555.1	5-EAS
Nicotiana attenuata	A	V	V	T	Y	S	I	I	V	AAP05760.1	5-EAS
Nicotiana tabacum	A	V	V	T	Y	S	I	I	V	AAP79448.1	5-EAS
Nicotiana tabacum	A	V	V	T	Y	S	I	I	V	AAG17667.1	5-EAS
Capsicum annuum	A	V	L	T	Y	C	T	I	V	CAA06614.1	5-EAS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	BAA82092.1	PSDS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	BAA82141.1	PSDS
Lycopersicon esculentum [c,e]	T	A	I	S	L	N	T	L	I	AAG09949.1	PSDS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	AAD02269.1	putative PSDS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	AAD02268.1	putative PSDS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	BAA82108.1	PSDS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	AAD02270.1	putative PSDS
Solanum tuberosum [c]	T	A	I	S	L	N	T	L	I	AAD02223.1	PSDS
Lycopersicon esculentum [c,e]	T	A	I	S	L	N	T	L	I	AAG09950.1	PSDS
Hyoscyamus muticus [c]	T	A	I	S	L	N	T	L	I	AAA86340.1	PSDS
Hyoscyamus muticus HPS [c]	T	A	I	S	L	N	T	L	I	AAA86337.1	PSDS
HPS M9	A	V	V	T	Y	S	I	I	V		

[a] Residue numbering according to TEAS reference sequence.

[b] Note: see UniProtKB/Swiss-Prot entry Q40577 for corrections to the originally published sequence.

[c] Originally annotated as "vetispiradiene" synthase in the old nomenclature, since changed and referred to here as “premnaspirodiene.”

[d] Activity is classified according 5-epi-aristolochene synthase (5-EAS) or premnaspirodiene synthase (PSDS).

[e] The previous taxonomic classification *Lycopersicon esculentum* is used here, consistent with the database entries for their respective proteins. However, the revised taxonomic nomenclature has been changed to *Solanum esculentum*.

Table 2.3. SCOPE library construction statistics

M9 Library Statistics	
Library complexity	512
Clones sequenced	3,047
Fold oversampling	5.9
Unique clones identified	432
Additional mutations:	
silent	52
frame-shift	91
point	65
total	208
mutation rate	6.80%

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins. Residues of HPS origin are indicated by shading and numbering is according to TEAS.

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
1	Ala	Ala	Ile	Ser	Leu	Asn	Ile	Ile	Ile	6	23.06	7.41	22.09	47.43	
2	Thr	Ala	Ile	Ser	Leu	Asn	Ile	Ile	Ile	7	13.50	2.40	11.84	72.26	
3	Ala	Val	Ile	Ser	Leu	Asn	Ile	Ile	Ile	5	22.81	9.72	29.76	37.71	
4	Thr	Val	Ile	Ser	Leu	Asn	Ile	Ile	Ile	6	18.82	5.47	18.83	56.88	
5	Ala	Ala	Val	Ser	Leu	Asn	Ile	Ile	Ile	5	25.77	18.65	42.88	12.71	
6	Thr	Ala	Val	Ser	Leu	Asn	Ile	Ile	Ile	6	18.88	11.70	35.70	33.71	
7	Ala	Val	Val	Ser	Leu	Asn	Ile	Ile	Ile	4	17.68	28.46	44.83	9.02	
8	Thr	Val	Val	Ser	Leu	Asn	Ile	Ile	Ile	5	15.30	25.91	41.71	17.08	
9	Ala	Ala	Ile	Thr	Leu	Asn	Ile	Ile	Ile	5	ND				
10	Thr	Ala	Ile	Thr	Leu	Asn	Ile	Ile	Ile	6	22.86	13.62	7.75	55.76	
11	Ala	Val	Ile	Thr	Leu	Asn	Ile	Ile	Ile	4	36.76	27.76	14.27	21.21	
12	Thr	Val	Ile	Thr	Leu	Asn	Ile	Ile	Ile	5	23.28	22.68	14.54	39.50	
13	Ala	Ala	Val	Thr	Leu	Asn	Ile	Ile	Ile	4	12.75	74.38	9.21	3.67	
14	Thr	Ala	Val	Thr	Leu	Asn	Ile	Ile	Ile	5	ND				
15	Ala	Val	Val	Thr	Leu	Asn	Ile	Ile	Ile	3	12.68	59.71	23.42	4.18	
16	Thr	Val	Val	Thr	Leu	Asn	Ile	Ile	Ile	4	15.11	48.38	17.66	18.85	
17	Ala	Ala	Ile	Ser	Tyr	Asn	Ile	Ile	Ile	5	ND				
18	Thr	Ala	Ile	Ser	Tyr	Asn	Ile	Ile	Ile	6	11.00	20.22	11.02	57.76	
19	Ala	Val	Ile	Ser	Tyr	Asn	Ile	Ile	Ile	4	19.28	19.28	39.76	21.69	M4[1]
20	Thr	Val	Ile	Ser	Tyr	Asn	Ile	Ile	Ile	5	19.08	17.88	10.71	52.33	
21	Ala	Ala	Val	Ser	Tyr	Asn	Ile	Ile	Ile	4	25.16	47.86	11.97	15.02	
22	Thr	Ala	Val	Ser	Tyr	Asn	Ile	Ile	Ile	5	22.88	21.01	36.21	19.90	
23	Ala	Val	Val	Ser	Tyr	Asn	Ile	Ile	Ile	3	25.04	52.45	16.72	5.80	
24	Thr	Val	Val	Ser	Tyr	Asn	Ile	Ile	Ile	4	5.10	65.60	18.64	10.65	
25	Ala	Ala	Ile	Thr	Tyr	Asn	Ile	Ile	Ile	4	ND				
26	Thr	Ala	Ile	Thr	Tyr	Asn	Ile	Ile	Ile	5	12.07	20.67	10.56	56.70	
27	Ala	Val	Ile	Thr	Tyr	Asn	Ile	Ile	Ile	3	0.78	57.31	30.32	11.59	
28	Thr	Val	Ile	Thr	Tyr	Asn	Ile	Ile	Ile	4	2.04	70.55	16.49	10.91	
29	Ala	Ala	Val	Thr	Tyr	Asn	Ile	Ile	Ile	3	22.03	55.90	12.19	9.88	
30	Thr	Ala	Val	Thr	Tyr	Asn	Ile	Ile	Ile	4	18.34	67.63	7.35	6.68	
31	Ala	Val	Val	Thr	Tyr	Asn	Ile	Ile	Ile	2	15.83	52.49	30.12	1.57	
32	Thr	Val	Val	Thr	Tyr	Asn	Ile	Ile	Ile	3	37.88	45.89	9.39	6.84	
33	Ala	Ala	Ile	Ser	Leu	Ser	Ile	Ile	Ile	5	13.46	28.97	24.96	32.61	
34	Thr	Ala	Ile	Ser	Leu	Ser	Ile	Ile	Ile	6	13.25	7.30	21.40	58.05	
35	Ala	Val	Ile	Ser	Leu	Ser	Ile	Ile	Ile	4	22.12	24.40	43.08	10.39	
36	Thr	Val	Ile	Ser	Leu	Ser	Ile	Ile	Ile	5	4.11	33.41	31.39	31.09	
37	Ala	Ala	Val	Ser	Leu	Ser	Ile	Ile	Ile	4	19.98	40.30	24.91	14.81	
38	Thr	Ala	Val	Ser	Leu	Ser	Ile	Ile	Ile	5	11.61	14.80	52.31	21.28	
39	Ala	Val	Val	Ser	Leu	Ser	Ile	Ile	Ile	3	14.41	22.84	58.55	4.19	
40	Thr	Val	Val	Ser	Leu	Ser	Ile	Ile	Ile	4	15.14	28.99	28.32	27.55	
41	Ala	Ala	Ile	Thr	Leu	Ser	Ile	Ile	Ile	4	8.95	50.19	31.46	9.41	
42	Thr	Ala	Ile	Thr	Leu	Ser	Ile	Ile	Ile	5	23.10	12.87	8.56	55.48	
43	Ala	Val	Ile	Thr	Leu	Ser	Ile	Ile	Ile	3	19.23	45.94	23.37	11.46	
44	Thr	Val	Ile	Thr	Leu	Ser	Ile	Ile	Ile	4	16.40	30.66	27.46	25.47	
45	Ala	Ala	Val	Thr	Leu	Ser	Ile	Ile	Ile	3	41.37	23.71	9.04	25.88	
46	Thr	Ala	Val	Thr	Leu	Ser	Ile	Ile	Ile	4	ND				
47	Ala	Val	Val	Thr	Leu	Ser	Ile	Ile	Ile	2	17.93	50.74	25.94	5.39	
48	Thr	Val	Val	Thr	Leu	Ser	Ile	Ile	Ile	3	9.99	59.18	26.67	4.16	
49	Ala	Ala	Ile	Ser	Tyr	Ser	Ile	Ile	Ile	4	19.42	41.70	28.49	10.39	
50	Thr	Ala	Ile	Ser	Tyr	Ser	Ile	Ile	Ile	5	17.19	12.23	32.70	37.88	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
51	Ala	Val	Ile	Ser	Tyr	Ser	Ile	Ile	Ile	3	11.92	47.69	31.40	9.00	M3[1]
52	Thr	Val	Ile	Ser	Tyr	Ser	Ile	Ile	Ile	4	14.44	27.93	37.17	20.46	
53	Ala	Ala	Val	Ser	Tyr	Ser	Ile	Ile	Ile	3	10.10	46.13	31.21	12.56	
54	Thr	Ala	Val	Ser	Tyr	Ser	Ile	Ile	Ile	4	19.09	17.87	52.31	10.73	
55	Ala	Val	Val	Ser	Tyr	Ser	Ile	Ile	Ile	2	17.43	25.82	54.49	2.27	
56	Thr	Val	Val	Ser	Tyr	Ser	Ile	Ile	Ile	3	28.70	20.41	45.85	5.04	
57	Ala	Ala	Ile	Thr	Tyr	Ser	Ile	Ile	Ile	3	24.53	48.75	13.42	13.31	
58	Thr	Ala	Ile	Thr	Tyr	Ser	Ile	Ile	Ile	4	21.74	35.14	24.45	18.67	
59	Ala	Val	Ile	Thr	Tyr	Ser	Ile	Ile	Ile	2	32.91	42.76	21.19	3.13	
60	Thr	Val	Ile	Thr	Tyr	Ser	Ile	Ile	Ile	3	3.03	42.91	24.61	29.46	
61	Ala	Ala	Val	Thr	Tyr	Ser	Ile	Ile	Ile	2	20.46	58.46	17.08	4.00	
62	Thr	Ala	Val	Thr	Tyr	Ser	Ile	Ile	Ile	3	ND				
63	Ala	Val	Val	Thr	Tyr	Ser	Ile	Ile	Ile	1	11.32	64.79	18.26	5.63	M1b[1]
64	Thr	Val	Val	Thr	Tyr	Ser	Ile	Ile	Ile	2	12.54	54.58	30.49	2.39	
65	Ala	Ala	Ile	Ser	Leu	Asn	Thr	Ile	Ile	7	18.30	6.49	24.61	50.60	
66	Thr	Ala	Ile	Ser	Leu	Asn	Thr	Ile	Ile	8	16.03	2.66	11.48	69.83	
67	Ala	Val	Ile	Ser	Leu	Asn	Thr	Ile	Ile	6	12.56	27.44	16.29	43.71	
68	Thr	Val	Ile	Ser	Leu	Asn	Thr	Ile	Ile	7	10.61	52.36	26.97	10.06	
69	Ala	Ala	Val	Ser	Leu	Asn	Thr	Ile	Ile	6	19.59	21.30	46.12	12.99	
70	Thr	Ala	Val	Ser	Leu	Asn	Thr	Ile	Ile	7	13.79	16.43	36.67	33.11	
71	Ala	Val	Val	Ser	Leu	Asn	Thr	Ile	Ile	5	12.15	26.05	51.11	10.70	
72	Thr	Val	Val	Ser	Leu	Asn	Thr	Ile	Ile	6	10.89	44.59	26.20	18.32	
73	Ala	Ala	Ile	Thr	Leu	Asn	Thr	Ile	Ile	6	ND				
74	Thr	Ala	Ile	Thr	Leu	Asn	Thr	Ile	Ile	7	19.31	22.14	11.22	47.33	
75	Ala	Val	Ile	Thr	Leu	Asn	Thr	Ile	Ile	5	40.27	25.75	12.17	21.82	
76	Thr	Val	Ile	Thr	Leu	Asn	Thr	Ile	Ile	6	21.27	31.55	15.64	31.54	
77	Ala	Ala	Val	Thr	Leu	Asn	Thr	Ile	Ile	5	36.13	38.65	17.32	7.89	
78	Thr	Ala	Val	Thr	Leu	Asn	Thr	Ile	Ile	6	ND				
79	Ala	Val	Val	Thr	Leu	Asn	Thr	Ile	Ile	4	12.11	5.80	12.73	69.35	
80	Thr	Val	Val	Thr	Leu	Asn	Thr	Ile	Ile	5	23.28	35.99	13.63	27.10	
81	Ala	Ala	Ile	Ser	Tyr	Asn	Thr	Ile	Ile	6	16.87	27.40	19.95	35.78	
82	Thr	Ala	Ile	Ser	Tyr	Asn	Thr	Ile	Ile	7	12.19	39.47	6.37	41.98	
83	Ala	Val	Ile	Ser	Tyr	Asn	Thr	Ile	Ile	5	3.18	38.71	29.71	28.40	
84	Thr	Val	Ile	Ser	Tyr	Asn	Thr	Ile	Ile	6	29.69	40.99	17.79	11.54	
85	Ala	Ala	Val	Ser	Tyr	Asn	Thr	Ile	Ile	5	21.28	21.66	43.61	13.45	
86	Thr	Ala	Val	Ser	Tyr	Asn	Thr	Ile	Ile	6	16.11	38.63	16.61	28.65	
87	Ala	Val	Val	Ser	Tyr	Asn	Thr	Ile	Ile	4	9.11	83.79	4.91	2.18	
88	Thr	Val	Val	Ser	Tyr	Asn	Thr	Ile	Ile	5	36.54	29.97	13.48	20.02	
89	Ala	Ala	Ile	Thr	Tyr	Asn	Thr	Ile	Ile	5	ND				
90	Thr	Ala	Ile	Thr	Tyr	Asn	Thr	Ile	Ile	6	18.24	51.47	10.55	19.74	
91	Ala	Val	Ile	Thr	Tyr	Asn	Thr	Ile	Ile	4	3.43	64.51	21.28	10.78	
92	Thr	Val	Ile	Thr	Tyr	Asn	Thr	Ile	Ile	5	9.25	48.68	29.74	12.34	
93	Ala	Ala	Val	Thr	Tyr	Asn	Thr	Ile	Ile	4	38.05	41.04	15.11	5.80	
94	Thr	Ala	Val	Thr	Tyr	Asn	Thr	Ile	Ile	5	ND				
95	Ala	Val	Val	Thr	Tyr	Asn	Thr	Ile	Ile	3	31.77	48.82	15.69	3.72	
96	Thr	Val	Val	Thr	Tyr	Asn	Thr	Ile	Ile	4	9.60	57.91	8.93	23.56	
97	Ala	Ala	Ile	Ser	Leu	Ser	Thr	Ile	Ile	6	24.52	8.70	35.41	31.37	
98	Thr	Ala	Ile	Ser	Leu	Ser	Thr	Ile	Ile	7	14.50	3.91	21.06	60.53	
99	Ala	Val	Ile	Ser	Leu	Ser	Thr	Ile	Ile	5	14.07	25.79	46.98	13.16	
100	Thr	Val	Ile	Ser	Leu	Ser	Thr	Ile	Ile	6	25.29	54.30	14.36	6.05	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
101	Ala	Ala	Val	Ser	Leu	Ser	Thr	Ile	Ile	5	19.65	22.91	48.39	9.05	
102	Thr	Ala	Val	Ser	Leu	Ser	Thr	Ile	Ile	6	13.82	14.42	50.95	20.81	
103	Ala	Val	Val	Ser	Leu	Ser	Thr	Ile	Ile	4	13.55	24.43	57.71	4.31	
104	Thr	Val	Val	Ser	Leu	Ser	Thr	Ile	Ile	5	11.83	22.30	55.81	10.05	
105	Ala	Ala	Ile	Thr	Leu	Ser	Thr	Ile	Ile	5	ND				
106	Thr	Ala	Ile	Thr	Leu	Ser	Thr	Ile	Ile	6	ND				
107	Ala	Val	Ile	Thr	Leu	Ser	Thr	Ile	Ile	4	ND				
108	Thr	Val	Ile	Thr	Leu	Ser	Thr	Ile	Ile	5	16.08	30.32	28.02	25.59	
109	Ala	Ala	Val	Thr	Leu	Ser	Thr	Ile	Ile	4	21.65	44.71	30.05	3.59	
110	Thr	Ala	Val	Thr	Leu	Ser	Thr	Ile	Ile	5	14.16	39.91	34.19	11.75	
111	Ala	Val	Val	Thr	Leu	Ser	Thr	Ile	Ile	3	14.07	51.27	32.54	2.12	
112	Thr	Val	Val	Thr	Leu	Ser	Thr	Ile	Ile	4	13.37	49.66	32.69	4.28	
113	Ala	Ala	Ile	Ser	Tyr	Ser	Thr	Ile	Ile	5	17.74	29.40	36.60	16.26	
114	Thr	Ala	Ile	Ser	Tyr	Ser	Thr	Ile	Ile	6	12.50	49.61	28.04	9.85	
115	Ala	Val	Ile	Ser	Tyr	Ser	Thr	Ile	Ile	4	11.78	35.77	26.82	25.63	
116	Thr	Val	Ile	Ser	Tyr	Ser	Thr	Ile	Ile	5	14.33	16.59	42.81	26.27	
117	Ala	Ala	Val	Ser	Tyr	Ser	Thr	Ile	Ile	4	ND				
118	Thr	Ala	Val	Ser	Tyr	Ser	Thr	Ile	Ile	5	16.34	52.18	13.85	17.63	
119	Ala	Val	Val	Ser	Tyr	Ser	Thr	Ile	Ile	3	17.07	27.46	53.20	2.26	
120	Thr	Val	Val	Ser	Tyr	Ser	Thr	Ile	Ile	4	11.66	42.49	36.07	9.79	
121	Ala	Ala	Ile	Thr	Tyr	Ser	Thr	Ile	Ile	4	9.60	7.88	22.41	60.10	
122	Thr	Ala	Ile	Thr	Tyr	Ser	Thr	Ile	Ile	5	ND				
123	Ala	Val	Ile	Thr	Tyr	Ser	Thr	Ile	Ile	3	ND				
124	Thr	Val	Ile	Thr	Tyr	Ser	Thr	Ile	Ile	4	10.02	39.42	39.01	11.55	
125	Ala	Ala	Val	Thr	Tyr	Ser	Thr	Ile	Ile	3	28.99	38.94	10.42	21.64	
126	Thr	Ala	Val	Thr	Tyr	Ser	Thr	Ile	Ile	4	13.32	44.98	31.25	10.46	
127	Ala	Val	Val	Thr	Tyr	Ser	Thr	Ile	Ile	2	16.80	53.53	27.92	1.75	
128	Thr	Val	Val	Thr	Tyr	Ser	Thr	Ile	Ile	3	7.36	49.05	30.96	12.62	
129	Ala	Ala	Ile	Ser	Leu	Asn	Ile	Leu	Ile	7	8.13	75.26	1.14	15.47	
130	Thr	Ala	Ile	Ser	Leu	Asn	Ile	Leu	Ile	8	14.38	3.59	10.31	71.72	
131	Ala	Val	Ile	Ser	Leu	Asn	Ile	Leu	Ile	6	18.13	25.96	47.18	8.72	
132	Thr	Val	Ile	Ser	Leu	Asn	Ile	Leu	Ile	7	11.33	13.74	16.95	57.99	
133	Ala	Ala	Val	Ser	Leu	Asn	Ile	Leu	Ile	6	22.40	24.50	35.32	17.79	
134	Thr	Ala	Val	Ser	Leu	Asn	Ile	Leu	Ile	7	15.12	14.66	28.78	41.45	
135	Ala	Val	Val	Ser	Leu	Asn	Ile	Leu	Ile	5	17.49	33.71	35.89	12.91	
136	Thr	Val	Val	Ser	Leu	Asn	Ile	Leu	Ile	6	10.12	35.68	30.99	23.20	
137	Ala	Ala	Ile	Thr	Leu	Asn	Ile	Leu	Ile	6	ND				
138	Thr	Ala	Ile	Thr	Leu	Asn	Ile	Leu	Ile	7	ND				
139	Ala	Val	Ile	Thr	Leu	Asn	Ile	Leu	Ile	5	22.44	45.41	23.93	8.22	
140	Thr	Val	Ile	Thr	Leu	Asn	Ile	Leu	Ile	6	13.56	49.03	27.94	9.46	
141	Ala	Ala	Val	Thr	Leu	Asn	Ile	Leu	Ile	5	32.89	42.93	13.08	11.11	
142	Thr	Ala	Val	Thr	Leu	Asn	Ile	Leu	Ile	6	19.40	36.05	17.59	26.96	
143	Ala	Val	Val	Thr	Leu	Asn	Ile	Leu	Ile	4	ND				
144	Thr	Val	Val	Thr	Leu	Asn	Ile	Leu	Ile	5	17.89	52.55	15.38	14.18	
145	Ala	Ala	Ile	Ser	Tyr	Asn	Ile	Leu	Ile	6	19.25	21.60	27.44	31.70	
146	Thr	Ala	Ile	Ser	Tyr	Asn	Ile	Leu	Ile	7	ND				
147	Ala	Val	Ile	Ser	Tyr	Asn	Ile	Leu	Ile	5	2.04	69.67	14.24	14.05	
148	Thr	Val	Ile	Ser	Tyr	Asn	Ile	Leu	Ile	6	11.15	30.83	26.46	31.57	
149	Ala	Ala	Val	Ser	Tyr	Asn	Ile	Leu	Ile	5	11.68	38.16	39.15	11.01	
150	Thr	Ala	Val	Ser	Tyr	Asn	Ile	Leu	Ile	6	11.32	23.71	38.44	26.53	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
151	Ala	Val	Val	Ser	Tyr	Asn	Ile	Leu	Ile	4	6.81	68.88	14.61	9.69	
152	Thr	Val	Val	Ser	Tyr	Asn	Ile	Leu	Ile	5	15.35	13.73	17.58	53.34	
153	Ala	Ala	Ile	Thr	Tyr	Asn	Ile	Leu	Ile	5	ND				
154	Thr	Ala	Ile	Thr	Tyr	Asn	Ile	Leu	Ile	6	ND				
155	Ala	Val	Ile	Thr	Tyr	Asn	Ile	Leu	Ile	4	12.66	67.42	16.46	3.47	
156	Thr	Val	Ile	Thr	Tyr	Asn	Ile	Leu	Ile	5	9.32	50.51	27.94	12.23	
157	Ala	Ala	Val	Thr	Tyr	Asn	Ile	Leu	Ile	4	19.69	49.35	7.68	23.29	
158	Thr	Ala	Val	Thr	Tyr	Asn	Ile	Leu	Ile	5	ND				
159	Ala	Val	Val	Thr	Tyr	Asn	Ile	Leu	Ile	3	16.87	59.99	15.38	7.76	
160	Thr	Val	Val	Thr	Tyr	Asn	Ile	Leu	Ile	4	9.02	62.59	19.35	9.03	
161	Ala	Ala	Ile	Ser	Leu	Ser	Ile	Leu	Ile	6	17.74	14.18	32.21	35.87	
162	Thr	Ala	Ile	Ser	Leu	Ser	Ile	Leu	Ile	7	12.22	14.43	13.63	59.72	
163	Ala	Val	Ile	Ser	Leu	Ser	Ile	Leu	Ile	5	12.08	70.94	13.28	3.70	
164	Thr	Val	Ile	Ser	Leu	Ser	Ile	Leu	Ile	6	1.89	14.52	24.91	58.68	
165	Ala	Ala	Val	Ser	Leu	Ser	Ile	Leu	Ile	5	15.13	26.77	43.52	14.58	
166	Thr	Ala	Val	Ser	Leu	Ser	Ile	Leu	Ile	6	13.77	19.20	35.47	31.57	
167	Ala	Val	Val	Ser	Leu	Ser	Ile	Leu	Ile	4	17.31	31.29	33.09	18.32	
168	Thr	Val	Val	Ser	Leu	Ser	Ile	Leu	Ile	5	5.57	25.55	42.91	25.97	
169	Ala	Ala	Ile	Thr	Leu	Ser	Ile	Leu	Ile	5	24.02	31.10	14.35	30.52	
170	Thr	Ala	Ile	Thr	Leu	Ser	Ile	Leu	Ile	6	ND				
171	Ala	Val	Ile	Thr	Leu	Ser	Ile	Leu	Ile	4	26.32	37.20	14.52	21.96	
172	Thr	Val	Ile	Thr	Leu	Ser	Ile	Leu	Ile	5	18.93	30.70	14.80	35.57	
173	Ala	Ala	Val	Thr	Leu	Ser	Ile	Leu	Ile	4	24.94	49.69	18.59	6.78	
174	Thr	Ala	Val	Thr	Leu	Ser	Ile	Leu	Ile	5	12.41	48.29	20.21	19.09	
175	Ala	Val	Val	Thr	Leu	Ser	Ile	Leu	Ile	3	14.64	28.79	40.93	15.63	
176	Thr	Val	Val	Thr	Leu	Ser	Ile	Leu	Ile	4	11.97	60.13	18.99	8.91	
177	Ala	Ala	Ile	Ser	Tyr	Ser	Ile	Leu	Ile	5	ND				
178	Thr	Ala	Ile	Ser	Tyr	Ser	Ile	Leu	Ile	6	ND				
179	Ala	Val	Ile	Ser	Tyr	Ser	Ile	Leu	Ile	4	4.87	33.83	30.99	30.31	
180	Thr	Val	Ile	Ser	Tyr	Ser	Ile	Leu	Ile	5	1.43	64.12	14.11	20.34	
181	Ala	Ala	Val	Ser	Tyr	Ser	Ile	Leu	Ile	4	ND				
182	Thr	Ala	Val	Ser	Tyr	Ser	Ile	Leu	Ile	5	11.60	25.24	42.88	20.28	
183	Ala	Val	Val	Ser	Tyr	Ser	Ile	Leu	Ile	3	11.41	44.94	38.54	5.11	
184	Thr	Val	Val	Ser	Tyr	Ser	Ile	Leu	Ile	4	10.55	37.34	41.88	10.23	
185	Ala	Ala	Ile	Thr	Tyr	Ser	Ile	Leu	Ile	4	ND				
186	Thr	Ala	Ile	Thr	Tyr	Ser	Ile	Leu	Ile	5	ND				
187	Ala	Val	Ile	Thr	Tyr	Ser	Ile	Leu	Ile	3	31.58	33.55	13.85	21.02	
188	Thr	Val	Ile	Thr	Tyr	Ser	Ile	Leu	Ile	4	13.73	10.96	22.21	53.10	
189	Ala	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Ile	3	16.00	59.40	18.08	6.52	
190	Thr	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Ile	4	13.24	53.00	21.54	12.23	
191	Ala	Val	Val	Thr	Tyr	Ser	Ile	Leu	Ile	2	18.20	63.87	15.00	2.92	
192	Thr	Val	Val	Thr	Tyr	Ser	Ile	Leu	Ile	3	28.68	38.79	17.98	14.56	
193	Ala	Ala	Ile	Ser	Leu	Asn	Thr	Leu	Ile	8	26.01	7.71	15.26	51.02	
194	Thr	Ala	Ile	Ser	Leu	Asn	Thr	Leu	Ile	9	15.16	6.74	9.32	68.77	M9[1]
195	Ala	Val	Ile	Ser	Leu	Asn	Thr	Leu	Ile	7	18.25	27.09	18.02	36.65	M7[1]
196	Thr	Val	Ile	Ser	Leu	Asn	Thr	Leu	Ile	8	13.48	9.93	16.18	60.42	
197	Ala	Ala	Val	Ser	Leu	Asn	Thr	Leu	Ile	7	27.54	6.78	44.20	21.48	
198	Thr	Ala	Val	Ser	Leu	Asn	Thr	Leu	Ile	8	13.65	15.23	29.24	41.87	
199	Ala	Val	Val	Ser	Leu	Asn	Thr	Leu	Ile	6	12.30	56.70	10.64	20.36	
200	Thr	Val	Val	Ser	Leu	Asn	Thr	Leu	Ile	7	14.91	30.54	33.78	20.77	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
201	Ala	Ala	Ile	Thr	Leu	Asn	Thr	Leu	Ile	7	20.43	25.53	10.76	43.28	
202	Thr	Ala	Ile	Thr	Leu	Asn	Thr	Leu	Ile	8	19.90	49.19	6.31	24.59	
203	Ala	Val	Ile	Thr	Leu	Asn	Thr	Leu	Ile	6	6.83	83.71	6.60	2.87	
204	Thr	Val	Ile	Thr	Leu	Asn	Thr	Leu	Ile	7	24.60	32.28	8.27	34.85	
205	Ala	Ala	Val	Thr	Leu	Asn	Thr	Leu	Ile	6	35.83	44.00	10.85	9.32	
206	Thr	Ala	Val	Thr	Leu	Asn	Thr	Leu	Ile	7	ND				
207	Ala	Val	Val	Thr	Leu	Asn	Thr	Leu	Ile	5	ND				
208	Thr	Val	Val	Thr	Leu	Asn	Thr	Leu	Ile	6	8.65	66.95	6.95	17.46	
209	Ala	Ala	Ile	Ser	Tyr	Asn	Thr	Leu	Ile	7	17.30	30.59	22.12	29.99	
210	Thr	Ala	Ile	Ser	Tyr	Asn	Thr	Leu	Ile	8	13.05	17.03	11.87	58.06	
211	Ala	Val	Ile	Ser	Tyr	Asn	Thr	Leu	Ile	6	18.12	28.35	25.96	27.57	M6[1]
212	Thr	Val	Ile	Ser	Tyr	Asn	Thr	Leu	Ile	7	18.93	17.42	21.35	42.29	
213	Ala	Ala	Val	Ser	Tyr	Asn	Thr	Leu	Ile	6	16.00	38.14	33.25	12.61	
214	Thr	Ala	Val	Ser	Tyr	Asn	Thr	Leu	Ile	7	21.51	61.59	10.94	5.96	
215	Ala	Val	Val	Ser	Tyr	Asn	Thr	Leu	Ile	5	11.49	44.74	36.37	7.40	
216	Thr	Val	Val	Ser	Tyr	Asn	Thr	Leu	Ile	6	8.56	31.80	35.11	24.53	
217	Ala	Ala	Ile	Thr	Tyr	Asn	Thr	Leu	Ile	6	ND				
218	Thr	Ala	Ile	Thr	Tyr	Asn	Thr	Leu	Ile	7	7.81	58.49	23.88	9.82	
219	Ala	Val	Ile	Thr	Tyr	Asn	Thr	Leu	Ile	5	45.89	38.83	5.64	9.64	
220	Thr	Val	Ile	Thr	Tyr	Asn	Thr	Leu	Ile	6	7.82	61.56	20.55	10.08	
221	Ala	Ala	Val	Thr	Tyr	Asn	Thr	Leu	Ile	5	32.32	50.83	10.73	6.12	
222	Thr	Ala	Val	Thr	Tyr	Asn	Thr	Leu	Ile	6	ND				
223	Ala	Val	Val	Thr	Tyr	Asn	Thr	Leu	Ile	4	49.70	33.69	5.71	10.90	
224	Thr	Val	Val	Thr	Tyr	Asn	Thr	Leu	Ile	5	6.34	70.32	15.60	7.74	
225	Ala	Ala	Ile	Ser	Leu	Ser	Thr	Leu	Ile	7	17.96	9.73	25.00	47.31	
226	Thr	Ala	Ile	Ser	Leu	Ser	Thr	Leu	Ile	8	9.01	2.59	15.45	72.95	
227	Ala	Val	Ile	Ser	Leu	Ser	Thr	Leu	Ile	6	20.46	39.81	16.08	23.65	
228	Thr	Val	Ile	Ser	Leu	Ser	Thr	Leu	Ile	7	1.96	15.39	25.27	57.37	
229	Ala	Ala	Val	Ser	Leu	Ser	Thr	Leu	Ile	6	17.65	32.85	38.22	11.27	
230	Thr	Ala	Val	Ser	Leu	Ser	Thr	Leu	Ile	7	8.85	39.97	24.37	26.81	
231	Ala	Val	Val	Ser	Leu	Ser	Thr	Leu	Ile	5	12.77	35.02	42.69	9.51	
232	Thr	Val	Val	Ser	Leu	Ser	Thr	Leu	Ile	6	ND				
233	Ala	Ala	Ile	Thr	Leu	Ser	Thr	Leu	Ile	6	21.09	34.80	5.48	38.63	
234	Thr	Ala	Ile	Thr	Leu	Ser	Thr	Leu	Ile	7	19.00	17.13	12.30	51.57	
235	Ala	Val	Ile	Thr	Leu	Ser	Thr	Leu	Ile	5	26.93	37.55	13.32	22.20	
236	Thr	Val	Ile	Thr	Leu	Ser	Thr	Leu	Ile	6	19.17	9.11	16.70	55.02	
237	Ala	Ala	Val	Thr	Leu	Ser	Thr	Leu	Ile	5	14.77	39.74	20.71	24.78	
238	Thr	Ala	Val	Thr	Leu	Ser	Thr	Leu	Ile	6	ND				
239	Ala	Val	Val	Thr	Leu	Ser	Thr	Leu	Ile	4	ND				
240	Thr	Val	Val	Thr	Leu	Ser	Thr	Leu	Ile	5	13.57	59.25	18.05	9.12	
241	Ala	Ala	Ile	Ser	Tyr	Ser	Thr	Leu	Ile	6	ND				
242	Thr	Ala	Ile	Ser	Tyr	Ser	Thr	Leu	Ile	7	23.25	45.57	6.74	24.44	
243	Ala	Val	Ile	Ser	Tyr	Ser	Thr	Leu	Ile	5	18.95	30.10	29.52	21.42	
244	Thr	Val	Ile	Ser	Tyr	Ser	Thr	Leu	Ile	6	8.61	67.28	10.93	13.19	
245	Ala	Ala	Val	Ser	Tyr	Ser	Thr	Leu	Ile	5	12.49	38.75	41.11	7.65	
246	Thr	Ala	Val	Ser	Tyr	Ser	Thr	Leu	Ile	6	11.86	27.22	39.81	21.10	
247	Ala	Val	Val	Ser	Tyr	Ser	Thr	Leu	Ile	4	ND				
248	Thr	Val	Val	Ser	Tyr	Ser	Thr	Leu	Ile	5	ND				
249	Ala	Ala	Ile	Thr	Tyr	Ser	Thr	Leu	Ile	5	ND				
250	Thr	Ala	Ile	Thr	Tyr	Ser	Thr	Leu	Ile	6	25.81	36.32	12.84	25.03	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
251	Ala	Val	Ile	Thr	Tyr	Ser	Thr	Leu	Ile	4	23.70	61.41	13.16	1.73	
252	Thr	Val	Ile	Thr	Tyr	Ser	Thr	Leu	Ile	5	5.14	59.20	13.85	21.81	
253	Ala	Ala	Val	Thr	Tyr	Ser	Thr	Leu	Ile	4	23.70	56.93	14.40	4.96	
254	Thr	Ala	Val	Thr	Tyr	Ser	Thr	Leu	Ile	5	ND				
255	Ala	Val	Val	Thr	Tyr	Ser	Thr	Leu	Ile	3	18.40	63.50	13.91	4.19	
256	Thr	Val	Val	Thr	Tyr	Ser	Thr	Leu	Ile	4	14.53	31.97	41.02	12.49	
257	Ala	Ala	Ile	Ser	Leu	Asn	Ile	Ile	Val	5	16.33	20.38	16.78	46.51	
258	Thr	Ala	Ile	Ser	Leu	Asn	Ile	Ile	Val	6	7.14	14.31	11.38	67.17	
259	Ala	Val	Ile	Ser	Leu	Asn	Ile	Ile	Val	4	16.89	33.04	15.72	34.35	
260	Thr	Val	Ile	Ser	Leu	Asn	Ile	Ile	Val	5	14.44	42.80	16.20	26.56	
261	Ala	Ala	Val	Ser	Leu	Asn	Ile	Ile	Val	4	15.52	48.15	24.51	11.82	
262	Thr	Ala	Val	Ser	Leu	Asn	Ile	Ile	Val	5	10.65	26.67	23.74	38.95	
263	Ala	Val	Val	Ser	Leu	Asn	Ile	Ile	Val	3	14.82	64.06	14.97	6.16	
264	Thr	Val	Val	Ser	Leu	Asn	Ile	Ile	Val	4	11.99	30.74	54.76	2.51	
265	Ala	Ala	Ile	Thr	Leu	Asn	Ile	Ile	Val	4	ND				
266	Thr	Ala	Ile	Thr	Leu	Asn	Ile	Ile	Val	5	ND				
267	Ala	Val	Ile	Thr	Leu	Asn	Ile	Ile	Val	3	ND				
268	Thr	Val	Ile	Thr	Leu	Asn	Ile	Ile	Val	4	13.73	32.31	13.00	40.97	
269	Ala	Ala	Val	Thr	Leu	Asn	Ile	Ile	Val	3	22.57	56.54	6.72	14.18	
270	Thr	Ala	Val	Thr	Leu	Asn	Ile	Ile	Val	4	ND				
271	Ala	Val	Val	Thr	Leu	Asn	Ile	Ile	Val	2	15.08	75.83	5.03	4.05	
272	Thr	Val	Val	Thr	Leu	Asn	Ile	Ile	Val	3	12.31	69.02	6.98	11.69	
273	Ala	Ala	Ile	Ser	Tyr	Asn	Ile	Ile	Val	4	ND				
274	Thr	Ala	Ile	Ser	Tyr	Asn	Ile	Ile	Val	5	ND				
275	Ala	Val	Ile	Ser	Tyr	Asn	Ile	Ile	Val	3	6.61	83.40	7.47	2.52	
276	Thr	Val	Ile	Ser	Tyr	Asn	Ile	Ile	Val	4	15.13	23.40	57.85	3.62	
277	Ala	Ala	Val	Ser	Tyr	Asn	Ile	Ile	Val	3	ND				
278	Thr	Ala	Val	Ser	Tyr	Asn	Ile	Ile	Val	4	6.81	34.35	31.69	27.14	
279	Ala	Val	Val	Ser	Tyr	Asn	Ile	Ile	Val	2	11.28	60.93	12.40	15.38	
280	Thr	Val	Val	Ser	Tyr	Asn	Ile	Ile	Val	3	13.34	35.20	20.63	30.83	
281	Ala	Ala	Ile	Thr	Tyr	Asn	Ile	Ile	Val	3	ND				
282	Thr	Ala	Ile	Thr	Tyr	Asn	Ile	Ile	Val	4	23.90	48.13	7.15	20.82	
283	Ala	Val	Ile	Thr	Tyr	Asn	Ile	Ile	Val	2	9.42	15.38	13.21	61.98	
284	Thr	Val	Ile	Thr	Tyr	Asn	Ile	Ile	Val	3	9.02	65.38	10.72	14.87	
285	Ala	Ala	Val	Thr	Tyr	Asn	Ile	Ile	Val	2	ND				
286	Thr	Ala	Val	Thr	Tyr	Asn	Ile	Ile	Val	3	8.00	68.07	8.80	15.12	
287	Ala	Val	Val	Thr	Tyr	Asn	Ile	Ile	Val	1	13.00	71.41	8.31	7.29	
288	Thr	Val	Val	Thr	Tyr	Asn	Ile	Ile	Val	2	11.07	43.98	17.84	27.11	
289	Ala	Ala	Ile	Ser	Leu	Ser	Ile	Ile	Val	4	15.36	29.87	22.64	32.12	
290	Thr	Ala	Ile	Ser	Leu	Ser	Ile	Ile	Val	5	9.77	53.79	11.71	24.74	
291	Ala	Val	Ile	Ser	Leu	Ser	Ile	Ile	Val	3	17.52	35.60	23.21	23.68	
292	Thr	Val	Ile	Ser	Leu	Ser	Ile	Ile	Val	4	18.06	61.37	17.20	3.36	
293	Ala	Ala	Val	Ser	Leu	Ser	Ile	Ile	Val	3	10.54	44.94	35.57	8.95	
294	Thr	Ala	Val	Ser	Leu	Ser	Ile	Ile	Val	4	10.35	34.19	27.99	27.47	
295	Ala	Val	Val	Ser	Leu	Ser	Ile	Ile	Val	2	11.04	56.31	24.05	8.60	
296	Thr	Val	Val	Ser	Leu	Ser	Ile	Ile	Val	3	2.79	53.31	13.61	30.30	
297	Ala	Ala	Ile	Thr	Leu	Ser	Ile	Ile	Val	3	ND				
298	Thr	Ala	Ile	Thr	Leu	Ser	Ile	Ile	Val	4	ND				
299	Ala	Val	Ile	Thr	Leu	Ser	Ile	Ile	Val	2	12.89	65.29	8.24	13.59	
300	Thr	Val	Ile	Thr	Leu	Ser	Ile	Ile	Val	3	13.22	54.22	29.74	2.81	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
301	Ala	Ala	Val	Thr	Leu	Ser	Ile	Ile	Val	2	9.93	79.24	7.79	3.04	
302	Thr	Ala	Val	Thr	Leu	Ser	Ile	Ile	Val	3	7.45	63.27	13.38	15.91	
303	Ala	Val	Val	Thr	Leu	Ser	Ile	Ile	Val	1	7.08	84.04	7.40	1.48	
304	Thr	Val	Val	Thr	Leu	Ser	Ile	Ile	Val	2	5.95	77.92	9.86	6.28	
305	Ala	Ala	Ile	Ser	Tyr	Ser	Ile	Ile	Val	3	10.54	51.41	21.55	16.51	
306	Thr	Ala	Ile	Ser	Tyr	Ser	Ile	Ile	Val	4	ND				
307	Ala	Val	Ile	Ser	Tyr	Ser	Ile	Ile	Val	2	5.07	64.09	18.43	12.41	
308	Thr	Val	Ile	Ser	Tyr	Ser	Ile	Ile	Val	3	25.60	23.55	42.25	8.60	
309	Ala	Ala	Val	Ser	Tyr	Ser	Ile	Ile	Val	2	10.09	47.99	31.19	10.73	
310	Thr	Ala	Val	Ser	Tyr	Ser	Ile	Ile	Val	3	ND				
311	Ala	Val	Val	Ser	Tyr	Ser	Ile	Ile	Val	1	5.64	71.75	20.07	2.54	M1a[1]
312	Thr	Val	Val	Ser	Tyr	Ser	Ile	Ile	Val	2	8.74	55.08	23.40	12.78	
313	Ala	Ala	Ile	Thr	Tyr	Ser	Ile	Ile	Val	2	20.98	66.81	5.79	6.42	
314	Thr	Ala	Ile	Thr	Tyr	Ser	Ile	Ile	Val	3	13.04	55.55	7.98	23.42	
315	Ala	Val	Ile	Thr	Tyr	Ser	Ile	Ile	Val	1	17.72	71.98	5.14	5.15	
316	Thr	Val	Ile	Thr	Tyr	Ser	Ile	Ile	Val	2	8.64	53.41	27.00	10.95	
317	Ala	Ala	Val	Thr	Tyr	Ser	Ile	Ile	Val	1	14.68	74.60	6.36	4.36	
318	Thr	Ala	Val	Thr	Tyr	Ser	Ile	Ile	Val	2	4.14	78.93	8.70	8.23	
319	Ala	Val	Val	Thr	Tyr	Ser	Ile	Ile	Val	0	7.97	84.48	6.09	1.46	
320	Thr	Val	Val	Thr	Tyr	Ser	Ile	Ile	Val	1	4.53	86.10	6.76	2.62	
321	Ala	Ala	Ile	Ser	Leu	Asn	Thr	Ile	Val	6	15.43	20.24	15.71	48.62	
322	Thr	Ala	Ile	Ser	Leu	Asn	Thr	Ile	Val	7	18.95	37.09	21.31	22.65	
323	Ala	Val	Ile	Ser	Leu	Asn	Thr	Ile	Val	5	30.28	46.48	13.64	9.61	
324	Thr	Val	Ile	Ser	Leu	Asn	Thr	Ile	Val	6	7.83	68.42	6.69	17.06	
325	Ala	Ala	Val	Ser	Leu	Asn	Thr	Ile	Val	5	13.41	49.54	24.40	12.66	
326	Thr	Ala	Val	Ser	Leu	Asn	Thr	Ile	Val	6	13.90	49.09	27.45	9.56	
327	Ala	Val	Val	Ser	Leu	Asn	Thr	Ile	Val	4	13.97	52.71	20.20	13.12	
328	Thr	Val	Val	Ser	Leu	Asn	Thr	Ile	Val	5	8.58	56.90	18.60	15.92	
329	Ala	Ala	Ile	Thr	Leu	Asn	Thr	Ile	Val	5	ND				
330	Thr	Ala	Ile	Thr	Leu	Asn	Thr	Ile	Val	6	8.74	27.34	10.20	53.72	
331	Ala	Val	Ile	Thr	Leu	Asn	Thr	Ile	Val	4	22.06	47.79	6.47	23.68	
332	Thr	Val	Ile	Thr	Leu	Asn	Thr	Ile	Val	5	11.10	47.39	6.76	34.75	
333	Ala	Ala	Val	Thr	Leu	Asn	Thr	Ile	Val	4	11.79	50.37	28.17	9.68	
334	Thr	Ala	Val	Thr	Leu	Asn	Thr	Ile	Val	5	ND				
335	Ala	Val	Val	Thr	Leu	Asn	Thr	Ile	Val	3	ND				
336	Thr	Val	Val	Thr	Leu	Asn	Thr	Ile	Val	4	13.39	62.36	10.08	14.16	
337	Ala	Ala	Ile	Ser	Tyr	Asn	Thr	Ile	Val	5	25.31	43.47	22.77	8.45	
338	Thr	Ala	Ile	Ser	Tyr	Asn	Thr	Ile	Val	6	ND				
339	Ala	Val	Ile	Ser	Tyr	Asn	Thr	Ile	Val	4	10.17	50.65	16.79	22.39	
340	Thr	Val	Ile	Ser	Tyr	Asn	Thr	Ile	Val	5	13.38	73.72	10.18	2.73	
341	Ala	Ala	Val	Ser	Tyr	Asn	Thr	Ile	Val	4	8.48	52.97	20.60	17.95	
342	Thr	Ala	Val	Ser	Tyr	Asn	Thr	Ile	Val	5	ND				
343	Ala	Val	Val	Ser	Tyr	Asn	Thr	Ile	Val	3	22.00	19.92	12.02	46.06	
344	Thr	Val	Val	Ser	Tyr	Asn	Thr	Ile	Val	4	14.31	5.42	12.09	68.18	
345	Ala	Ala	Ile	Thr	Tyr	Asn	Thr	Ile	Val	4	ND				
346	Thr	Ala	Ile	Thr	Tyr	Asn	Thr	Ile	Val	5	ND				
347	Ala	Val	Ile	Thr	Tyr	Asn	Thr	Ile	Val	3	11.97	39.68	14.17	34.17	
348	Thr	Val	Ile	Thr	Tyr	Asn	Thr	Ile	Val	4	25.84	38.73	23.34	12.09	
349	Ala	Ala	Val	Thr	Tyr	Asn	Thr	Ile	Val	3	16.26	71.24	7.15	5.35	
350	Thr	Ala	Val	Thr	Tyr	Asn	Thr	Ile	Val	4	8.73	65.80	9.85	15.63	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
351	Ala	Val	Val	Thr	Tyr	Asn	Thr	Ile	Val	2	11.89	43.77	13.97	30.38	
352	Thr	Val	Val	Thr	Tyr	Asn	Thr	Ile	Val	3	5.57	66.34	19.82	8.26	
353	Ala	Ala	Ile	Ser	Leu	Ser	Thr	Ile	Val	5	13.57	28.80	24.29	33.34	
354	Thr	Ala	Ile	Ser	Leu	Ser	Thr	Ile	Val	6	6.40	15.95	14.53	63.12	
355	Ala	Val	Ile	Ser	Leu	Ser	Thr	Ile	Val	4	ND				
356	Thr	Val	Ile	Ser	Leu	Ser	Thr	Ile	Val	5	15.23	24.36	17.57	42.84	
357	Ala	Ala	Val	Ser	Leu	Ser	Thr	Ile	Val	4	9.96	57.86	26.00	6.19	
358	Thr	Ala	Val	Ser	Leu	Ser	Thr	Ile	Val	5	10.00	36.87	30.54	22.60	
359	Ala	Val	Val	Ser	Leu	Ser	Thr	Ile	Val	3	12.92	40.65	19.35	27.09	
360	Thr	Val	Val	Ser	Leu	Ser	Thr	Ile	Val	4	9.17	44.97	33.89	11.97	
361	Ala	Ala	Ile	Thr	Leu	Ser	Thr	Ile	Val	4	20.05	51.93	9.15	18.88	
362	Thr	Ala	Ile	Thr	Leu	Ser	Thr	Ile	Val	5	ND				
363	Ala	Val	Ile	Thr	Leu	Ser	Thr	Ile	Val	3	16.59	60.66	7.72	15.03	
364	Thr	Val	Ile	Thr	Leu	Ser	Thr	Ile	Val	4	20.52	31.61	20.01	27.86	
365	Ala	Ala	Val	Thr	Leu	Ser	Thr	Ile	Val	3	10.11	77.99	7.56	4.34	
366	Thr	Ala	Val	Thr	Leu	Ser	Thr	Ile	Val	4	7.74	63.39	13.29	15.58	
367	Ala	Val	Val	Thr	Leu	Ser	Thr	Ile	Val	2	28.34	46.85	7.96	16.85	
368	Thr	Val	Val	Thr	Leu	Ser	Thr	Ile	Val	3	7.79	76.39	8.60	7.22	
369	Ala	Ala	Ile	Ser	Tyr	Ser	Thr	Ile	Val	4	16.57	28.45	23.73	31.25	
370	Thr	Ala	Ile	Ser	Tyr	Ser	Thr	Ile	Val	5	ND				
371	Ala	Val	Ile	Ser	Tyr	Ser	Thr	Ile	Val	3	9.93	58.66	18.63	12.78	
372	Thr	Val	Ile	Ser	Tyr	Ser	Thr	Ile	Val	4	26.52	35.03	12.89	25.56	
373	Ala	Ala	Val	Ser	Tyr	Ser	Thr	Ile	Val	3	6.21	67.12	23.23	3.43	
374	Thr	Ala	Val	Ser	Tyr	Ser	Thr	Ile	Val	4	6.18	66.47	12.62	14.73	
375	Ala	Val	Val	Ser	Tyr	Ser	Thr	Ile	Val	2	9.77	51.70	17.45	21.08	
376	Thr	Val	Val	Ser	Tyr	Ser	Thr	Ile	Val	3	5.29	65.62	21.25	7.85	
377	Ala	Ala	Ile	Thr	Tyr	Ser	Thr	Ile	Val	3	ND				
378	Thr	Ala	Ile	Thr	Tyr	Ser	Thr	Ile	Val	4	ND				
379	Ala	Val	Ile	Thr	Tyr	Ser	Thr	Ile	Val	2	ND				
380	Thr	Val	Ile	Thr	Tyr	Ser	Thr	Ile	Val	3	ND				
381	Ala	Ala	Val	Thr	Tyr	Ser	Thr	Ile	Val	2	11.88	79.53	6.32	2.26	
382	Thr	Ala	Val	Thr	Tyr	Ser	Thr	Ile	Val	3	21.41	9.76	21.21	47.62	
383	Ala	Val	Val	Thr	Tyr	Ser	Thr	Ile	Val	1	9.40	58.13	19.29	13.17	
384	Thr	Val	Val	Thr	Tyr	Ser	Thr	Ile	Val	2	12.58	55.01	29.45	2.96	
385	Ala	Ala	Ile	Ser	Leu	Asn	Ile	Leu	Val	6	9.12	30.10	24.32	36.46	
386	Thr	Ala	Ile	Ser	Leu	Asn	Ile	Leu	Val	7	8.64	6.71	10.45	74.20	
387	Ala	Val	Ile	Ser	Leu	Asn	Ile	Leu	Val	5	9.82	32.10	21.37	36.71	
388	Thr	Val	Ile	Ser	Leu	Asn	Ile	Leu	Val	6	27.07	21.80	16.69	34.44	
389	Ala	Ala	Val	Ser	Leu	Asn	Ile	Leu	Val	5	18.20	43.68	23.11	15.01	
390	Thr	Ala	Val	Ser	Leu	Asn	Ile	Leu	Val	6	9.38	35.45	29.62	25.55	
391	Ala	Val	Val	Ser	Leu	Asn	Ile	Leu	Val	4	18.28	49.40	19.73	12.59	
392	Thr	Val	Val	Ser	Leu	Asn	Ile	Leu	Val	5	7.39	78.19	9.34	5.08	
393	Ala	Ala	Ile	Thr	Leu	Asn	Ile	Leu	Val	5	ND				
394	Thr	Ala	Ile	Thr	Leu	Asn	Ile	Leu	Val	6	ND				
395	Ala	Val	Ile	Thr	Leu	Asn	Ile	Leu	Val	4	25.45	41.99	5.78	26.78	
396	Thr	Val	Ile	Thr	Leu	Asn	Ile	Leu	Val	5	20.32	29.58	6.02	44.08	
397	Ala	Ala	Val	Thr	Leu	Asn	Ile	Leu	Val	4	21.31	61.65	7.42	9.61	
398	Thr	Ala	Val	Thr	Leu	Asn	Ile	Leu	Val	5	ND				
399	Ala	Val	Val	Thr	Leu	Asn	Ile	Leu	Val	3	ND				
400	Thr	Val	Val	Thr	Leu	Asn	Ile	Leu	Val	4	13.67	61.92	8.22	16.18	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
401	Ala	Ala	Ile	Ser	Tyr	Asn	Ile	Leu	Val	5	17.48	34.85	20.17	27.51	
402	Thr	Ala	Ile	Ser	Tyr	Asn	Ile	Leu	Val	6	ND				
403	Ala	Val	Ile	Ser	Tyr	Asn	Ile	Leu	Val	4	12.40	48.03	15.44	24.13	
404	Thr	Val	Ile	Ser	Tyr	Asn	Ile	Leu	Val	5	8.38	78.01	8.39	5.21	
405	Ala	Ala	Val	Ser	Tyr	Asn	Ile	Leu	Val	4	ND				
406	Thr	Ala	Val	Ser	Tyr	Asn	Ile	Leu	Val	5	11.66	38.61	22.88	26.85	
407	Ala	Val	Val	Ser	Tyr	Asn	Ile	Leu	Val	3	5.75	83.41	6.61	4.23	
408	Thr	Val	Val	Ser	Tyr	Asn	Ile	Leu	Val	4	8.62	55.64	17.81	17.92	
409	Ala	Ala	Ile	Thr	Tyr	Asn	Ile	Leu	Val	4	ND				
410	Thr	Ala	Ile	Thr	Tyr	Asn	Ile	Leu	Val	5	15.88	42.07	17.14	24.91	
411	Ala	Val	Ile	Thr	Tyr	Asn	Ile	Leu	Val	3	ND				
412	Thr	Val	Ile	Thr	Tyr	Asn	Ile	Leu	Val	4	10.93	51.52	23.48	14.07	
413	Ala	Ala	Val	Thr	Tyr	Asn	Ile	Leu	Val	3	16.08	70.69	7.54	5.70	
414	Thr	Ala	Val	Thr	Tyr	Asn	Ile	Leu	Val	4	16.25	47.07	12.97	23.72	
415	Ala	Val	Val	Thr	Tyr	Asn	Ile	Leu	Val	2	6.02	65.96	21.85	6.16	
416	Thr	Val	Val	Thr	Tyr	Asn	Ile	Leu	Val	3	18.68	39.34	18.34	23.64	
417	Ala	Ala	Ile	Ser	Leu	Ser	Ile	Leu	Val	5	11.89	39.32	7.68	41.11	
418	Thr	Ala	Ile	Ser	Leu	Ser	Ile	Leu	Val	6	8.50	9.72	11.47	70.31	
419	Ala	Val	Ile	Ser	Leu	Ser	Ile	Leu	Val	4	ND				
420	Thr	Val	Ile	Ser	Leu	Ser	Ile	Leu	Val	5	10.72	24.42	13.14	51.72	
421	Ala	Ala	Val	Ser	Leu	Ser	Ile	Leu	Val	4	12.61	52.36	23.02	12.01	
422	Thr	Ala	Val	Ser	Leu	Ser	Ile	Leu	Val	5	10.13	37.14	26.49	26.24	
423	Ala	Val	Val	Ser	Leu	Ser	Ile	Leu	Val	3	15.47	55.17	21.98	7.38	
424	Thr	Val	Val	Ser	Leu	Ser	Ile	Leu	Val	4	9.84	47.20	21.56	21.40	
425	Ala	Ala	Ile	Thr	Leu	Ser	Ile	Leu	Val	4	27.07	37.87	6.55	28.51	
426	Thr	Ala	Ile	Thr	Leu	Ser	Ile	Leu	Val	5	ND				
427	Ala	Val	Ile	Thr	Leu	Ser	Ile	Leu	Val	3	11.62	52.14	13.04	23.20	
428	Thr	Val	Ile	Thr	Leu	Ser	Ile	Leu	Val	4	26.05	21.05	14.28	38.62	
429	Ala	Ala	Val	Thr	Leu	Ser	Ile	Leu	Val	3	21.83	63.76	7.32	7.09	
430	Thr	Ala	Val	Thr	Leu	Ser	Ile	Leu	Val	4	16.41	46.61	9.46	27.52	
431	Ala	Val	Val	Thr	Leu	Ser	Ile	Leu	Val	2	12.77	58.25	24.70	4.28	
432	Thr	Val	Val	Thr	Leu	Ser	Ile	Leu	Val	3	10.21	68.64	5.86	15.29	
433	Ala	Ala	Ile	Ser	Tyr	Ser	Ile	Leu	Val	4	11.87	43.56	18.69	25.88	
434	Thr	Ala	Ile	Ser	Tyr	Ser	Ile	Leu	Val	5	ND				
435	Ala	Val	Ile	Ser	Tyr	Ser	Ile	Leu	Val	3	14.24	52.02	14.65	19.09	
436	Thr	Val	Ile	Ser	Tyr	Ser	Ile	Leu	Val	4	10.01	11.67	16.14	62.19	
437	Ala	Ala	Val	Ser	Tyr	Ser	Ile	Leu	Val	3	13.51	40.57	38.95	6.97	
438	Thr	Ala	Val	Ser	Tyr	Ser	Ile	Leu	Val	4	ND				
439	Ala	Val	Val	Ser	Tyr	Ser	Ile	Leu	Val	2	6.95	67.22	12.79	13.04	
440	Thr	Val	Val	Ser	Tyr	Ser	Ile	Leu	Val	3	5.86	65.36	18.24	10.54	
441	Ala	Ala	Ile	Thr	Tyr	Ser	Ile	Leu	Val	3	22.74	58.80	6.43	12.04	
442	Thr	Ala	Ile	Thr	Tyr	Ser	Ile	Leu	Val	4	12.75	44.06	12.49	30.70	
443	Ala	Val	Ile	Thr	Tyr	Ser	Ile	Leu	Val	2	ND				
444	Thr	Val	Ile	Thr	Tyr	Ser	Ile	Leu	Val	3	11.30	62.56	6.32	19.81	
445	Ala	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Val	2	20.05	51.93	9.15	18.88	
446	Thr	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Val	3	7.21	71.94	8.56	12.29	
447	Ala	Val	Val	Thr	Tyr	Ser	Ile	Leu	Val	1	19.79	68.84	6.38	4.99	
448	Thr	Val	Val	Thr	Tyr	Ser	Ile	Leu	Val	2	11.79	63.17	18.04	7.01	
449	Ala	Ala	Ile	Ser	Leu	Asn	Thr	Leu	Val	7	ND				
450	Thr	Ala	Ile	Ser	Leu	Asn	Thr	Leu	Val	8	7.84	6.40	9.21	76.55	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
451	Ala	Val	Ile	Ser	Leu	Asn	Thr	Leu	Val	6	15.08	40.02	20.33	24.58	
452	Thr	Val	Ile	Ser	Leu	Asn	Thr	Leu	Val	7	11.43	15.60	11.69	61.28	
453	Ala	Ala	Val	Ser	Leu	Asn	Thr	Leu	Val	6	12.16	52.85	18.07	16.91	
454	Thr	Ala	Val	Ser	Leu	Asn	Thr	Leu	Val	7	12.76	27.83	19.05	40.35	
455	Ala	Val	Val	Ser	Leu	Asn	Thr	Leu	Val	5	13.55	24.42	40.68	21.34	
456	Thr	Val	Val	Ser	Leu	Asn	Thr	Leu	Val	6	13.08	41.33	19.74	25.86	
457	Ala	Ala	Ile	Thr	Leu	Asn	Thr	Leu	Val	6	30.17	30.78	6.02	33.03	
458	Thr	Ala	Ile	Thr	Leu	Asn	Thr	Leu	Val	7	ND				
459	Ala	Val	Ile	Thr	Leu	Asn	Thr	Leu	Val	5	18.10	50.53	7.27	24.11	
460	Thr	Val	Ile	Thr	Leu	Asn	Thr	Leu	Val	6	15.26	37.18	6.97	40.59	
461	Ala	Ala	Val	Thr	Leu	Asn	Thr	Leu	Val	5	19.59	64.04	6.46	9.90	
462	Thr	Ala	Val	Thr	Leu	Asn	Thr	Leu	Val	6	22.97	55.30	21.73	0.00	
463	Ala	Val	Val	Thr	Leu	Asn	Thr	Leu	Val	4	17.82	70.78	4.87	6.53	
464	Thr	Val	Val	Thr	Leu	Asn	Thr	Leu	Val	5	11.94	64.49	7.02	16.55	
465	Ala	Ala	Ile	Ser	Tyr	Asn	Thr	Leu	Val	6	25.51	17.69	43.20	13.60	
466	Thr	Ala	Ile	Ser	Tyr	Asn	Thr	Leu	Val	7	17.58	15.97	23.33	43.11	
467	Ala	Val	Ile	Ser	Tyr	Asn	Thr	Leu	Val	5	14.18	44.60	15.88	25.35	
468	Thr	Val	Ile	Ser	Tyr	Asn	Thr	Leu	Val	6	11.57	33.38	15.68	39.37	
469	Ala	Ala	Val	Ser	Tyr	Asn	Thr	Leu	Val	5	27.42	46.68	12.66	13.24	
470	Thr	Ala	Val	Ser	Tyr	Asn	Thr	Leu	Val	6	12.06	38.70	22.20	27.04	
471	Ala	Val	Val	Ser	Tyr	Asn	Thr	Leu	Val	4	9.02	39.53	24.39	27.06	
472	Thr	Val	Val	Ser	Tyr	Asn	Thr	Leu	Val	5	9.55	56.86	18.99	14.60	
473	Ala	Ala	Ile	Thr	Tyr	Asn	Thr	Leu	Val	5	ND				
474	Thr	Ala	Ile	Thr	Tyr	Asn	Thr	Leu	Val	6	17.35	36.11	5.97	40.58	
475	Ala	Val	Ile	Thr	Tyr	Asn	Thr	Leu	Val	4	ND				
476	Thr	Val	Ile	Thr	Tyr	Asn	Thr	Leu	Val	5	6.04	83.46	6.90	3.60	
477	Ala	Ala	Val	Thr	Tyr	Asn	Thr	Leu	Val	4	16.71	70.28	7.09	5.91	
478	Thr	Ala	Val	Thr	Tyr	Asn	Thr	Leu	Val	5	ND				
479	Ala	Val	Val	Thr	Tyr	Asn	Thr	Leu	Val	3	16.02	23.59	13.81	46.58	
480	Thr	Val	Val	Thr	Tyr	Asn	Thr	Leu	Val	4	14.53	51.54	31.56	2.37	
481	Ala	Ala	Ile	Ser	Leu	Ser	Thr	Leu	Val	6	8.58	24.39	19.85	47.18	
482	Thr	Ala	Ile	Ser	Leu	Ser	Thr	Leu	Val	7	8.82	8.74	10.30	72.14	
483	Ala	Val	Ile	Ser	Leu	Ser	Thr	Leu	Val	5	ND				
484	Thr	Val	Ile	Ser	Leu	Ser	Thr	Leu	Val	6	10.38	21.21	14.21	54.20	
485	Ala	Ala	Val	Ser	Leu	Ser	Thr	Leu	Val	5	9.68	56.44	21.01	12.87	
486	Thr	Ala	Val	Ser	Leu	Ser	Thr	Leu	Val	6	7.29	31.01	23.60	38.10	
487	Ala	Val	Val	Ser	Leu	Ser	Thr	Leu	Val	4	12.25	57.61	21.20	8.93	
488	Thr	Val	Val	Ser	Leu	Ser	Thr	Leu	Val	5	7.96	48.85	18.97	24.22	
489	Ala	Ala	Ile	Thr	Leu	Ser	Thr	Leu	Val	5	17.41	45.92	9.39	27.28	
490	Thr	Ala	Ile	Thr	Leu	Ser	Thr	Leu	Val	6	ND				
491	Ala	Val	Ile	Thr	Leu	Ser	Thr	Leu	Val	4	17.97	52.07	5.77	24.19	
492	Thr	Val	Ile	Thr	Leu	Ser	Thr	Leu	Val	5	11.72	41.12	6.89	40.26	
493	Ala	Ala	Val	Thr	Leu	Ser	Thr	Leu	Val	4	13.69	32.53	4.93	48.85	
494	Thr	Ala	Val	Thr	Leu	Ser	Thr	Leu	Val	5	ND				
495	Ala	Val	Val	Thr	Leu	Ser	Thr	Leu	Val	3	ND				
496	Thr	Val	Val	Thr	Leu	Ser	Thr	Leu	Val	4	10.72	68.37	7.85	13.06	
497	Ala	Ala	Ile	Ser	Tyr	Ser	Thr	Leu	Val	5	9.58	52.92	21.49	16.01	
498	Thr	Ala	Ile	Ser	Tyr	Ser	Thr	Leu	Val	6	ND				
499	Ala	Val	Ile	Ser	Tyr	Ser	Thr	Leu	Val	4	10.88	39.63	22.57	26.92	
500	Thr	Val	Ile	Ser	Tyr	Ser	Thr	Leu	Val	5	19.56	33.93	13.41	33.09	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.4. Gas chromatography – mass spectrometry data of M9 mutant proteins (cont.)

	Mutated positions										Product output[b]				
Number	274	291	372	402*	406	436	438	439	516*	M[a]	MP	5-EA (2)	4-EE (4)	PSD (3)	Notes
501	Ala	Ala	Val	Ser	Tyr	Ser	Thr	Leu	Val	4	8.46	67.08	18.30	6.16	
502	Thr	Ala	Val	Ser	Tyr	Ser	Thr	Leu	Val	5	ND				
503	Ala	Val	Val	Ser	Tyr	Ser	Thr	Leu	Val	3	5.97	73.62	16.09	4.32	
504	Thr	Val	Val	Ser	Tyr	Ser	Thr	Leu	Val	4	15.74	29.96	27.55	26.74	
505	Ala	Ala	Ile	Thr	Tyr	Ser	Thr	Leu	Val	4	ND				
506	Thr	Ala	Ile	Thr	Tyr	Ser	Thr	Leu	Val	5	24.22	3.42	17.67	54.68	
507	Ala	Val	Ile	Thr	Tyr	Ser	Thr	Leu	Val	3	14.32	74.80	5.65	5.22	
508	Thr	Val	Ile	Thr	Tyr	Ser	Thr	Leu	Val	4	8.90	51.33	28.68	11.09	
509	Ala	Ala	Val	Thr	Tyr	Ser	Thr	Leu	Val	3	24.18	45.61	0.00	30.21	
510	Thr	Ala	Val	Thr	Tyr	Ser	Thr	Leu	Val	4	8.10	68.09	8.42	15.39	
511	Ala	Val	Val	Thr	Tyr	Ser	Thr	Leu	Val	2	10.24	58.62	11.78	19.37	
512	Thr	Val	Val	Thr	Tyr	Ser	Thr	Leu	Val	3	6.67	79.61	6.51	7.21	

* Active site residues

[1] Reference for Greenhagen et al (2006)[12]

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (MP) remaining minor products. ND indicates no data.

Table 2.5. Kinetic measurements of selected library mutants

	Mutated positions										Products[b]							
Number	274	291	372	402*	406	436	438	439	516*	M[a]	GA	5-EA	4-EE	PSD	Kcat app (min-1)	STDEV	Fold difference[c]	Notes
13	Ala	Ala	Val	Thr	Leu	Asn	Ile	Ile	Ile	4	9.27	77.34	9.58	3.81	12.91	4.28	1.71	
189	Ala	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Ile	3	8.11	63.84	20.83	7.22	12.75	3.52	1.69	
492	Thr	Val	Ile	Thr	Leu	Ser	Thr	Leu	Val	5	4.05	44.70	7.49	43.76	10.28	2.09	1.36	
109	Ala	Ala	Val	Thr	Leu	Ser	Thr	Ile	Ile	4	16.67	47.55	31.96	3.82	9.93	3.32	1.32	
319	Ala	Val	Val	Thr	Tyr	Ser	Ile	Ile	Val	0	6.76	81.11	9.49	2.63	7.54	2.22	1.00	TEAS wt
445	Ala	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Val	2	16.77	41.14	16.01	26.09	5.56	2.92	1.36	
397	Ala	Ala	Val	Thr	Leu	Asn	Ile	Leu	Val	4	16.49	65.43	7.88	10.20	5.39	3.22	1.40	
240	Thr	Val	Val	Thr	Leu	Ser	Thr	Leu	Ile	5	6.48	64.11	19.53	9.87	4.91	2.38	1.54	
262	Thr	Ala	Val	Ser	Leu	Asn	Ile	Ile	Val	5	3.72	23.25	16.42	56.61	4.88	1.46	1.55	
260	Thr	Val	Ile	Ser	Leu	Asn	Ile	Ile	Val	5	9.10	45.47	17.21	28.22	3.85	1.45	1.96	
190	Thr	Ala	Val	Thr	Tyr	Ser	Ile	Leu	Ile	4	8.60	55.83	22.69	12.88	3.49	1.05	2.16	
458	Thr	Ala	Ile	Thr	Leu	Asn	Thr	Leu	Val	7	10.01	32.41	6.18	51.40	2.82	0.33	2.67	
305	Ala	Ala	Ile	Ser	Tyr	Ser	Ile	Ile	Val	3	6.44	53.77	22.54	17.26	2.79	0.96	2.70	
381	Ala	Ala	Val	Thr	Tyr	Ser	Thr	Ile	Val	2	9.43	81.75	6.50	2.33	2.28	1.26	3.31	
400	Thr	Val	Val	Thr	Leu	Asn	Ile	Leu	Val	4	6.93	66.76	8.86	17.44	2.09	1.24	3.61	
497	Ala	Ala	Ile	Ser	Tyr	Ser	Thr	Leu	Val	5	5.53	55.29	22.45	16.73	1.82	0.55	4.15	
263	Ala	Val	Val	Ser	Leu	Asn	Ile	Ile	Val	3	7.09	69.87	16.33	6.71	1.69	1.02	4.46	
496	Thr	Val	Val	Thr	Leu	Ser	Thr	Leu	Val	4	5.52	72.35	8.31	13.82	1.68	0.35	4.48	
44	Thr	Val	Ile	Thr	Leu	Ser	Ile	Ile	Ile	4	11.61	32.42	29.04	26.93	1.48	0.53	5.08	
173	Ala	Ala	Val	Thr	Leu	Ser	Ile	Leu	Ile	4	20.19	52.84	19.77	7.21	1.36	0.32	5.56	
465	Ala	Ala	Ile	Ser	Tyr	Asn	Thr	Leu	Val	6	10.61	21.23	51.84	16.32	1.18	0.43	6.37	
347	Ala	Val	Ile	Thr	Tyr	Asn	Thr	Ile	Val	3	4.81	22.48	19.15	53.56	1.15	0.35	6.53	
269	Ala	Ala	Val	Thr	Leu	Asn	Ile	Ile	Val	3	16.60	60.89	7.23	15.28	0.96	0.20	7.89	
486	Thr	Ala	Val	Ser	Leu	Ser	Thr	Leu	Val	6	1.04	31.85	26.33	40.78	0.85	0.48	8.90	
395	Ala	Val	Ile	Thr	Leu	Asn	Ile	Leu	Val	4	16.27	47.16	6.49	30.08	0.69	0.21	10.89	
349	Ala	Ala	Val	Thr	Tyr	Asn	Thr	Ile	Val	3	13.47	73.61	7.39	5.53	0.67	0.20	11.29	
55	Ala	Val	Val	Ser	Tyr	Ser	Ile	Ile	Ile	2	6.96	29.09	61.40	2.55	0.63	0.19	11.97	EES
195	Ala	Val	Ile	Ser	Leu	Asn	Thr	Leu	Ile	7	10.19	29.76	19.79	40.26	0.61	0.18	12.31	
104	Thr	Val	Val	Ser	Leu	Ser	Thr	Ile	Ile	5	4.20	24.24	60.64	10.92	0.51	0.15	14.72	
395	Ala	Val	Ile	Thr	Leu	Asn	Ile	Leu	Val	4	16.27	47.16	6.49	30.08	0.49	0.15	15.33	
437	Ala	Ala	Val	Ser	Tyr	Ser	Ile	Leu	Val	3	6.88	43.68	41.94	7.51	0.34	0.10	22.01	
245	Ala	Ala	Val	Ser	Tyr	Ser	Thr	Leu	Ile	5	5.75	38.08	42.54	13.63	0.32	0.07	23.24	
209	Ala	Ala	Ile	Ser	Tyr	Asn	Thr	Leu	Ile	7	12.98	32.19	23.28	31.56	0.29	0.23	26.28	
341	Ala	Ala	Val	Ser	Tyr	Asn	Thr	Ile	Val	4	4.36	64.84	24.27	6.53	0.29	0.09	26.28	
221	Ala	Ala	Val	Thr	Tyr	Asn	Thr	Leu	Ile	5	29.36	53.05	11.20	6.39	0.24	0.04	31.93	
294	Thr	Ala	Val	Ser	Leu	Ser	Ile	Ile	Val	4	3.83	37.62	31.43	27.12	0.14	0.04	54.10	
								HPS wt		0	1.74	0.00	1.42	96.84	16.70	5.03	2.21	
								HPS_M8		8	16.39	72.45	5.61	5.55	1.47	1.32	5.11	

* Active site residues

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene 5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (GA, 5) germacrene A.

[c] Fold difference refers to comparison of the $k_{cat\ app}$ versus TEAS wt reference sequence, where numbers in blue or red are above and below, respectively.

Table 2.6. Average chemical distances for each position

Position	Average distance[a]	STDEV
274	32.92	17.75
291	31.08	20.80
372	28.64	21.30
402	31.56	20.60
406	32.23	20.07
436	35.85	18.19
438	40.71	18.94
439	35.81	22.26
516	34.27	18.15

[a] Distances from pairwise alignment of GC-MS quantified products (Materials and methods) were tabulated and averaged for each position throughout the library.

Table 2.7. Influence of active site substitutions on product specificity

		Mutated positions										Products[b]			
Activity	Number	274	291	372	402*	406	436	438	439	516*	M[a]	GA	5-EA	4-EE	PSD
HPS-like	226	Thr	Ala	Ile	Ser	Leu	Ser	Thr	Leu	Ile	8	2.75	4.22	11.82	81.21
	450	Thr	Ala	Ile	Ser	Leu	Asn	Thr	Leu	Val	8	7.85	2.92	12.60	76.62
	79	Ala	Val	Val	Thr	Leu	Asn	Thr	Ile	Ile	4	3.60	6.46	14.33	75.60
	283	Ala	Val	Ile	Thr	Tyr	Asn	Ile	Ile	Val	2	5.06	16.13	13.85	64.96
EES-like	119	Ala	Val	Val	Ser	Tyr	Ser	Thr	Ile	Ile	3	2.76	32.21	62.39	2.65
	264	Thr	Val	Val	Ser	Leu	Asn	Ile	Ile	Val	4	4.84	33.23	59.22	2.71
	256	Thr	Val	Val	Thr	Tyr	Ser	Thr	Leu	Ile	4	2.44	36.38	47.51	13.68
	480	Thr	Val	Val	Thr	Tyr	Asn	Thr	Leu	Val	4	3.65	56.28	36.27	3.80
TEAS-like	87	Ala	Val	Val	Ser	Tyr	Asn	Thr	Ile	Ile	4	5.20	81.63	8.35	4.82
	407	Ala	Val	Val	Ser	Tyr	Asn	Ile	Leu	Val	3	3.63	85.28	6.76	4.33
	203	Ala	Val	Ile	Thr	Leu	Asn	Thr	Leu	Ile	6	3.90	81.60	10.07	4.43
	319	Ala	Val	Val	Thr	Tyr	Ser	Ile	Ile	Val	0	5.16	87.05	6.28	1.50
GAS-like[1]	37	Ala	Ala	Val	Ser	Leu	Ser	Ile	Ile	Ile	4	38.04	34.41	20.74	6.81
	274	Thr	Ala	Ile	Ser	Tyr	Asn	Ile	Ile	Val	5	23.45	41.84	19.67	15.04
	219	Ala	Val	Ile	Thr	Tyr	Asn	Thr	Leu	Ile	5	44.47	39.85	5.79	9.89
	457	Ala	Ala	Ile	Thr	Leu	Asn	Thr	Leu	Val	6	23.99	33.51	6.55	35.95

[*] Active site residues

[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)

[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (GA, 5) germacrene A.

[c] Fold difference refers to comparison of the Kcat app versus TEAS wt reference sequence, where numbers in blue or red are above and below, respectively.

Table 2.8. Minimal combinations of mutations converting TEAS to HPS-like product specificity.

	Mutated positions										Products[b]			
Number	274	291	372	402*	406	436	438	439	516*	M[a]	GA	5-EA	4-EE	PSD
194	Thr	Ala	Ile	Ser	Leu	Asn	Thr	Leu	Ile	9	11.09	7.07	9.77	72.07
283	Ala	Val	Ile	Thr	Tyr	Asn	Ile	Ile	Val	2	5.06	16.13	13.85	64.96
479	Ala	Val	Val	Thr	Tyr	Asn	Thr	Leu	Val	3	7.85	25.88	15.16	51.11
493	Ala	Ala	Val	Thr	Leu	Ser	Thr	Leu	Val	4	6.35	35.29	5.35	53.00
188	Thr	Val	Ile	Thr	Tyr	Ser	Ile	Leu	Ile	4	3.46	12.27	24.85	59.42
436	Thr	Val	Ile	Ser	Tyr	Ser	Ile	Leu	Val	4	3.52	12.51	17.30	66.67
344	Thr	Val	Val	Ser	Tyr	Asn	Thr	Ile	Val	4	6.94	5.89	13.13	74.04
79	Ala	Val	Val	Thr	Leu	Asn	Thr	Ile	Ile	4	3.43	6.38	13.99	76.20
257	Ala	Ala	Ile	Ser	Leu	Asn	Ile	Ile	Val	5	7.41	22.55	18.57	51.47
20	Thr	Val	Ile	Ser	Tyr	Asn	Ile	Ile	Ile	5	15.53	18.66	11.18	54.63
420	Thr	Val	Ile	Ser	Leu	Ser	Ile	Leu	Val	5	1.65	26.90	14.48	56.97
152	Thr	Val	Val	Ser	Tyr	Asn	Ile	Leu	Ile	5	8.64	14.82	18.98	57.56
42	Thr	Ala	Ile	Thr	Leu	Ser	Ile	Ile	Ile	5	19.01	13.55	9.01	58.43
506	Thr	Ala	Ile	Thr	Tyr	Ser	Thr	Leu	Val	5	17.24	3.73	19.30	59.72
26	Thr	Ala	Ile	Thr	Tyr	Asn	Ile	Ile	Ile	5	6.38	22.01	11.24	60.37

[*] Active site residues
[a] M refers to the number of mutated positions of TEAS (the sum of shaded positions)
[b] Product output, expressed as a percentage of total ion chromatogram from GC-MS data, is composed of 5-epiaristolochene (5-EA, 2), 4-epi-eremophilene (4-EE, 4), premnaspirodiene (PSD, 3), and (GA, 5) germacrene A.

ACKNOWLEDGEMENTS

The text of chapter 2, in full, is a reprint of the material as it appears in Nature Chemical Biology 2008, Vol. 18, pp. 3039-3042. Permission was obtained from the co-authors. I was the third author of this work. As mentioned in the manuscript, Paul O'maille designed the study, conducted experiments, analyzed data and wrote the manuscript, Arthur Malone conducted experiments and developed small-scale protein purification, I conducted experiments, analyzed data and contributed revision to the manuscript, B Andes Hess Jr conducted quantum mechanics calculations and contributed revisions to the manuscript, Lidia Smentek conducted quantum mechanics calculations, Iseult Sheehan conducted experiments, Bryan Greenhagen and Joseph Chappell designed the study and contributed revisions to the manuscript, Gerard Manning analyzed data, developed the biosynthetic tree and chemical

distance analysis, and contributed revisions to the manuscript, and Joseph P. Noel designed the study, analyzed the data and wrote the manuscript. This research was performed under the supervision of Joseph P. Noel.

REFERENCES

1. Grayer, R. J.; Kokubun, T., Plant-fungal interactions: the search for phytoalexins and other antifungal compounds from higher plants. *Phytochemistry* **2001,** *56*, 253-263.

2. Pedras, M. S.; Okanga, F. I.; Zaharia, I. L.; Khan, A. Q., Phytoalexins from crucifers: synthesis, biosynthesis, and biotransformation. *Phytochemistry* **2000,** *53*, 161-176.

3. Harborne, J. B., The comparative biochemistry of phytoalexin induction in plants. *Biochem. Syst. Ecol.* **1999,** *27*, 335-367.

4. Akiyama, K.; Matsuzaki, K.; Hayashi, H., Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **2005,** *435*, 824-827.

5. Mumm, R.; Hilker, M., The significance of background odour for an egg parasitoid to detect plants with host eggs. *Chem. Senses* **2005,** *30*, 337-343.

6. Feeny, P., Herbivores: Their Interactions with Secondary Plant Metabolites. Academic Press: 1992.

7. Gershenzon, J.; Dudareva, N., The function of terpene natural products in the natural world. *Nature chemical biology* **2007,** *3* (7), 408-414.

8. O'Maille, P. E.; Bakhtina, M.; Tsai, M. D., Structure-based combinatorial protein engineering (SCOPE). *Journal of Molecular Biology* **2002,** *321* (4), 677-691.

9. O'Maille, P. E.; Chappell, J.; Noel, J. P., A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases. *Anal. Biochem.* **2004,** *335*, 210-217.

10. Back, K.; He, S.; Kim, K. U.; Shin, D. H., Cloning and bacterial expression of sesquiterpene cyclase, a key branch point enzyme for the synthesis of sesquiterpenoid phytoalexin capsidiol in UV-challenged leaves of Capsicum annuum. *Plant Cell Physiol.* **1998,** *39*, 899-904.

11. Facchini, P. J.; Chappell, J., Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. *Proc. Natl. Acad. Sci. USA* **1992,** *89*, 11088-11092.

12. Greenhagen, B. T.; O'Maille, P. E.; Noel, J. P.; Chappell, J., Identifying and manipulating structural determinates linking catalytic specificities in terpene

synthases. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (26), 9826-9831.

13. Dudareva, N., (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell* **2003,** *15*, 1227-1241.

14. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R., Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **1998,** *95*, 4126-4133.

15. O'Maille, P. E.; Tsai, M. D.; Greenhagen, B. T.; Chappell, J.; Noel, J. P., Gene library synthesis by structure-based combinatorial protein engineering. *Methods Enzymol.* **2004,** *388*, 75-91.

16. O'Maille, P. E.; Chappell, J.; Noel, J. P., Biosynthetic potential of sesquiterpene synthases: Alternative products of tobacco 5-epi-aristolochene synthase. *Archives of Biochemistry and Biophysics* **2006,** *448* (1-2), 73-82.

17. Copley, S. D., Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* **2003,** *7*, 265-272.

18. Jensen, R. A., Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **1976,** *30*, 409-425.

19. O'Brien, P. J.; Herschlag, D., Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **1999,** *6*, R91-R105.

20. Wilderman, P. R.; Peters, R. J., A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *J. Am. Chem. Soc.* **2007,** *129*, 15736-15737.

21. Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D., Designed divergent evolution of enzyme function. *Nature* **2006,** *440* (7087), 1078-1082.

22. Hyatt, D. C.; Croteau, R., Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (-)-pinene synthase from Abies grandis. *Arch. Biochem. Biophys.* **2005,** *439*, 222-233.

23. Kampranis, S. C.; Ioannidis, D.; Purvis, A.; Mahrez, W.; Ninga, E.; Katerelos, N. A.; Anssour, S.; Dunwell, J. M.; Degenhardt, J.; Makris, A. M.; Goodenough, P. W.; Johnson, C. B., Rational conversion of substrate and product specificity in a Salvia monoterpene synthase: structural insights into the evolution of terpene synthase function. *The Plant Cell* **2007,** *19* (6), 1994-2005.

24. Kollner, T. G.; Schnee, C.; Gershenzon, J.; Degenhardt, J., The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* **2004,** *16*, 1115-1131.

25. Weinreich, D. M.; Delaney, N. F.; Depristo, M. A.; Hartl, D. L., Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **2006,** *312*, 111-114.

26. Ortlund, E. A.; Bridgham, J. T.; Redinbo, M. R.; Thornton, J. W., Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **2007,** *317*, 1544-1548.

27. Bershtein, S.; Segal, M.; Bekerman, R.; Tokuriki, N.; Tawfik, D. S., Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **2006,** *444*, 929-932.

28. Miller, S. P.; Lunzer, M.; Dean, A. M., Direct demonstration of an adaptive constraint. *Science* **2006,** *314*, 458-461.

29. Thulasiram, H. V.; Erickson, H. K.; Poulter, C. D., Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science* **2007,** *316*, 73-76.

30. Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. T.; Benkovic, S. J.; Hammes-Schiffer, S., Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **2002,** *99*, 2794-2799.

31. Rajagopalan, P. T.; Lutz, S.; Benkovic, S. J., Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: mutational effects on hydride transfer rates. *Biochemistry* **2002,** *41*, 12618-12628.

32. Lockless, S. W.; Ranganathan, R., Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **1999,** *286*, 295-299.

33. Austin, M. B.; O'Maille, P. E.; Noel, J. P., Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nat. Chem. Biol.* **2008,** *4*, 217-222.

34. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S., MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **2007,** *24*, 1596-1599.

35. Ellis, K. J.; Morrison, J. F., Buffers of constant ionic strength for studying pH-dependent processes. *Methods Enzymol* **1982,** *87*, 405-26.

36. Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **1989,** *182* (2), 319-26.

Chapter 3

Structural Elucidation of Cisoid and Transoid Cyclization Pathways of a Sesquiterpene Synthase Using 2-Fluorofarnesyl Diphosphates

3.1. ABSTRACT

Sesquiterpene skeletal complexity in nature originates from the enzyme-catalyzed ionization of (trans,trans)-farnesyl diphosphate (FPP) (1a) and subsequent cyclization along either 2,3-transoid or 2,3-cisoid farnesyl cation pathways. Tobacco 5-epi-aristolochene synthase (TEAS), a transoid synthase, produces cisoid products as a component of its minor product spectrum. To investigate the cryptic cisoid cyclization pathway in TEAS, we employed (cis,trans)-FPP (1b) as an alternative substrate. Strikingly, TEAS was catalytically robust in the enzymatic conversion of (cis,trans)-FPP (1b) to exclusively (≥99.5%) cisoid products. Further, crystallographic characterization of wild-type TEAS and a catalytically promiscuous mutant (M4 TEAS) with 2-fluoro analogues of both all-trans FPP (1a) and (cis,trans)-FPP (1b) revealed binding modes consistent with preorganization of the farnesyl chain. These results provide a structural glimpse into both cisoid and transoid cyclization pathways efficiently templated by a single enzyme active site, consistent with the recently elucidated stereochemistry of the cisoid products. Further, computational studies using density functional theory calculations reveal concerted, highly asynchronous cyclization pathways leading to the major cisoid cyclization products. The implications of these discoveries for expanded sesquiterpene diversity in nature are discussed.

3.2. INTRODUCTION

Terpenes comprise the most structurally diverse class of natural products, playing essential ecological roles by mediating communication between plants and insects, by providing antimicrobial defenses for plants, and likely acting in additional undefined capacities (as reviewed previously[1]. Terpenoids originate from primary isoprenoid metabolism, wherein iterative condensation of 5-carbon isoprene units (isopentenyl

diphosphate and dimethylallyl diphosphate) catalyzed by prenyltransferases produce polyisoprenoid diphosphate substrates of varying lengths (for a review, see Liang et al 2002[2]). Terpene synthases, in turn, often referred to as cyclases given the cyclic nature of many of their products, transform the polyisoprenoid diphosphate substrates, e.g., geranyl diphosphate C10, farnesyl diphosphate C15, or geranylgeranyl diphosphate C20, into structurally diverse mono-, sesqui-, and diterpene products, respectively.

The structural complexity of these molecules underlies their diverse biological activities. Ruzicka formulated the biogenetic isoprene rule, which predicted the formation of sesquiterpenes arising from the head-to-tail connection of three isoprene units (5 carbons each) where the skeletal complexity can be formally deduced from farnesol[3]. Cane later developed a general stereochemical model for sesquiterpene biogenesis involving the idealized fold of the farnesyl chain in the active site posing the reacting carbons to direct a sequence of electrophilic cyclizations and rearrangements following pyrophosphate loss/ionization[4]. Moreover, a limited number of conformations of the farnesyl substrate give rise to much greater product diversity. Product specificity or in many cases product diversity arises from a limited number of farnesyl chain conformations, wherein the reacting double bonds reside mutually perpendicular to a common plane. Thus, there is a direct correspondence between the absolute stereochemical configuration of the sesquiterpene product and the inferred conformation of the precursor[4]. A central challenge for the structural enzymologist is to define how individual terpene synthases statically or dynamically discriminate between alternative polyprenyl cation conformational modes or selectively favor particular conformations to shepherd reactive intermediates along distinct cyclization cascades.

Structural biology provides a framework for addressing the evolutionary origins of complex terpenoid metabolites and their biosynthetic pathways. Terpene synthases comprise a structurally conserved enzyme family, which adopt a common α-helical architecture termed the class I terpenoid cyclase fold, first revealed from the crystal structures of tobacco 5-epi-aristolochene synthase (TEAS) from Nicotiana tobaccum[5] and pentalenene synthase from Streptomyces UC5319[6]. The lyase function of these enzymes stems from two conserved metal binding motifs: the "aspartate-rich" DDxxD motif that coordinates two Mg^{2+} ions and the "NSE/DTE" motif that coordinates a third Mg^{2+}. These static X-ray crystallographic studies show that the binding of $(Mg^{2+})_3$-PPi stabilizes the active site in a closed conformation that is sequestered from bulk solvent[7]. In addition to multiple divalent cation coordination bonds, the PPi anion accepts hydrogen bonds from conserved basic residues when bound to the closed synthase conformation, while a hydrophobic pocket, lined by a number of aromatic residues, cradles the farnesyl chain and most likely templates the cyclization reaction by enforcing particular substrate conformations and stabilizing carbocations through π-stacking interactions.

While the wealth of structural diversity among terpene hydrocarbons arises from bifurcations along multistep cyclization pathways, divergence at the earliest mechanistic step defines two major classes of terpene synthases and hence distinct product families. The "transoid" synthases ionize (trans,trans)-farnesyl diphosphate (FPP) (1a) to generate the 2,3-transoid farnesyl cation (trans along the C2−C3 bond) followed by initial C1 attack on the distal C10−C11 double bond, prior to further downstream cyclizations (Figure 3.1). By contrast, the "cisoid" synthases conduct an initial C2−C3 double bond isomerization prior to cyclization, wherein the nascent farnesyl cation is recaptured at C3 by pyrophosphate to form the neutral, enzyme-bound (3R)- or (3S)-nerolidyl diphosphate (NPP), thus allowing rotation

about the C2−C3 bond from trans to cis. Reionization of NPP generates the 2,3-cisoid farnesyl cation, which undergoes further cyclization via initial C1 attack either on the proximal C6−C7 or on the distal C10−C11 double bonds prior to further transformations. This reaction mechanism has been invoked to account for the formation of β-macrocarpene[8], amorpha-4,11-diene[9], trichodiene[10], and cedranes such as isocedrol[11]. Moreover, the biosynthesis of epi-isozizaene was recently described in connection with the isolation and functional characterization of epi-isozizaene synthase from Streptomyces coelicolor[12]. The cisoid mechanistic class in sesquiterpene cyclases is akin to the majority of monoterpene cyclases that proceed via the cis (neryl) allylic cation to form the corresponding cyclic monoterpene products (for a review, see Davis et al 2000[13]).

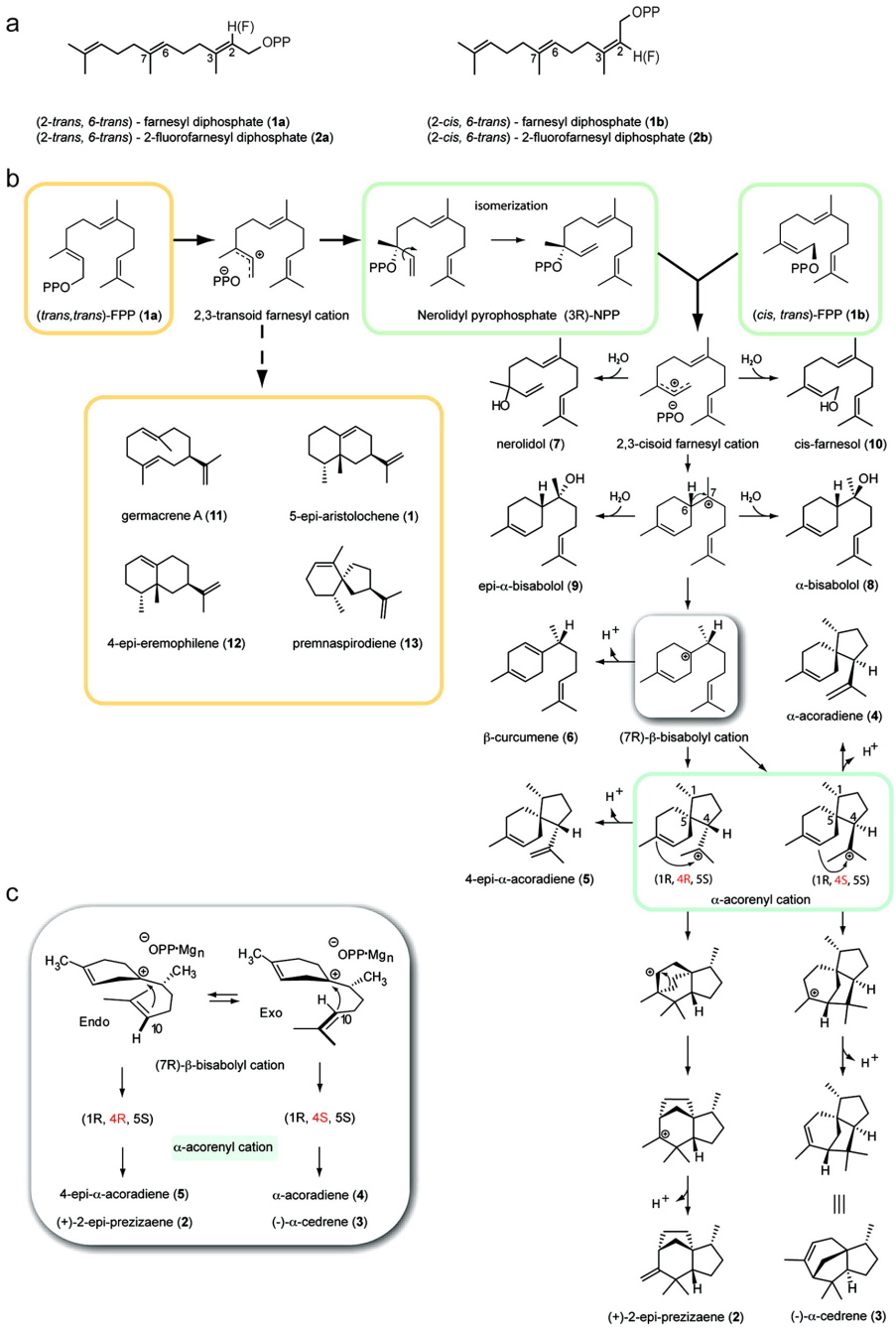


Figure 3.1. Mechanism of TEAS-catalyzed cyclization of (cis,trans)-FPP to (+)-2-epi-prezizaene (2). a) The structure and semisystematic nomenclature for isomers of the FPP substrate and fluorinated analogues used in this study are indicated. b) Based on biochemical and stereochemical information governing the nature of the cisoid products, a cyclization mechanism is proposed to account for all identified products along this multistep pathway.[18] c) The configuration of the terminal isopropenyl tail of the (7R)-β-bisabolyl cation relates to the final cyclization products of the cisoid pathway.

Though most sesquiterpene synthases can be classified as belonging to either the cisoid or transoid classes, some display cryptic activities associated with the other pathway. Notably, TEAS catalyzes the formation of (+)-5-epi-aristolochene (1), the first committed step in the biosynthesis of the phytoalexin capsidiol, the principal component of tobacco's antifungal chemical defense[14]. Aside from its major product, TEAS generates an additional 24 minor products, some of which are derived from the cisoid cyclization pathway[15]. The structural variations of these cisoid products result from a multistep mechanism of cyclizations and rearrangements, suggesting that TEAS templates the cisoid cation pathway with fidelity and enables the formation of a distinct set of complex skeletal structures. These unexpected observations give rise to several confounding questions. Does a single parental fold of the farnesyl chain give rise to products along both the cisoid and transoid pathways in TEAS? On a structural level, how are both pathways templated within a single active site? How can the cryptic cisoid pathway in TEAS become activated? Does this "vestigial" activity portend an unanticipated new function for TEAS in tobacco?

To address these questions, we investigated the cisoid cyclization recently discovered in TEAS using synthetically derived (cis,trans)-FPP (1b), a geometrical isomer of the native all-trans substrate (Figure 3.1, panel a). [The descriptors "cis" and "trans" in (cis,trans)-FPP and the fluoroFPP isomers refer to the longest carbon chain about the 2,3 and 6,7 double bonds, respectively, as defined in Figure 3.1, panel a. For more formalized nomenclature, see Fox et al 2001[16] and Rigaudy et al 1979[17]] Remarkably, TEAS efficiently converts this alternative substrate into predominantly (+)-2-epi-prezizaene (2), a novel sesquiterpene hydrocarbon related to the naturally occurring alcohol jinkohol[18], along with other cisoid cation-derived products. Large-scale enzyme reactions produced sufficient amounts of

hydrocarbon products for stereochemical elucidation and positive identification of nine compounds[18].

In the present investigation, crystallographic analyses of wild-type TEAS and a previously reported promiscuous mutant (TEAS M4)[19], with unreactive 2-fluoro analogues (2a and 2b) of (trans,trans)- and (cis,trans)-FPP (1a and 1b, respectively) revealed catalytically relevant binding modes and distinct farnesyl chain topologies that are consistent with preorganization by the active-site for cisoid or transoid cyclization, and hence, the predicted stereochemical course of the reaction. Further, key transition state geometries calculated using density functional theory revealed concerted, highly asynchronous cyclization pathways. Thus, combining biochemical, computational, and crystallographic analyses with the recently elucidated stereochemistry of the cisoid products, we pictorially reconstruct herein the TEAS-catalyzed transoid and cisoid cyclization pathways. Further, comparison of wild-type and mutant TEAS-analogue complexes provides structural snapshots and insights into product specificity/diversity reflected in the preorganization of the farnesyl chain along 2 major cyclization pathways.

3.3. RESULTS AND DISCUSSION

3.3.1. TEAS-Directed Cisoid Cyclization with (cis,trans)-FPP

To investigate the cryptic cyclization activity in TEAS, we synthesized the 2,3-cis geometrical isomer of farnesyl diphosphate (cis,trans)-FPP (1b)[18]. This substrate analogue is effectively "preisomerized", and hence its ionization by TEAS would be expected to generate the cisoid farnesyl cation, which in turn should feed directly into the cisoid cyclization pathway (Figure 3.1, panel b). Indeed, our pilot experiments revealed that TEAS generated a near exclusive spectrum of cisoid products when incubated with (cis,trans)-FPP (1b) as

substrate, including the previously reported iso-prezizaene ((+)-2-epi-prezizaene, 2) as the dominant reaction product (Table 3.1, Figure 3.2). This result demonstrated the ability of TEAS to template the cisoid cyclization pathway with a high degree of product specificity and catalytic efficiency.

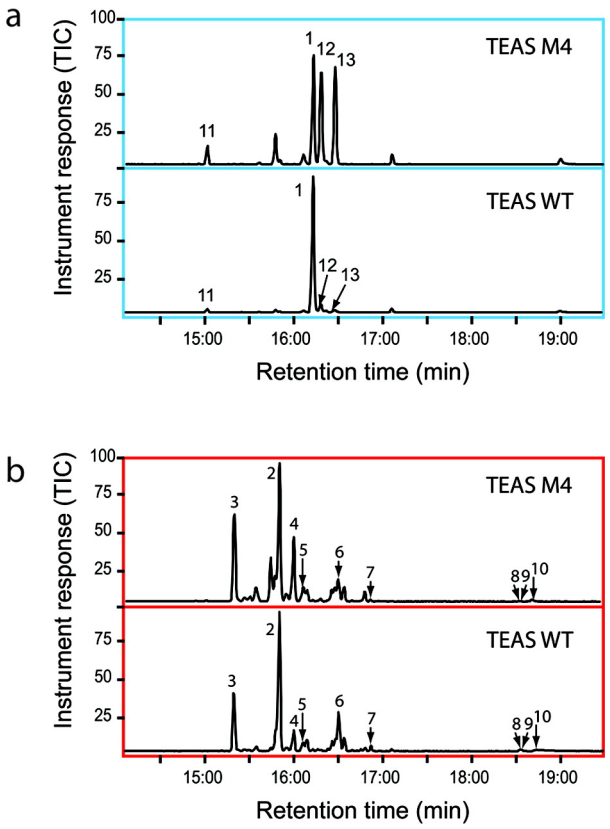


Figure 3.2. Gas chromatograms of products from incubations of wild-type TEAS and the M4 mutant with (cis,trans)- and (trans,trans)-FPP. TEAS or its M4 mutant were incubated with either (a) (trans,trans)-FPP (1a) or (b) (cis,trans)-FPP (1b) using the vial assay followed with analysis by GC–MS as described in Methods. Major product peaks are labeled according to identified products listed in Table 3.1.

As detailed in a concurrent report, the structure, stereochemistry, and enantio-purity was determined for nine cisoid products of TEAS isolated from large-scale enzyme incubations with (cis,trans)-FPP (1b), an achievement enabling the formulation of a mechanistic proposal for their biosynthetic origin (Figure 3.1, panel b)[18]. Chromatographic

separations or enrichments of five hydrocarbon and three alcohol fractions, together with comparative NMR spectral data, chiral GC analyses, optical rotation measurements, and chemical correlations, allowed assignment of structures for (+)-2-epi-prezizaene (2), (−)-α-cedrene (3), α-acoradiene (4), 4-epi-α-acoradiene (5), (−)-β-curcumene (6), nerolidol (7), the α-bisabalol epimers (8 and 9), and 2,3-cis-farnesol (10) shown in Figure 3.1, panel b and listed in Table 3.1. Importantly, knowledge of the relative and absolute stereochemistry of the cisoid hydrocarbon products provided vital information for guiding computational studies and accurately describing the stereochemical course of the TEAS cisoid cyclization reaction mechanism (Figure 3.1, panel b and below).

To further characterize the product specificity of wild-type TEAS and a previously described promiscuous mutant (A274T V372I Y406L V516I) referred to here as M4 TEAS[19], we performed GC−MS analyses via the vial assay[20] to examine product mass spectra derived from native (1a) and (cis,trans)-FPP (1b) substrates (Figure 3.2, Table 3.1). TEAS generated 18 distinct terpene products from the 2-cis substrate, including (+)-2-epi-prezizaene (2), which constitutes nearly half the product spectrum (46% by TIC). With the native all-trans FPP (1a) substrate, M4 TEAS exhibits relaxed product specificity in the terminal steps of the transoid cyclization pathway, producing roughly equal amounts of 5-epi-aristolochene (1), 4-epi-eremophilene (12), and premnaspirodiene (13). With (cis,trans)-FPP (1b), M4 TEAS produced the same repertoire of hydrocarbons as the wild-type enzyme, again displaying relaxed product specificity with only a third of the turnovers producing (+)-2-epi-prezizaene (2). This result indicates that the catalytic promiscuity of M4 TEAS extends to the cisoid pathway.

To assess the efficiency of the enzymatic conversion of (cis,trans)-FPP (1b) to (+)-2-epi-prezizaene (2) by wild-type TEAS and M4 TEAS, we conducted steady-state kinetic experiments (Table 3.2). The experiments using (trans,trans)-FPP (1a) reveal that both

enzymes display comparable catalytic efficiencies (k_{cat}/K_m), with the higher K_m of M4 TEAS (13.3 μM) offset by a higher overall turnover number (9.4 min^{-1}). Both enzymes utilize (cis,trans)-FPP (1b) with similar catalytic efficiencies, with the promiscuous TEAS M4 possessing a K_m (7.9 μM) lower than that of wild-type but also a slower overall turnover (4.6 min^{-1}), the reverse of the trend observed for (trans,trans)-FPP (1a). Additionally, enzymatic activities were assessed using the 2-fluoro analogues of (trans,trans)-FPP (2a) and (cis,trans)-FPP (2b) (Figure 3.1, panel a). Wild-type TEAS and several of its catalytically active mutants fail to turn over either (2-trans,6-trans)-2-fluorofarnesyl diphosphate (trans-2F-FPP, 2a) or (2-cis,6-trans)-2-fluorofarnesyl diphosphate (cis-2F-FPP, 2b) after 24-h incubation periods. The lack of measurable catalytic activity when using the fluoro-substituted FPPs is most likely due to a strong electron-withdrawing inductive effect due to the presence of the fluoro substituent at position 2 that prevents ionization and pyrophosphate loss. By contrast, the fungal sesquiterpene cyclase aristolochene synthase converted trans-2F-FPP (2a) cleanly to 2-fluorogermacrene A after extended incubation times[21], and trichodiene synthase produced several fluorinated sesquiterpene hydrocarbons of unknown structure[22]. Although unreactive fluoro analogues have been useful for crystallography with limonene synthase[23] and more recently 2F-FPP complexes with aristolochene synthase[24], the inertness of the 2-fluoro-FPPs proved useful for our crystallographic experiments.

Table 3.1. Enzymatic products from incubations of TEAS wild-type and the M4 mutant with (*cis*,*trans*)- or (*trans*,*trans*)-FPP

			% by mass	% by total ion count (TIC)	
Peak	**Product**	**RT (min)**	**TEAS wt**	**TEAS wt**	**TEAS M4**
		(*cis*,*trans*)-FPP products			
3	(−)-α-cedrene	15.3	18.95	18.85	18.48
2	(+)-2-epi-prezizaene	15.8	37.89	46.52	30.54
4	α-acoradiene	16	3.44	6.80	13.40
5	4-epi-α-acoradiene	16.1	1.03	2.54	4.06
6	(−)-β-curcumene	16.5	13.78	12.30	4.65
7	nerolidol	17.1	3.61	0.90	1.81
8	α-bisabolol	18.6	1.80	0.11	0.05
9	epi-α-bisabolol	18.6	1.80	0.11	0.05
10	*cis*-farnesol	18.8	5.69	0.29	0.50
	remaining	n.a.	11.99	11.58	26.46
		(*trans*,*trans*)-FPP products			
11	germacrene A	15.1		3.65	10.98
1	5-epi-aristolochene	16.2		78.90	30.66
12	4-epi-eremophilene	16.3		6.21	27.46
13	premnaspirodiene	16.48		1.66	25.47
	remaining	n.a.		9.58	5.43

Table 3.2. Kinetic parameters of TEAS wild-type and the M4 enzyme determined using either (*cis*,*trans*)- or (*trans*,*trans*)-FPP

	(*trans*,*trans*)-FPP			(*cis*,*trans*)-FPP		
	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (μM^{-1} min^{-1})	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (μM^{-1} min^{-1})
TEAS wt	2.5±0.7	8.4±0.89	0.3	5.71±0.08	14.03±2.59	0.41
TEAS M4	9.43±0.21	13.31±1.04	0.71	4.62±0.07	7.97±1.26	0.58

3.3.2. Stereochemical mechanism of cyclization

On the basis of the elucidated stereochemistry of the major cisoid products, a reaction mechanism for the TEAS-catalyzed cyclization of the cisoid farnesyl cation is proposed (Figure 3.1, panel b)[18]. Catalysis begins with divalent cation-assisted ionization of (cis,trans)-FPP generating the cisoid farnesyl cation. The ensuing 1,6 cyclization involves C1 attack on the re face of C6 of the C6−C7 double bond to produce the (6S)-α-bisabolyl cation. This step is followed by a 120° CW rotation and 6,7 hydride shift to form the (7R)-β-bisabolyl cation. The (7R)-β-bisabolyl cation is a key reaction intermediate, lying at the intersection of the majority of cisoid hydrocarbon products. The orientation of the terminal isoprene unit at this stage directs the subsequent divergence of reaction trajectories at the C6−C10 cyclization step (Figure 3.1, panel c). When the isoprene unit is oriented endo, the C6−C10 cyclization produces the (1R,4R,5S)-α-acorenyl cation. This intermediate undergoes a further C3–C11 cyclization and then a Wagner–Meerwein rearrangement to a tertiary carbocation prior to proton elimination to produce (+)-2-epi-prezizaene (2). Conversely, the exo configuration leads to the (1R 4S,5S)-α-acorenyl cation, possessing the opposite stereochemistry at C4 relative to the aforementioned prezizaene pathway. A C2–C11 cyclization of this cation followed by proton elimination terminates the reaction pathway at (−)-α-cedrene (3).

The remaining stereochemically defined products comprise roughly equal amounts of sesquiterpene hydrocarbons and alcohols, and their formation can be rationalized as branches off the main reaction pathway (Figure 3.1, panel b). Early in the mechanism, water quenching on C1 or C3 of the nascent cis-farnesyl cation accounts for cis-farnesol (10) and nerolidol (7), respectively. Immediately following the initial C1−C6 cyclization to the α-bisabolyl cation, water quenching again intercepts the cyclization path by indiscriminant attack on either face of the cation to produce equal amounts of α- and epi-α-bisabolol (8 and 9), comprising

approximately one-third of the alcohol products. Alternative proton eliminations from C5 of the (7R)-β-bisabolyl cation account for the third most abundant product in the TEAS cisoid spectrum of products, namely, (−)-β-curcumene (6), representing 16% of total hydrocarbon product. Finally, α- and 4-epi-α-acoradienes (4 and 5) stem from proton elimination from the terminal isopropenyl tail of the acorenyl cations, representing the remaining products observed at 4% and 1.2% total hydrocarbon, respectively.

3.3.3. Computational analysis of the TEAS cisoid mechanism

The intrinsic reactivity, conformation, and energy of carbocation intermediates define physically allowable cyclization pathways, which ultimately pass through the selectivity filter of active site geometry and electrostatics, most likely modulated by enzyme dynamics. To computationally examine the conformation and intrinsic energetics of the cisoid cyclization pathway, we conducted density functional theory (DFT) calculations. While numerous transition states were identified which connect consecutive intermediates in the proposed reaction mechanism (Figure 3.1), alternative connectivities were discovered that bypass adjacent carbocations, thereby directly linking more distal steps in the cisoid cyclization pathway (Figure 3.3, panel a). For example, a transition structure (14) was found that bypasses the (7R)-β-bisabolyl cation in a concerted, highly asynchronous reaction by directly connecting the (6S)-α-bisabolyl cation to the (4S)-α-acorenyl cation along the pathway to (−)-α-cedrene. Following this, a transition structure was found for the next step, linking the (4S)-α-acorenyl cation to the final carbocation, thereby completing this pathway (via 14) from the (6S)-α-bisabolyl cation to (−)-α-cedrene (3) (Figure 3.3, panel b). Interestingly, the (6S)-α-bisabolyl cation to the (4R)-α-acorenyl connection was not uncovered due to steric occlusion,

indicating the importance of the (7R)-β-bisabolyl cation along the pathway to (+)-2-epi-prezizaene (2).

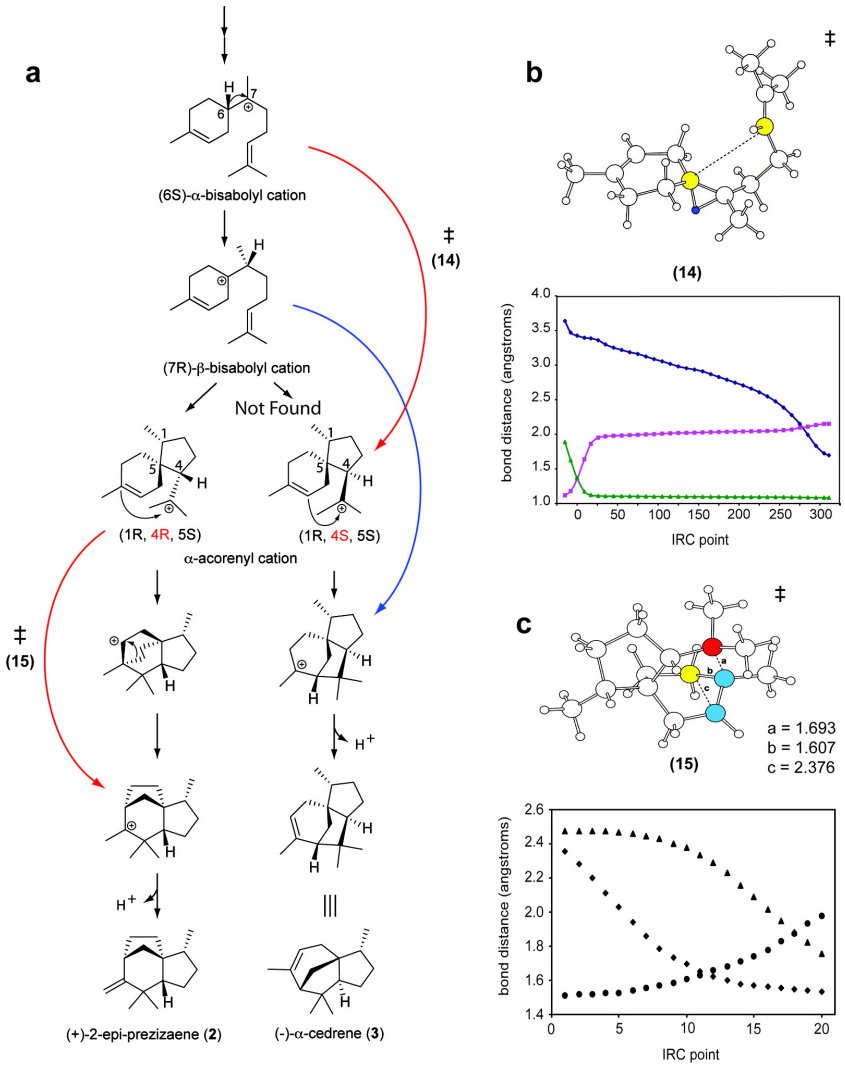


Figure 3.3. Computational analysis of the TEAS cisoid cyclization pathway. a) Density functional theory (DFT) calculations were performed on the TEAS cisoid pathway and revealed concerted, highly asynchronous reactions with a transition state (14 or 15) specific to the formation of (−)-α-cedrene (3) or (+)-2-epi-prezizaene (2), respectively (red arrows). Hong and Tantillo (26) previously located an alternative concerted highly asynchronous transition in the cedrene pathway (blue arrow). b) A transition state structure (14) was discovered that connects the (6S)-α-bisabolyl cation to the (4S)-α-acorenyl cation corresponding to point 50 on the intrinsic reaction coordinate plot. The migrating hydrogen (dark blue) and the two carbons (yellow) forming a nascent σ-bond (dashed line) are depicted. The plot shows the change in the dashed bond distance (), the original () and new (▲) C−H bond distances of the migrating hydrogen in 14 during the course of the reaction taken from an IRC calculation. The transition state structure is point 0. c) A transition state structure (15) along the (+)-2-epi-prezizaene (2) pathway is shown. Change in bond distances shown during the course of the IRC calculations bond (, a bond; ●, b bond; ▲, c bond) indicates the highly asynchrous nature of this step. The transition state structure (15) is point 10.

Formally, the formation of (+)-2-epi-prezizaene (2) involves a high-energy secondary carbocation from the anti-Markovnikov C3−C11 cyclization. It has been suggested that such high-energy secondary carbocations in terpene biosynthesis can be avoided in the gas phase via concerted, highly asynchronous mechanisms, in analogy to the formation of the C and D rings in the cyclization of squalene oxide to lanosterol[25]. These mechanisms have also been discovered in related computational studies of sesquiterpene cyclization. This was indeed the case here as we located a transition structure (15) and demonstrated with intrinsic reaction coordinate (IRC) calculations that it connected the distal cyclization events leading to (+)-2-epi-prezizaene (2) in a concerted, highly asynchronous step (Figure 3.3, panel c). Hong and Tantillo have independently identified this same transition state.[26]

3.3.4. Structure of wild-type TEAS and M4 TEAS with 2-fluoro analogues

We expected that the three-dimensional structures of TEAS-2F-FPP complexes would be informative regarding the static templating of both cisoid and transoid pathways in the TEAS active site. To investigate the structural basis for substrate preorganization and catalytic promiscuity along the transoid and cisoid cyclization pathways, we carried out crystal soaks with the nonionizable substrate analogues trans-2F-FPP (2a) and cis-2F-FPP (2b), respectively. These experiments yielded protein−small molecule complexes diffracting to resolutions ranging from 2.1 to 2.6 Å (Table 3.3).

Table 3.3. Crystallographic data and refinement statistics[a]

	wt TEAS-*trans*-2F-FPP	**wt TEAS-*cis*-2F-FPP**	**M4 TEAS-*trans*-2F-FPP**	**M4 TEAS-*cis*-2F-FPP**
pdb code	3M01	3M0	3LZ9	3M00
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit-cell parameters:				
a(Å)	125.5	125.5	126.3	126.1
b(Å)	125.5	125.5	126.3	126.1
c(Å)	122.7	121.3	121.9	122.4
α-β-γ°	90	90	90	90
Monomers per Asymm unit	1	1	1	1
Resolution range (Å)	500.0−2.6	500.0−2.5	500.0−2.28	500.0−2.1
No. reflections measured	219537	324161	377780	248702
Merging *R*-factor	0.093 (0.325)	0.077 (0.282)	0.090 (0.346)	0.080 (0.358)
I/σ	17.28 (4.63)	23.53 (5.12)	16.04 (3.59)	16.45 (3.55)
Completeness	0.963 (0.914)	0.979 (0.938)	0.992 (0.903)	0.956 (0.872)
Redundancy	7.36 (7.48)	8.54 (8.55)	7.74 (6.53)	4.02 (3.92)
No. reflections used	30227	30047	22460	57483
R-factor	0.2065	0.1976	0.1935	0.2205
Free *R*-factor	0.2423	0.2257	0.2464	0.2586
No. amino acid residues	547	547	547	547
No. water molecules	150	168	174	412

[a] Values in parentheses represent highest resolution shell.

Global comparison of all structures by superpositioning C-α carbons revealed a high degree of similarity with root mean square deviation (rmsd) values ranging from 0.22 to 0.37 Å for all atoms (Table 3.4). Annotating structures according to B-factors suggested a common pattern of dynamic regions across all structures refined (Figure 3.6). In contrast to the

originally published TEAS·farnesyl hydroxyphosphonate (FHP) structure, all complexes described here exhibit disorder in a portion of the J−K catalytic loop, a region encompassing amino acids 521−533 that completes the enclosure of the active site during catalysis[5]. Several residues were excised from both the wild-type and M4 TEAS models during refinement due to a lack of clearly observable electron density and the attendant poor refinement of these regions (Figure 3.7). As previously noted, the mutations in the M4 TEAS protein reside either in the active site (V516I) or distribute more peripherally around the active site surface (A274T, V372I, and Y406L) with distances from the active site center ranging from 7 to 14 Å (Figure 3.8). While each mutated side chain was readily discernible in the electron density, no significant backbone distortions were evident, strongly hinting at dynamic, not static, modulation of the active site contour for templating transformations of farnesyl cations in TEAS. However, the V516I mutation directly affects the active site contour with implications for substrate binding as discussed below.

Observable electron density is present in the active site regions for all complexes, and the positions of ligand-binding residues were clearly established with the exception of Y527 on the J−K loop. In all the structures, contiguous electron density stretches from the DDxxD motif through the diphosphate moiety into the NSE/DTE motif enshrouding the catalytically essential Mg^{2+} ions (Figure 3.4, panel c). Although three Mg^{2+} ions are visible in each complex, a complete octahedral coordination sphere of waters is only discernible in the highest resolution M4 TEAS-cis-2F-FPP complex. Electron density surrounding the diphosphate appendage is the most prominent feature in the calculated electron density (without ligands modeled) with large σ values in the SIGMAA-weighted $2F_o - F_c$ electron density maps (Figure 3.4, panel b). Clear electron density extends from the diphosphate through the first isoprene unit containing the fluoro substituent in all complexes but trails off

through the center of the chain and picks up again at the distal isoprene unit (Figure 3.4, panel e). Despite the waning electron density for the distal isoprene units, the farnesyl chain clearly curls into a U-shape in all complexes, particularly at lower σ where continuous density is apparent in the wild-type complexes (Figure 3.4, panel e). Taken together, these complexes display near complete occupancy (based upon the unmistakable diphosphate and first isoprene unit electron density), and aside from an incomplete J−K loop, Mg^{2+} ions and ligands are bound with the farnesyl chain folded in a manner consistent with the formation of major cyclization products along both the transoid and cisoid mechanistic pathways.

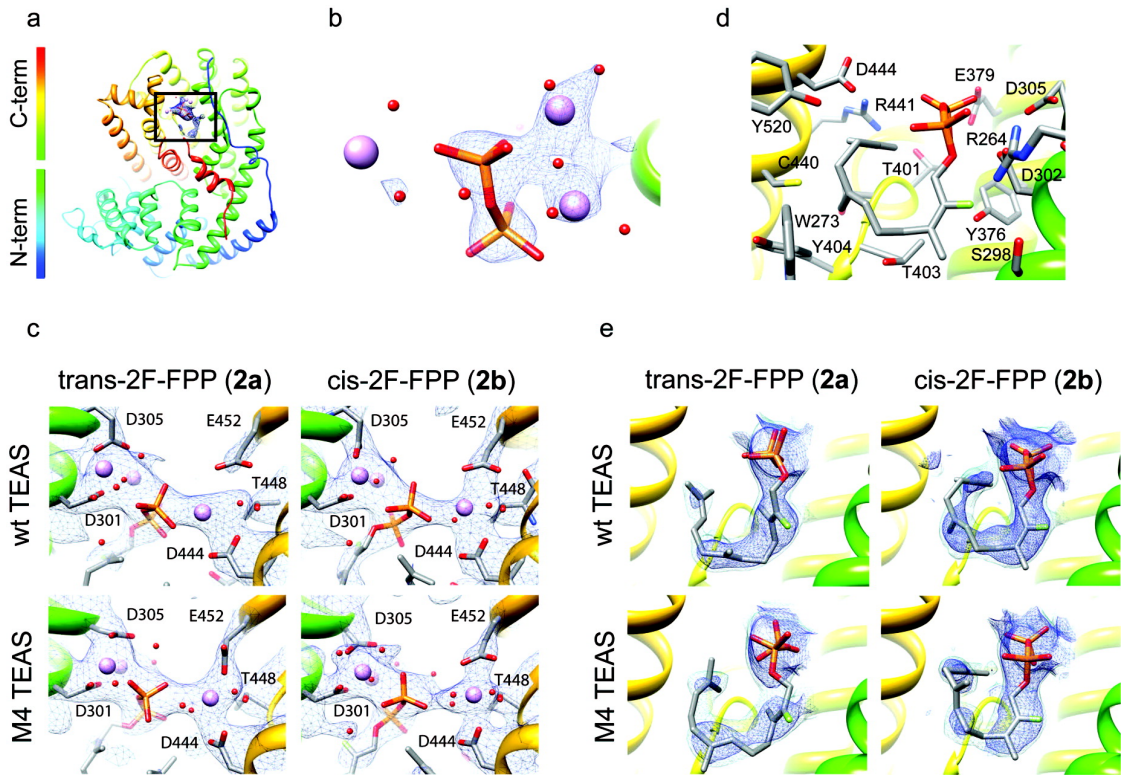


Figure 3.4. Crystallographic analysis of wild-type and M4 TEASs bound to fluoro-FPPs. a) Global structure of TEAS is illustrated as a rainbow-colored ribbon with the active site region boxed. b) Zoomed-in view of the Mg^{2+}-diphosphate coordination complex of the M4 TEAS-cis-2F-FPP complex with the $2F_o - F_c$ map contoured at 3σ. c) Close-up view of the DDxxD motif (residues 301, 302 (not shown), and 305), neighboring NSE/DTE motif (residues 444, 448, and 452), coordinating Mg^{2+} and diphosphate in the indicated fluoro-farnesyl diphosphate complexes contoured at 1σ in the $2F_o - F_c$ SIGMAA-weighted electron density map. d) Close-up of the TEAS-cis-2F-FPP complex active site showing the bound ligand and the neighboring TEAS residues. e) Ligand density for the respective complexes with the SIGMAA-weighted $2F_o - F_c$ electron density map contoured to either 1σ (dark blue) or 0.6σ (light blue).

Comparison of ligand binding modes between the cis- and trans-2F-FPP complexes reveals important differences relating to catalysis. While the orientation of the C−O bond in both trans-2F-FPP structures is nearly perpendicular to the plane of the C2−C3 double bond as required for maximum activity, the C−O bond adopts a parallel position in the cis-2F-FPP structures and hence represents an inactive conformation. If this conformation were reflective of the (cis,trans)-FPP binding, then rotation of the C1−C2 bond would be required to form a

catalytically active complex. This inactive conformation may be promoted by the 2-fluoro moieties in cis-2F-FPP (2b) through its electrostatic interaction with Arg 264 residing 3 Å away (Figure 3.4, panel d).

3.3.5. Spatial reconstruction of cisoid and transoid reaction pathways in TEAS

Multiple substrate binding modes discerned during building and refinement for the extended farnesyl chain could potentially satisfy, and likely contribute to, the observed electron densities (Figure 3.9). Despite these ambiguities, the general topology of the farnesyl chain is clear and consistent with the anticipated parental fold inferred from the elucidated stereochemistry of the final products. Importantly, electron density for the cis-2F-FPP complexes reveals that the terminal isoprene unit curls into a helical (endo) fold in accordance with the anticipated conformation (Figure 3.1, panel c). This orients the plane of the C10–C11 double bond parallel to a potentially attacking carbocation at C1. In contrast, the plane of the C10–C11 double bond of the terminal isoprene unit is perpendicular to a nascent C1 carbocation in the trans-2F-FPP complexes, in accord with an initial C1–C10 cyclization along the transoid pathway.

To spatially reconstruct the two major cyclization pathways in TEAS, we manually docked transition state structures and models of the major products into the respective active sites of wild-type trans-2F-FPP and cis-2F-FPP complexes (Figure 3.5, panel a). Restraining the placement of products/intermediates such that cyclization to a specific product most likely proceeds with the minimal amount of conformational distortion for the nascent farnesyl cation en route to the final products results in a mechanistically plausible transition state geometry en route to the observed dominant product.

Superposition of the transition state structure (15) on the farnesyl chain indicates that substantial contraction of the substrate must occur to produce the compact (+)-2-epi-prezizaene (2) final product. The crucial elements of preorganization are the juxtaposition of C1 and C6 together with the endo orientation of the farnesyl tail, both consistent with the observed electron density of the TEAS·cis-2F-FPP complex. Therefore, the static picture drawn from these observations reveals a catalytically relevant substrate binding conformation and substrate/intermediate preorganization very early along the cisoid pathway catalyzed by TEAS. Based on this model, the pyrophosphate ion would reside close by but suitably sequestered by neighboring interactions to stabilize the developing positive charge of the secondary carbocation while limiting recapture probability prior to the final proton elimination yielding (+)-2-epi-prezizaene (2).

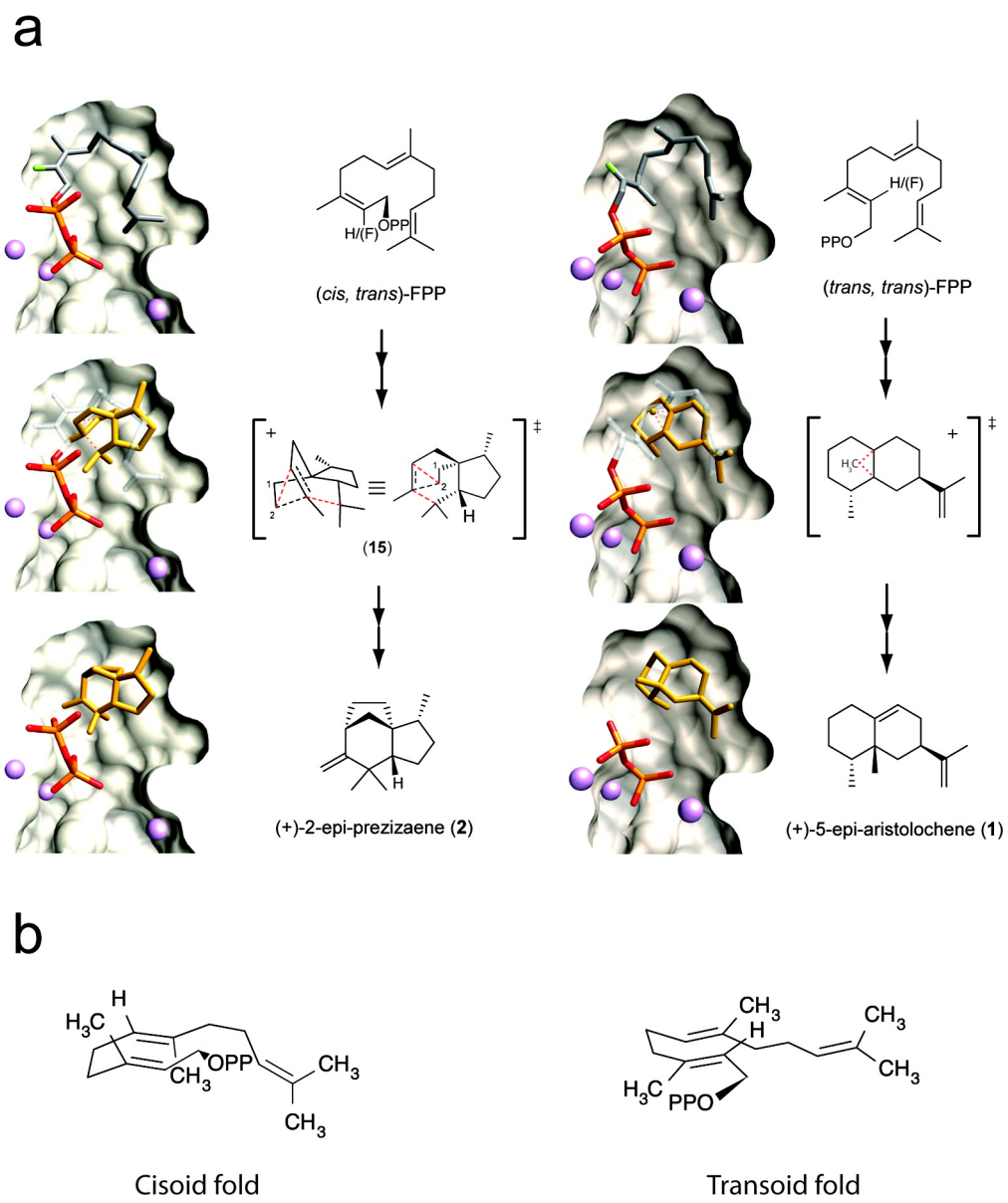


Figure 3.5. Spatial reconstruction of the transoid and cisoid cyclization pathways in TEAS. a) Refined conformations for the trans-2F-FPP or cis-2F-FPP·TEAS complexes are displayed in the binding pocket (clipped surface), and models of indicated reaction intermediates or products were manually positioned relative to the refined conformations. An accompanying schematic of chemical structures designates the 2-fluoro positions in each substrate as H(F). Images were rendered with UCSF Chimera (57). Transition state structures are shown alongside their corresponding rendered figures; dashed lines are used to indicate bond breakage and formation. b) Proposed substrate folds leading to the cisoid and transoid cyclization pathways in TEAS.

For the transoid pathway, the key transition state structure leading to the major product (+)-5-epi-aristolochene (1) involves a methyl migration atop the decalin ring system of the eremophilyl carbocation (Figure 3.5, panel a, right) as previously reported[19]. To achieve this energetically favorable alkyl migration, substrate folding must preorganize an initial electrophilic attack of C1 on C10 of the distal double bond. This requires substantial movement of the chain following ionization, as these atoms are 5 Å apart in the ground state complexes. However, judging by the degree of overlap between the transition state model and the farnesyl chain, this motion can be accommodated largely within the first isoprene unit with minimal conformational adjustments of the more distal isoprene units. Therefore, this conformation of the farnesyl chain is consistent with TEAS preorganizing FPP (or more likely the resultant acyclic farnesyl cation) for cyclization to 5-epi-aristolochene (1), in contrast to the original TEAS·FHP complex(5). The phosphate moiety of FHP overlaps with the β-phosphates of 2F-FPPs, although the farnesyl chain is more extended and folds in essentially the opposite direction (Figure 3.9).

3.3.6. Cisoid cyclase activities with (trans,trans)-FPP

On the basis of our spatial reconstruction of the cisoid cyclization pathway, we propose a model describing the "cisoid fold" of (trans,trans)-FPP (1a) that is representative of the preorganization of the farnesyl chain leading to its conversion to cisoid-cation-derived hydrocarbons. Accordingly, an alternative, catalytically productive binding mode of FPP is populated in which the farnesyl chain curls into its typical U-shaped topology, but with the first two isoprenoid units inverted relative to the "transoid fold" configuration (Figure 3.5, panel b). The cisoid binding mode therefore possesses the DU configuration, opposite to that described for germacrene A[27], and importantly, with the distal isoprenoid unit curled below

this plane into an alternative binding pocket formed by T401, T402, C440, R441, D444, and the diphosphate moiety (Figure 3.4, panel a). We posit that the positioning and anchoring of the terminal isoprenoid unit is an essential stereochemical feature for triggering ionization, perhaps through both steric and electronic effects. Upon ionization, a kinetically slow initial isomerization occurs as the C10–C11 double bond is rotated out of position for electrophilic attack by C1; in turn, a re face capture by the pyrophosphate ion on C3 of the nascent farnesyl cation generates the neutral (3S)-NPP intermediate. Rotation around the C2–C3 bond followed by reionization generates the 2,3-cis-farnesyl cation and entry into the cisoid cyclization pathway.

3.3.7. Structural picture of catalytic promiscuity

The ligand–protein structures of a promiscuous TEAS mutant offer a glimpse into the structural underpinnings of product specificity or lack thereof. To discern the structural basis for product specificity in both cisoid and transoid cyclization pathways, we conducted a comparative analysis of wild-type TEAS and M4 mutant structures with particular attention focused on the active site contour and farnesyl chain binding modes. The most obvious surface distortion, whether statically or dynamically derived, is contributed by the V516I mutation, which introduces a methyl group into the active site cavity (Figure 3.10). While no drastic distortion is evident in the comparative models of the farnesyl chain, the electron density for the farnesyl chain in the M4 TEAS structures is discontinuous, indicative of increased dynamic motion and/or local disorder (Figure 3.4, panel e). Interestingly, only the M4 TEAS-cis-2F-FPP complex exhibits a significant shift in the position of Y520, which additionally alters the active site surface features. The ligand-dependency for the Y520 shift may reflect the interaction between the farnesyl chain, wherein the central isoprene unit is inverted

relative to the corresponding wild type complex, and active site residues in defining the preorganized binding state. Despite the higher resolution of the M4 structures, the density for the ligand is discontinuous, even when the electron density maps are viewed at low σ, in contrast to the wild-type structures (Figure 3.4, panel e). It is probable that a more dynamic farnesyl chain in M4-TEAS-cis/trans-2F-FPP structures explains the lack of electron density for the entire farnesyl chain consistent with this M4 TEAS's promiscuous catalytic activity along both pathways.

3.3.8. Conclusions

(cis,trans)-FPP proved effective in directing reactions along the cisoid cyclization pathway in TEAS. The isolation and stereochemical elucidation of the products lead to formulation of reasonable reaction pathways to the cisoid-derived sesquiterpene skeletons[18]. Traditionally, chemical tools have been a vital part of defining the stereochemical course of terpene biosynthesis, as elegantly exemplified by the use of (1R)-[1-^{3}H]- and (1S)-[1-^{3}H]-geranyl diphosphate to provide direct experimental confirmation that cyclization along the cisoid pathway results in net retention of configuration at C1 of the substrate[28]. Fluoro isoprenoid diphosphate substrate analogues also have been instrumental in elucidating mechanistic aspects of terpene biosynthesis, most notably through the interception of reaction intermediates such as 6-fluorogermacrene A[29] and 7-fluoroverticillenes[30] in TEAS and taxadiene synthase enzymes, respectively. Most recently, 2F-FPP and 12,13-difluorofarnesyl diphosphate (DF-FPP) were instrumental in deciphering the probable order of metal-ion binding and conformational changes required for catalysis by aristolochene synthase from Aspergillus terrus[24].

To date, crystallographic analyses of terpene cyclases have yielded important insights into how these enzymes function on the atomic scale. Most notably, anchoring the diphosphate moiety of the substrate and metal coordination by the DDxxD and NSE/DTE motifs shown in crystal structures paints a picture of the fundamental role of these events in terpene synthase catalysis. Structures containing inorganic pyrophosphate or a substrate analogue bound in the active site display ordering of various loops proximal to the active site, consistent with a closed protein conformation that shields reactive carbocation intermediates from solvent[5, 23, 31]. Further, alterations in pyrophosphate binding are thought to aid in the modulation of prenyl chain orientation within the active site and most likely modulate the fate of the early intermediates along prescribed mechanistic pathways[32]. All structures reported in the current study contain the full complement of Mg^{2+} ions coordinating the diphosphate of the fluoro-farnesyl analogues. However, despite this clear coordination geometry, elements of the J−K loop remain disordered in both wild-type and mutant structures bound to diphosphate containing ligands with Y527 electron density missing from the active site. This lack of observable density for Y527 stands in contrast to the original TEAS·FHP complex where this residue is clearly discernible (Figure 3.7).

These observations hint at a greater role of dynamics in terpene chain cyclization than evident from the early structural work based upon static crystal structures. The wild-type ligand complexes in the current study revealed density for the farnesyl chain folded in a manner consistent with catalysis, and this can be interpreted in light of the established stereochemistry for all TEAS products. Moreover, these static observations enabled the positing of 2 distinct parental folds each of which gives rise to either cisoid or transoid cyclization pathways for TEAS (Figure 3.5, panel b). The structure of a complex of limonene synthase with 2-fluorolinalyl diphosphate captured another instance where the isoprenoid

chain conformation is consistent with the geometry of the final product[23]. However, it has been noted that most reported crystal structures of terpene synthases complexed with isoprenoid substrate analogues, including 2F-FPP used here, reveal isoprenoid tail conformations that are not catalytically relevant[5, 23, 24, 31]. Considering the ideal case, where there is unambiguous density for every atom of the farnesyl chain, a central challenge in the field remains to resolve structural features responsible for product specificity or lack thereof from a static picture alone, given the degeneracy of possible products arising from a single parental substrate fold. Future progress toward defining the origins of sesquiterpene skeletal complexity will undoubtedly benefit from integrating dynamic information from NMR and time-resolved fluorescence (in progress) with computational approaches and protein crystallography to develop a much clearer and time-resolved biophysical picture of terpene synthase directed cyclization.

What possible relevance does the cryptic cisoid cyclization pathway of TEAS have in the natural world? Although (cis,trans)-FPP has not been identified as a metabolite in tobacco or related Solanaceous plants, a (cis,trans)-farnesyl diphosphate synthase has been identified in Mycobacterium tuberculosis involved in bacterial cell wall synthesis[33-34], suggesting the potential relevance of this compound in other biological systems. Moreover, while often observed, the biological significance of small amounts (3–14% of total product) of (cis,trans)-FPP formation by FPP synthases[35] has been ignored to date. Is it possible then that TEAS possesses a "moonlighting" role in vivo by gathering up what we would normally consider biosynthetic "waste" and recycling it into a bioactive product? While TEAS produces cisoid terpenes in vitro, the presence of these metabolites has yet to be confirmed in planta. Nonetheless, TEAS clearly possesses an efficient catalytic potential to access presently unanticipated in vivo chemical diversity from lengthy branches of the cisoid reaction pathway,

a property that may have been naturally selected for and that can also be immediately exploited for biotechnological applications starting with (cis,trans)-FPP.

3.4. METHODS

3.4.1. Organic synthesis

(cis,trans)-FPP was available from the concurrent investigation[18]. (trans, trans)- and (cis,trans)-2-FluoroFPPs were accessed from the corresponding 2-fluorofarnesol isomers[30] by conversion to the respective 2-fluorofarnesyl chlorides and S_N2 displacements with $(n\text{Bu})_4\text{N}$/diphosphate (HOPP)[36] with complete retention of the 2,3-double bond configurations by means of procedures similar to those reported previously (see Supporting Information)[30]. (cis,trans)-2-FluoroFPP has not been previously described in the literature. Characterization data for ammonium salt of 2b are as follows: white solid (51 mg, 68%); ^{1}H NMR (CD_3OD, 400 MHz) δ 5.17–5.11 (m, 1H, vinyl *H*), 5.11–5.05 (m, 1H, vinyl *H*), 4.59 (dd, 2H, J = 23.3, 5.4 Hz, C*H*$_2$OPP), 2.16–2.11 (m, 4H, C*H*$_2$), 2.09–2.04 (m, 2H, C*H*$_2$), 2.00–1.95 (m, 2H, C*H*$_2$), 1.68 (d, 3H, J = 3.5 Hz, C*H*$_3$), 1.66 (q, 3H, J = 1.2 Hz, C*H*$_3$), 1.61 (d, 3H, J = 1.2 Hz, C*H*$_3$), 1.60 (br d, 3H, J = 0.6 Hz, C*H*$_3$); ^{31}P NMR (CD_3OD, 162 MHz) δ −7.99 (br d, J = 14.9 Hz), −9.30 (br d, J = 14.1 Hz); ^{19}F NMR (CD_3OD, 376 MHz) δ −118.9 (td, J = 23.2, 3.5 Hz).

3.4.2. Protein expression and purification

pH9GW expression vectors (an in-house Gateway destination vector) were transformed into E. coli BL21(λDE3) and plated on LB agar containing 50 μg/mL kanamycin for selection. Colonies were transferred to 100 mL of liquid media (LB with kanamycin) followed by 16-h growth with shaking at 37°C at 275 rpm. Cultures were diluted 50-fold into 1 L of Terrific Broth with kanamycin, followed by growth with shaking at 37°C at 275 rpm

until cultures reached OD600 ≥ 1.5. Protein expression was induced by addition of isopropyl β-D-thiogalactoside (IPTG) to 0.1 mM followed by growth with shaking at 20°C at 275 rpm for 5 h. Cells were harvested by centrifugation and cell pellets frozen at -20°C. Frozen pellets were re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, pH 8.0, 10% [v/v] glycerol, 10 mM β- mercaptoethanol, and 1% [v/v] Tween-20) containing 1 mg/mL lysozyme followed by stirring at 4°C for 1 h. After sonication and centrifugation, the clarified supernatant was passed over a column of Ni^{2+}–NTA resin (Qiagen), washed with 10 bed volumes of lysis buffer and 10 bed volumes of wash buffer (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, pH 8.0, 20 mM β-mercaptoethanol, and 10% [v/v] glycerol), and the His-tagged protein was eluted with elution buffer (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 250 mM imidazole, pH 8.0, 20 mM β- mercaptoethanol, and 10% [v/v] glycerol). N-terminal His-tags were removed via proteolysis with thrombin as follows: thrombin was added to a ratio of 1:1,000 [w/w] directly to the eluted protein fraction and dialyzed against two changes of buffer (50 mM Tris–HCl, pH 8.0, 100 mM NaCl, and 10 mM β-mercaptoethanol) over 24 h at 4 °C. Following digestion, samples were passed over a column containing 0.5 mL benzamidine sepharose to remove thrombin and 0.5 mL Ni2+-NTA resin to capture undigested protein. The resulting protein solutions were collected and concentrated to approximately 10 mg/mL or greater by centrifugation using 30,000 Da molecular weight cut-off concentrators (Millipore, Bedford, MA). Concentrated samples were injected onto a Sephacryl S-200 column equilibrated with buffer (25 mM Tris–HCl, pH 8.0, 50 mM NaCl and 1 mM DTT). Fractions corresponding to digested protein were verified by SDS-PAGE, pooled and concentrated (as described above) to approximately 20 mg/mL and aliquoted for freezing at -80°C. Samples were judged to be ~99% pure by Coomassie stained SDS-PAGE gels.

3.4.3. Kinetic measurement

Kinetic characterization of purified wild-type and M4 mutant TEASs were conducted as previously described[20]. Briefly, 500-μL scale reactions using a 3-component buffer system (25 mM 2-(N-morpholino)ethanesulfonic acid (MES), 50 mM Tris, and 25 mM 3-(cyclohexylamino)propanesulfonic acid (CAPS) at pH 7.0 with 10 mM MgCl2) were conducted in triplicate at room temperature (25°C) with 15 nM protein and variable concentrations of (cis,trans)-FPP. Reaction products were analyzed using a Hewlett–Packard 6890 gas chromatograph (GC) coupled to a 5973 mass selective detector (MSD) equipped with an HP- 5MS capillary column (0.25 mm i.d. 30 m with 0.25 μm film thickness) (Agilent Technologies). Product quantification was performed using SIM mode, set to detect ions with m/z = 91, 133, and 189. The GC was operated at a He flow rate of 2 mL/min, and the MSD was operated at 70 eV. Split-less injections (2 μL) were performed with an inlet temp of 250°C, a temp that drives the Cope rearrangement of germacrene A (11) to completion. The GC was programmed with an initial oven temp of 50°C (5-min hold), which was then increased 10°C/min up to 180°C (4-min hold), followed by a 100°C/min ramp until 240°C (1-min hold). A solvent delay of 8.5 min was allowed prior to the acquisition of the MS data. (+)-2-Epi-prezizaene (2) was quantified by integration of peak areas using Enhanced Chemstation (version B.01.00, Agilent Technologies).

The GC–MS instrument was calibrated with an authentic (+)-2-epi-prezizaene standard[16]. Corrected velocity data (Table 3.1) were fitted to the Michaelis–Menten equation using GraphPad Prism (version 4.00 for Windows, GraphPad Software).

3.4.4. Protein crystallization and data collection

Crystallization of purified proteins was conducted using hanging drops over a 0.5-mL reservoir (15% w/v PEG8K, 200 mM $Mg(OAc)_2$, 100 mM 3-(N-morpholino)-2-hydroxypropanesulfonic acid (MOPSO)-Na^+, pH 7.0). Crystal soaks were conducted overnight in reservoir solution containing 10 mM fluoro-farnesyl diphosphate ligand. Crystals were frozen in soak solution also containing 20% v/v ethylene glycol as cryoprotectant. Data collection was performed at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 1–5 for wild-type TEAS-2F-FPP and cis-2F-FPP complexes, while data for the M4 TEAS mutant structures were collected at the Advanced Light Source (ALS) beamline 8.2.1. All data were processed using XDS software[37]. The initial crystallographic structure solutions were obtained through molecular replacement analyses using the TEAS-FHP complex (PDB id 5eat) as the search model with Molrep in Collaborative Computational Project No. 4 (CCP4)[38]. Model building was performed using COOT[39] and rounds of refinement were conducted using Crystallography NMR System (CNS)[40]. To refine the position of the farnesyl chain, the first isoprene unit containing the 2-flouro group was built and refined, followed by sequential addition, building and refinement of remaining isoprene units. For the wild type cis-2F-FPP complex, additional multi-conformer refinement was undertaken using the program RefMac[41-48] in CCP4[38]. The fold inferred from the known stereochemistry of the final products was built into the density followed by a final round of refinement using CNS to produce the current models of the farnesyl chain (Figure 3.10).

3.4.5. Computational methods

All calculations were performed with GAUSSIAN 98W and GAUSSIAN 09W[49] and the density functional method using B3LYP, Becke's three-parameter hybrid method[50] with

the Lee–Yang–Parr correlation functional[51] and the 6-31G* basis set[52]. All stationary points were confirmed with second derivative calculations. Energies reported here include zero-point energy corrections calculated with unscaled B3LYP/6-31G* frequencies obtained analytically with G98W. Intrinsic reaction coordinate calculations[53-54] were used to determine reaction pathways. Single point mpw1pw91/6-311+G(2d,p)//B3LYP/6-31G* calculations as recommended by Matsuda[55] were carried out for all stationary points reported.[56]

3.4.6. Product elucidation

A preparative-scale incubation was carried out using 127 mg (310 mmol) of (cis,trans)-FPP and a total of 18 mg of recombinant wild-type TEAS in order to accumulate sufficient material for chromatographic fractionations, NMR analyses, and optical rotation measurements of the major products as described[18].

3.5. SUPPORTING INFORMATION

3.5.1. Preparation and Characterization of (2-cis, 6-trans)-2-Fluorofarnesyl Diphosphate

General Aspects:

^{1}H and ^{13}C NMR spectra were recorded in $CDCl_3$ (^{1}H, 7.26; ^{13}C, 77.0) or CD_3OD [^{1}H, 3.31 (quintet); ^{13}C, 49.2 (septet)] with U400 and U500 spectrometers in SCS NMR Spectroscopy Facility at the University of Illinois. Chemical shifts are in ppm and coupling constants are in Hertz. The abbreviation ‘app’ is used to describe the apparent multiplicity of the peak and may or may not be a valid first-order analysis.

All chemical reactions were performed in flame-dried glassware under nitrogen. THF and Et_2O were dried and distilled from Na/benzophenone; benzene and CH_2Cl_2 were dried and distilled from CaH_2. Hexane and ethyl acetate were freshly distilled from CaH_2. DMF,

acetonitrile, and $CDCl_3$ were dried over molecular sieves (4 Å) prior use. TLC analyses were performed on silica gel 60 F254 precoated-plates 250 μm. All retention factors (R_f) are on silica gel TLC plates until otherwise noted. TLC visualizations were performed with 5% phosphomolybdic acid (0.2 M in 2.5% concd. H_2SO_4/EtOH (v/v)), I_2 vapor, or UV light. Commercial reagents were used without further purification unless specifically noted. Column chromatography was performed according to Still's procedure[58] using 100-700 times excess 32- 64 μm grade silica gel. Products separated by chromatography are specified in elution order.

(2Z, 6E)- 1-Chloro-3,7,11-trimethylundeca-2,6,10-triene ((2-cis, 6-trans)-2-Fluorofarnesyl Chloride)

(2-cis, 6-trans)-2-Fluorofarnesol[30] was converted to the allylic chloride under Meyers' conditions[59] as previously describe for (2-trans, 6-trans)-2-fluorofarnesol.[24] Reaction of the alcohol (44 mg, 0.18 mmol) with LiCl (77 mg, 1.8 mmol), s-collidine (222 mg, 1.8 mmol), andMsCl (67 mg, 0.54 mmol) in dry DMF provided the chloride as a yellow oil (47 mg, 99%). The chloride was converted to the diphosphate directly without purification. Product characterization data: TLC R_f 0.83 (15% EtOAc in hexane); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 5.09 (m, 2H, vinyl H), 4.18 (dd, 2H, J = 22.5, 0.5 Hz, CH2Cl), 1.95-2.18 (m, 8H, 4CH2), 1.72 (app d, 3H, Japp = 3.5 Hz, CH3), 1.68 (d, 3H, J = 1.0 Hz, CH3), 1.60 (s, 6H, 2CH3); 19F NMR (CDCl3, 376 MHz) δ –116.7 (td, J = 23.2, 2.8 Hz).

(2E, 6E)-2-Fluoro-3,7,11-trimethyl undeca-2,6,10-trien-1-yl Diphosphate, Trisammonium Salt (2b, (2-cis, 6-trans)-2-Fluorofarnesyl Diphosphate).

The diphosphorylation was carried out as previously described for the trans,trans isomer4 using Poulter's methodology.[60] The reaction of the chloride (47 mg, 0.18 mmol), $HOPP(NBu_4)_3$ (320 mg, 0.36 mmol) and 3 Å molecular sieves (400 mg) in CH_3CN (2.0 mL) provided the crude tetrabutylammonium diphosphate as a yellow oil (366 mg). Based on the ^{31}P NMR spectrum, it was a 1: 0.81 mixture of inorganic pyrophosphate and organic diphosphate (corrected yield 91%). Ion exchange chromatography on BioRad $(NH_4)^+$cation exchange resin (40 mL of 25 mM NH_4HCO_3 in 2% v/v 1-propanol/D.I. water) and lyophilization followed by washing with MeOH (3 x 5 mL) to remove the inorganic pyrophosphate afforded the $(NH_4)^+$ salt of diphosphate 2b as a white solid (51 mg, 68 %): ^1H NMR (CD_3OD, 400 MHz) δ 5.17-5.11 (m, 1H, vinyl *H*), 5.11-5.05 (m, 1H, vinyl *H*), 4.59 (dd, 2H, J = 23.3, 5.4 Hz, *CH*$_2$OPP), 2.16-2.11 (m, 4H, *CH*$_2$), 2.09-2.04 (m, 2H, *CH*$_2$), 2.00-1.95 (m, 2H, *CH*$_2$), 1.68 (d, 3H, *J* = 3.5 Hz, *CH*$_3$), 1.66 (q, 3H, *J* = 1.2 Hz, *CH*$_3$), 1.61 (d, 3H, *J* = 1.2 Hz, *CH*$_3$), 1.60 (br d, 3H, *J* = 0.6 Hz, *CH*$_3$); ^{31}P NMR (CD_3OD, 162 MHz) δ -7.99 (br d, *J* = 14.9 Hz), -9.30 (br d, *J* = 14.1 Hz); ^{19}F NMR (CD_3OD, 376 MHz) δ -118.9 (td, *J* = 23.2, 3.5 Hz).

Table 3.4. Global Comparison of TEAS WT and M4 crystal structures[a]

	M4 TEAS cis-2F-FPP	M4 TEAS trans-2F-FPP	wt TEAS cis-2F-FPP	wt TEAS trans-2F-FPP
M4 TEAS·cis-2F-FPP	-	-	-	-
M4 TEAS·trans-2F-FPP	0.242	-	-	-
wt TEAS·cis-2F-FPP	0.282	0.321	-	-
wt TEAS-trans-2F-FPP	0.29	0.328	0.219	-
5EAT	0.334	0.369	0.294	0.335

[a]Global comparisons were performed by superpositioning all C-alpha carbons to derive root mean square deviation (rmsd) values expressed in the unit angstroms.

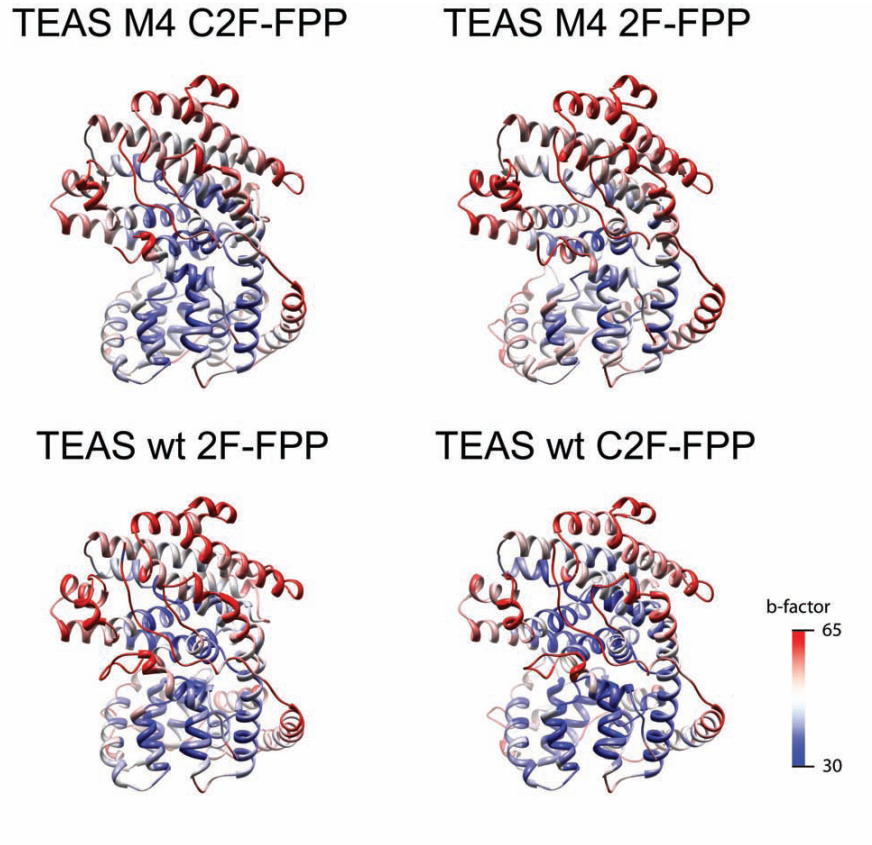


Figure 3.6. Annotation of global structure using B-factors reveals a similar pattern of dynamically accessible polypeptide segments. All structures were colored according to their refined isotropic by B-factors, with the corresponding color values of the blue to red gradient shown in the legend at the bottom right.

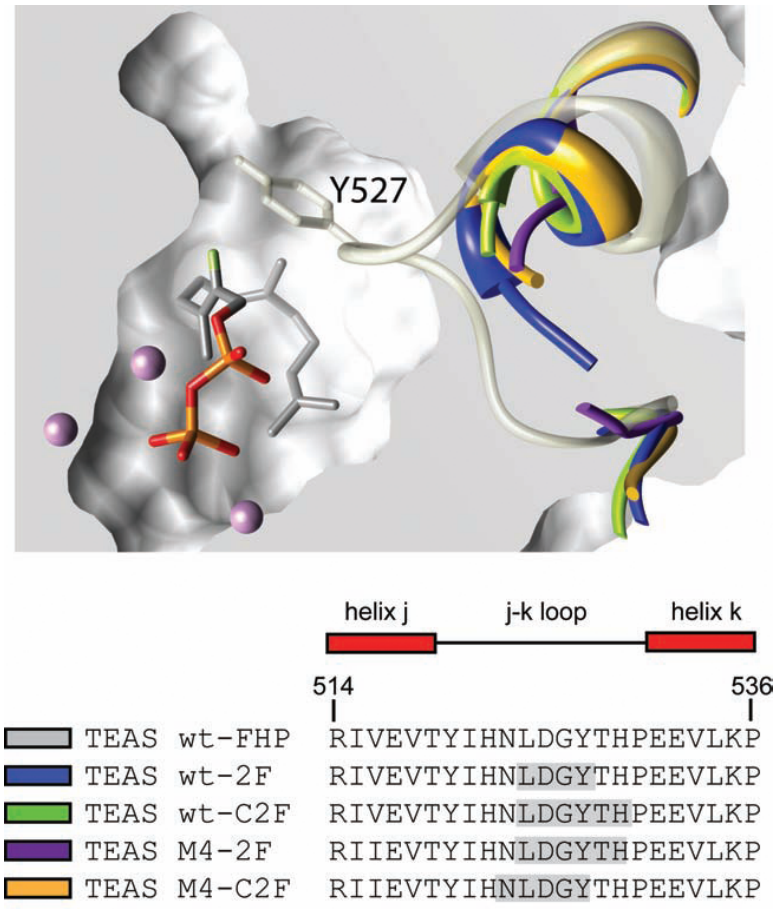


Figure 3.7. Disorder in the J-K loop of experimental crystal structures. An active site model for the wild-type TEAS trans-2F-FPP is shown as a van der Waals surface clipped to reveal the bound substrate analogue and helices J and K with the intervening loops. All experimental structures are overlaid on the original TEAS-FHP structure (pdb id 5eat) shown in a grey semitransparent trace. Each structure is colored as indicated in the legend below, with the omitted J-K loop regions highlighted in grey.

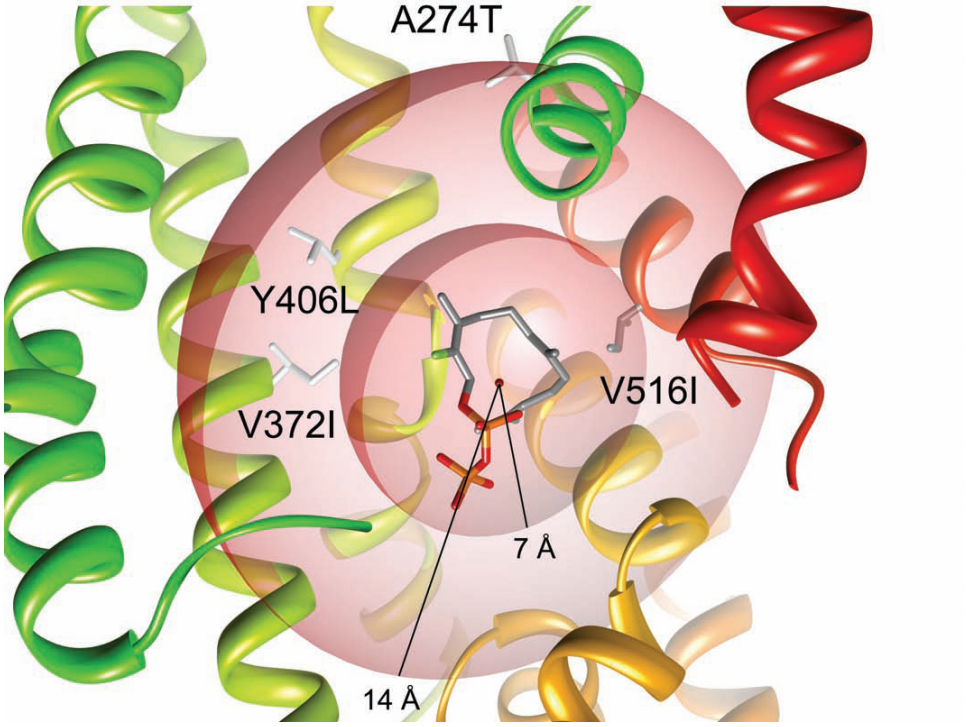


Figure 3.8. Spatial distribution of M4 mutations and closest distances to the farnesyl chain. a. The global structure of M4 TEAS with bound cis-2F-FPP ligand modeled into the active site and the protein backbone is depicted as rainbow colored ribbons. Distances from the active sited center to the side-chains of the M4 mutations are shown as dashed lines.

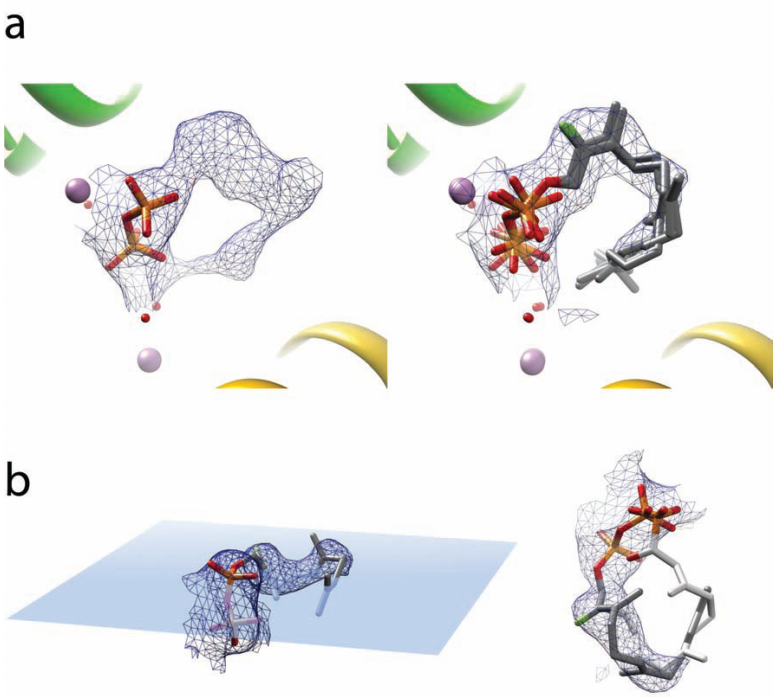


Figure 3.9. Farnesyl chain topology of wild-type TEAS from fluorofarnesyl analogues. a. Observable electron density from the wild-type complex with cis-2F-FPP reveals a U-shaped curl (left panel) possibly contributed to by four distinct binding modes of the farnesyl chain (right panel). b. Calculated electron density contoured at 1σ in the SIGMAA-weighted $2F_o$-F_c map with the modeled trans-2F-FPP shown with a plane passing through the U-shape curl of the farnesyl chain (left panel). An overlay of trans-2F-FPP (silver chain) with farnesylhydroxy phosphonate (FHP, white chain) in the calculated electron density for the trans-2F-FPP ligand from the left panel.

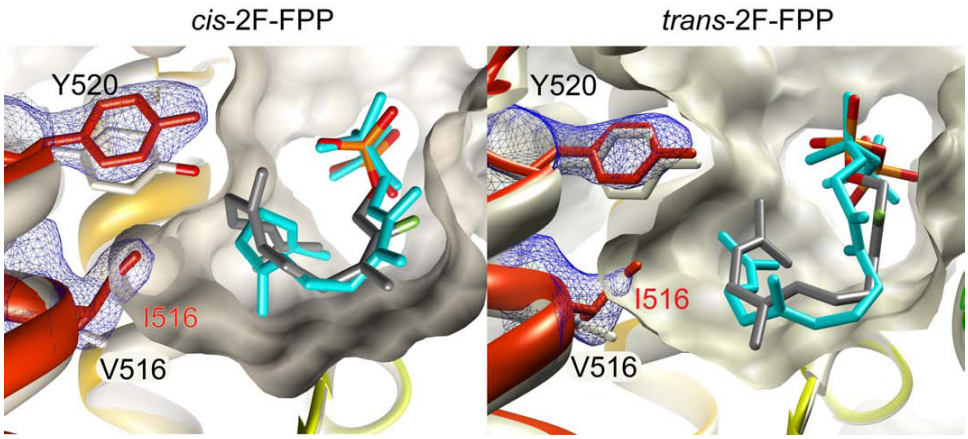


Figure 3.10. Spatial depiction of mutational effects in M4 TEAS on the active site contour and substrate-binding mode in the trans-2F-FPP and cis-2F-FPP complexes. The ribbon and active site surface (cream) of wild-type TEAS wild is superimposed on the corresponding M4 TEAS 2F-FPP complex, with ribbons and side chains rendered with rainbow coloration (as in Fig. 3a and 4a). The ligand from wild-type TEAS (cyan) and M4 TEAS (gray) is overlaid and electron density from the SIGMAA-weighted $2F_o$-F_c electron density maps at 1σ is shown for Y520 and I516 for the M4 TEAS structures.

ACKNOWLEDGEMENTS

The text of chapter 3, in full, is a reprint of material as it appears in ACS Chemical Biology 2010, 5 (4), pp 377–392, with the exception of the section under supporting information entitled “computational details” which was excluded. Permission was obtained from all co-authors. I am second author of this work. Paul O’Maille wrote the manuscript, and was also involved with protein purification, GCMS data analysis, crystallization experiments, and crystallographic data processing, structure solution and refinement. I was responsible for protein purification, GCMS data analysis, crystallization experiments, crystallographic data processing, structure solution, refinement, and contributed revisions to the manuscript. Juan Faraldos was responsible for organic synthesis, NMR characterization of sesquiterpenes, and contributed revisions to the manuscript. Yuxin (Marilyn) Zhao was responsible for chemical synthesis of cis-FPP. B. Andes Hess Jr. and Lidia Smentek were responsible for all

computational studies. The research included in the manuscript was performed under the supervision of Robert Coates and Joseph P. Noel (who also contributed revisions and helped write the manuscript).

REFERENCES

1. Gershenzon, J. and Dudareva, N. (2007) The function of terpene natural products in the natural world Nat. Chem. Biol. 3, 408– 414.

2. Liang, P., Ko, T., and Wang, A. (2002) Structure, mechanism and function of prenyltransferases Eur. J. Biochem. 269, 3339– 3354.

3. Ruzicka, L., Eschenmoser, A., and Heusser, H. (1953) The isoprene rule and the biogenesis of terpenic compounds Experentia 9, 357– 367.

4. Cane, D. (1985) Isoprenoid biosynthesis. Stereochemistry of the cyclization of allylic pyrophosphates Acc. Chem. Res. 18, 220– 226.

5. Starks, C., Back, K., Chappell, J., and Noel, J. (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase Science 277, 1815– 1820.

6. Lesburg, C., Zhai, G., Cane, D., and Christianson, D. (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology Science 277, 1820– 1824.

7. Shishova, E., Di Costanzo, L., Cane, D., and Christianson, D. (2007) X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate Biochemistry 46, 1941– 1951.

8. Kllner, T., Schnee, C., Li, S., Svatos, A., Schneider, B., Gershenzon, J., and Degenhardt, J. (2008) Protonation of a neutral (S)-β-bisabolene intermediate is involved in (S)-β-macrocarpene formation by the maize sesquiterpene synthases TPS6 and TPS11 J. Biol. Chem. 283, 20779– 20788.

9. Picaud, S., Mercke, P., He, X., Sterner, O., Brodelius, M., Cane, D., and Brodelius, P. (2006) Amorpha-4,11-diene synthase: mechanism and stereochemistry of the enzymatic cyclization of farnesyl diphosphate Arch. Biochem. Biophys. 448, 150– 155.

10. Cane, D. and Ha, H. (1988) Trichodiene biosynthesis and the role of nerolidyl pyrophosphate in the enzymatic cyclization of farnesyl pyrophosphate J. Am. Chem. Soc. 110, 6865– 6870.

11. Mercke, P., Crock, J., Croteau, R., and Brodelius, P. (1999) Cloning, expression, and characterization of epi-cedrol synthase, a sesquiterpene cyclase from Artemisia annua L Arch. Biochem. Biophys. 369, 213– 222.

12. Lin, X. and Cane, D. (2009) Biosynthesis of the sesquiterpene antibiotic albaflavenone in Streptomyces coelicolor. Mechanism and stereochemistry of the enzymatic formation of epi-isozizaene J. Am. Chem. Soc. 131, 6332– 6333.

13. Davis, E. M. and Croteau, R. (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes Top. Curr. Chem. 209, 53– 95.

14. Gordon, M., Stoessl, A., and Stothers, J. (1973) Post-infectional inhibitors from plants. 4. Structure of capsidiol - antifungal sesquiterpene from sweet peppers Can. J. Chem. 51, 748– 752.

15. O'Maille, P. E., Chappell, J., and Noel, J. (2006) Biosynthetic potential of sesquiterpene synthases: alternative products of tobacco 5-epi-aristolochene synthase Arch. Biochem. Biophys. 448, 73– 82.

16. Fox, R. B. and Powell, W. H. (2001) Nomenclature of Organic Compounds, 2nd ed., pp 306− 308, American Chemical Society and Oxford University Press, Oxford.

17. Rigaudy, J. and Klesney, S. P. (1979) IUPAC Nomenclature of Organic Chemistry: Sections A, B, C, D, E, F and H, pp 475− 477, Pergamon Press, Oxford.

18. Faraldos, J. A., O'Maille, P. E., Dellas, N., Noel, J., and Coates, R. M. (2009) Bisabolyl-derived sesquiterpenes from tobacco 5-epi-aristolochene synthase-catalyzed cyclization of (2Z, 6E)-farnesyl diphosphate. J. Am. Chem. Soc., accepted for publication.

19. O'Maille, P. E., Malone, A., Dellas, N., Andes Hess, B. J., Smentek, L., Sheehan, I., Greenhagen, B., Chappell, J., Manning, G., and Noel, J. (2008) Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases Nat. Chem. Biol. 4, 617– 623.

20. O'Maille, P. E., Chappell, J., and Noel, J. (2004) A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases Anal. Biochem. 335, 210– 217.

21. Miller, D., Yu, F., and Allemann, R. (2007) Aristolochene synthase-catalyzed cyclization of 2-fluorofarnesyl-diphosphate to 2-fluorogermacrene A ChemBiochem 8, 1819– 1825.

22. Vedula, L., Zhao, Y., Coates, R., Koyama, T., Cane, D., and Christianson, D. (2007) Exploring biosynthetic diversity with trichodiene synthase Arch. Biochem. Biophys. 466, 260– 266.

23. Hyatt, D., Youn, B., Zhao, Y., Santhamma, B., Coates, R., Croteau, R., and Kang, C. (2007) Structure of limonene synthase, a simple model for terpenoid cyclase catalysis Proc. Natl. Acad. Sci. U.S.A. 104, 5360– 5365.

24. Shishova, E., Yu, F., Miller, D. J., Faraldos, J., Zhao, Y., Coates, R., Allemann, R., Cane, D., and Christianson, D. (2008) X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal binding sequence for catalysis J. Biol. Chem. 283, 15431– 15439.

25. Hess, B. (2002) Concomitant C-ring expansion and D-ring formation in lanosterol biosynthesis from squalene without violation of Markovnikov's rule J. Am. Chem. Soc. 124, 10286– 10287.

26. Hong, Y. and Tantillo, D. (2009) Consequences of conformational preorganization in sesquiterpene biosynthesis: theoretical studies on the formation of the bisabolene, curcumene, acoradiene, zizaene, cedrene, duprezianene, and sesquithuriferol sesquiterpenes J. Am. Chem. Soc. 131, 7999– 8015.

27. Faraldos, J. A., Wu, S., Chappell, J., and Coates, R. M. (2007) Conformational analysis of (+)-germacrene A by variable-temperature NMR and NOE spectroscopy Tetrahedron 63, 7733– 7742.

28. Croteau, R., Felton, N., and Wheeler, C. (1985) Stereochemistry at C-1 of geranyl pyrophosphate and neryl pyrophosphate in the cyclization to (−)-bornyl pyrophosphate J. Biol. Chem. 260, 5956– 5962.

29. Faraldos, J. A., Zhao, Y., O'Maille, P. E., Noel, J., and Coates, R. M. (2007) Interception of the enzymatic conversion of farnesyl diphosphate to 5-epi-aristolochene by using a fluoro substrate analogue: 1-fluorogermacrene A from (2E,6Z)-6-fluorofarnesyl diphosphate ChemBioChem 8, 1826– 1833.

30. Jin, Y. H., Williams, D., Croteau, R., and Coates, R. M. (2005) Taxadiene synthase-catalyzed cyclization of 6-fluorogeranylgeranyl diphosphate to 7-fluoroverticillenes J. Am. Chem. Soc. 127, 7834– 7842.

31. Whittington, D., Wise, M., Urbansky, M., Coates, R., Croteau, R., and Christianson, D. (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase Proc. Natl. Acad. Sci. U.S.A. 99, 15375– 15380.

32. Vedula, L., Cane, D., and Christianson, D. (2005) Role of arginine-304 in the diphosphate-triggered active site closure mechanism of trichodiene synthase Biochemistry 44, 12719– 12727.

33. Schulbach, M., Brennan, P., and Crick, D. (2000) Identification of a short (C15) chain Z-isoprenyl diphosphate synthase and a homologous long (C50) chain isoprenyl diphosphate synthase in Mycobacterium tuberculosis J. Biol. Chem. 275, 22876– 22881.

34. Schulbach, M., Mahapatra, S., Macchia, M., Barontini, S., Papi, C., Minutolo, F., Bertini, S., Brennan, P., and Crick, D. (2001) Purification, enzymatic characterization, and inhibition of the Z-farnesyl diphosphate synthase from Mycobacterium tuberculosis J. Biol. Chem. 276, 11624– 11630.

35. Thulasiram, H. and Poulter, C. D. (2006) Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity J. Am. Chem. Soc. 128, 15819– 15823.

36. Woodside, A., Huang, Z., and Poulter, C. D. (1993) Trisammonium geranyl diphosphate, in Organic Synthesis, Collect. Vol. 8, pp 616− 620, Wiley, New York.

37. Kabsch, W. (1993) Automated processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants J. Appl. Crystallogr. 26, 795– 800.

38. (1994) The CCP4 suite: programs for protein crystallography, Acta Crystallogr. D 50, 760– 763.

39. Emsley, P. and Cowton, K. (2004) Coot: model-building tools for molecular graphics Acta Crystallogr. D 60, 2126– 2132.

40. Brunger, A., Adams, P., Clore, G., Delano, W., Gros, P., Grosse-Kunstleve, R. W., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T., and Warren, G. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination Acta Crystallogr. D 54, 905– 921.

41. Murshudov, G., Vagin, A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT Acta Crystallog.r D 55, 247– 255.

42. Murshudov, G., Vagin, A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method Acta Crystallogr. D 53, 240– 255.

43. Pannu, N., Murshudov, G., Dodson, E., and Read, R. (1998) Incorporation of prior phase information strengthens maximum-likelihood structure refinement Acta Crystallogr. D 54, 1285– 1294.

44. Skubak, P., Murshudov, G., and Pannu, N. (2004) Direct incorporation of experimental phase information in model refinement Acta Crystallogr. D 60, 2196– 2201.

45. Steiner, R., Lebedev, A., and Murshudov, G. (2003) Fisher's information in maximum-likelihood macromolecular crystallographic refinement Acta Crystallogr. D 59, 2114– 2124.

46. Vagin, A., Steiner, R., Lebedev, A., Potterton, L., Mcnicholas, S., Long, F., and Murshudov, G. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use Acta Crystallogr. D 60, 2184– 2195.

47. Winn, M., Isupov, M., and Murshudov, G. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement Acta Crystallogr. D 57, 122–133.

48. Winn, M., Murshudov, G., and Papiz, M. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions Acta Crystallogr. D 374, 300– 321.

49. Frisch, M. et al. (1998) Gaussian, Inc., Pittsburgh, PA.

50. Becke, A. (1993) Density-functional thermochemistry 3. The role of exact exchange J. Chem. Phys. 98, 5648-5652

51. Lee, C., Yang, W., and Parr, R. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density Phys. Rev. B 37, 785.

52. Hariharan, P. and Pople, J. (1973) The influence of polarization functions on molecular orbital hydrogenation energies Theor. Chim. Acta 28, 213.

53. Gonzalez, C. and Schlegel, H. (1989) An improved algorithm for reaction path following J. Chem. Phys. 90, 2154.

54. Gonzalez, C. and Schlegel, H. (1990) Reaction path following in mass-weighted internal coordinates J. Phys. Chem. 94, 5523.

55. Matsuda, S. and Wilson, W. (2006) Mechanistic insights into triterpene synthesis from quantum mechanical calculations. Detection of systematic errors in B3LYP cyclization energies Org. Biomol. Chem. 4, 530.

56. Adamo, C. and Barone, V. (1998) Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models J. Chem. Phys. 108, 664.

57. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis J. Comput. Chem. 25, 1605– 1612.

58. Still, W. C., Kahn, M., and Mitra, A. (1978) Rapid chromatographic technique for preparative separations with moderate resolution, Journal of Organic Chemistry 43, 2923-2925.

59. Collington, E. W., and Meyers, A. I. (1971) Facile and specific conversion of allylic alcohols to allylic chlorides without rearrangement Journal of Organic Chemistry 36, 3044-&.

60. Woodside, A. B., Zheng, H., and Poulter, C. D. (1988) Trisammonium geranyl diphosphate, Organic Syntheses 66, 211-219.

Chapter 4

A Conserved Amino Terminal Motif in Patchouli Alcohol Synthase Controls Product Distribution

4.1. ABSTRACT

Terpene cyclases are a class of enzymes in specialized metabolism that utilize the universal building blocks isopentenyl diphosphate and dimethylallyl diphosphate of primary metabolism to synthesize of broad array of downstream isoprenoid products. Terpene synthases cyclize C_{10}, C_{15}, or C_{20} isoprenoid diphosphates into one or more terpenes products. Here, we demonstrate the importance of a pseudo-conserved Arg-Pro (RP) motif at the amino terminal regions of a selection of both (product) diverse and more specific sesquiterpene cyclases including patchoulol synthase (PAS), 5-epi aristolochene synthase (TEAS), amorpha-4,11-diene synthase (ADS), and premnaspirodiene synthase (HPS). The corresponding motif in monoterpene cyclases (an Arg pair termed the RR motif) has been proposed to be involved with the isomerization event in monoterpene cyclases, although a clearly defined role has not yet been articulated. We find that mutation of the RP motif in PAS causes extreme product profile shifts to mechanistically simpler products upon mutation to any residue at Arg or to bulkier or more flexible residues at Pro, suggesting a newfound role for the RP motif in modulating product profile. However, TEAS, ADS, and HPS show only slight changes in product profile upon mutation. Additionally, mutational studies of a conserved salt bridge interaction between Arg of the RP motif and a C-terminal Glu provide additional evidence to suggest that certain plant terpene cyclases may "cap" their C-terminal active sites with their N-terminal domains, thereby aiding in the exclusion of water from the terpene cyclase active site, which houses highly reactive carbocation intermediates throughout the reaction. These results suggest that the RP motif has varied importance in product profile control among sesquiterpene cyclases and contributes to active site capping to facilitate the terpene cyclase reaction.

4.2. INTRODUCTION

Terpene cyclases (synthases) encompass a family of enzymes serving critical roles in the secondary metabolism and chemical ecology of plants, bacteria, fungi and marine organisms[1]. These enzymes participate in isoprenoid biosynthesis, catalyzing the conversion of geranyl diphosphate (GPP), farnesyl diphosphate (FPP), or geranygeranyl diphosphate (GGPP) into monoterpenes, sesquiterpenes, or diterpenes, respectively. In plants, terpene cyclases contain both an N-terminal and C-terminal domain. The C-terminal domain, which encompasses the last two-thirds of the protein, catalyzes ionization-initiated cyclization reactions. This domain contains two magnesium binding motifs (DDXXD and NSE/DTE), functioning to position the diphosphate moiety of the substrate[2] and also a sphere of hydrophobic residues to aid in the stabilization of carbocationic intermediates that the enzyme manipulates throughout the reaction.

The function of the N-terminal domain in most plant terpene cyclases remains unknown. Published crystal structures of plant monoterpene and sesquiterpene cyclases reveal that the N-terminal domain structurally resembles the active site region in glycosyl hydrolases, although does not exhibit a similar functionality[3]. A function for the amino-terminal region has however been assigned in certain plant diterpene cyclases. For example, in both abietadiene synthase[4] and *ent*-kaurene synthase[5, 6], the N-terminal domain participates in a proton-initiated reaction to cyclize GGPP into the intermediate copalyl diphosphate before proceeding on to each C-terminal ionization-dependent product. Evidence supporting the idea that these diterpene cyclases resemble ancestral terpene cyclases suggests that this domain may be vestigial in the more recent cyclases.[7]

Although the N-terminal domain has not been clearly assigned a functional role in plant monoterpene or sesquiterpene cyclases, it does appear to serve an indirect role in

catalysis for all cyclases. One proposal, based on several observations of crystal structures within this family, is that the N-terminal domain more closely associates with the C-terminal domain upon substrate binding. This association is thought to aide in the formation of a hydrophobic cavity within which carbocationic intermediates could survive without risk of premature solvent quench. For example, the crystal structure of (+)-bornyl diphosphate synthase reveals many hydrogen bonding interactions between the N- and C-terminal domains, two of which occur with aspartates of the DDXXD motif[8]. This same notion that the amino-terminal domain caps the C-terminal domain upon substrate or substrate analog binding was also recognized in crystal structures of 5-epi-aristolochene synthase[9]. Another proposal with regard to the monoterpene synthases is that a highly conserved Arg pair (called the RR motif) is responsible for substrate isomerization, a necessary requirement for cyclization catalyzed by monoterpene cyclases[10]. Additional amino terminal studies including mutations at this pair of residues in two sesquiterpene cyclases (δ-selinene and γ-humulene synthase, containing the RR motif) and a diterpene cyclase (abietadiene synthase, containing a similar KR-motif) have demonstrated a variety of affects on both the enzymatic activity and product profile.[4, 11] Many plant sesquiterpene cyclases contain an RP motif in place of the Arg pair, although mutational analyses have yet to be performed on this variation of the motif.

In an effort to further understand the structural and functional role of the RP motif variant, we have performed a thorough mutational analysis at the RP motif in patchoulol synthase from *Pogostemon cablin* (PAS). Patchoulol synthase is a moderately promiscuous sesquiterpene cyclase, producing 13 or more sesquiterpenes with its major product being an alcohol known as (-)-patchoulol[12] (Figure 4.1).

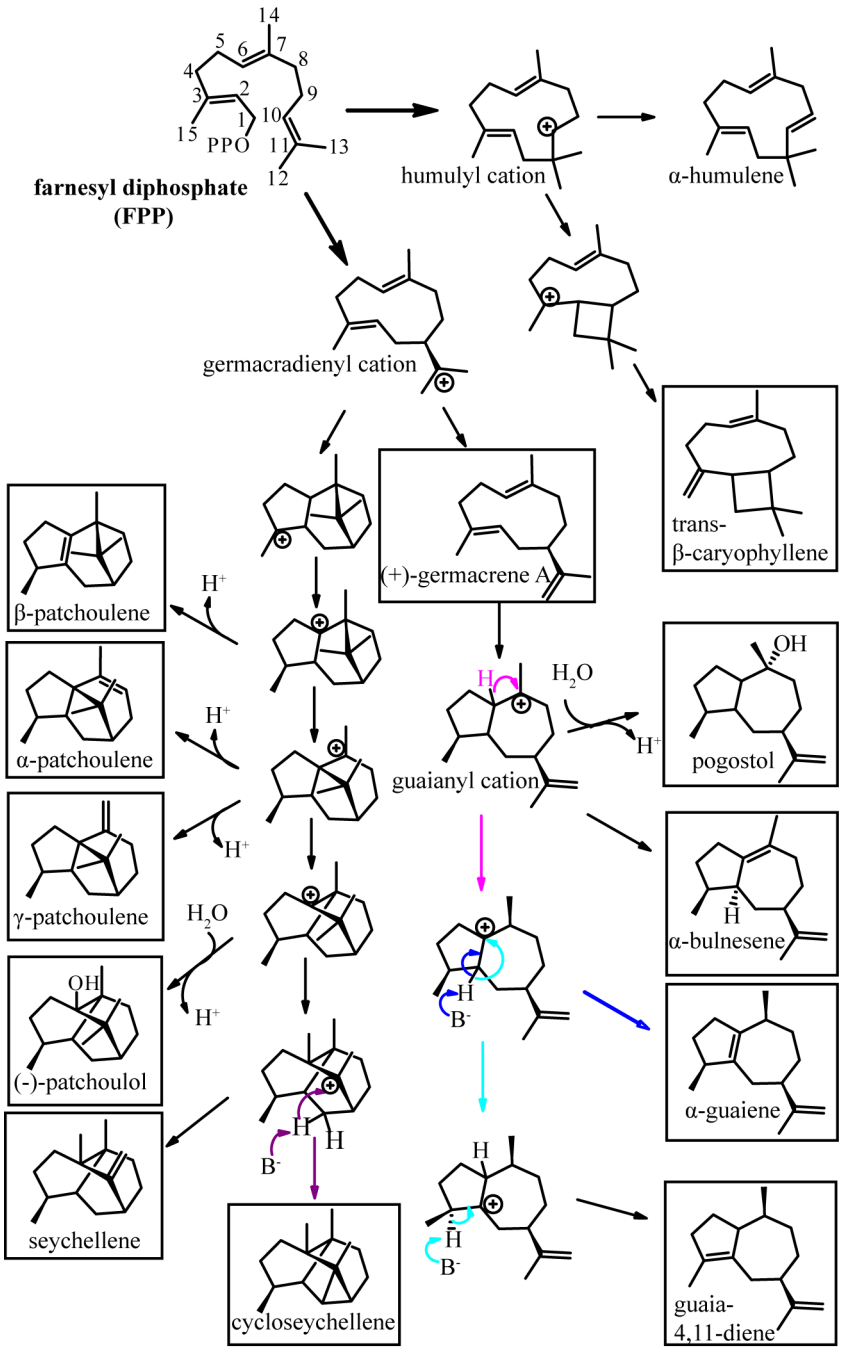


Figure 4.1. Reaction Mechanism of patchoulol synthase accounting for all thirteen sesquiterpene products. Mechanism was constructed with reference to Deguerry *et al* (2006).[12]

Given that the RP motif is somewhat conserved among the plant sesquiterpene cyclases, we have also mutated these positions in three other sesquiterpene synthases (5-epi-aristolochene synthase from *Nicotiana tabacum*: TEAS; amorpha-4,11-diene synthase from *Artemisia annua*: ADS; premnaspirodiene synthase from *Hyoscyamus muticus*: HPS) in an attempt to define a broader role for this variant of the motif. Recently published crystal structures of certain TEAS mutants[13] reveal important interactions between the RP motif and other regions of the protein, which are assumed to be essential for normal RP motif function. This analysis reveals a specific and cooperative structural and functional role for the RP motif in PAS, and also demonstrates that the motif has varying degrees of importance in other plant sesquiterpene cyclases.

4.3. RESULTS AND DISCUSSION

4.3.1. RP motif in PAS

Wild type PAS utilizes FPP to synthesize 13 identifiable products that can be grouped into four categories according to their mechanistic complexity: 1) Mechanistically simple products, such as β-elemene (a cope-rearrangement of germacrene A[14]) β-caryophyllene, and α-humulene. These molecules have carbocation precursors that are easily created through cyclization followed by only one or two rearrangements before the final elimination step. 2) Mechanistically intermediate products derived from the guaianyl cation, including pogostol, α-guaiene, α-bulnesene, and guaia-4,11-diene. The sesquiterpene α-bulnesene is considered the simplest of this group because its formation requires no further hydride rearrangements. 3) Mechanistically intermediate products derived from the patchoulene cation, including α-, β-, and γ-patchoulene. 4) Mechanistically complex products such as seychellene,

cycloseychellene and (-)-patchoulol, which are derived from the most intricate series of catalytic events (Figure 4.1).

In order to initially assess the effect of amino-terminal truncation on PAS product profile, a series of amino terminal deletions were constructed, including N6, N10, N15, N16, and N17, which represent 5, 9, 14, 15, and 16 amino acids deleted from the amino terminus, respectively (Figure 4.2). Of these constructs, only N17 significantly alters the product profile and is unable to produce the majority of products including (-)-patchoulol. Notably, N15, which is two amino acids longer than N17, completely restores the product distribution (Figure 4.3, Table 4.1, Table 4.2). The two amino acids that appear to play a role in this restoration are Arg15-Pro16 (termed the RP motif).

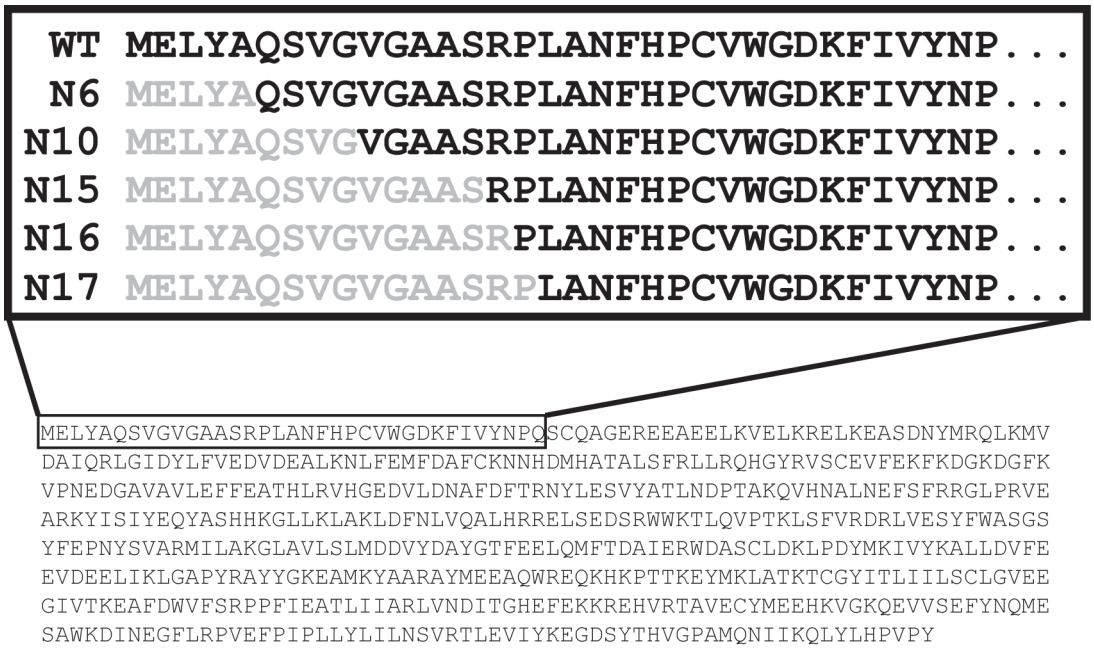


Figure 4.2. Truncation mutant constructs in patchoulol synthase. The lower sequence represents the full PAS sequence, shown above this in the box the amino terminal sequences of each construct; truncated portions of each construct are shown in gray.

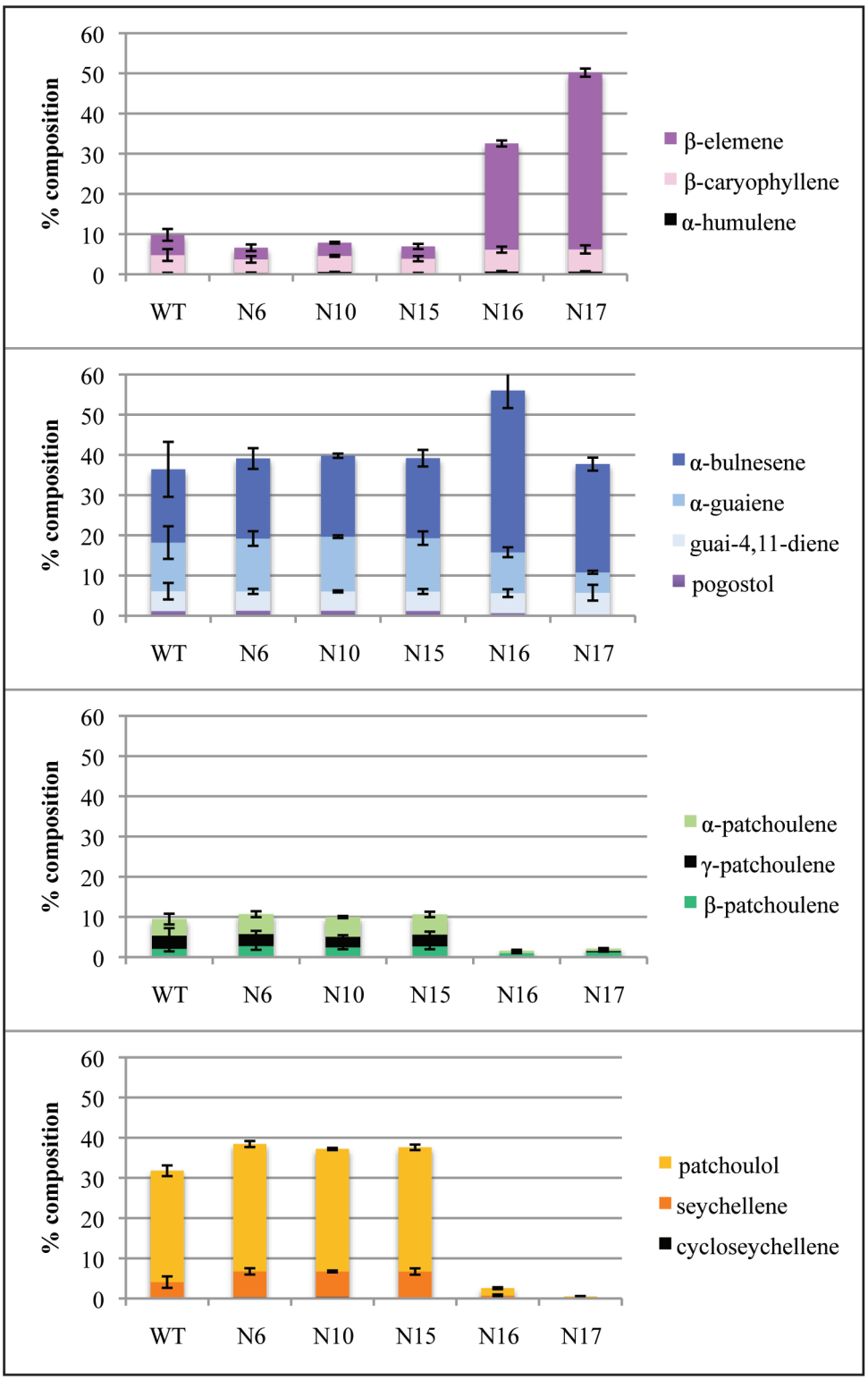


Figure 4.3. Percent compositions of all products in the truncation mutants of PAS. Products are divided into four categories based on their mechanistic complexity of formation.

Mutations at both Arg15 and Pro16 in the full-length PAS construct were subsequently made in order to further define a functional role for this pair of amino acids. These mutations include R15K, R15Q, R15E, P16A, P16S, P16I, P16R, and P16G.

The three mutations made at Arg15 (R15K, R15Q, and R15E) have significantly compromised activities and cause dramatic shifts of the PAS product profile toward mechanistically simpler products (Figure 4.4, Table 4.3, Table 4.4). All three mutations at this position produce heightened levels of germacrene A (a mechanistically simple product) and α-bulnesene, and negligible levels of the more complex products. These results lead to the conclusion that Arg15 is critical for formation of complex products and for maintenance of the PAS product profile.

In general, mutation at Pro16 progressively derails the product profile toward mechanistically simpler products in going from wild type to P16A to P16S to P16I to P16R to P16G (Figure 4.4, Table 4.5, Table 4.6). The bulkiness of the amino acid substituted at this position negatively affects its product profile: a bulkier residue produces larger amounts of germacrene A and α-bulnesene and smaller amounts of the mechanistically complex products. The Gly substitution is an exception to this trend: of all the Pro16 mutants, it is the smallest substitution yet it displays the simplest product profile (Figure 4.4). This result highlights the importance of structural rigidity at the Pro16 position in the RP motif. Assuming dynamics play a role in maintenance of the PAS product profile, the flexible Gly residue at this position may disrupt the enzyme's ability to effectively cap the active site with its N-terminus. Notably, conversion of the RP motif to a monoterpene synthase-like RR motif by means of the P16R mutant greatly reduces the mechanistic complexity of the product profile (Figure 4.4).

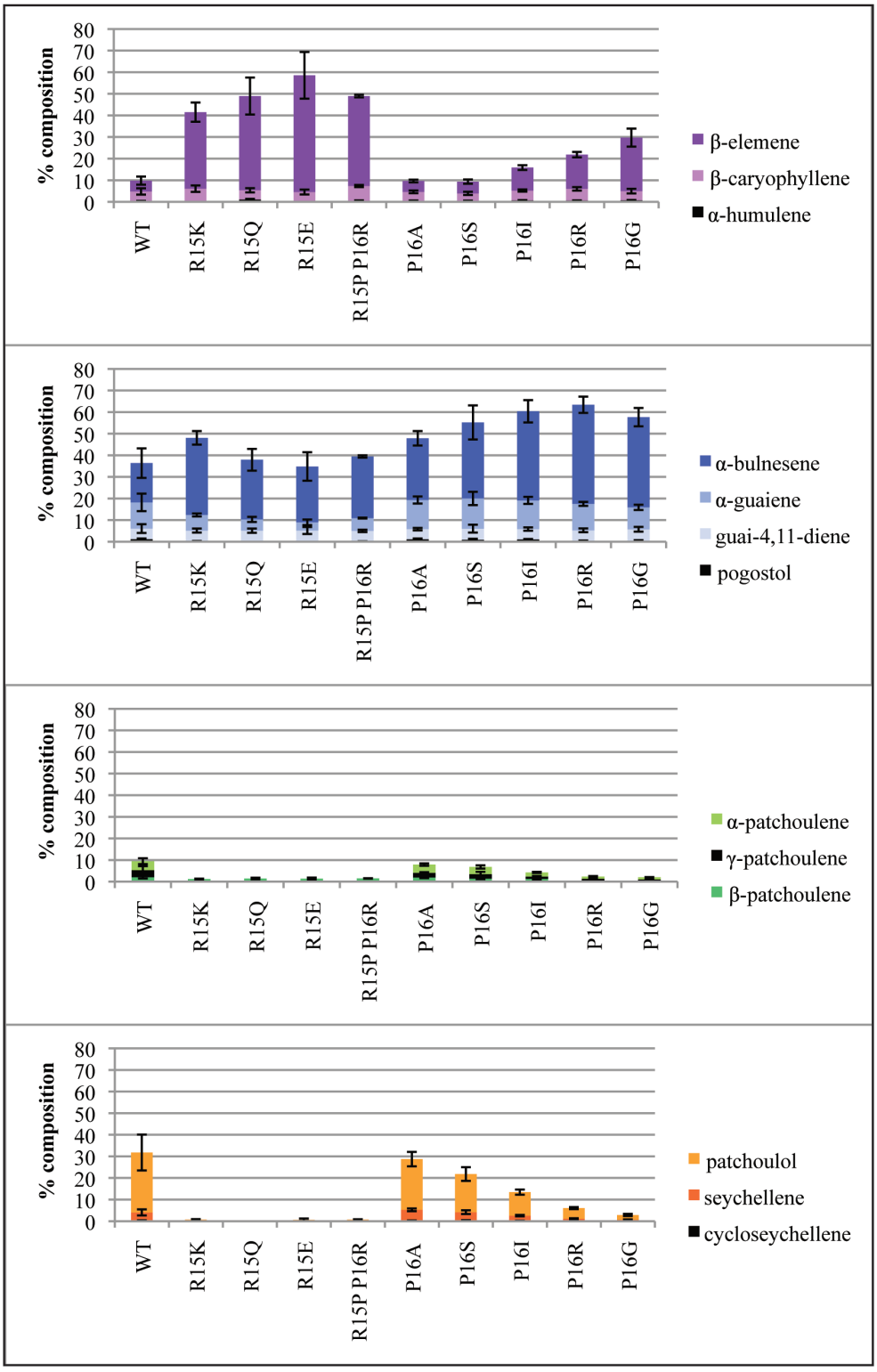


Figure 4.4. Percent compositions of all products in PAS RP motif mutants. Products are divided into four categories based on their mechanistic complexity of formation.

4.3.2. RP motif mutants in other sesquiterpene cyclases

Three other sesquiterpene cyclases were mutated at their RP motifs, including premnaspirodiene synthase from *Hyoscyamus muticus* (HPS), amorpha-4,11-diene synthase from *Artemisia annua* (ADS), and 5-epi-aristolochene synthase from *Nicotiana tabacum* (TEAS). These cyclases are highly specific, producing one major product and many minor products at very low, sometimes undetectable levels[15, 16]. Initially, all three enzymes were engineered to transform their RP motif into an RR motif with a single point mutation at the Pro position. HPS P18R and ADS P11R maintain wild type levels of their major products, with values of premnaspirodiene at 93.7(±0.5)% and 93.4(±0.3)% for HPS WT and HPS P11R, respectively, and values of amorpha-4,11-diene at 84.3(±3.2)% and 85.0(±1.6) for ADS WT and ADS P11R, respectively (Table 4.11, Table 4.12, Table 4.13, Table 4.14). There are no accompanying variations in percent composition of minor products for either enzyme. TEAS P16R, however, does show slight changes in its product profile compared to wild type, but only with respect to two of its products: in going from wild type to P16R, the % composition of 5-epi-aristolochene decreases from 82.6(±0.4)% to 79.6(±0.5)%, while the % composition of germacrene-A increases from 1.7(±0.1)% to 4.0(±0.2)% (Figure 4.7, Table 4.9, Table 4.10). These product profile changes observed for the TEAS mutant are easily accounted for compared to product profile changes for the PAS mutants discussed above.

Crystal structure data from TEAS enabled the identification of another feature that may be important for amino-terminal capping. The original crystal structure of wild type TEAS was solved in 1997, however the first 16 amino acids are missing from the PDB coordinates, likely due to a dynamic amino-terminus[9]. A more recently published crystal structure of wild type TEAS complexed with the non-hydrolyzable FPP analog trans,2-fluoro-farnesyl diphosphate (2F-FPP) shows several more residues built into the N-terminus,

including Arg15 and Pro16[13]. In this structure, Arg15 participates in a salt-bridge interaction with Glu312 of the C-terminal domain. This interaction can also be observed in monoterpene cyclases; one example is the corresponding Arg68-Glu368 salt bridge in the crystal structure of limonene synthase from *Mentha spicata* complexed with a fluorinated derivative of geranyl diphosphate[17]. Given that the two amino acids comprising the salt bridge are conserved among sequences of monoterpene and sesquiterpene cyclases (Figure 4.5) and that the salt bridge itself is noted in several reports of plant terpene cyclase crystal structures[8, 17], it is very possible that this electrostatic interaction assumes some structural or functional significance. Both Glu312 in TEAS and Glu368 in limonene synthase from *Mentha spicata* are located in the center of a helix that shields one side of the active site, which means that the salt bridge may help enclose this area by effectively capping it with the amino terminal region in both enzymes, which has been previously suggested[17] (Figure 4.5). From this observation made in TEAS, a series of additional mutations were made at Glu312 (Glu315 in PAS) and Arg15 in both TEAS and PAS to explore the significance of this salt bridge interaction.

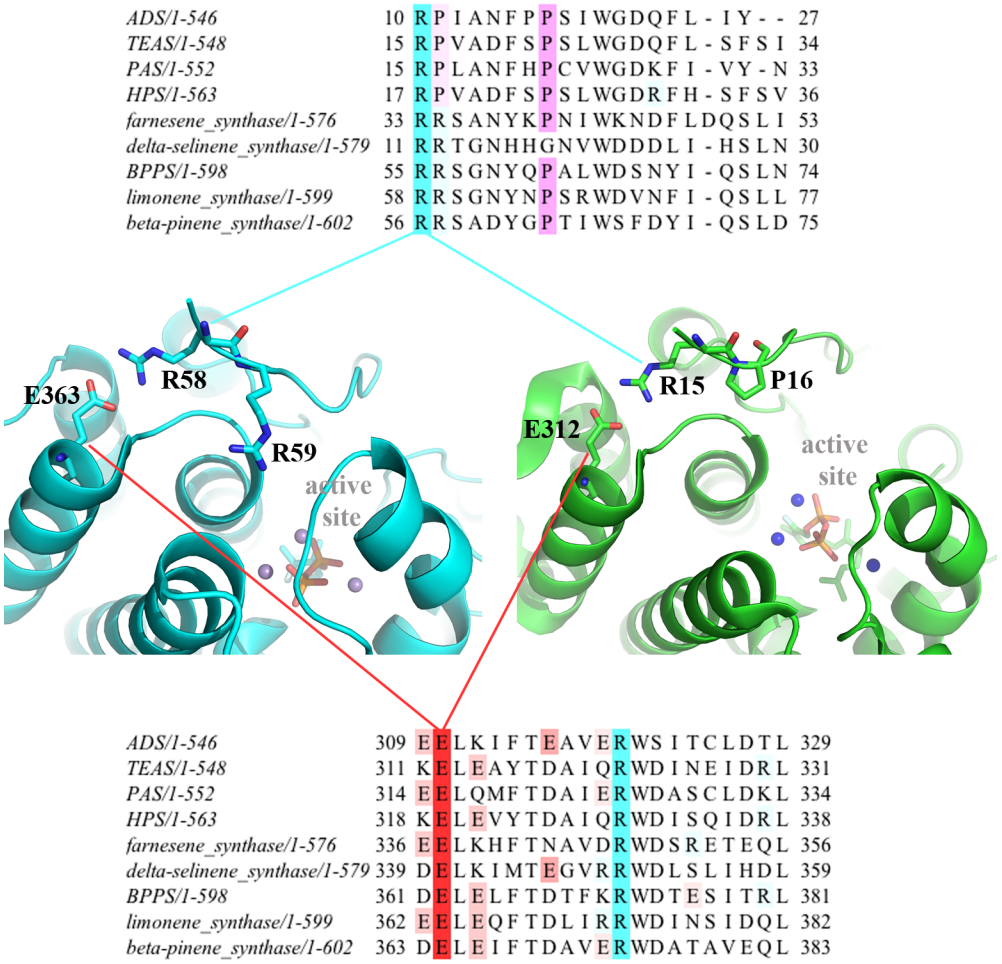


Figure 4.5. Conservation of a salt bridge interaction with the RP or RR motif. Shown in the center are two images from the crystal structures of limonene synthase (left, pdb ID: 2ONG[17]) and TEAS (right, pdb ID: 3M00[13]). Shown above and below are sequence alignments demonstrating conservation of both moieties among the terpene cyclases discussed in this work. BPPS stands for (+)-bornyl diphosphate synthase.

4.3.3. Salt bridge mutants in PAS and TEAS

Mutants were constructed for both TEAS and PAS to determine not only the necessity of the salt bridge for normal catalysis but also whether the salt bridge is contextually dependent (whether Arg and Glu could swap positions and still maintain wild type activity).

The mutations in PAS that are relevant for the salt bridge interaction include E315D, E315Q, E315R, R15E, and R15E/E315R, which have total product peak areas at 73%, 8%,

6%, 6%, and 3% of the wild type value, respectively (Figure 4.8). When comparing E315D to E315Q, the drastic drop in product peak area is most likely due to the absence of a salt bridge in the E315Q mutant (Table 4.7, Table 4.8). This result highlights the importance of the salt bridge in PAS for stability and/or catalysis. Additionally, R15E/E315R behaves very poorly compared to wild type, which suggests that the salt bridge is contextual. All mutant product profiles, regardless of whether or not the salt bridge is present, show increased percent production of mechanistically simpler compounds and α-bulnesene compared to wild type (Figure 4.6, Table 4.7, Table 4.8). However, both E315D and R15E/E315R produce measurable levels of patchoulol while the other mutants do not, suggesting that the salt bridge may contribute to the enzyme's ability to make this mechanistically complex product (Figure 4.6).

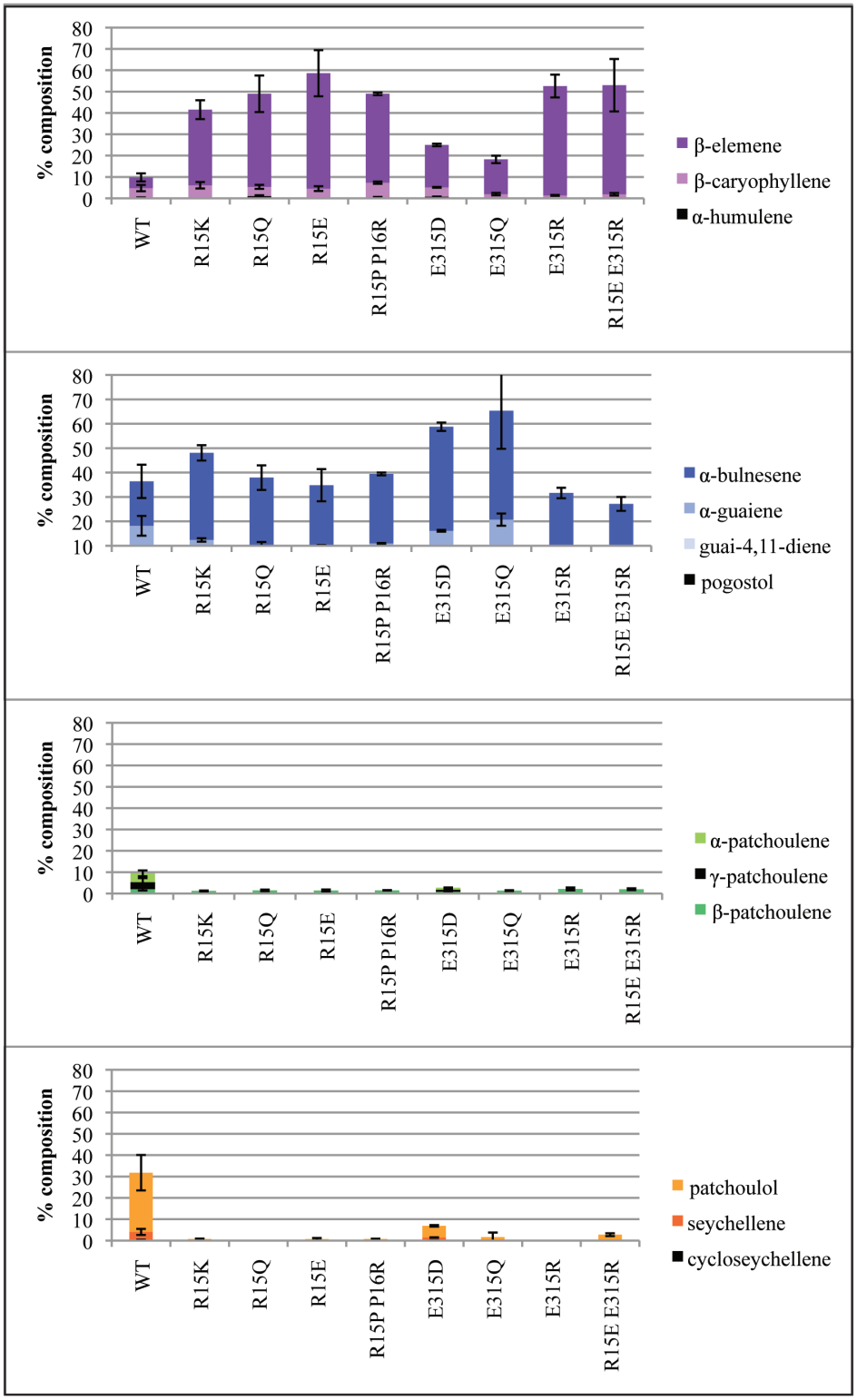


Figure 4.6. Percent compositions of all products in PAS salt bridge mutants. Products are divided into four categories based on their mechanistic complexity of formation.

The E315D mutant can also produce significant amounts of the later patchoulene products (α-patchoulene and γ-patchoulene) while the E315Q mutant cannot, which is to be expected considering patchoulol is derived from a rearranged patchoulene carbocation. One would therefore also expect that the patchoulene products would be apparent in the profile of the double mutant R15E/E315R, however this is not the case. The most likely reason for this is that α-patchoulene and γ-patchoulene are more difficult to resolve on the GC chromatogram due to the large number of unique sesquiterpene hydrocarbons that elute in this region. Patchoulol, on the other hand, has a much longer retention time than any of the other PAS products, and is both easily resolvable and contains the unique m/z ion at 222, characteristic of a sesquiterpene alcohol. These results from all salt bridge mutants present strong evidence that in PAS, 1) the salt bridge is important for stability and catalysis, 2) the context of the amino acid pair participating in the salt bridge is important, and therefore this pair cannot be switched, and 3) the salt bridge is necessary to observe the production of patchoulol and the later patchoulenes (α-patchoulene and γ-patchoulene).

The relevant mutants constructed in TEAS include E312R, R15E/E312R, E312D, E312Q, and R15E. TEAS E312R and R15E/312R are unstable mutants that produce large aggregation peaks when run on a gel filtration column and have negligible activity after overnight incubations with FPP. The calculated percent product compositions for these mutants are therefore considered unreliable and this data was not used for subsequent analysis. TEAS E312D, E312Q, and R15E all produce higher levels of germacrene A than wild type, with values of 6.9(±0.3)%, 6.7(±0.3)%, and 9.1(±0.2)%, respectively, compared to a wild type value of 1.7(±0.1)% (Table 4.9, Table 4.10). These percentages trade off with the percent compositions of 5-epi-aristolochene, which are 77.3(±0.6)%, 77.4(±0.4)%, and 74.5(±0.5)% for TEAS E312D, D312Q, and R15E, respectively, compared to a wild type value of

82.6(±0.4) (Figure 4.7, Table 4.9, Table 4.10). Minimal and often insignificant variations are observed between mutants and wild type for all remaining detectable minor products. In addition to having matching product profiles, TEAS E312D and E312Q have total product peak areas at 89% of the wild type value (Figure 4.8). It is therefore clear that unlike in PAS, the absence of this salt bridge in TEAS does not affect product profile or activity. However, the fact that both E312R and R15E/E312R are highly unstable and virtually inactive is an indication that the context of the salt bridge is highly important.

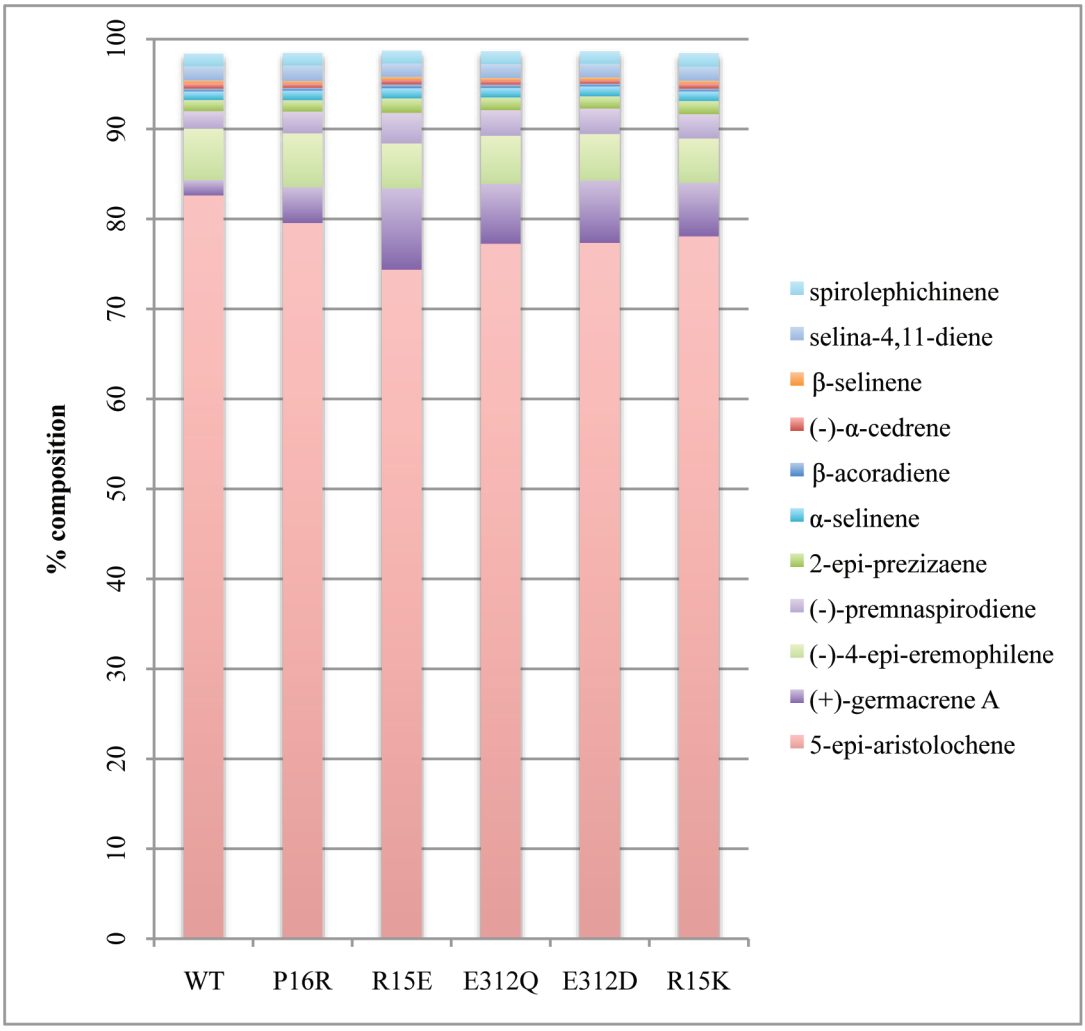


Figure 4.7. Percent compositions of all products in TEAS mutants

4.3.4. Conclusions

From these results, it is clear the RP motif is crucial for normal activity in PAS. Any mutation at Arg15 and progressively bulkier substitutions at Pro16 are detrimental toward production of mechanistically complex PAS products. One of the reasons that Arg15 is necessary for the maintenance of wild type activity is most likely due to an electrostatic interaction that exists between this residue and Glu315, as also seen in crystal structures of wild type TEAS complexed with fluoro-FPP analogs[13], a crystal structure of limonene synthase complexed with a fluoro-GPP analog[17], and a crystal structure of (+)-bornyl diphosphate synthase complexed with pyrophosphate[8]. Bulkier mutations at Pro16 produce product profiles that are increasingly abundant in mechanistically simpler products such as germacrene A and α-bulnesene, with the exception of P16G, which displays one of the simplest product profiles. The behavior of the PAS P16G mutant suggests that the restricted conformational space explored by Pro residues is important in an otherwise flexible amino terminal segment. Therefore, in PAS, Arg15 and Pro16 cooperate to achieve dynamic regulation of the amino-terminal region; the structural rigidity of Pro16 aides in the positioning of Arg15 such that it can interact with a residue in the C-terminal domain of the protein, thereby capping the active site with the amino-terminal region. This capping mechanism probably helps shield reactive carbocation intermediates present in the active site from bulk solvent throughout the course of the reaction. The rigidity of Pro16 in PAS is analogous to the structural stabilization provided by an additional salt bridge observed in (+)-bornyl diphosphate synthase between the second Arg in the RR motif (R56) and Asp355.[8]

There are undoubtedly other amino-terminal residues involved in hydrogen bonding, van der Waals, and electrostatic interactions that contribute to amino-terminal "active site capping" in plant terpene cyclases. However, the observation that these two residues can

restore the wild type product profile in going from the N17 truncation mutant to N15 is strong evidence for the importance of these specific residues in the amino-terminal region of PAS.

Of the three other sesquiterpenes that have been mutated at their RP motifs (TEAS, HPS, and ADS), only TEAS shows a mutant product profile that differs from wild type. Unlike PAS, product fluctuations observed in the TEAS product profile are quite easily accounted for: in TEAS P16R and in all active salt bridge mutants (TEAS E312D, TEAS E312Q, and TEAS R15E), a loss in the major product 5-epi-aristolochene directly corresponds to a gain in the mechanistically simple product germacrene A; these losses and gains are at most 7-8%. In contrast, PAS P16R loses almost 20% patchoulol and gains 17% α-bulnesene compared to wild type, in addition to loss and gain of other products. The fact that PAS P16R appears to have a much more dramatic affect on the wild type product profile compared to TEAS P16R suggests that perhaps enzyme promiscuity and RP motif functionality are somehow correlated. This hypothesis also makes sense with the results observed for the two other relatively non-promiscuous enzymes HPS and ADS, where mutant product profiles either did not deviate from wild type or exhibit undetectable product derailment. However, in previous work by Little *et al* (2002) on two promiscuous sesquiterpene cyclases, δ-selinene synthase shows dramatic changes in product profile when mutated in this region while γ-humulene synthase does not[11]. This result suggests that the level of promiscuity does not necessarily correlate with RP motif functional trends.

In comparison to PAS, mutation of the salt bridge in TEAS does not appear to be as important for maintenance of the product profile. This was expected to a certain extent due to the fact that the TEAS RP motif mutants also do not alter the product profiles to a large degree. However, Glu312 in TEAS does appear to be important for enzyme stability and activity, given the E312R behaves poorly and the double mutant R1E/E312R (that reverses the

salt bridge) is very unstable and almost completely inactive. This result suggests that the salt bridge between the Glu and Arg is indeed contextually dependent. For all active salt bridge mutants, losses in the percentage of 5-epi-aristolochene correspond almost exactly to gains in % production of germacrene A. This result may be obvious considering germacrene A is the second most produced sesquiterpene in the TEAS profile. However, it means that mutations in the RP motif or the salt bridge in TEAS result in derailment towards one very simple monocyclic product, indicating premature carbocation release from the active site.

In PAS, given that the P16G mutant and mutants at the Arg15 position cause dramatic product profile shifts, it appears that an increase in flexibility within this region has a similar effect as a disruption in the salt bridge interaction. It is therefore likely that the dynamics of this amino-terminal region of PAS are in part controlled by Pro16, which offers the structural rigidity required to position Arg15 appropriately for the salt bridge interaction with Glu315. Variations of the RR motif appear to exist in terpene cyclases that are thought to be the ancestors of this family of enzymes. For example, abietadiene synthase, a diterpene cyclases from *Artemisia annua*, contains a KR-motif that is located in a similar position as the RR motif in monoterpene cyclases. When both residues are mutated to Ala, this enzyme can no longer efficiently perform the ionization-dependent reaction, although there is no report of product derailment[18]. The fact that this motif is present in a cyclase that is thought to be ancestral to present day terpene cyclases is an indication of its rooted importance within the family. The problem is indeed more complex than it looks, because some diterpene cyclases, such as levopimaradiene synthase from Ginkgo biloba, do not contain this motif.

4.4. METHODS

4.4.1. Mutant Construction, Overexpression, and purification

All truncation mutants and point mutations were constructed using the QuikChange protocol with PfuTurbo® DNA Polymerase (Stratagene) together with a 7 min PCR extension time. The plasmid pHis9Gateway (a pet28-based gateway vector containing an N-temrinal nine-histidine tag) containing the mutated terpene cyclase gene insert was transformed into *E. coli* Bl21(DE3) competent cells (Novagen). One colony was grown in LB media (75 ml) overnight at 37°C, 25 ml of the overnight culture was transferred into one liter of TB media and grown at 37°C until an OD600 of 1.2. Isopropyl-β-D-thiogalactoside (0.2 mM final concentration) was then added, cells were shaken for 5-6 hours at 22°C, harvested by centrifugation and lysed using lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1% (v/v) Tween-20, 10% (v/v) glycerol, 10 mM 2-mercaptoethanol) containing lysozyme (1 mg ml-1). The lysate was stirred at 4°C for 1 hr, sonicated, and centrifuged at 21,000 rpm for 45 min at 4°C. The supernatant was loaded onto a column containing Ni-NTA agarose S3 resin (Qiagen), washed with lysis buffer and wash buffer (lysis buffer without Tween- 20), and then eluted with elution buffer (wash buffer containing 250 mM imidazole). The protein was digested with thrombin overnight during dialysis in buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl) containing 10 mM 2-mercaptoethanol. The dialyzed solution was passed through a column containing Benzamidine Sepharose 4 Fast Flow (high sub) (GE Healthcare) and Ni-NTA agarose. The eluant concentrated and passed through a HiLoad™ 16/16 Superdex™ 200 prep grade (GE Healthcare) gel filtration column using dialysis buffer containing 2 mM DTT. IPK fractions were combined, concentrated, and frozen at –80°C.

4.4.2. Specific activity measurements and product profile quantification by GC-MS

All wild type and mutant terpene cyclases samples were prepared in triplicate using the vial assay protocol.[19] 1μM enzyme was injected into a 500μl aqueous layer containing

10mM MgCl2, 100μM FPP, and a 3-component buffer system (25 mM 2-(N-morpholino)ethanesulfonic acid (MES), 50 mM Tris, and 25 mM 3-(cyclohexylamino)propanesulfonic acid (CAPS) at pH 7.0). The aqueous layer was immediately overlayed with an equal volume of ethyl acetate and the reactions were incubated overnight, vortexed, and then run on the Gas Chromatograph Mass Spectrometer (GCMS) using a previous established protocol[13].

SUPPORTING INFORMATION

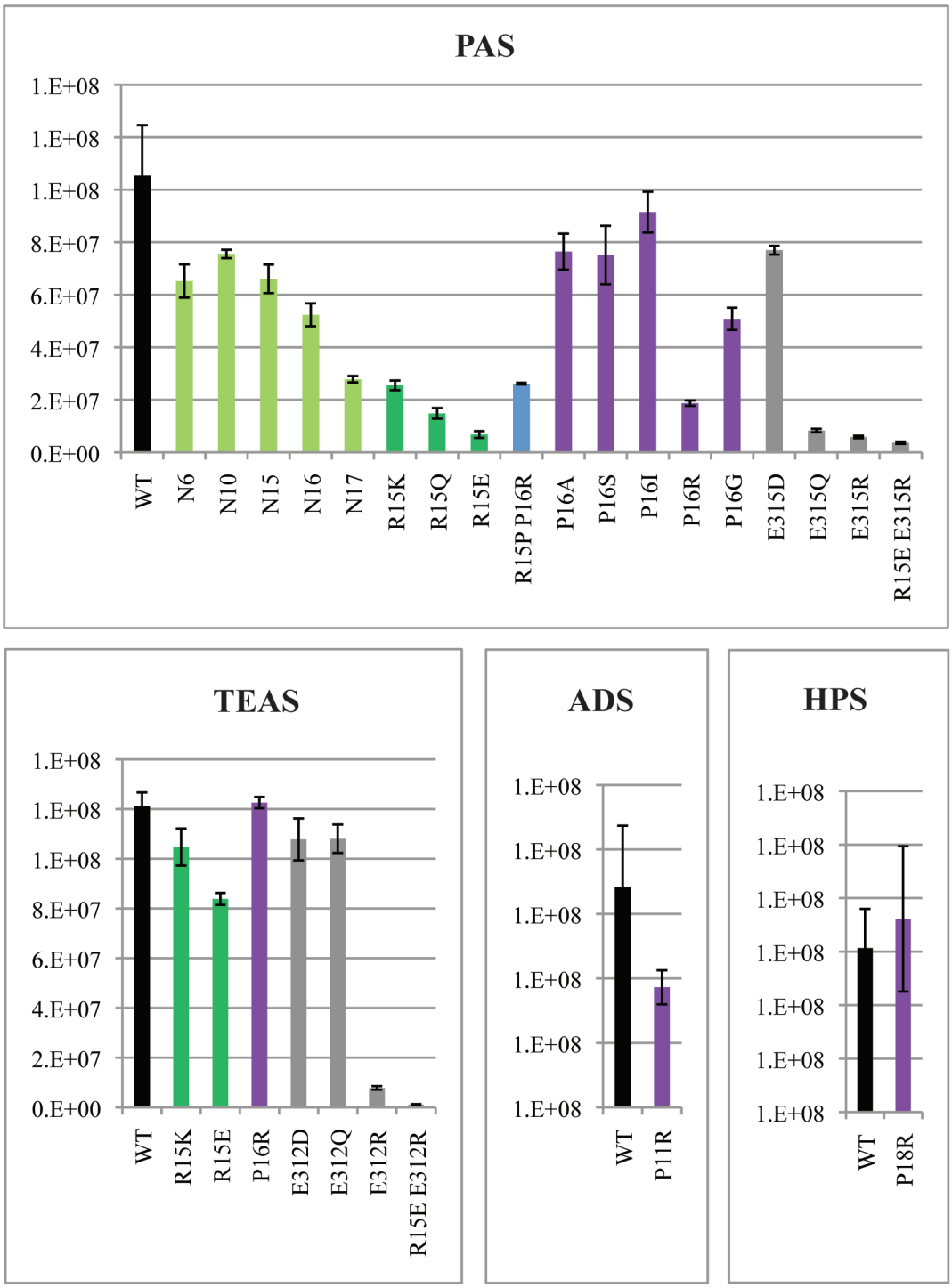


Figure 4.8. Total peak areas for all mutant and wild type enzymes.

Table 4.1. PAS truncation mutant % compositions

	WT	N6	N10	N15	N16	N17
β-patchoulene	2.05	2.73	2.41	2.68	1.12	1.52
β-elemene	5.01	2.86	3.28	3.05	26.39	43.95
cycloseychellene	0.28	0.37	0.39	0.25	0.00	0.00
β-caryophyllene	4.49	3.35	3.97	3.55	5.43	5.51
α-guaiene	12.10	13.14	13.61	13.27	10.17	5.07
seychellene	3.79	6.39	6.32	6.46	0.74	0.00
humulene	0.31	0.38	0.61	0.33	0.71	0.70
α-patchoulene	4.13	4.95	4.90	5.02	0.47	0.30
γ-patchoulene	3.28	3.00	2.65	2.92	0.00	0.29
guai-4,11-diene	4.92	4.76	4.77	4.81	4.96	5.66
α-bulnesene	18.20	19.91	20.13	19.89	40.22	26.93
pogostol	1.18	1.29	1.29	1.22	0.65	0.06
patchoulol	27.72	31.68	30.49	30.91	1.83	0.50
other	7.54	5.19	5.18	5.63	7.30	9.51

Table 4.2. PAS truncation mutant standard deviations for % compositions

	WT	N6	N10	N15	N16	N17
β-patchoulene	0.62	0.90	0.45	0.73	0.15	0.18
β-elemene	1.88	0.63	0.14	0.48	2.61	2.13
cycloseychellene	0.12	0.06	0.11	0.02	0.00	0.00
β-caryophyllene	1.47	0.82	0.24	0.64	0.75	1.01
α-guaiene	4.06	1.83	0.34	1.67	1.25	0.37
seychellene	1.42	0.78	0.23	0.80	0.27	0.00
humulene	0.10	0.08	0.02	0.05	0.11	0.06
α-patchoulene	1.34	0.74	0.26	0.68	0.24	0.13
γ-patchoulene	1.85	0.80	0.37	0.73	0.00	0.07
guai-4,11-diene	2.04	0.64	0.25	0.63	0.96	1.96
α-bulnesene	6.83	2.59	0.54	2.07	4.36	1.62
pogostol	0.33	0.70	0.16	0.46	0.16	0.11
patchoulol	8.29	6.15	1.10	4.01	0.22	0.06
other	2.83	1.96	1.88	2.13	3.02	4.23

Table 4.3. PAS Arg15 mutant % compositions

	WT	R15K	R15Q	R15E	R15P/ P16R
β-patchoulene	2.05	1.21	1.45	1.41	1.50
β-elemene	5.01	35.41	43.60	54.04	41.62
cycloseychellene	0.28	0.00	0.00	0.00	0.00
β-caryophyllene	4.49	6.10	4.29	4.54	6.82
α-guaiene	12.10	7.24	5.23	3.81	5.92
seychellene	3.79	0.00	0.00	0.00	0.00
humulene	0.31	0.00	1.06	0.00	0.50
α-patchoulene	4.13	0.00	0.00	0.00	0.00
γ-patchoulene	3.28	0.00	0.00	0.00	0.00
guai-4,11-diene	4.92	5.04	5.11	5.19	4.97
α-bulnesene	18.20	35.68	27.61	25.81	28.51
pogostol	1.18	0.11	0.00	0.00	0.06
patchoulol	27.72	0.75	0.00	0.65	0.75
other	7.54	8.46	11.64	10.75	9.36

Table 4.4. PAS Arg15 mutant standard deviations for % compositions

	WT	R15K	R15Q	R15E	R15P/ P16R
β-patchoulene	0.62	0.18	0.36	0.47	0.10
β-elemene	1.88	4.44	8.57	10.81	0.60
cycloseychellene	0.12	0.00	0.00	0.00	0.00
β-caryophyllene	1.47	1.48	0.96	1.13	0.53
α-guaiene	4.06	0.69	1.20	1.29	0.20
seychellene	1.42	0.00	0.00	0.00	0.00
humulene	0.10	0.00	0.35	0.00	0.14
α-patchoulene	1.34	0.00	0.00	0.00	0.00
γ-patchoulene	1.85	0.00	0.00	0.00	0.00
guai-4,11-diene	2.04	0.94	0.99	1.59	0.56
α-bulnesene	6.83	3.15	5.04	6.58	0.55
pogostol	0.33	0.19	0.00	0.00	0.10
patchoulol	8.29	0.19	0.00	0.59	0.12
other	2.83	3.64	4.54	5.90	4.03

Table 4.5. PAS Pro16 mutant % compositions

	WT	P16A	P16S	P16I	P16R	P16G	R15K
β-patchoulene	2.05	1.94	1.40	1.25	0.86	1.03	1.21
β-elemene	5.01	5.02	5.43	10.64	15.86	24.76	35.41
cycloseychellene	0.28	0.28	0.30	0.18	0.00	0.00	0.00
β-caryophyllene	4.49	4.07	3.47	4.53	5.41	4.28	6.10
α-guaiene	12.10	13.48	13.98	13.27	12.25	10.16	7.24
seychellene	3.79	4.99	3.88	2.37	1.18	0.72	0.00
humulene	0.31	0.56	0.45	0.68	0.60	0.70	0.00
α-patchoulene	4.13	3.84	3.30	1.90	1.03	0.59	0.00
γ-patchoulene	3.28	2.13	2.08	1.10	0.45	0.41	0.00
guai-4,11-diene	4.92	4.68	5.17	4.81	4.98	5.30	5.04
α-bulnesene	18.20	28.53	35.16	41.24	45.85	41.77	35.68
pogostol	1.18	1.16	0.88	1.05	0.29	0.44	0.11
patchoulol	27.72	23.48	17.63	10.90	4.88	2.14	0.75
other	7.54	5.85	5.90	6.07	6.37	7.69	8.46

Table 4.6. PAS Pro16 mutant standard deviations for % compositions

	WT	P16A	P16S	P16I	P16R	P16G	R15K
β-patchoulene	0.62	0.32	0.23	0.29	0.10	0.18	0.18
β-elemene	1.88	0.63	0.96	1.11	1.27	4.19	4.44
cycloseychellene	0.12	0.05	0.12	0.03	0.00	0.00	0.00
β-caryophyllene	1.47	0.62	0.73	0.50	0.81	1.08	1.48
α-guaiene	4.06	1.63	3.09	1.65	0.94	1.22	0.69
seychellene	1.42	0.62	0.88	0.32	0.19	0.12	0.00
humulene	0.10	0.07	0.07	0.09	0.05	0.13	0.00
α-patchoulene	1.34	0.54	0.76	0.27	0.23	0.09	0.00
γ-patchoulene	1.85	0.29	1.07	0.26	0.39	0.19	0.00
guai-4,11-diene	2.04	0.56	1.81	0.79	0.85	1.10	0.94
α-bulnesene	6.83	3.35	7.87	5.15	3.77	4.24	3.15
pogostol	0.33	0.32	0.52	0.22	0.26	0.30	0.19
patchoulol	8.29	3.35	3.15	1.20	0.46	0.51	0.19
other	2.83	2.15	2.16	2.32	2.73	3.14	3.64

Table 4.7. PAS salt bridge mutant % compositions

	WT	E315D	E315Q	E315R	R15E/ E315R
β-patchoulene	2.05	1.08	1.35	2.14	2.01
β-elemene	5.01	19.86	16.25	51.21	51.14
cycloseychellene	0.28	0.15	0.00	0.00	0.00
β-caryophyllene	4.49	4.33	1.97	1.38	1.87
α-guaiene	12.10	10.81	13.96	0.46	0.00
seychellene	3.79	1.29	0.00	0.00	0.00
humulene	0.31	0.77	0.00	0.00	0.00
α-patchoulene	4.13	0.95	0.00	0.00	0.00
γ-patchoulene	3.28	0.72	0.00	0.00	0.00
guai-4,11-diene	4.92	4.54	6.54	6.20	5.45
α-bulnesene	18.20	42.64	44.61	24.97	21.71
pogostol	1.18	0.79	0.22	0.00	0.00
patchoulol	27.72	5.45	1.60	0.00	2.74
other	7.54	6.62	13.50	13.63	15.07

Table 4.8. PAS salt bridge mutant standard deviations of % compositions

	WT	E315D	E315Q	E315R	R15E/ E315R
β-patchoulene	0.62	0.05	0.18	0.68	0.43
β-elemene	1.88	0.59	1.77	5.36	12.29
cycloseychellene	0.12	0.00	0.00	0.00	0.00
β-caryophyllene	1.47	0.18	0.54	0.19	0.59
α-guaiene	4.06	0.35	2.50	0.07	0.00
seychellene	1.42	0.06	0.00	0.00	0.00
humulene	0.10	0.03	0.00	0.00	0.00
α-patchoulene	1.34	0.07	0.00	0.00	0.00
γ-patchoulene	1.85	0.08	0.00	0.00	0.00
guai-4,11-diene	2.04	0.65	2.41	1.44	2.11
α-bulnesene	6.83	1.70	15.58	2.14	2.84
pogostol	0.33	0.18	0.39	0.00	0.00
patchoulol	8.29	0.36	2.13	0.00	0.67
other	2.83	2.64	7.09	7.89	9.14

Table 4.9. TEAS mutant % compositions

	WT	R15E	E312Q	E312D	P16R	R15E/ E312R	R15K	E312R
β-acoradiene	0.30	0.41	0.35	0.31	0.25	0.00	0.30	0.00
(-)-α-cedrene	0.64	0.65	0.55	0.46	0.51	0.00	0.64	3.73
2-epi-prezizaene	1.24	1.60	1.42	1.34	1.27	0.00	1.47	0.00
(+)-germacrene A	1.69	9.07	6.66	6.94	4.00	8.88	5.97	15.11
α-selinene	0.96	1.15	1.07	1.11	1.12	0.00	1.07	0.91
β-selinene	0.27	0.16	0.19	0.21	0.24	10.33	0.23	1.94
selina-4,11-diene	1.60	1.52	1.57	1.55	1.72	0.00	1.57	0.56
5-epi-aristolochene	82.63	74.37	77.26	77.34	79.55	55.19	78.08	58.59
(-)-4-epi-eremophilene	5.72	4.97	5.34	5.18	5.96	0.00	4.92	5.12
(-)-premnaspirodiene	1.95	3.40	2.84	2.84	2.43	25.59	2.71	5.78
spirolephichinene	1.39	1.41	1.41	1.38	1.39	0.00	1.50	8.25

Table 4.10. TEAS standard deviations for % compositions

	WT	R15E	E312Q	E312D	P16R	R15E/ E312R	R15K	E312R
β-acoradiene	0.03	0.02	0.01	0.06	0.06	0.00	0.10	0.00
(-)-α-cedrene	0.09	0.12	0.06	0.04	0.09	0.00	0.05	1.89
2-epi-prezizaene	0.02	0.08	0.04	0.01	0.03	0.00	0.04	0.00
(+)-germacrene A	0.13	0.20	0.34	0.27	0.20	2.02	0.21	1.26
α-selinene	0.08	0.16	0.17	0.12	0.08	0.00	0.12	0.26
β-selinene	0.01	0.10	0.06	0.08	0.01	3.18	0.12	0.75
selina-4,11-diene	0.01	0.02	0.01	0.02	0.03	0.00	0.03	0.51
5-epi-aristolochene	0.41	0.52	0.36	0.58	0.48	2.51	0.59	1.85
(-)-4-epi-eremophilene	0.01	0.02	0.04	0.02	0.07	0.00	0.02	0.63
(-)-premnaspirodiene	0.06	0.10	0.08	0.11	0.09	1.66	0.07	3.41
spirolephichinene	0.10	0.29	0.08	0.14	0.12	0.00	0.20	1.00

Table 4.11. ADS mutant % compositions

	WT	P16R
β-farnesene	1.49	1.37
amorpha-4,11-diene	84.16	85.12
amorpha-4,7(11)-diene	4.23	4.08
γ-humulene	1.78	2.15
β-sesquipellandrene	1.39	1.10
amorpha-4-en-11-ol	0.43	0.40
amorpha-4-en-7-ol	2.36	2.49
α-bisabolol	0.39	0.27
other	3.78	3.01

Table 4.12. ADS mutant standard deviations of % compositions

	WT	P16R
β-farnesene	0.14	0.13
amorpha-4,11-diene	0.55	0.59
amorpha-4,7(11)-diene	0.33	0.12
γ-humulene	0.02	0.05
β-sesquipellandrene	0.06	0.06
amorpha-4-en-11-ol	0.03	0.03
amorpha-4-en-7-ol	0.06	0.05
α-bisabolol	0.03	0.02
other	0.12	0.10

Table 4.13. HPS mutant % compositions

	WT	P16R
(+)-germacrene A	0.16	0.20
(-)-α-cedrene	0.09	0.09
spirolephichinene	1.48	1.51
isoprezizaene	0.12	0.12
selina-4,11-diene	0.78	0.70
(-)-4-epi-eremophilene	2.14	2.55
α-selinene	0.35	0.25
(-)-premnaspirodiene	93.71	93.40
β-selinene	0.15	0.16
other	1.02	1.02

Table 4.14. HPS mutant standard deviations of % compositions

	WT	P16R
(+)-germacrene A	0.03	0.03
(-)-a-cedrene	0.02	0.00
spirolephichinene	0.18	0.12
isoprezizaene	0.03	0.01
selina-4,11-diene	0.02	0.01
(-)-4-epi-eremophilene	0.24	0.13
a-selinene	0.13	0.02
(-)-premnaspirodiene	0.46	0.32
b-selinene	0.01	0.03
other	0.02	0.02

ACKNOWLEDGEMENTS

The text of chapter 4, in part, is currently being prepared for submission for publication of the material. Dellas, Nikki; Noel, Joseph P. I am the first author of this material. All experiments were performed under the supervision of Joseph P. Noel.

REFERENCES

1. Gershenzon, J.; Dudareva, N., The function of terpene natural products in the natural world. *Nature chemical biology* **2007,** *3* (7), 408-414.

2. Christianson, D. W., Structural biology and chemistry of the terpenoid cyclases. *Chemical reviews* **2006,** *106* (8), 3412-3442.

3. Wendt, K. U.; Schulz, G. E., Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure (London, England : 1993)* **1998,** *6* (2), 127-133.

4. Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H. J.; Coates, R. M.; Croteau, R. B., Abietadiene synthase from grand fir (Abies grandis): characterization and mechanism of action of the "pseudomature" recombinant enzyme. *Biochemistry (John Wiley & Sons)* **2000,** *39* (50), 15592-15602.

5. Kawaide, H.; Sassa, T.; Kamiya, Y., Functional analysis of the two interacting cyclase domains in ent-kaurene synthase from the fungus Phaeosphaeria sp. L487 and a comparison with cyclases from higher plants. *The Journal of biological chemistry* **2000,** *275* (4), 2276-2280.

6. Toyomasu, T.; Kawaide, H.; Ishizaki, A.; Shinoda, S.; Otsuka, M.; Mitsuhashi, W.; Sassa, T., Cloning of a full-length cDNA encoding ent-kaurene synthase from Gibberella fujikuroi: functional analysis of a bifunctional diterpene cyclase. *Bioscience, biotechnology, and biochemistry* **2000,** *64* (3), 660-664.

7. Trapp, S. C.; Croteau, R. B., Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **2001,** *158* (2), 811-832.

8. Whittington, D. A.; Wise, M. L.; Urbansky, M.; Coates, R. M.; Croteau, R. B.; Christianson, D. W., Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (24), 15375-15380.

9. Starks, C. M.; Back, K.; Chappell, J.; Noel, J. P., Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **1997,** *277* (5333), 1815-1820.

10. Williams, D. C.; McGarvey, D. J.; Katahira, E. J.; Croteau, R., Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* **1998,** *37* (35), 12213-12220.

11. Little, D. B.; Croteau, R. B., Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases delta-selinene synthase and gamma-humulene synthase. *Archives of Biochemistry and Biophysics* **2002,** *402* (1), 120-135.

12. Deguerry, F.; Pastore, L.; Wu, S.; Clark, A.; Chappell, J.; Schalk, M., The diverse sesquiterpene profile of patchouli, Pogostemon cablin, is correlated with a limited number of sesquiterpene synthases. *Archives of Biochemistry and Biophysics* **2006,** *454* (2), 123-136.

13. Noel, J. P.; Dellas, N.; Faraldos, J. A.; Zhao, M.; Hess, B. A., Jr.; Smentek, L.; Coates, R. M.; O'Maille, P. E., Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. *ACS chemical biology* **2010,** *5* (4), 377-392.

14. Prosser, I.; Phillips, A. L.; Gittings, S.; Lewis, M. J.; Hooper, A. M.; Pickett, J. A.; Beale, M. H., (+)-(10R)-Germacrene A synthase from goldenrod, Solidago canadensis; cDNA isolation, bacterial expression and functional analysis. *Phytochemistry* **2002,** *60* (7), 691-702.

15. O'Maille, P. E.; Chappell, J.; Noel, J. P., Biosynthetic potential of sesquiterpene synthases: Alternative products of tobacco 5-epi-aristolochene synthase. *Archives of Biochemistry and Biophysics* **2006,** *448* (1-2), 73-82.

16. Mercke, P.; Bengtsson, M.; Bouwmeester, H. J.; Posthumus, M. A.; Brodelius, P. E., Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in Artemisia annua L. *Archives of Biochemistry and Biophysics* **2000,** *381* (2), 173-180.

17. Hyatt, D. C.; Youn, B.; Zhao, Y.; Santhamma, B.; Coates, R. M.; Croteau, R. B.; Kang, C., Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (13), 5360-5365.

18. Peters, R. J.; Carter, O. A.; Zhang, Y.; Matthews, B. W.; Croteau, R. B., Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. *Biochemistry* **2003,** *42* (9), 2700-7.

19. O'Maille, P. E.; Chappell, J.; Noel, J. P., A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases. *Analytical Biochemistry* **2004,** *335* (2), 210-217.

Chapter 5

Mutation of Archaeal Isopentenyl Phosphate Kinase Highlights Mechanism and Guides Phosphorylation of Additional Isoprenoid Monophosphates

5.1. ABSTRACT

The biosynthesis of isopentenyl diphosphate (IPP) from either the mevalonate (MVA) or the 1-deoxy-d-xylulose 5-phosphate (DXP) pathway provides the key metabolite for primary and secondary isoprenoid biosynthesis. Isoprenoid metabolism plays crucial roles in membrane stability, steroid biosynthesis, vitamin production, protein localization, defense and communication, photoprotection, sugar transport, and glycoprotein biosynthesis. Recently, an alternative branch of the MVA pathway was discovered in the archaeon *Methanocaldococcus jannaschii* involving a small molecule kinase, isopentenyl phosphate kinase (IPK). IPK belongs to the amino acid kinase (AAK) superfamily. In vitro, IPK phosphorylates isopentenyl monophosphate (IP) in an ATP and Mg^{2+}-dependent reaction producing IPP. Here, we describe crystal structures of IPK from *M. jannaschii* refined to nominal resolutions of 2.0−2.8 Å. Notably, an active site histidine residue (His60) forms a hydrogen bond with the terminal phosphate of both substrate and product. This His residue serves as a marker for a subset of the AAK family that catalyzes phosphorylation of phosphate or phosphonate functional groups; the larger family includes carboxyl-directed kinases, which lack this active site residue. Using steady-state kinetic analysis of H60A, H60N, and H60Q mutants, the protonated form of the $N\varepsilon_2$ nitrogen of His60 was shown to be essential for catalysis, most likely through hydrogen bond stabilization of the transition state accompanying transphosphorylation. Moreover, the structures served as the starting point for the engineering of IPK mutants capable of the chemoenzymatic synthesis of longer chain isoprenoid diphosphates from monophosphate precursors.

5.2. INTRODUCTION

Isopentenyl diphosphate (IPP) and its isomeric partner dimethylallyl diphosphate (DMAPP) are precursors for a diverse collection of primary and secondary isoprenoid metabolites in all organisms. Following its formation, successive units of IPP are used together either with DMAPP, formed by the action of types I or II IPP isomerases, or with the IPP extended isoprenoid diphosphate chain, to biosynthesize C10, C15, or C20 oligoprenyl diphosphates known as geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), respectively, as well as larger isoprenoid diphosphates. In plants and some microorganisms, GPP, FPP, and GGPP also serve as starting materials for the biosynthesis of a large class of specialized and often cyclic terpene hydrocarbons[1]. FPP is the most ubiquitous of the three isoprenoid diphosphate building blocks, as it resides at the juncture of bifurcating branches of the general isoprenoid biosynthetic pathway leading to both primary and secondary metabolites. Squalene, hopanoids, and steroids, serve as critical components of cellular membranes and, in the case of steroids, also serve as transcription modulators through nuclear hormone receptor engagement[2, 3]. Moreover, dolichols play essential roles in N-glycosylation and membrane anchorage of sugars in eukaryotes and archaea[4]. The 20-carbon GGPP molecule functions as the precursor to all carotenoids, the latter of which provides photoprotection in plants, fungi, algae, bacteria, and some archaea[5, 6]. Interestingly, GGPP also is a precursor to the isoprenoid-derived hydrocarbon moiety of lipids that is present exclusively in archaea (see Koga *et al* (2007) for a review on archaeal lipids[7]).

Over the last two decades, two distinct pathways have been characterized that biosynthesize IPP and DMAPP, namely the mevalonate (MVA) pathway and the more recently discovered 1-deoxy-d-xylulose 5-phosphate (DXP) pathway[8]. The MVA pathway is utilized by animals, plants (cytosol), fungi, and certain bacteria, while the DXP pathway resides in plant plastids, a

number of eubacteria, cyanobacteria, and certain parasitic organisms[9]. In archaea, orthologs for almost all of the genes encoding the MVA pathway are present except for two; interestingly, the last two genes encoding phosphomevalonate kinase and diphosphomevalonate decarboxylase appear to be missing from the genomes of almost all archaea. For this reason, the isoprenoid pathway in archaea is referred to as "The Lost Pathway"[10]. In 2006, Grochowski et al. discovered an enzyme and its associated gene in the archaeon *Methanocaldococcus jannaschii* that belongs to the larger family of amino acid kinases (AAK) but catalyzes the ATP-dependent phosphorylation of IP, thereby producing IPP[11]. This enzyme, named isopentenyl phosphate kinase (IPK), appears to be a starting point for the functional reconstruction of The Lost Pathway, representative of a completely unexpected biosynthetic variation of the MVA pathway.

IPK shares significant sequence homology with proteins in the AAK superfamily (Pf000696, Figure 5.1). Members of this family employ Mg^{2+}–ATP to catalyze phosphorylation of carboxylate, carbamate, phosphonate, or phosphate functional groups. Here, crystal structures of IPK from *M. jannaschii* are presented in the 'apo' form and in complex with substrate (IP) and product (IPP). These structures allow for rational mutagenesis and biochemical analyses of residue(s) near the phosphate moiety and the isopentenyl tail of IP, respectively. Mutation of a residue near the phosphate of IP demonstrates a key role for His-directed hydrogen bonding in the phosphorylation of phosphate or phosphonate groups. Mutation of residues near the isoprenyl moiety of IP establishes IPK as a starting point for engineering the phosphorylation of alternative phosphate/phosphonate bearing small molecules, including geranyl monophosphate (GP) and farnesyl monophosphate (FP). This sets the stage for a multitude of applications in chemoenzymatic syntheses including diphosphate analog synthesis, low cost radio-labeling of isoprenoid diphosphates with ^{32}P or

^{35}S containing β-phosphates, and the possible in vivo recycling of isoprenoid monophosphates formed upon FPP up-regulation and degradation in heterologous hosts.

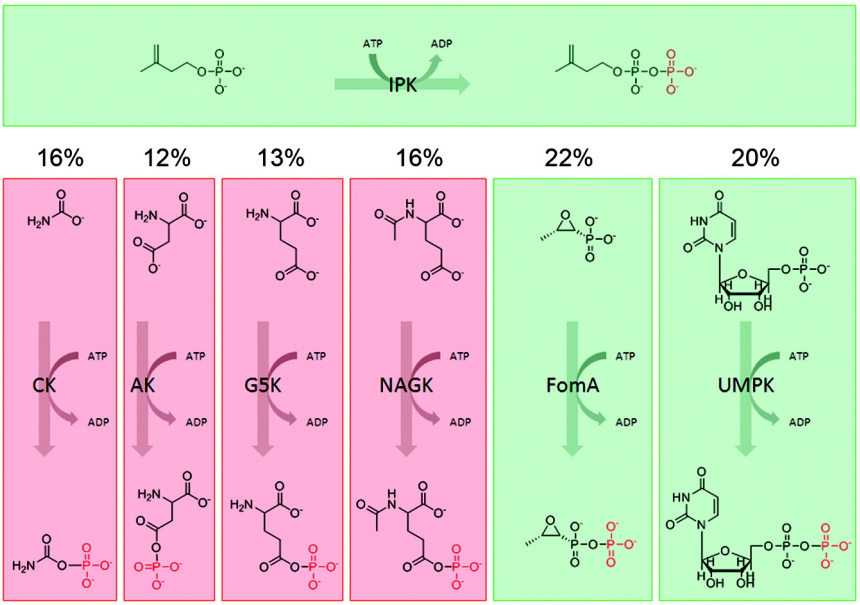


Figure 5.1. The amino acid kinase (AAK) family members. Isopentenyl phosphate kinase (IPK) reaction depicted across the top. Representative family members displayed from left to right: carbamate kinase (CK), aspartokinase (AK), glutamate-5-kinase (G5K), N-acetyl-l-glutamate kinase (NAGK), fosfomycin resistance kinase (FomA), and uridine monophosphate kinase (UMPK). The percent sequence identities relative to IPK are listed above each enzyme. Reactions shaded green utilize a phosphate or phosphonate phosphoryl acceptor, while the reactions shaded red utilize carbamate or carboxylate groups as phosphate acceptors.

5.3. RESULTS AND DISCUSSION

5.3.1. Three-dimensional architecture

IPK represents the newest member of the AAK superfamily to be structurally determined. The overall fold, commonly referred to as the open αβα sandwich, was first discovered in carbamate kinase from *E. faecalis*[12]. IPK is architecturally most similar to fosfomycin resistance kinase (FomA) from *S. wedmorensis* with a root-mean-square deviation (rmsd) of 2.0 Å for superimposed backbone atoms (NH−Cα−C) and a sequence identity of 22%. However, it shares the highest sequence identity, 25%, with uridine monophosphate

kinase (UMPK) from *A. fulgidus*. Two subdivisions of the AAK superfamily exist, referred to here as the phosphate and carboxylate subdivisions, respectively. Enzymes in the phosphate subdivision, including IPK, FomA, and UMPK, catalyze phosphorylation of a phosphate or phosphonate moiety. Enzymes in the carboxylate subdivision, including carbamate kinase, N-acetylglutamate kinase (NAGK), aspartokinase, and glutamate-5-kinase, catalyze phosphorylation of a carbamate or carboxylate group (Figure 5.1).

Like all other AAKs, IPK adopts a dimeric quaternary structure, and each monomer folds into structurally distinct N- and C-terminal domains (Figure 5.2). The N-terminal domain, spanning residues 1−171, binds the nucleophilic phosphate group (IP in IPK). The C-terminal domain, spanning residues 171−260, coordinates a Mg^{2+} ion and binds the phosphate donor ATP. Although all attempts to crystallize *M. jannaschii* IPK with ATP, ADP, AMPPNP, and a variety of other analogs have thus far been unsuccessful, the location of the nucleotide-binding site is structurally conserved among all family members, affording a reasonable model for ATP binding. Each monomer of IPK contains 16 β-strands, 8 α-helices, and 1 3_{10} helix. The core of the open αβα sandwich, represented by 8 β-strands, β14, β16, β15, β11, β1, β2, β8, and β5, resides between 4 α-helices on one side, αF, αH, αE, and αD, and 3 on the other, αG, αA, and αC (Figure 5.2). Four β-hairpins, one α-helix, and one 3_{10} helix (η1) decorate the periphery of the central β-sandwich. Three of the hairpins, β3−β4, β6−β7, and β9−β10, lie in the N-terminal domain and surround the back wall and one side of the IP binding pocket; the αB helix covers the remaining side of the isopentenyl-binding surface. The fourth β-hairpin, β12−β13, located within the C-terminal domain, resides in close proximity to the expected location for the adenine ring of ATP. Finally, the 3_{10} helix links one end of the central β5 strand and the β6−β7 hairpin. Although the β1−αA junction is depicted as a loop in Figure 5.2, it also adopts a helical structure in some cases.

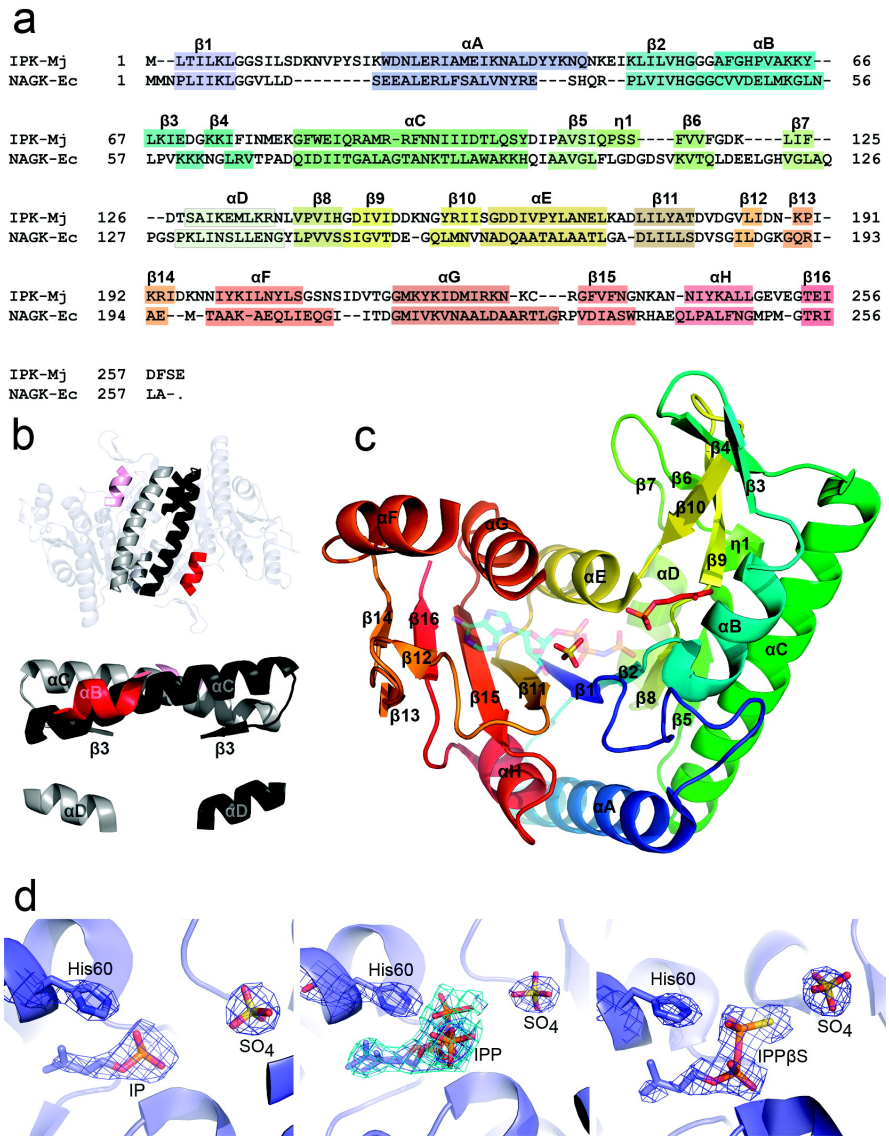


Figure 5.2. Primary sequence, tertiary architecture, and active site snapshots of IPK. a) Primary sequence of IPK from *M. jannaschii* aligned with *E. coli* NAGK. The color coding of each motif correlates with its color shown on the three-dimensional model. b) Global view of the IPK dimer (top) and a close-up view of the dimerization interface (bottom). Motifs positioned near the dimerization interface are gray (or pink) for one monomer and black (or red) for the other. c) Ribbon diagram of the IPK monomer. The structure is colored using a blue to red gradient from the N- to C-terminus. The C-terminal ATP-binding domain contains a β-sulfate residing in a location coinciding with the β-phosphate of ATP. The ATP analog AMPPNP is faintly colored and blended into the background (modeled from PDBID: 1gs5) and serves as a reference for the putative location of ATP in IPK. The crystallographically observed isopentenyl phosphate (IP) substrate is shown bound within the N-terminal domain. d) The active sites of IPK complexed with IP (left), IPP (middle), and IPPβS (right). Electron density surrounding each ligand (dark and light blue are contoured to 1σ and 0.6σ, respectively) shown as 2Fo−Fc omit electron density maps, where the ligands were removed before a round of refinement and subsequent phase and map calculations.

The IPK crystalline dimer is consistent with its oligomeric state deduced by gel filtration chromatography. The dimer orients around a noncrystallographic two-fold axis. This dyadic axis sits perpendicular to the extended β-sheet spanning the length of the dimer (16 β-strands with 8 β-strands per monomer). Although every AAK family member utilizes a similar dimerization interface, each dimer is unique in that its monomers orient differently with respect to one another[13]. The IPK dimer closely resembles that of UMPK, with the αC helices from each monomer crossing one another at an angle of 190°[13] (Figure 5.2, panel b). In IPK, this interface is comprised of 8 charged hydrogen bonds and 29 noncharged hydrogen bonds with 14 residues participating in van der Waals interactions[14]. The majority of the hydrogen-bonding interactions stitch together three structural motifs: (i) the αC helices of each monomer; (ii) the αD helix of one monomer and the β9−β10 hairpin of the dyad related monomer; and (iii) the 310 helix of one monomer and β5 of its dyadic partner. Hydrophobic interactions between the two monomers include residues from the αC and αD helices, the 3_{10} helix, and the β4, β5, β6, β8, β9, and β10 strands. These residues form an intimate hydrophobic interface further cementing the monomers together and burying a considerable amount of accessible surface area (1869 $Å^2$ of buried surface area per monomer).

5.3.2. Active site architecture

The refined 'apo' structure contains two active site sulfate molecules bound per monomer. One sulfate superimposes onto the position of the monophosphate of IP in the IP-bound structure. The second sulfate is present in all structures determined thus far and lies in the approximate location of the β-phosphate of ATP observed in other structures of AAK family members (15-17). This second sulfate ion is in close proximity to Gly9, Lys6, Lys221,

and Thr179 (Figure 5.3). The equivalent residues in other AAK family members stabilize the β-phosphate of ADP or ATP analogues (PDBIDs: 2hmf[15], 1ohb[16], 2j0w[17], 2bri[18], 3c1m[19], 3d41[20], 1gs5[21]). Therefore, the sulfate ion appears to serve as a reasonable spatial mimic of the β-phosphate of ATP and is referred to as the β-sulfate ion.

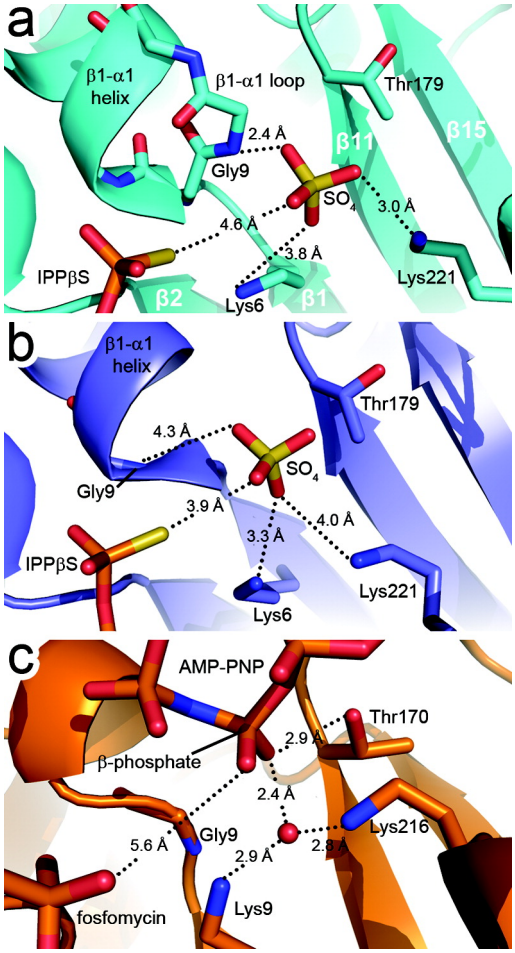


Figure 5.3. Comparative close-up views of the nucleotide phosphate-binding region of the IPK and FomA active sites. a) Monomer A of the IPK−IPPβS complex depicting the β-sulfate ion and the surrounding residues. b) Monomer B of the IPK−IPPβS complex oriented as in panel a. c) FomA complexed with the ATP analog AMPPNP and fosfomycin (PDB ID: 3d41)[20]. As depicted here, the β-sulfate ion in both IPK monomers shares a similar position and interacts with the same residues as does the β-phosphate group on AMPPNP in FomA.

The structures of IPK in complex with IP and IPP define the secondary structural elements comprising the IP-binding pocket and include the β2−αB glycine-rich loop, the αB helix, the β3−β4 hairpin, the β4−αC loop, the N-terminal section of the αC helix, and the β9−β10 hairpin (Figure 5.4, panel a). The β2−αB loop is one of two conserved glycine-rich loops present throughout the AAK family. It is thought to contribute to charge neutralization in the transition state during phosphoryl transfer[16, 20]. Notably, the orientation of the αB helix is conserved only in the phosphate division of the AAK superfamily (including IPK, FomA, and UMPK). In FomA, the αB helix orders when substrate is present but is otherwise disordered[20]. In contrast, the αB helix in IPK is ordered in both 'apo'- and IP-bound structures. Of even more limited familial distribution, the β3−β4 hairpin is present only in IPK and NAGK. In NAG-bound NAGK structures, the hairpin often exists in a closed conformation[21, 22]; in contrast, all structures of IPK reveal the motif in an open conformation. Regardless, the hairpin may play a role in shielding the substrate-binding pocket from the surrounding solvent in both enzymes.

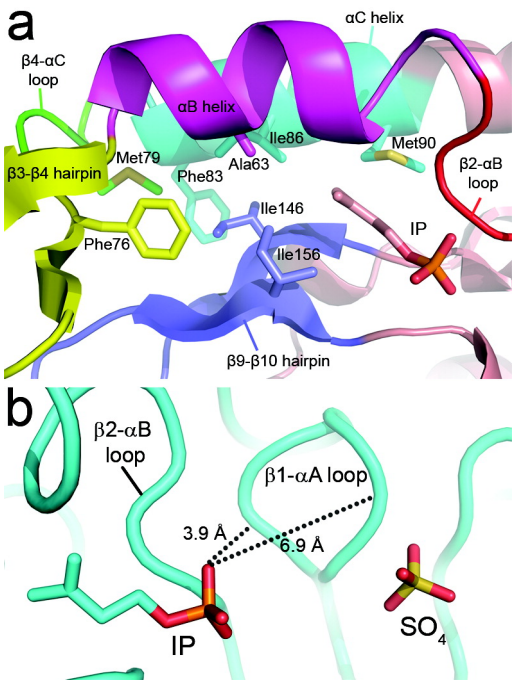


Figure 5.4. N-terminal domain and dual loop conformations in IPK. a) Close-up view of the N-terminal domain depicting the isopentenyl tail and the surrounding hydrophobic residues. The motifs surrounding the active site are colored as follows: β2−αB glycine-rich loop (red), αB helix (magenta), β3−β4 hairpin (yellow), β4−αC loop (green), N-terminal portion of the αC helix (cyan), and the β9−β10 hairpin (blue). Residues within van der Waals contact of the isopentenyl chain include Ile86, Met90, and Ile156. b) Dual conformation of the β1−αA loop in monomer A of the IPK−IP complex. One conformation places the loop close to the β2−αB loop and the IP substrate, while the other conformation places the loop in close proximity to the β-sulfate ion.

The branched C5 tail of the substrate resides in a pocket surrounded principally by hydrophobic residues, including Ala63, Phe76, Met79, Phe83, Ile86, Met90, Ile146, and Ile156, (Figure 5.4, panel a). The arrangement of residues within the cavity suggests that transmutation of the isopentenyl binding pocket to accommodate longer hydrocarbon chains may be relatively facile. The phosphate moiety of IP occupies the active site region between three short motifs including His60 and the β2−αB (residues 54−56) and β10−αE (residues 157−159) loops (Figure 5.5, panels a−c). In both monomers, the $N\varepsilon_2$ atom of His60 and the N atoms of Gly55 and Gly159 when protonated could stabilize the three nonbridging O atoms on the phosphate of IP through hydrogen bonds (Figure 5.5, panels b−c). In monomer B, Lys6

also interacts with one of these nonbridging O-atoms indirectly through hydrogen-bonding interactions with an intervening water molecule; in monomer A, there is no water molecule to facilitate this interaction.

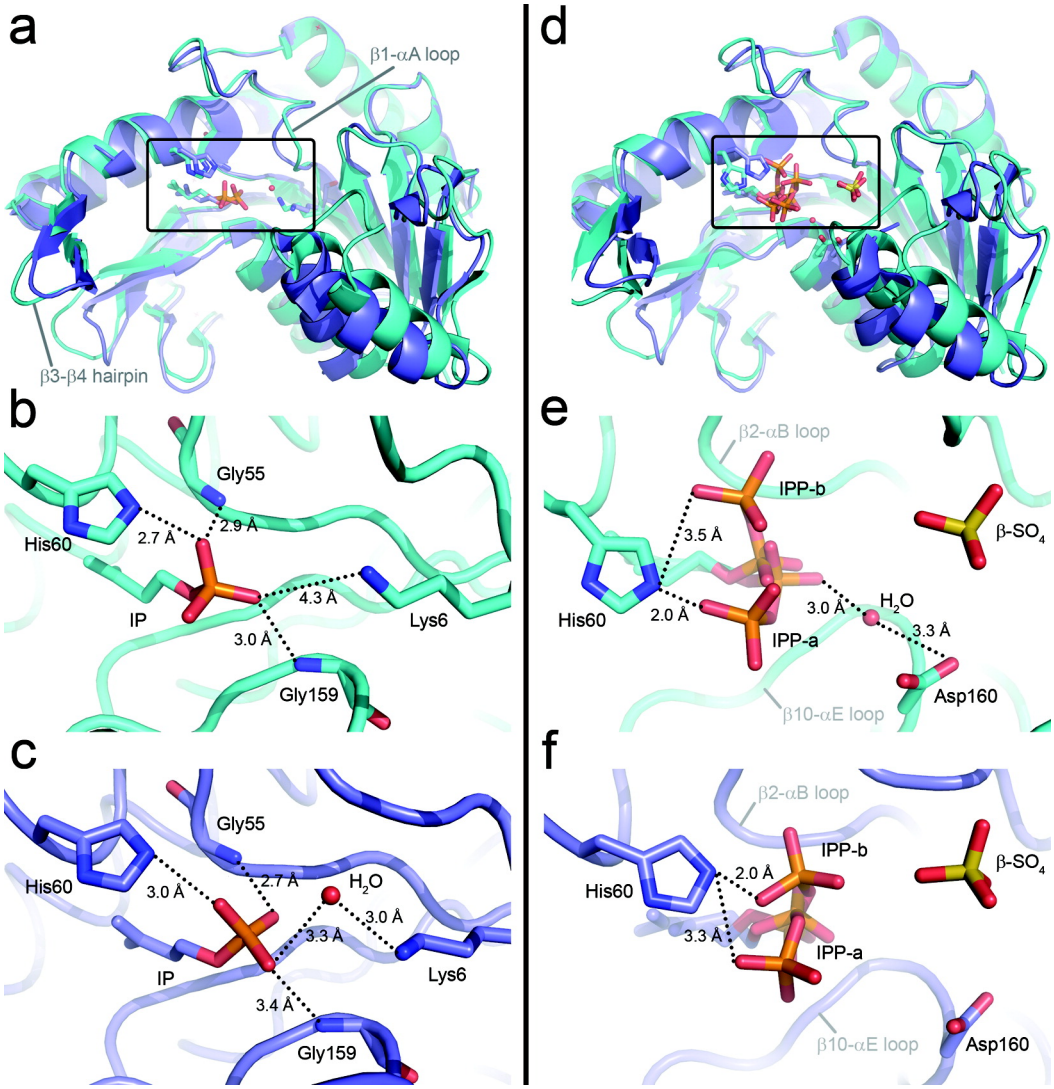


Figure 5.5. IPK in complex with IP and IPP. a) Tertiary structure superposition of monomers A and B of the IPK−IP complex. The rmsd between the two monomers is 1.31 Å. b−c) Close-up views of residues proximal to and hydrogen bonding with the α-phosphate of IP in monomers A (panel b) and B (panel c). In monomer B, a water molecule bridges the side-chain amino group of Lys6 and a nonbridging oxygen atom of the IP phosphate. d) Tertiary structure superposition of monomers A and B of the IPK−IPP complex. The rmsd between the two monomers is 1.39 Å. e−f) Views of the multiple conformers of IPP (labeled as IPP-a and IPP-b) in both monomers A (panel e) and B (panel f).

In monomer A, a loop at the β1−αA junction (Gly8−Leu12), residing near the active site, can adopt two distinct substrate-binding conformations based upon refinement of alternative conformations and the observed electron density. In one conformation, the loop lies near the active site β-sulfate ion, while in the other, the loop lies closer to the β2−αB loop (Figure 5.4, panel b). None of the residues in this loop participate in hydrogen-bonding interactions with IP; however, the dual binding conformations are not observed for the 'apo' structure, suggesting that loop movement is partially dependent on the presence of substrate. In monomer B, the loop adopts one binding conformation roughly equidistant between the two modes present in monomer A.

5.3.3. Multiple conformations of IPP in a single active site

The crystal structure of IPK with its product bound reveals that IPP adopts two distinct conformers designated conformers A and B (Figure 5.5, panels d−f). These conformers are similar except for the orientation of the β-phosphate group and the adjacent bridging O atom. In conformer A, these two moieties sit closer to the β10−αE loop, while in conformer B, they reside closer to the β2−αB loop (Figure 5.5, panels e−f). In both conformers, a nonbridging O atom of the β-phosphate group hydrogen bonds to the protonated $N\varepsilon_2$ atom of His60. A superposition of the two monomers (Figure 5.5, panel d), reveals that His60 sits in a different location in each monomer, which may reflect conformations of this residue that are dynamically accessible as the phosphorylation reaction proceeds.

In monomer A only, a water molecule rests between a nonbridging O atom from the α-phosphate of IPP and the carboxylate moiety of Asp160 (Figure 5.5, panel e). This water molecule is also found in substrate-bound structures of FomA kinase (PDBID 3d41)[20], E. coli NAGK (pdb ID 1gs5)[21], P. furiosus UMPK (pdb ID 2bmu)[18], and E. coli UMPK (pdb ID

2bne)[23], and it is stabilized in a similar fashion in each of these structures. Asp160 of IPK is highly conserved among the AAK family and has been suggested to function as an active site base and a central organizing residue[20, 24].

As discussed previously, the β1−αA loop occupies two conformations in monomer A of the IPK−IP complex: one that places it in close proximity to the β2−αB loop, and another that interacts with the β-sulfate ion (Figure 5.4, panel b). In monomer A of the IPK−IPP complex, the latter conformation is the major binding mode observed. The former conformation can also be seen in the electron density, although this binding mode is so subtle that it did not refine well and was not built into the final structure. This minor binding mode may, however, hold some significance, as it places Gly9 of the β1−αA loop in close proximity to the β-phosphate group in conformer B of IPP. The β1−αA loop is often reported to interact with the β- and γ-phosphate groups from ATP analogs; however, there are also examples of this loop interacting with the β-phosphate of the product (in UMPK from E. coli)[16, 23].

The catalytically relevant conformer for IPP is most likely conformer B. This hypothesis is supported by three pieces of information: (i) The β1−αA loop, which is thought to play a key role during phosphoryl transfer, accesses a minor binding mode that is in close proximity to the β-phosphate of conformer B[16]; (ii) A superposition of UDP-bound UMPK from E. coli and IPP-bound IPK demonstrates that the phosphate moiety of UDP superimposes with conformer B of IPP[23]; and (iii) The ATPγS/IP/Mg2+ complex structure (discussed below) exhibits clear electron density for a IPPβS molecule bound in a single conformation that superimposes with conformer B of IPP from the IPP-bound structure. Conformer A of IPP may, therefore, represent a post-reaction enzyme−product (EP) complex.

5.3.4. Product-bound active site containing IPPβS

When a crystal of IPK was soaked in a stabilizing solution containing IP, Mg^{2+}, and ATPγS, the product IPPβS was observed bound in the active site. This product resembles IPP except one of the nonbridging O atoms on the β-phosphate is clearly replaced with an S atom, as evidenced by the additional electron density associated with the β-thiophosphate. Notably, no electron density for the second product ADP is seen. This is the only structure determined where both substrates, IP and ATPγS, were soaked into the crystal resulting in a catalyzed reaction in the crystal lattice. Interestingly, this structure reveals only one binding mode for IPPβS consistent with the orientation of conformer B in the IPP-bound structure. In both monomers, Gly159 from the β10−αE loop stabilizes the α-phosphate group of IPPβS; the β-thiophosphate group remains in close proximity to His60. However, in monomer B (compared to monomer A), the substrate migrates to a position that shifts away from this residue and resides closer to the β-sulfate ion (Figure 5.3). The intermediate location of the β-phosphate group in monomer B coupled with the inferred heightened dynamics of certain loops within this monomer suggests that the monomer B structure depicts an earlier phase in the transphosphorylation reaction compared to monomer A.

5.3.5. His60 plays a key role in binding and catalysis

From the results discussed above, it is evident that His60 plays an important role in both substrate and product sequestration. This binding role is accomplished through a hydrogen-bonding interaction between the protonated $N\varepsilon_2$ atom of His60 and a nonbridging O atom from the terminal phosphate group on either the substrate (IP) or the product (IPP). His60 was mutated to Asn, Gln, and Ala; the Asn and Gln mutations are isosteric with the protonated $N\varepsilon_1$ and $N\varepsilon_2$ groups on His, respectively. The three mutants were assayed at 25 °C

using the pyruvate kinase/lactate dehydrogenase coupled reaction to detect kinase activity. Turnover for the H60A and H60N mutants was not detected. In contrast, the H60Q mutant, whose $-NH_2$ side-chain moiety mimics the protonated $N\varepsilon_2$ nitrogen of His60, catalyzed a measurable transphosphorylation of IP. Notably, the apparent $K_{m,IP}$ values for H60Q and wild-type IPK are 34.5 and 4.3 μM, respectively, while the apparent k_{cat} values are 0.04 s^{-1} and 1.46 s^{-1}, respectively. These experimental values yield an apparent catalytic efficiency, k_{cat}/K_m, 340 times lower for the H60Q IPK mutant (Table 5.1).

Table 5.1. Kinetic Data for IPK-Mj Wild-Type and H60Q at 25°C

Protein Name	$K_{m,\ ATP}$	$K_{m,\ IP}$ (μM)	k_{cat} (s^{-1})	$k_{cat}/K_{m,IP}$ ($s^{-1}\mu M^{-1}$)
IPK-*Mjannaschii*	198.2 ± 32.7	4.30 ± 0.58	1.46 ± 0.03	0.34
IPK-*Mjannaschii* H60Q	559.3±116.9	34.5 ± 7.2	0.040 ± 0.002	0.001

These steady-state kinetic results suggest several interpretations for the unique role of His60 in catalysis: (i) Since both H60A and H60N exhibit no measurable activity, while H60Q remains active (albeit at a fraction of wild-type activity), binding and/or catalysis appears to be dependent on the presence of a hydrogen bond donor that is isosteric with the protonated $N\varepsilon_2$ nitrogen of His60; 2) Given that the H60Q mutant possesses a higher apparent K_m than wild-type, His60 also appears to be important for ground-state binding. Additional flexibility in the Gln side chain relative to the imidazole group of His60 may hinder its ability to bind substrate as effectively as wild-type IPK; 3) The k_{cat}/K_m value is more than 300 times higher for wild-type compared to that of the H60Q mutant, which again suggests that His60, through its added charge and lowered conformational flexibility, plays a role in stabilization of the more negatively charged transition state accompanying phosphoryl transfer.

Through comparison of the solved IPK structures, it is evident that His60 shifts from stabilizing the α-phosphate on the substrate IP to stabilizing the β-phosphate on the product IPP. In FomA, His58 (the equivalent residue to His60 of IPK) indirectly stabilizes the substrate through an intervening water molecule that is within hydrogen-bonding distance of both His58 and fosfomycin[20]. In UMPKs, an arginine residue that aligns with His60 appears to serve two roles: (i) in bacterial UMPKs, this Arg interacts with GTP, which is an allosteric activator for all bacterial UMPKs[25]; and (ii) in both bacterial and archaeal UMPKs, this Arg stabilizes the phosphate intermediate throughout the course of the reaction[23, 26]. In comparisons of crystal structures of E. coli UMPK, Arg62 points in opposite directions to fulfill these two roles[25], indicating that the length and the conformational freedom of this residue are important for its functional versatility. It is therefore not surprising that a mutation to histidine at this equivalent position results in loss of GTP activation, thereby reducing the enzyme's catalytic activity[27].

The phosphate division of the AAK family encompasses the three enzymes discussed above, which are the only currently known family members that contain a residue aligning with His60 of IPK; the four family members in the carboxylate division do not possess a residue or a motif that structurally aligns with this region of IPK. Regardless of other roles that this residue may play, His60 in IPK, His58 in FomA, and the aligning arginine in all UMPKs most likely play a role in substrate/product binding and/or transition-state stabilization.

5.3.6. IPK mutants can phosphorylate oligoprenyl monophosphates

As mentioned previously, the tail of the IP substrate is sequestered within a hydrophobic binding pocket (Figure 5.4, panel a). At the back of the pocket, several residues,

Ile86, Ile146, and Ile156, can be mutated to smaller amino acids to accommodate the binding of ligands with extended carbon chains (such as GP and FP). With this idea in mind, several point mutants were generated including I86A, I86G, I146A, I146G, and I156A, and most of these were able to convert FP to FPP while the wild-type enzyme lacked an equivalent activity (Figure 5.6). These observations support the idea that mutation to smaller residues widens the cavity to allow for the binding of an extended isoprenoid tail, while mutation to bulkier residues hinders this ability. It was found that several double and triple mutant combinations displayed improved FP to FPP conversion by an order of magnitude compared to the single mutants, providing evidence that a deeper or larger cavity is more effective for FP binding and catalysis (Figure 5.6). It is also reassuring that these mutations are contextually dependent, meaning that mutations at the very back of the active site (at position 83) are not effective unless they are present with mutations closer to the front of the active site (at positions 86, 146, or 156).

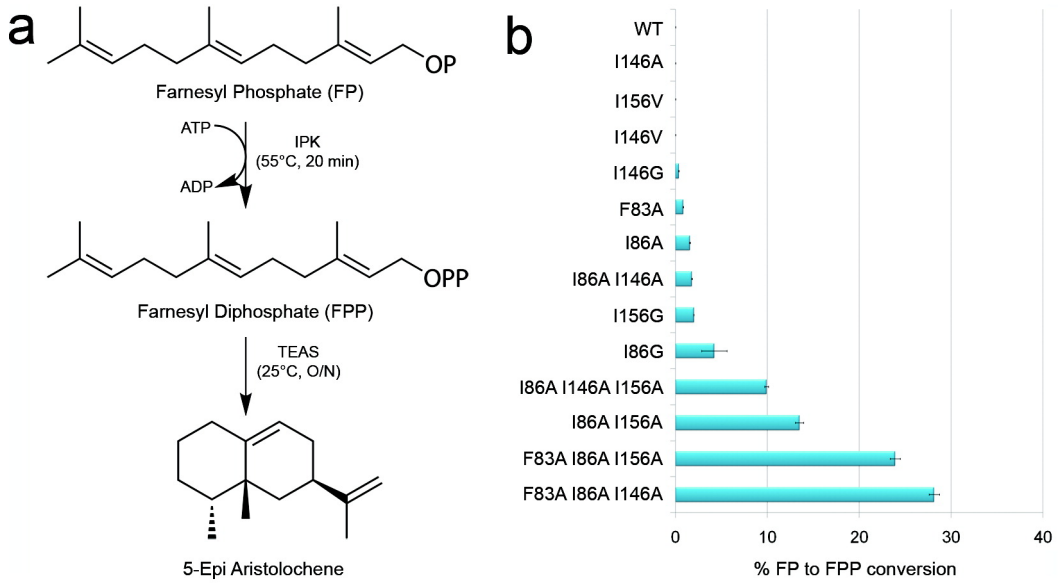


Figure 5.6. Farnesyl phosphate (FP) phosphorylation by IPK chain length mutants. a) The coupled IPK−sesquiterpene synthase reaction used to test for FP transphosphorylation. b) Comparative bar graph depicting several IPK tunnel mutants qualitatively tested for their ability to convert FP to FPP (expressed as a percentage of maximal production of 5-epi-aristolocene produced from IPK-generated FPP using wild-type IPK and identical concentrations of wild-type tobacco 5-epi-aristolochene synthase incubated for equivalent lengths of time).

Modeling of an FP molecule within the active site of IPK suggests that a C15 isoprenyl tail can orient in several different directions without introducing a large number of steric clashes. These various orientations could be explored by mutating the appropriate amino acid side chains (to relieve putative steric clashes) and by utilizing a high-throughput coupled assay to test each mutant for its ability to convert FP to FPP. The coupled kinase/terpene synthase assay (Figure 5.6) is ideal for rapid and qualitative analysis, which is necessary here since a large number of mutants must be screened to explore all possible FP orientations and mutant combinations. Thus far, we have obtained active mutants that were designed to accommodate one putative FP orientation. Crystal structures of FP-bound IPK mutants will assist in the future design of more robust mutants. Finally, exploration of the above mutations

in the context of a mesophilic ortholog should afford improvement in kinetic activity for new phosphate-bearing substrates; a more accurate picture of dynamic behavior can be painted for a mesophilic IPK being studied at ambient temperature rather than thermophilic IPK (used here) being studied at ambient temperature.

5.3.7. Conclusions

Isopentenyl phosphate kinase (IPK) from *M. jannaschii* is the newest member of the large amino acid kinase (AAK) family to be structurally characterized. The phosphate division of the AAK family is comprised of three proteins [IPK, fosfomycin resistance kinase (FomA), and uridine monophosphate kinase (UMPK)] that exclusively align with one another along their αB helices. More importantly, they all contain a superimposable residue at position 60 (in IPK) that indirectly or directly stabilizes the terminal phosphate group of the substrate or product. Using the His60 marker, we have been able to identify putative IPK homologues from a number of phylogenetically diverse eukarya. Work to be presented elsewhere is focused on in vitro and in vivo analyses of these latter proteins, given that, in most cases, the organisms in question also contain the full complement of predicted mevalonate (MVA) pathway enzymes.

Finally, we have shown that our initial goal to rationally engineer IPK to accept longer chain substrates is possible. These mutants along with an expanded mutational analysis in IPKs from mesophilic hosts can serve a variety of applications. For example, they can be used to recycle isoprenyl monophosphates, which are thought to be one possible in vivo byproduct of farnesyl diphosphate (FPP) degradation (through the action of an alkaline phosphatase)[28, 29]. This recycling mechanism would be useful in an in vivo system designed to overproduce isoprenoids (such as terpenes) by means of the MVA pathway in a fungal or bacterial host.

The IPK chain-length mutants would also be useful in the chemo-enzymatic synthesis of radio-labeled geranyl diphosphate (GPP) or FPP as well as a variety of other analogs including fluorescently tagged isoprenyl tails[30, 31]. We have demonstrated that IPK can be rationally engineered to accept and phosphorylate oligoprenyl monophosphate substrates, such as FP. Future work with IPK from *M. jannaschii* (and with orthologous IPKs) will focus on redesigning the enzyme(s) such that they can bind and more efficiently turn over a variety of bulky GP and FP analogs.

5.4. METHODS

5.4.1. Activity assays and steady-state kinetic analyses

All specific activity and kinetic measurements were performed using the pyruvate kinase−lactate dehydrogenase coupled assay[32]. The reaction in 200 μL includes 7 U of pyruvate kinase, 10 U of lactate dehydrogenase, 2 mM of phosphoenolpyruvate, 0.16 mM of NADH, 50 mM of Tris−HCl, pH of 8.0, 100 mM of KCl, 8 mM of $MgCl_2$, and varying concentrations of ATP or IP (purchased from Larodan Fine Chemicals and Isoprenoids, LLC). When IP was varied, the concentration of ATP was fixed at 4 mM. When ATP was varied, the concentration of IP was fixed at 100 μM for wild-type IPK and 500 μM for the H60Q mutant. The reaction was initiated by the addition of IPK (0.15 μg mL^{-1} final concentration) and followed by observing the depletion of NADH at 340 nm, expressed as Δ(AU340)/Δt and converted to Δ(ADP)/Δt. These values were plotted against substrate concentration in GraphPad Prism (Version 5.01 for Windows) to compute the kinetic parameters k_{cat} and K_m, using the "nonlinear regression enzyme kinetic analysis" option.

5.4.2. Kinase/terpene synthase coupled Assay for chain-length mutants

The coupled assay consists of two steps: the kinase reaction followed by the terpene synthase reaction (Figure 5.6). The 50 μL of kinase reaction includes 4 mM of ATP, 8 mM of $MgCl_2$, 1 mM of FP (or GP), and 50 mM of Tris−HCl, pH of 8.0. The reaction was initiated by the injection of IPK (1 μM of the final concentration), incubated at 55 °C for 20 min, and then cooled on ice for 10 min. The subsequent 500 μL of terpene cyclase reaction includes 10 μL of the aforementioned IPK reaction, 8 mM of $MgCl_2$, 30 μg of tobacco 5-epi aristolochene synthase (TEAS), and 50 mM of Tris−HCl, pH 8.0. Each sample was overlaid with an equal volume of ethyl acetate, incubated at RT overnight, and vortexed for terpene extraction and subsequent injection onto the GC-MS (Hewlett-Packard, 6890/5973 system) equipped with a HP-5MS column (0.25 mm × 30 m × 0.25 μm). The method employed is similar to that reported in O'Maille et al.[33] with an injection temperature reduced from 250 to 200 °C.

5.4.3. Structure solution and refinement

Data were processed and scaled with XDS[34]. The reduced single anomalous diffraction (SAD) data from the IPK 'apo' crystal treated with ethyl mercuric phosphate was used in SOLVE[35] to locate and refine the positions of two Hg atoms per asymmetric unit, followed by phasing (mean figure of merit: 0.33). The program RESOLVE[35] was then employed to build 424 out of 520 residues into the SAD-derived model. Model building and phase improvement were accomplished using ARP/wARP[36, 37]. The refined model was used as the starting model for the structure determination of IPK in complex with IP, IPP, and IPPβS. Simulated annealing and rigid-body refinements were performed in CNS version 1.2 for each structure before additional rounds of refinement in CNS version 1.2[38] and CCP4[36]. COOT was used for map-model visualization and manual model building[39]. Areaimol was used for a

calculation of the buried surface area per monomer[36]. PROCHECK was used to assess the quality of all models[40]. The data and refinement statistics are listed in Table 5.2.

Table 5.2. X-ray Diffraction Data Processing and Refinement Statistics

	IPK apo	**IPK-IP complex**	**IPK-IPP complex**	**IPK-IPPβS complex**
PDB ID	3K4O	3K52	3K4Y	3K56
Ligand (crystal drop)	none	IP + $MgCl_2$	IPP + $MgCl_2$	IP + ATPγS + $MgCl_2$
Ligand (observed)	none	IP	IPP	IPPβS
Data Collection and Processing				
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Resolution (Å)	2.0	2.7	2.55	2.35
Cell dimensions				
a (Å)	76.05	77.86	78.09	77.68
b (Å)	99.61	100.80	99.23	100.24
c (Å)	87.60	87.32	87.41	87.79
α = β = γ (°)	90	90	90	90
Molecules in asymmetric unit	2	2	2	2
No. measured reflections[a]	280086 (41535)	109680 (15190)	128795 (17570)	195455 (25991)
No. unique reflections[a]	43888 (6469)	18614 (2598)	21910 (3104)	27728 (3786)
Redundancy	6.38 (6.42)	5.89 (5.85)	5.87 (5.66)	7.05 (6.87)
R_{merge} (%)[a,b]	6.9 (32.9)	8.4 (38.5)	7.9 (33.99)	7.5 (39.8)
Completeness (%)[a]	95.6 (88.6)	95.5 (84.5)	95.5 (85.2)	93.9 (80.9)
I/σ(I)[a]	16.23 (4.58)	16.79 (3.70)	16.52 (4.34)	15.54 (3.84)

[a] values in parentheses represent data from the highest resolution shell

[b] Rmerge = Σhkl Σi|Ii(hkl) - <I(hkl)>|/Σhkl ΣiIi(hkl)

[c] Rfactor = Σ||Fobs|-|Fcalc||/Σ|Fobs|. Rwork is the Rfactor calculated using all diffraction data included in the refinement. Rfree is the Rfactor calculated using the randomly chosen 5% of diffraction data that were not included in the refinement.

[d] rmsd = root mean square deviation

Table 5.2. X-ray Diffraction Data Processing and Refinement Statistics (cont.)

Refinement				
Resolution range (Å)	50.0-2.0	50.0-2.7	50.0-2.55	50.0-2.35
No. reflections:				
Working set	40472	18323	21505	27724
Test set	2105	983	1141	1462
R_{work}/R_{free} [c]	0.223/0.241	0.224/0.286	0.224/0.287	0.220/0.259
No. atoms:				
Protein	4079	4079	4074	2076+1998
Ligand	0	20	28	28
Water	153	61	96	68
rmsd bond lengths (Å)[d]	0.008	0.022	0.022	0.022
rmsd bond angles (deg)	1.3	1.98	1.99	1.99
Refinement program	CNS	CNS, Refmac	CNS, Refmac	CNS, Refmac

[a] values in parentheses represent data from the highest resolution shell
[b] Rmerge = Σhkl Σi|Ii(hkl) - <I(hkl)>|/Σhkl ΣiIi(hkl)
[c] Rfactor = Σ||Fobs|-|Fcalc||/Σ|Fobs|. Rwork is the Rfactor calculated using all diffraction data included in the refinement. Rfree is the Rfactor calculated using the randomly chosen 5% of diffraction data that were not included in the refinement.
[d] rmsd = root mean square deviation

Additional programs used to view, analyze, and manipulate structure information include SSM Superpose, a program within COOT that superimposes the Cα atoms of one structure onto another generating an rmsd value[41], PyMOL, a molecular graphics program used to create images of the protein structure[42], and Adobe Photoshop CS4, used to label and manipulate images created with PyMOL.

5.4.4. Accession codes

3K4O: IPK 'apo'; 3K52: IPK−IP complex; 3K4Y: IPK−IPP complex; and 3K56: IPK−IPPβS complex. Gene cloning, protein expression, purification, crystallization and data collection are detailed in the Supporting Information.

5.5. SUPPORTING INFORMATION

5.5.1. Cloning of IPK genes and mutant construction

The IPK gene MJ0044[11] was amplified from *Methanocaldococcus jannaschii* genomic DNA (ATCC® 43067D-5™) by PCR. An IPK homolog from *Methanococcus maripaludis* was also amplified from genomic DNA (ATCC® BAA-1333D-5™) by PCR. Both genes were amplified using Phusion™ High-Fidelity DNA polymerase (New England Biolabs, Inc) employing a 60°C annealing temperature and a 30 s PCR extension time. The PCR products were digested with NcoI and XhoI (New England Biolabs, Inc), purified, and ligated into a NcoI/XhoI digested pHIS8 vector (a modified version of pET28a(+) containing an N-terminal 8-histidine tag) using T4 DNA ligase (New England Biolabs, Inc). All mutations in IPK were made using the QuikChange protocol with PfuTurbo® DNA Polymerase (Stratagene) together with a 6.5 min PCR extension time. The primer pairs used in all PCR reactions are listed in Table 5.3.

5.5.2. Protein expression and purification

The plasmid containing the IPK gene (IPKpHIS8) was transformed into E. coli Bl21(DE3) competent cells (Novagen). One colony was grown in LB media (75 ml) overnight at 37°C, 25 ml of the overnight culture was transferred to one liter of TB media and grown at 37°C until an OD600 of 1.2. Isopropyl-β-D-thiogalactoside (0.2 mM final concentration) was then added, cells were shaken overnight at 37°C for approximately 12–14 hr, harvested by centrifugation and lysed using lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1% (v/v) Tween-20, 10% (v/v) glycerol, 10 mM 2-mercaptoethanol) containing lysozyme (1 mg ml^{-1}). The lysate was stirred at 4°C for 1 hr, sonicated, and centrifuged at 21,000 rpm for 45 min at 4°C. The supernatant was loaded onto a column containing Ni-NTA

agarose S3 resin (Qiagen), washed with lysis buffer and wash buffer (lysis buffer without Tween- 20), and then eluted with elution buffer (wash buffer containing 250 mM imidazole). The protein was digested with thrombin overnight during dialysis in buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl) containing 10 mM 2-mercaptoethanol. The dialyzed solution was passed through a column containing Benzamidine Sepharose 4 Fast Flow (high sub) (GE Healthcare) and Ni-NTA agarose. The eluant was heated at 80°C for 10 min to precipitate contaminating proteins and the supernatant passed through a HiLoad™ 16/16 Superdex™ 200 prep grade (GE Healthcare) gel filtration column using dialysis buffer containing 2 mM DTT. IPK fractions were combined, concentrated to approximately 15 mg ml^{-1}, and frozen at –80°C.

5.5.3. Crystallization and data collection

Crystals of IPK were grown by hanging-drop vapor-diffusion using a 2 µl drop containing 1 µl of IPK (15 mg ml^{-1}) and 1 µl of reservoir. Crystallization conditions were obtained using the Hampton Crystal Screen I (Hampton Research) and optimized to improve crystal morphology and size. IPK crystals formed large plates over a reservoir containing 1.5–1.6 M ammonium sulfate at 298 K. The plates were visible after 1–2 days and reached maximum size after 1 week. Crystal soaks were set up with heavy atoms (0.1–0.5 mM ethyl mercuric phosphate) or ligands (1 mM IP, 5 mM IPP or 1 mMIP / 5 mM ATPγS) in 2.0 M ammonium sulfate. After 1–2 days, crystals were placed in a cryo-protectant (2.0 M ammonium sulfate and 20% (v/v) ethylene glycol) for 10–30 sec then flash frozen in liquid nitrogen. X-ray data were collected at 110 K on ALS beamlines 8.2.1 and 8.2.2 (Lawrence Berkeley National Laboratory, Berkeley, CA) using an ADSC Q315 CCD detector. All x-ray diffraction data including single anomalous diffraction (SAD) data were collected at λ=1.0 Å.

Table 5.3. Primer pairs for PCR reactions

Protein Name	Forward primer	Reverse Primer
IPK-*Mjannaschii*	5'-cccatggcggatccatgctaaccatattaaaattaggagg-3'	5'-tggtggtgctcgagttattctgaaaaatcaatttctgttc-3'
IPK-*Mmaripaludis*	3'-tggttcccatggaatgtttgcaatcttaaaactaggcgggag-5'	5'-gtggtgctcgagttaatttattaatgttccttttacattttt-3'
IPK-*Mjannaschii* H60A	5'-cgtccatggaggaggagcttttggtgctccagtagctaaa-3'	5'-tttagctactggagcaccaaaagctcctcctccatggacg-3'
IPK-*Mjannaschii* H60N	5'-atggaggaggagcttttggtaatccagtagctaaaaaatac-3	5'-gtatttttagctactggattaccaaaagctcctcctccat-3'
IPK-*Mjannaschii* H60Q	5'-catggaggaggagcttttggtcagccagtagctaaa-3'	5'-tttagctactggctgaccaaaagctcctcctccatg-3
IPK-*Mjannaschii* F83A	5'-caaaaaaatatttataaacatggagaaaggagcttgggaaattcaaagagcaatgagaagattt-3'	5'-aaatcttctcattgctctttgaatttcccaagctcctttctccatgtttataaatatttttttg-3'
IPK-*Mjannaschii* F83A_2[a]	5'-aatatttataaacatggagaaaggagcttgggaagctcaaagagcaatgaga-3'	5'-tctcattgctctttgagcttcccaagctcctttctccatgtttataaatatt-3'
IPK-*Mjannaschii* I86G	5'-ttataaacatggagaaaggattttgggaaggtcaaagagcaatgagaagatttaac-3'	5'-gttaaatcttctcattgctctttgaccttcccaaaatcctttctccatgtttataa-3'
IPK-*Mjannaschii* I86A	5'-ttataaacatggagaaaggattttgggaagctcaaagagcaatgagaagatttaac-3'	5'-gttaaatcttctcattgctctttgagcttcccaaaatcctttctccatgtttataa-3'
IPK-*Mjannaschii* I146G	5'-ggaatttagttccagttattcatggagatggtgtaattgatgataaaaacggct-3'	5'-agccgtttttatcatcaattacaccatctccatgaataactggaactaaattcc-3'
IPK-*Mjannaschii* I146A	5'-ggaatttagttccagttattcatggagatgctgtaattgatgataaaaacggct-3'	5'-agccgtttttatcatcaattacagcatctccatgaataactggaactaaattcc-3'
IPK-*Mjannaschii* I146V	5'-gaatttagttccagttattcatggagatgttgtaattgatgataaaaacgg-3'	5'-ccgtttttatcatcaattacaacatctccatgaataactggaactaaattc-3'
IPK-*Mjannaschii* I156A	5'-gtaattgatgataaaaacggctatagagcaatttctggagatgacatagttccata-3'	5'-tatggaactatgtcatctccagaaattgctctatagccgtttttatcatcaattac-3'
IPK-*Mjannaschii* I156V	5'-gtaattgatgataaaaacggctatagagttatttctggagatgacatagttccata-3'	5'-tatggaactatgtcatctccagaaataactctatagccgtttttatcatcaattac-3'

[a]This primer pair is in the context of the I86A mutation

ACKNOWLEDGEMENTS

The text of chapter 5, in full, is a reprint of the material as it appears in ACS Chemical Biology 2010, 5(6), pp 589-601. I am the primary author of this paper. The research was performed under the supervision of Joseph P. Noel.

REFERENCES

1. Gershenzon, J.; Dudareva, N., The function of terpene natural products in the natural world. *Nature chemical biology* **2007,** *3* (7), 408-414.

2. Novakova, Z.; Surin, S.; Blasko, J.; Majernik, A.; Smigan, P., Membrane proteins and squalene-hydrosqualene profile in methanoarchaeon Methanothermobacter thermautotrophicus resistant to N,N'-dicyclohexylcarbodiimide. *Folia microbiologica* **2008,** *53* (3), 237-240.

3. Ourisson, G.; Rohmer, M.; Poralla, K., Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annual Review of Microbiology* **1987,** *41*, 301-333.

4. Eichler, J.; Adams, M. W., Posttranslational protein modification in Archaea. *Microbiology and molecular biology reviews : MMBR* **2005,** *69* (3), 393-425.

5. Sieiro, C.; Poza, M.; de Miguel, T.; Villa, T. G., Genetic basis of microbial carotenogenesis. *International microbiology : the official journal of the Spanish Society for Microbiology* **2003,** *6* (1), 11-16.

6. Hemmi, H.; Ikejiri, S.; Nakayama, T.; Nishino, T., Fusion-type lycopene beta-cyclase from a thermoacidophilic archaeon Sulfolobus solfataricus. *Biochemical and biophysical research communications* **2003,** *305* (3), 586-591.

7. Koga, Y.; Morii, H., Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiology and molecular biology reviews : MMBR* **2007,** *71* (1), 97-120.

8. Rohmer, M., The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural product reports* **1999,** *16* (5), 565-574.

9. Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R., Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97* (24), 13172-13177.

10. Smit, A.; Mushegian, A., Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome research* **2000,** *10* (10), 1468-1484.

11. Grochowski, L. L.; Xu, H.; White, R. H., Methanocaldococcus jannaschii uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *Journal of Bacteriology* **2006,** *188* (9), 3192-3198.

12. Marina, A.; Alzari, P. M.; Bravo, J.; Uriarte, M.; Barcelona, B.; Fita, I.; Rubio, V., Carbamate kinase: New structural machinery for making carbamoyl phosphate, the common precursor of pyrimidines and arginine. *Protein science : a publication of the Protein Society* **1999,** *8* (4), 934-940.

13. Marco-Marin, C.; Gil-Ortiz, F.; Perez-Arellano, I.; Cervera, J.; Fita, I.; Rubio, V., A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of Escherichia coli glutamate 5-kinase. *Journal of Molecular Biology* **2007,** *367* (5), 1431-1446.

14. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **2007,** *372* (3), 774-797.

15. Faehnle, C. R.; Liu, X.; Pavlovsky, A.; Viola, R. E., The initial step in the archaeal aspartate biosynthetic pathway catalyzed by a monofunctional aspartokinase. *Acta crystallographica.Section F, Structural biology and crystallization communications* **2006,** *62* (Pt 10), 962-966.

16. Gil-Ortiz, F.; Ramon-Maiques, S.; Fita, I.; Rubio, V., The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF(4)(-) transition state mimic. *Journal of Molecular Biology* **2003,** *331* (1), 231-244.

17. Kotaka, M.; Ren, J.; Lockyer, M.; Hawkins, A. R.; Stammers, D. K., Structures of R- and T-state Escherichia coli aspartokinase III. Mechanisms of the allosteric transition and inhibition by lysine. *The Journal of biological chemistry* **2006,** *281* (42), 31544-31552.

18. Marco-Marin, C.; Gil-Ortiz, F.; Rubio, V., The crystal structure of Pyrococcus furiosus UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis. *Journal of Molecular Biology* **2005,** *352* (2), 438-454.

19. Liu, X.; Pavlovsky, A. G.; Viola, R. E., The structural basis for allosteric inhibition of a threonine-sensitive aspartokinase. *The Journal of biological chemistry* **2008,** *283* (23), 16216-16225.

20. Pakhomova, S.; Bartlett, S. G.; Augustus, A.; Kuzuyama, T.; Newcomer, M. E., Crystal structure of fosfomycin resistance kinase FomA from Streptomyces wedmorensis. *The Journal of biological chemistry* **2008,** *283* (42), 28518-28526.

21. Ramon-Maiques, S.; Marina, A.; Gil-Ortiz, F.; Fita, I.; Rubio, V., Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the

amino acid kinase enzyme family, during catalysis. *Structure (London, England : 1993)* **2002,** *10* (3), 329-342.

22. Ramon-Maiques, S.; Fernandez-Murga, M. L.; Gil-Ortiz, F.; Vagin, A.; Fita, I.; Rubio, V., Structural bases of feed-back control of arginine biosynthesis, revealed by the structures of two hexameric N-acetylglutamate kinases, from Thermotoga maritima and Pseudomonas aeruginosa. *Journal of Molecular Biology* **2006,** *356* (3), 695-713.

23. Briozzo, P.; Evrin, C.; Meyer, P.; Assairi, L.; Joly, N.; Barzu, O.; Gilles, A. M., Structure of Escherichia coli UMP kinase differs from that of other nucleoside monophosphate kinases and sheds new light on enzyme regulation. *The Journal of biological chemistry* **2005,** *280* (27), 25533-25540.

24. Marco-Marin, C.; Ramon-Maiques, S.; Tavarez, S.; Rubio, V., Site-directed mutagenesis of Escherichia coli acetylglutamate kinase and aspartokinase III probes the catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase. *Journal of Molecular Biology* **2003,** *334* (3), 459-476.

25. Meyer, P.; Evrin, C.; Briozzo, P.; Joly, N.; Barzu, O.; Gilles, A. M., Structural and functional characterization of Escherichia coli UMP kinase in complex with its allosteric regulator GTP. *The Journal of biological chemistry* **2008,** *283* (51), 36011-36018.

26. Jensen, K. S.; Johansson, E.; Jensen, K. F., Structural and enzymatic investigation of the Sulfolobus solfataricus uridylate kinase shows competitive UTP inhibition and the lack of GTP stimulation. *Biochemistry* **2007,** *46* (10), 2745-2757.

27. Bucurenci, N.; Serina, L.; Zaharia, C.; Landais, S.; Danchin, A.; Barzu, O., Mutational analysis of UMP kinase from Escherichia coli. *Journal of Bacteriology* **1998,** *180* (3), 473-477.

28. Song, L., A soluble form of phosphatase in Saccharomyces cerevisiae capable of converting farnesyl diphosphate into E,E-farnesol. *Applied Biochemistry and Biotechnology* **2006,** *128* (2), 149-158.

29. Coleman, J. E., Structure and mechanism of alkaline phosphatase. *Annual Review of Biophysics and Biomolecular Structure* **1992,** *21*, 441-483.

30. Rose, M. W.; Rose, N. D.; Boggs, J.; Lenevich, S.; Xu, J.; Barany, G.; Distefano, M. D., Evaluation of geranylazide and farnesylazide diphosphate for incorporation of prenylazides into a CAAX box-containing peptide using protein farnesyltransferase. *The journal of peptide research : official journal of the American Peptide Society* **2005,** *65* (6), 529-537.

31. Hovlid, M. L.; Edelstein, R. L.; Henry, O.; Ochocki, J.; DeGraw, A.; Lenevich, S.; Talbot, T.; Young, V. G.; Hruza, A. W.; Lopez-Gallego, F.; Labello, N. P.; Strickland,

C. L.; Schmidt-Dannert, C.; Distefano, M. D., Synthesis, properties, and applications of diazotrifluropropanoyl-containing photoactive analogs of farnesyl diphosphate containing modified linkages for enhanced stability. *Chemical biology & drug design* **2010,** *75* (1), 51-67.

32. Lindsley, J. E., Use of a real-time, coupled assay to measure the ATPase activity of DNA topoisomerase II. *Methods in molecular biology (Clifton, N.J.)* **2001,** *95*, 57-64.

33. O'Maille, P. E.; Tsai, M. D.; Greenhagen, B. T.; Chappell, J.; Noel, J. P., Gene library synthesis by structure-based combinatorial protein engineering. *Methods Enzymol.* **2004,** *388*, 75-91.

34. Kabsch, W., Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **1993,** *26* (6), 795-800.

35. Terwilliger, T., SOLVE and RESOLVE: automated structure solution, density modification and model building. *Journal of synchrotron radiation* **2004,** *11* (Pt 1), 49-52.

36. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta crystallographica.Section D, Biological crystallography* **1994,** *50* (Pt 5), 760-763.

37. Perrakis, A.; Morris, R.; Lamzin, V. S., Automated protein model building combined with iterative structure refinement. *Nature structural biology* **1999,** *6* (5), 458-463.

38. Brunger, A. T., Version 1.2 of the Crystallography and NMR system. *Nature protocols* **2007,** *2* (11), 2728-2733.

39. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica.Section D, Biological crystallography* **1998,** *54* (Pt 5), 905-921.

40. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **1993,** *26*, 283-291.

41. Krissinel, E.; Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica.Section D, Biological crystallography* **2004,** *60* (Pt 12 Pt 1), 2256-2268.

42. DeLano, W. L., The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA, 2002.

Chapter 6

Isopentenyl Phosphate Kinase Homologs Outside of Archaea Suggest a Bifurcating Mevalonate Pathway in a Diversity of Eukaryotes

6.1. ABSTRACT

Archaea encode a variant of the canonical mevalonate pathway, using isopentenyl phosphate kinase (IPK) as part of a two-enzyme substitution in the final steps of isopentenyl diphosphate (IPP) biosynthesis. We found IPK homologs intermittently distributed in most major eukaryotic lineages. These homologs retain IPK activity, suggesting that many eukaryotes possess a bifurcating mevalonate pathway.

6.2. INTRODUCTION

IPP and its isomer, dimethylallyl diphosphate (DMAPP), are essential precursors to all isoprenoids including steroids, terpenoids, carotenoids, and numerous primary and secondary metabolites. IPP biosynthesis occurs through the classical mevalonate pathway (MVA) in eukaryotes and some bacteria or through the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway in plastid-bearing eukaryotes and bacteria. Recent work suggests that archaea use a variant of the MVA pathway (referred to as a modified or alternative MVA pathway)[1-3]. Archaea have genes for all but the last two enzymes of the classical MVA pathway, phosphomevalonate kinase (PMK) and diphosphomevalonate decarboxylase (DPM-DC) (Figure 6.1, Table 6.2).[1] Grochowski et al. characterized an isopentenyl phosphate kinase (IPK) in the thermophilic archaeon *Methanocaldococcus jannaschii*, which catalyzes the ATP-dependent phosphorylation of isopentenyl phosphate (IP) to IPP[2]. These authors proposed the alternative MVA pathway, which reverses the last two steps of the classical pathway by using an (unknown) phosphomevalonate decarboxylase followed by IPK. Very recent work has reviewed the phylogeny of the MVA pathway across all three domains of life[3]; work here focuses on the characterization of IPKs in both Archaea and Eukarya.

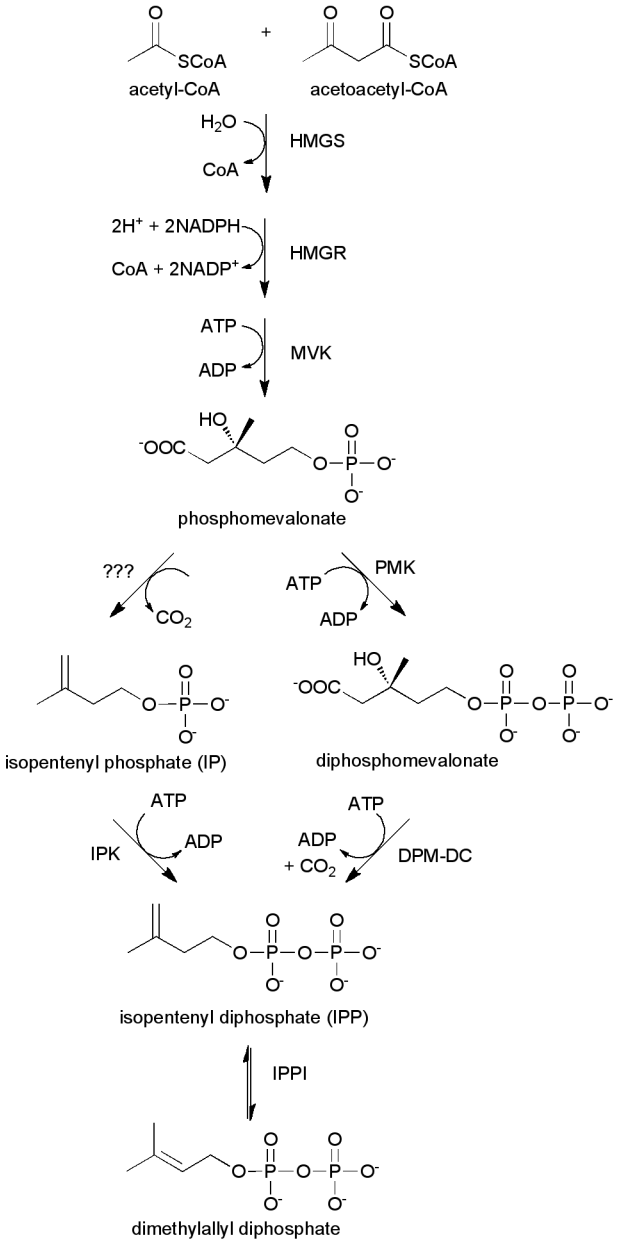


Figure 6.1. The bifurcating mevalonate pathway. The two pathways diverge following the production of phosphomevalonate.

The recently published crystal structures of IPK from *M. jannaschii* highlight an active site histidine residue that is critical for IP binding and catalysis throughout the course of the kinase reaction.[4] This residue is distinct from the equivalent residues of other members of

the amino acid kinase (AAK) superfamily to which IPK belongs, and it can therefore be used as a marker to identify putative IPK homologs.

6.3. RESULTS AND DISCUSSION

6.3.1. Phylogenetic diversity of IPK

We used overall sequence conservation coupled with the characteristic histidine to explore the phylogenetic diversity of IPK. Psi-blast and profile HMMs were used to detect IPK homologs in public protein, EST, and genome databases; other AAK profiles were used to distinguish ambiguous homologs. We find IPK in almost all archaea, a small cluster of GNS bacteria, and in an exceptionally sporadic distribution across most major eukaryotic lineages (Figure 6.2).

Within animals, the gene appears to have been independently lost many times in evolution (Figure 6.2, Supporting Information). Such scattered distribution suggests an unprecedented degree of gene loss or an equally unusual degree of horizontal gene transfer. For example, IPK is absent from choanoflagellates and sponges, but found in early branching animals such as *Trichoplax* and corals. It is found in a shark (*C. milii*) but not teleost fish, in an amphibian (the newt *N. viridescens*) but not in frogs, and in a lizard (*A. carolinensis*) and a snake (*P. olfersii*), but not in any bird or mammal. IPK is absent from most fungi; by contrast, it is found in every sequenced green plant genome. Bacterial IPK is restricted to a small cluster of 5 genomes within the class and phylum Chloroflexi (green non-sulfur bacteria). Bacteria also contain the closest structural homolog of IPK, fomA from *S. wedmorensis*, which phosphorylates and inactivates the antibiotic fosfomycin[5].

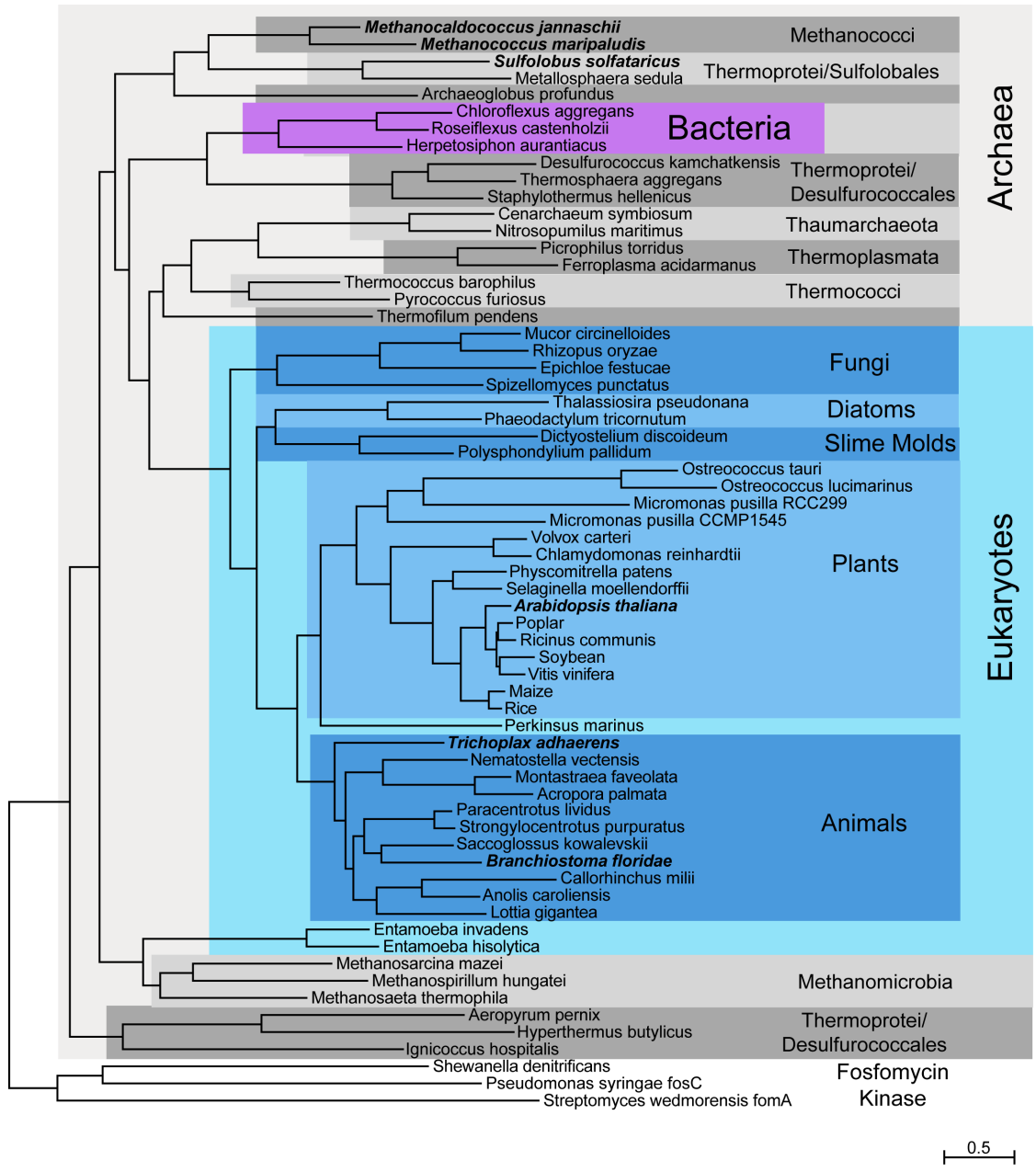


Figure 6.2. IPK phylogeny from eukaryotes (blue), selected archaea (gray), and a small group of bacteria (purple). Maximum likelihood tree calculated by PhyML with the major clades highlighted. Several bacterial species containing fosfomycin kinase form a separate branch at the bottom of the tree.

6.3.2. Catalytic activity of IPK homologs

We tested five homlogs for catalytic activity, including the characterized IPK *from M. jannaschii*[2, 4] two other archaea (*Methanococcus maripaludis* and *Sulfolobus solfataricus*) and three eukaryotes: *Trichoplax adhaerens* (early-branching metazoan), *Branchiostoma floridae* (chordate), and *Arabidopsis thaliana* (plant). Remarkably, all six IPK homologs catalyzed the phosphorylation of IP to IPP; kinetic constants are reported in Table 6.1.

Table 6.1. Kinetic constants for characterized IPKs

	$K_{m,\ IP}$ (μM)	k_{cat} (s^{-1})	$K_{i,\ IP}$	$k_{cat}/K_{m,IP}$ ($s^{-1}\mu M^{-1}$)	Goodness of fit (R^2)[a]
M. jannaschii	4.3 (±0.6)[b]	1.46 (±0.03)	-[c]	0.34 (±0.05)	0.90
M. maripaludis	21.4 (±4.3)	15.2 (±1.4)	877 (±550)	0.71 (±0.16)	0.99
S. solfataricus	23.6 (±4.8)	0.91 (±0.05)	-	0.04 (±0.01)	0.92
B. floridae	13.3 (±2.0)	27.2 (±1.2)	2820 (±1700)	2.05 (±0.32)	0.98
A. thaliana	0.79 (±0.35)	1.9 (±0.2)	522 (±381)	2.4 (±1.1)	0.95
T. adhaerens	3.1 (±1.6)	2.4 (±0.2)	-	0.77 (0.40)	0.79

[a] $R^2 = 1.0 – (SSreg/SStot)$, where SSreg = sum of squares value, SStot = sum of squares of the distances between each point and a horizontal line passing through the average of all y values
[b] Values in parentheses represent standard error (or propagation of error) for each kinetic constant
[c] K_i constant was not calculated

Although most archaea lack the last enzymes of the classical MVA pathway, the order Sulfolobales contains all of them in addition to IPK (Table 6.2). IPK from *S. solfataricus* has a much lower catalytic efficiency than the other IPKs tested (Table 6.1) and may therefore be losing function. This agrees with the observation that IPK has persisted across most eukaryotic lineages, but has been lost during many rare evolutionary events, probably due to partial redundancy with the MVA pathway. Green plants are the exception, in which IPK may have

gained an indispensable function. Subcellular compartmentalization is a precedent of plant isoprenoid biosynthesis[6], however no localization signals could be found in plant IPKs.

6.3.3. Role for IPK in other kingdoms of life

The unusual phylogeny of IPK and its membership in a family of kinases that phosphorylate such a broad range of substrates leave open the possibility that IPK may play a different physiological role, such as participating in the phosphorylation of an IP-like substrate or recycling IP that may accumulate in vivo as a consequence of phosphatase-dependent IPP degradation. Although certain archaeal IPKs demonstrate some ability to phosphorylate other isoprenoid substrates (such as dimethylallyl phosphate and geranyl phosphate), they prefer IP^7. Failure to date to identify the decarboxylase required to complete the alternative pathway is also reason to speculate on the role of IPK. The *M. jannaschii* gene MJ0403 which has sequence similarity to iron-binding dioxygenases has been proposed to serve this function[2], however attempts to show biochemical activity have not been successful, and the search for other possible decarboxylase candidates is under way.

6.3.4. Conclusions

Remote homology techniques and an active site histidine residue were used to successfully find and characterize IPK homologs among eukaryotes and some bacteria. In contrast to previous research which has only briefly described the existence of eukaryotic IPKs[3, 7], this detailed report includes a thorough phylogenetic analysis of this gene across all domains of life, and the first experimental evidence for the existence of IPK outside of archaea. The presence of active eukaryotic IPKs supports the idea that a bifurcating pathway may exist in plants, some animals, and several fungi and protists. Future work will involve the

complete biochemical characterization of both the classical and alternative MVA pathways (including the decarboxylase which has thus far been uncharacterized) in a given organism.

6.4. METHODS

6.4.1. Cloning of IPK homologs

IPK homologs from Archaea (*M. jannaschii, M. maripaludis* C5*, S. solfataricus* P2) were cloned from genomic DNA from American Type Cell Cultures (ATCC) as previously described for *M. jannaschii1* into a pET28a$^{(+)}$ vector containing an N-terminal 8-histidine tag using PCR primers as follows:

S. solfataricus P2 forward: 5'-tggttcCCATGGAttggaaatggatatgggatctgaattg-3'

S. solfataricus P2 reverse: 5 -gtggtgCTCGAGtcaggcattcggattacctcttactaaa-3'

M. maripaludis C5 forward: 5'-tggttcCCATGGaatgtttgcaatcttaaaactaggcgggag-3'

M. maripaludis C5 reverse: 5'-gtggtgCTCGAGttaatttattaatgttccttttacattttt-3'

The three IPK homologs from Eukarya (*A. thaliana, T. adhaerens, B. floridae*) were ordered as synthetic genes from Genscript (Piscataway, NJ, USA) and sub-cloned using Gateway technology from Invitrogen (San Diego, CA, USA) into pHIS9GW, an in-house vector modified to contain a 9-histidine tag.

6.4.2. Protein expression and purification

All proteins were expressed according to a previously described procedure with several modifications[4]. While all E. coli Bl21 (DE3) cells expressing archaeal proteins were induced with 0.2mM IPTG overnight at 37°C, all cells expressing eukaryotic proteins were induced with 1.0mM IPTG for five hours at 22°C. All proteins were purified similarly and as

previously described, however none of the proteins other than *M. jannaschii* were incubated at 80°C.

6.4.3. Steady-state kinetic analysis

Kinetic measurements were performed on IPK from *M. maripaludis, S. solfataricus, and B. floridae* using a coupled pyruvate kinase –lactate dehydrogenase assay as previously described for IPK from *M. jannaschii* that employs varying IP concentrations ranging from 2μM-1mM1. Steady-State kinetic curves were fitted using Prism (GraphPad Software Inc., San Diego, CA, USA) to compute Km, kcat, and where appropriate, Ki, IP. Activity measurements were performed for *T. adhaerens* and *A. thaliana* using the coupled assay at four different IP concentrations (2μM, 10μM, 50μM, and 100μM) in triplicate.

6.4.4. Bioinformatics

Public protein, cDNA, EST and genomic databases were searched for IPK homologs, using individual IPK protein sequences, and profile Hidden Markov models built from several individual IPK clades. Genes were predicted from genomic sequence using Genewise[8] and TimeLogic® GeneDetective™ (Active Motif Inc., Carlsbad, CA) programs, with manual editing. Protein sequences were aligned with Muscle[9], and edited with ClustalX[10] and in JalView[11]. Figure 6.2 was created using PhyML[11] using the SPR model and rooted with fosfomycin kinase sequences. Manual editing was used to merge EST sequences and gene predictions, correct frameshifts and fuse one gene split across two contigs. Discrepancies between individual ESTs were resolved to maximize sequence similarity to close homologs.

6.4.5. Phylogenetic Distribution of IPK

Archaea. IPK found in all but three of the 74 complete archaeal genomes found in the Integrated Microbial Genomes (IMG) database as of Mar 8, 20107. Exceptions are both *S. acidocaldarius* and *S. tokodaii*, and *Nanoarchaeum equitans*, a symbiont archaeon with a reduced genome.

Bacteria. Clear IPK homologs found only in all five sequenced genomes of the class *Chloroflexi*, but not within other classes of the phylum *Chloroflexi*. Divergent homologs found in *Streptomyces wedmorensis*, *Streptomyces fradiae* and one strain of *Pseudomonas syringae* (all probably fosfomycin kinases), and *Shewanella denitrificans*. The *P. syringae* gene is found only in a contig from strain PB-5123, and not several other sequenced strains. The sequence contains a frameshift within the ORF and lacks the H60 residue, both of which may be sequencing errors.

Eukaryotes. Searches were made of the non-redundant amino acid (NRAA) Genbank database8, the database of expressed sequence tags (dbEST)[12], and a wide variety of genome databases, including those at Ensembl (www.ensembl.org)[13], Joint Genome Institute (JGI, genome.jgi-psf.org/), Baylor College of Medicine (www.hgsc.bcm.tmc.edu), Sanger Institute (www.genedb.org/) and the Broad Institute (www.broadinstitute.org). Searches were with a series of IPK homologs (blastp against predicted peptides, tblastn against genome) and using a hidden Markov model profile searched against the genome, using Gene Detective.

6.5. SUPPORTING INFORMATION

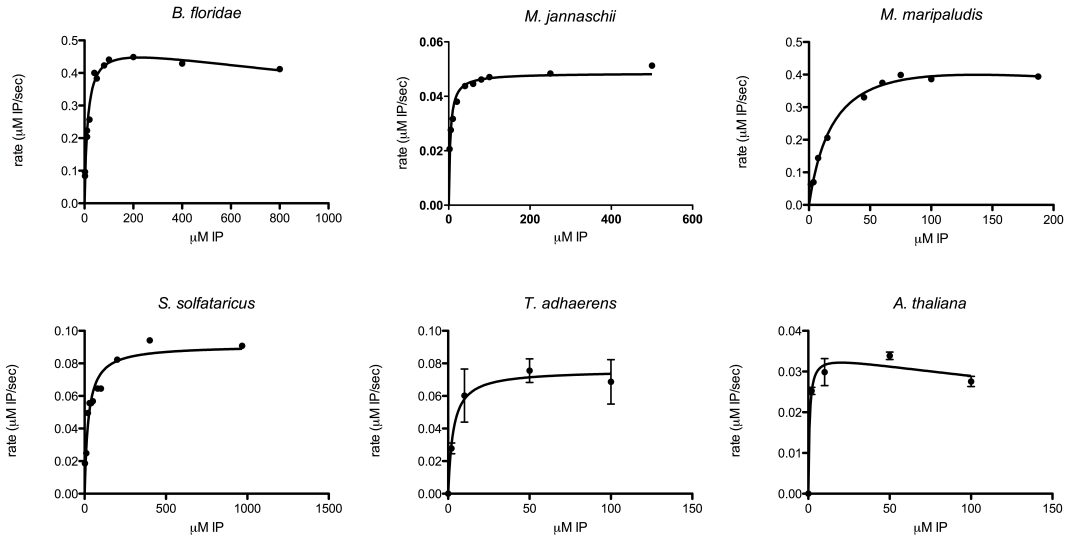


Figure 6.3. Steady-State Kinetics. The kinase reactions were performed with IPK at a fixed concentration ([E] << [S]) while the concentration of IP was varied from anywhere between 2μM and 1mM. The curves for IPK from each organism obeyed steady-state kinetics and were fitted accordingly, as shown in the graphs for these IPK homologs.

Table 6.2. Gene Identifier (GI) Numbers for MVA Pathway Gene Orthologs in Organisms with an Active IPK

GENE[a]	*M. jannaschii*	*M. maripaludis*	*S. solfataricus*	*B. floridae*	*T. adhaerens*	*A. thaliana*
HMGS	15669741	134045424	15897459	260792860	196008117	15234313
HMGR	15668887	134046615	15897456	260821882	196001137	79382641[b]
MVK	15669275	134045303	15897316	260803413	195999336	15240936
PMK	???	???	15899698	260829481	196002301	15222502
DPM-DC	???	???	15899699	260794527	196004226	15224931
IPK	15668214	134046789	15897030	260817561	195996013	22329798

[a] HMGS = 3-hydroxymethylglutaryl CoA (HMGCoA) synthase; HMGR = HMGCoA Reductase; MVK = mevalonate kinase; PMK = phosphomevalonate kinase; DPM-DC = diphosphomevalonate decarboxylase; IPK = isopentenyl phosphate kinase.
[b] *A. thaliana* contains two HMGRs, gi for HMG1 is shown in table, gi for HMG2 is 15227821

6.5.1. Supporting Information on the Phylogenetic Distribution of IPK

Mammals. No IPK found in NRAA genbank database or in 35 mammalian genomes at ensembl.org.

Birds and Reptiles. IPK found in the genome of *Anolis carolensis* (anole lizard) and in an EST from a colubrid snake (*Philodryas olfersii*). No IPK in the genomes of chicken, zebra finch or turkey.

Amphibians. IPK found in one EST from a newt (*Notophthalmus viridescens*), but not seen in the genome of *Xenopus tropicalis*.

Teleost fish. No IPK found in genomes or proteomes of 5 fish (*Danio rerio*, *Oryzias latipes*, *Takifugu rubripes*, *Tetraodon nigrovidris*, or *Gasterosteus aculeatus*). A partial IPK was assembled from the draft genome of the elephant shark, *Callorhinchus milii*.

Invertebrate chordates. IPK was found as a predicted gene in the lancelet, *Branchiostoma floridae*, and in the hemichordate acorn worm, *Saccoglossus kowalevskii*. No IPK was found in *Ciona intestinalis* or *Ciona savignyi*.

Echinoderms (Deuterostomes). An almost-complete IPK was assembled from ESTs of the sea star, *Paracentrotus lividus*, and an almost complete prediction was made from the related sea urchin *Strongylocentrotus purpuratus*.

Arthropods. No IPK was found in the genomes of insects: 12 *Drosophila sp.*, *Apis mellifera*, *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus* or *Pediculus humanus* or the arachnid, *Ixodes scapularis*. IPK was found in the EST from one crustacean, the lobster *Homarus americanus*, but not in the genome of another, *Daphnia pulex*.

Other Bilaterians. IPK was found in Annelids, in single ESTs from the leech, *Hirudo medicinalis* and the earthworm, *Eisenia fetida*, but was not found in the draft genome of the polychaete worm, *Capitella teleta* or the leech, *Helobdella robusta*. IPK was also found in the molluscs, in the form of a full length gene prediction from the limpet, *Lottia gigantea*, an EST from *Aplysia kurodai* and a partial prediction in the genome of *Aplysia californica*. IPK was

not found in the platyhelminthes genomes, *Schistosoma mansoni* or *S. japonicum*, or in 6 nematode genomes.

Cnidarians. ESTs for IPK were found in two corals (*Acropora palmata*, *Montastraea faveolata*) and an IPK was predicted in the sea anemone *Nematostella vectensis*, but could not be found in the *Hydra magnapilliata* genome.

Other early metazoans. An IPK was predicted in *Trichoplax adherens*, none could be found in the sponge *Amphimedon queenslandica*.

Pre-metazoans (within Holozoa). No IPK was found in the genomes of *Monosiga brevicollis*, *Salpingoeca rosetta* or *Capsaspora owczarzaki*.

Fungi. IPK was found in one of two chytrid genomes (*Spizellomyces punctatus*), and in the early branching fungi *Rhizopus oryzae* and *Mucor circinelloides*. No IPK was found in 34 other fungal genomes, though a complete IPK was surprisingly assembled from an EST library taken from *Epichloe festucae*, a member of the Pezizomycotina. Since several related genomes have been sequenced, this may be a case of horizontal transfer (or even EST library contamination).

Green plants. IPK is found in all green plant genomes surveyed, including flowering plants, the lycophyte (*Selaginella moellendorffii*), the moss (*Physcomitrella patens*), the two Ostreococcus genomes (*O. tauri*, *O. lucimarinus*), two strains of *Micromonas pusilla*, and both *Chlamydomonas reinhardtii* and *Volvox carteri*. There may be a duplication of IPK in *Physcomitrella*, but the two sequences in the genome are very close and might be allelic variants.

Amoebozoa. IPK is found in the slime molds *Dictyostelium discoideum* and *Polysphondylium pallidum*, and also in three species of *Entamoeba* (*E. hislolytica*, *E.*

invadens, *E. dispar*), though these three are more similar to IPK from a clade of archaea than to eukaryotes.

Alveolata. IPK is also found in *Perkinsus marinus*, but not in Apicomplexa (any of 6 *Plasmodium* genomes, or two Cryptosporidium genomes), or in any of 3 ciliates (*Tetrahymena thermophila*, *Paramecium tetraurelia*, *Ichthyophthirius multifilis*).

Diatoms. IPK is found in both *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*.

Kinetoplastida. No IPK is found in 4 species of *Trypanosoma* or 4 species of *Leishmania*.

Excavates. No IPK was found in three assemblages of *Giardia lamblia* or in *Trichomonas vaginalis*.

Others. IPK was found in ESTs from *Malawimonas jakobiformis*, and from the dinoflagellate *Alexandrium tamarense*, but not in the genome of *Naegleria gruberi* (Heterolobosea)

6.5.2. Ultra-conserved residues

All IPK sequences conserved the following 11 residues, or were missing these regions due to truncation (number by *M. jannaschii*): K6, G8, G9, K15, G54, H60, P140, G144, D213, T215, G216. K221 is conserved in all but one possible gene prediction error, and G253-T254 in the C-terminus may also be invariant, but mis-predicted in several sequences.

6.5.3. Additional sequences

Additional and fragmentary sequences omitted from the alignment file (omitted for being redundant or partial, degrading the tree).

>*Philodryas olfersii* (vertebrate: snake)

MAAAVDCILKLGGSALTQKNQLEMLKTESLQRAAALVSKLWEAGERRCIIVHGAGSF
GHFQAREYGVALGTSGRSAASDNLREGLCLTRLSVTKLNHLVTEQLISVGVPAVGISP
FGILANNKQECR*

>*Notophthalmus viridescens* (vertebrate: newt) From one EST

GHYGRGHKGQLETVRPDALQRAAAILKRMHAELKSCIVVHGAGSYGHFQAKDYGVS
KGTSGHSPIEMDHLRQGLCLTRLSVTKLNHLVTEQLVKEGVAAVGISPFGAWKMSGR
QVVQTGTEAVKDALISGYVPILHGDCALDADQHCCILSGDTIIEVLSKEFSPKQVVFLT
DVDGIYDQPPNCPGAQLLNSITVNPLWTLRACG*

>*Hirudo medicinalis* (medicinal leech: Annelida) big gap after H60 region, and poor alignment, weak alignment to other homologs

SIVLHGVGSFGHHLAARHRLNAGYSKDDKDFPLVLARIRSSLLKLNRQVVEEFIKRNV
PAVTVSVLCEYFQPRKVIFVLDCGCILSHPPNHPGSRPIRKLQVSEKDDSFNDLETGSL
VPDASGGMKAKVEAAISIVRKSRGSISMLFCPAGEVMRDLCNLREAKDGQLMTELVF
KASRDA*

>*Aplysia kurodai* One EST, upstream stop Eukaryota; Metazoa; Mollusca; Gastropoda;

MENLVIDVIIKFGGSAITDKNSLETLLPWQLEQAVRHVKRCTDAGLTCVIVHGAGSFG
HHQAKQYAVNAGLTGEQTDEETRRKPVGVLCHEASSHQIEQVXSLTLC*

>*Aplysia californica* from genome project sequence. Start only but verifies start of A. kurodai EST

MENLVIDVIIKFGGSAITDKNSLETLLPWQLEQAAKHVKRCTDAGLTCVIVHGAGVSA
YLGHIRPKYGKGGGFSLFRPHILKL

>*Homarus americanus* (American lobster: Arthopoda: Crustacea) from a single EST, has one frameshift, big gap after the H60 region (where the f/s is)

MMSIPVAVDLVIKLGGSAITEKSLPETFKEAAIRESVKLVKLCVNKGLKVVVVHGAGS
MxKVVEQLTDEGLPAVGLSPCGNWTTQDGKVVQSGVSAVSCFLGAGFVPVLHGDCV
LDTVTGCTILSGDTVIKVKSCLFIIYLFFRFNICL*

>*Eisenia fetida* (earthworm, annelid)

PGQVLVSTFPSWMTDNGKVVKSDFDVIERALADRFVPVLSGDVVFDVTKGCNVLSSD
VLLQAICERFDVKRAVFSMDCAGIFT*

>*Montastraea faveolata* (coral) From 2 ESTs

MYRGPWRGLQHTITAVFLENGIPALGISPCGSWLTSSGVVTRSAVTPIVELLEAGFVPI
LHGDCVLDDVQGCSILSGDKIIQRLAEELRPKRVVFLTDVDGIYDKPPEKEDSVLLRK
VYVKSDGQMNVTIATSNLHHDVTGGIREKLQTASNIIRISEQHSRVFVLNIMSETVAYS
VCSRGVLDGNGTEILAETQEFNRVNIQ*

>*Acropora palmata* (coral) From 1 EST

MHCGAWSRLQQMVVESFISNGIPAVGISPCGSWMTTEGVVTRSAIHPIVQCVAAGFVP
ILHGDCVLDTKQGCAALSGDKIIEKLVEELHPSRVVFLTDVGGIYNKPPDNEDATLIRT

VFVNPSGKMSVAIATSVLTHDVTGGVCEKLRTASNIVLISGGKTRVFVANVMAEANV
YS

>*Nicotiana tabacum* (plant) From 2 ESTs, C-terminal open

MEQNKASAPVPPTKRVRCIVKLGGAAITCKNKLETIDEENLTEVSSQLRQALIPNSDSA
KILGMDWSKRPGQSEPPSFVDEFSDQPVADSESFIVVHGAGSFGHFQASKSGVHKGGL
SRPLVKAGFVATRISVTSLNLEIVRALAREGIPSIGMSPFSCGWSTCQRNMTEADISMVI
KAIDAGFIPVLHGDAVLDTLQECTILSGDVIIRHLAAELKPEFVVFLTDVLGVYDRPPV
EPGAVLIREIAVREDGSWSVVKPRLEDTSKPVEFTVAAHDTTGGMVTKITEAAMIAKL
GIDVYITKAGTDHSVKAPP

>*Petunia axillaris* (plant) from 2 ESTs X's added based on alignment to other plants

MEQKATTVATKRVRCIVKLGGAAITCKNKLETIHEDNLRQVSSQLRQVLIPDSASAKV
LGMDWSKTPGHSEAPSIVDDFSYQPVATSETFIVVHGAGSFGHFQASKSSVHKGGLSR
PLVKAGFVATRISVTSLNLEIVRALAREGIPSIGMSPFSCGWSTCERNMTEADTSMVIK
ALDAGFIPVLHGDAVLDTLQDCTILSGDVIIRHXXXXXXXXXXXXXXXXXXXXSST
SGPGAVLIREIAVREDGSWSVVKPRLEDASKPVEFTVAAHDTTGGMVTKITEAAMIA
KLGIDVYITKAGTDHSVKALSGILQGGIPDDWLGTAIRYMS*

>*Alexandrium tamarense* (dinoflagellate) from two ESTs

VLDAAQGAAVLSGDVWMVELCKELKAKSAVFVTDVDGVFTRPPWEEGAELVREILV
DTKTGELELPGVSMSAASHDVTGGLKAELESAAEVLVRAPSVQAVYIVRAGSEGAAQ
ALRGEAPLRGTTLRRKPRD*

>*Malawimonas jakobiformis* (Malawimonadidae) from one EST

MVIIVKFGGSALTDKASFETLRSDALNRCSVAVSQALAAGHRVIIVHGAGSFGHHQA
KRFALSAGLLSHSAAAGSTVTVSAETERAWSAIAPHEQQLGLAHCRASVQRLNAHVV
HSLLRRNVPAVTMSAFPNWFTDGKRLVSDIVPPVLAALERGLVPVLHGDVVMDHAQ
GITVLSGDVIWRCFAVRCRRRRSAR

>*Entamoeba dispar* (Apicomplexan)

MNSIPNLIILKIGGSYLTEKNRVDGPPVLENIHVFSKCLAQFIHSHPKQPIILAHGAGSFG
HVPAAKYHLAEGFHKTGVIECEMAMQELSSVIVNSLIKEGVSAIPFHPFNFVVTENKRI
VDMYLQPLQMMINQGIIPVVHGDIAMDIIQGSCILSADQLVPELAIRFGCSRIGFICNTP
VLNDKGEVIPLINEQNYDSIKKFLHGCKGVDVTGGMAGKISELMIAAKKHMIQSYVF
EGTKECLELFLEGNDVGTKVCQ*

>*Emiliania huxleyi* (Haptophyceae)

MLITRHAAILAAGLLLLTPVAGAFLFLGRRLRRRALRSARCTLVVKLGGSAVTDKTCF
ETVRVAALRETACALSRSPLLAGTVLVHGAGSFGHFHAREHGVSRGTAHSAFSWRGF
ALTRSSVTRLNGIVLTALLEQGIAACGLPPFPRWVLCGGALTDTDEPLGEVRSLLSRG
VVPVLHGDAVFDEARGAAILSGDTLVEELRRLPPSARPAPSSLTDVAGIFHRPPGEDG
AALLRRIVVGPSGEVVDLPQMRTAAHDVTGGVAAERSAPAALHAFPSLIPAALLVCR
LRRS*

>*Chloroflexus aurantiacus* J-10-fl

MYTFIKFGGSVITDKTGREAADLVVIERLAHAVAEARAADPNLALVLGHGSGSFGHH
YAARYGVHRGIPLSADHTGFALTAAAALRLNRIVVDTLLAAQVPAVSFQPSASLQST

NGQIITWETAPIAEALQRRLVPVIHGDVAFDTAQGTAIISTEALLSFLALRSPLQPRRIIL
VGEAAVYTADPHRDPTAQPIPLINQENIAQVLVMTGGSRAADVTGGMRSKIELMWHL
IERLPELEVTLIGPDPALLTAALLGQSLAMGTVIKRW*

>*Roseiflexus sp*. RS-1

MIVFIKFGGSVITDKQQQERADIDTIRQLAEELRQALDAARDLCVIVGHGSGSFGHVY
AQRYGIHRGLAPDDDWMGFALTSGAALRLNRIVVDELLAAGIPALALQPSTTLLARG
GRLVHWETGSLERALERRMVPVIHGDVAFDDVQGSAIISTEQLLAHLATLPTLRPARI
VLVGEAGVYTADPRINPQAERIARIDRRNIANVLAGAGGSHGVDVTGGMRSKVELM
WQLVQTVPGLQVYLIGPKPGSLKRALLGDDTVEGTVIVGG*

>*Streptomyces fradiae* (fomA homolog)

MTPDFLAIKVGGSLFSRKDEPGSLDDDAVTRFARNFARLAETYRGRMVLISGGGAFG
HGAIRDHDTAHAFSLAGLTEATFEVKKRWAEKLRQIGVDAFPLQLAAMCTLRDGTPQ
LRSEVLRSVLDHGVLPVLAGDALFDEHGKLWAFSSDRVPEVLLPMVEGRLRVVTLTD
VDGIVTDGAGGDAILPEIDARSPQQAYAALWGSSEWDATGAMHTKLDALVTCARRG
AECFIMRGRPDSDLEFLTAPFSSWPAHVRSTRITTTASV*

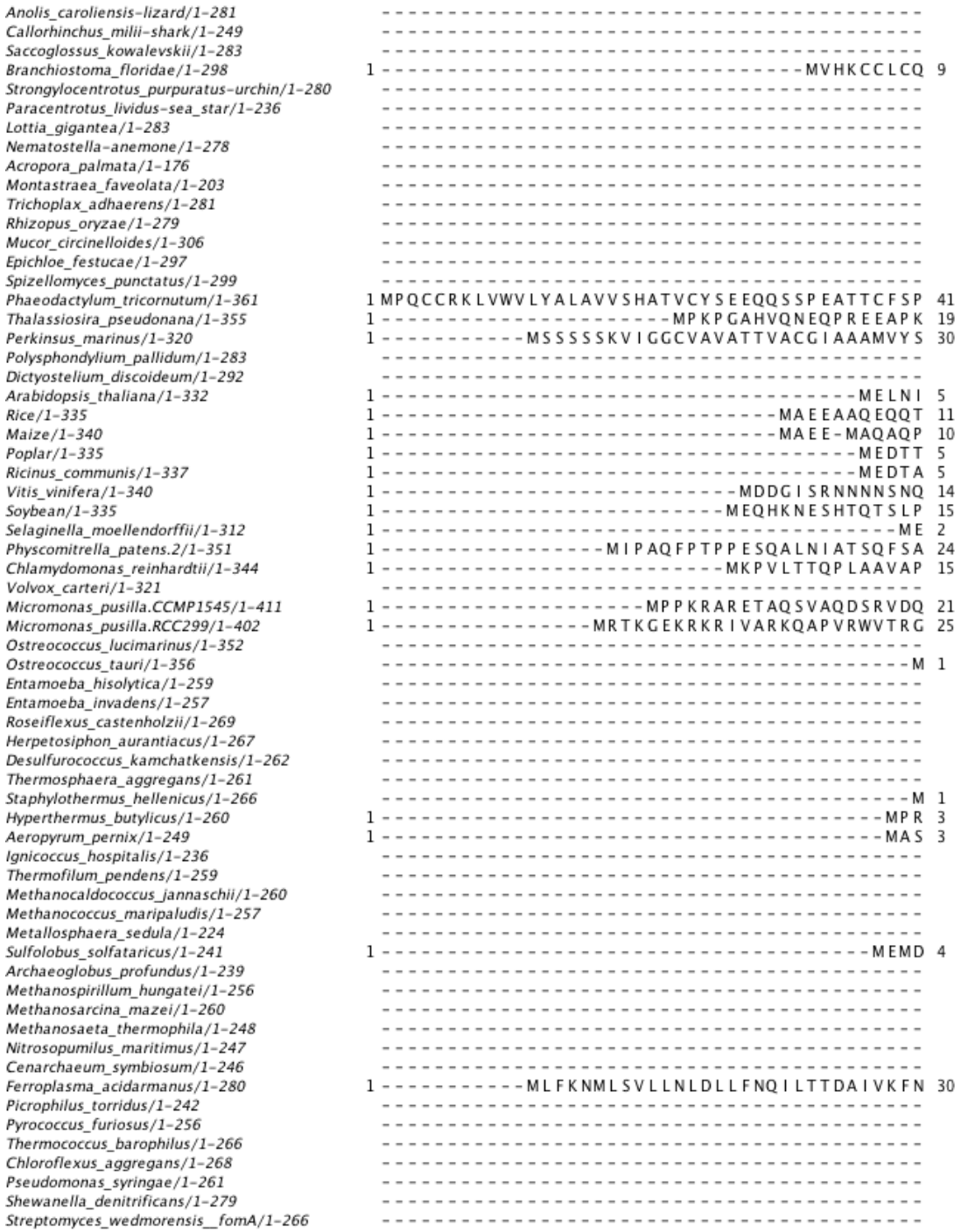


Figure 6.4. Alignment of IPKs from the three domains of life

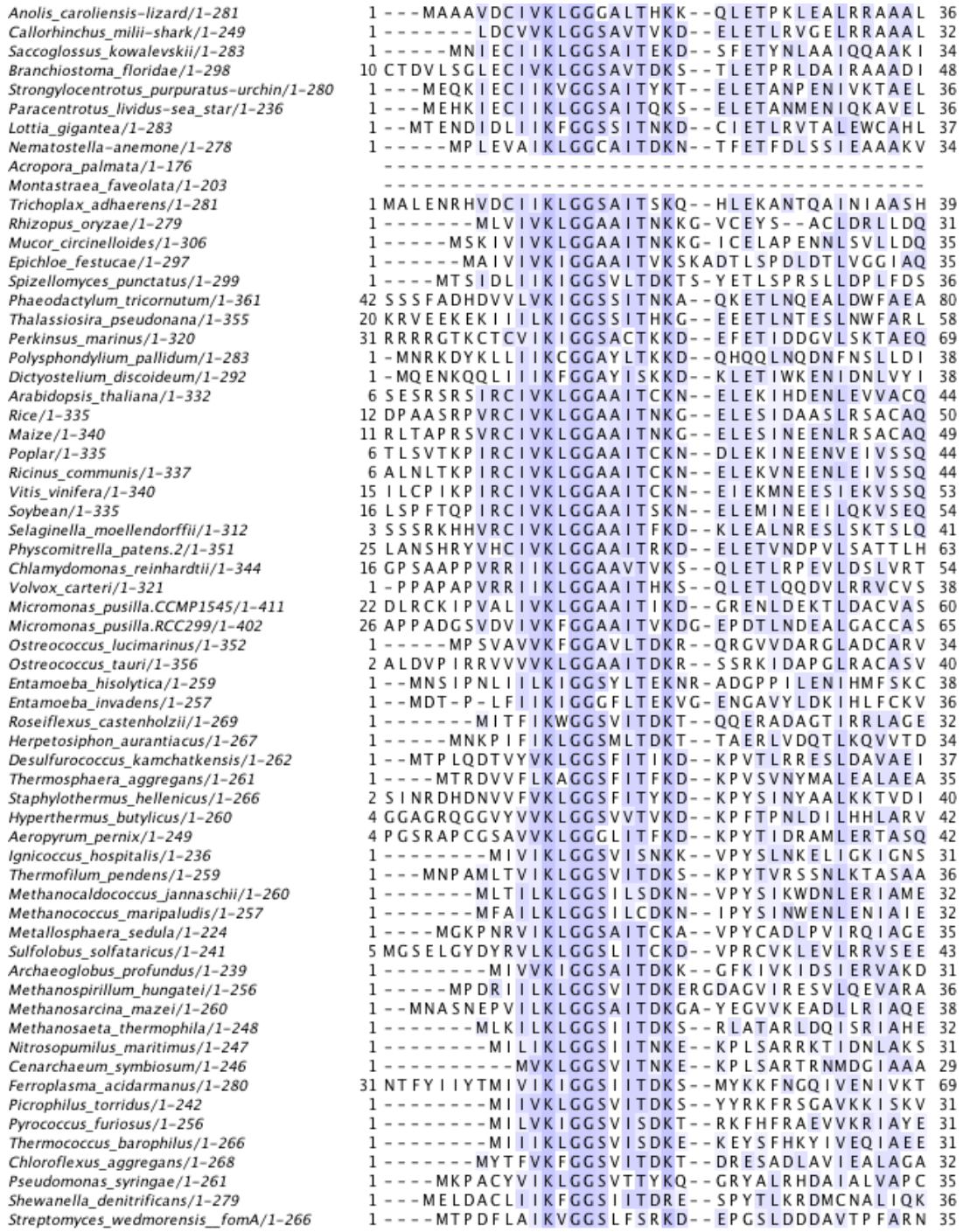


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

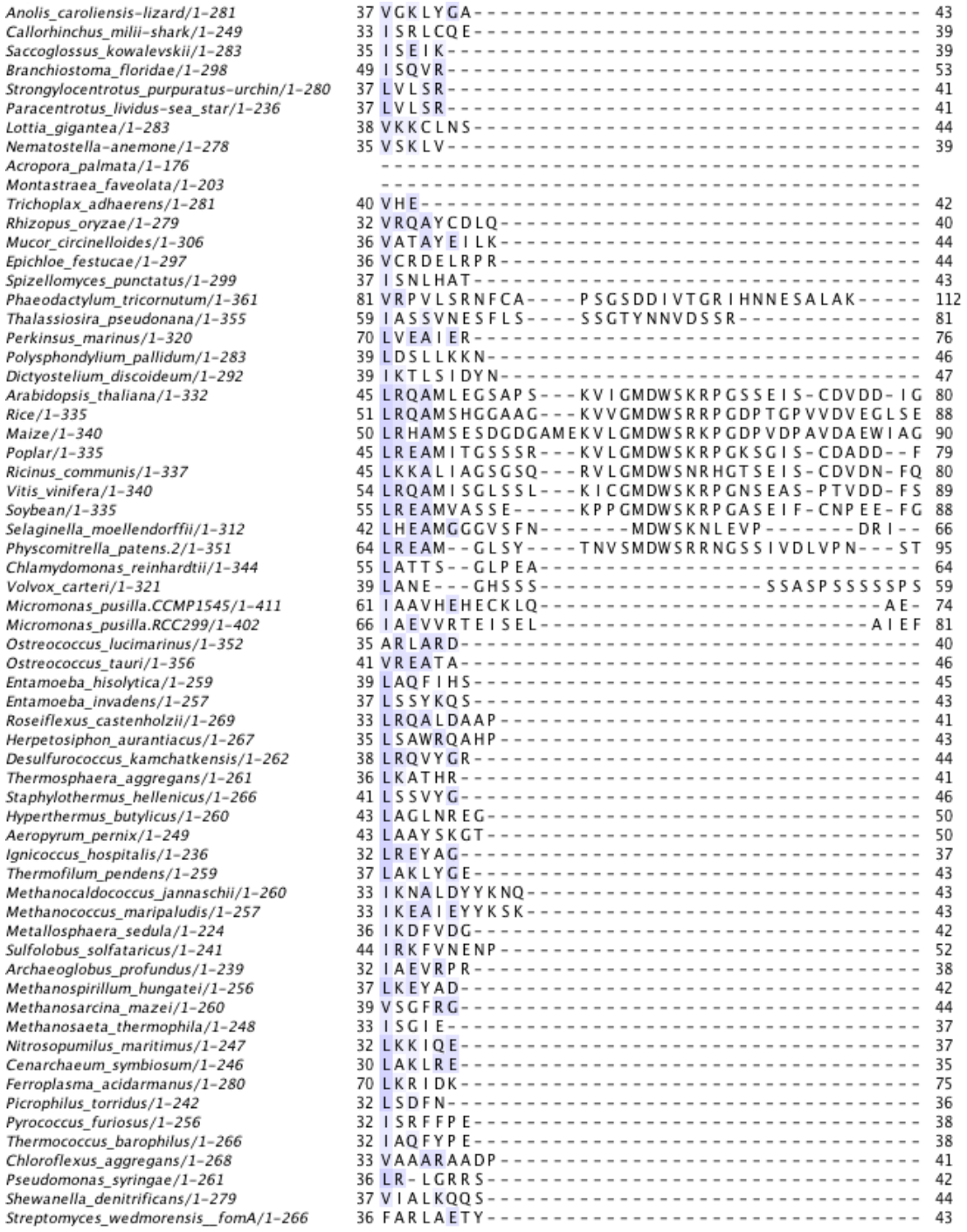


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

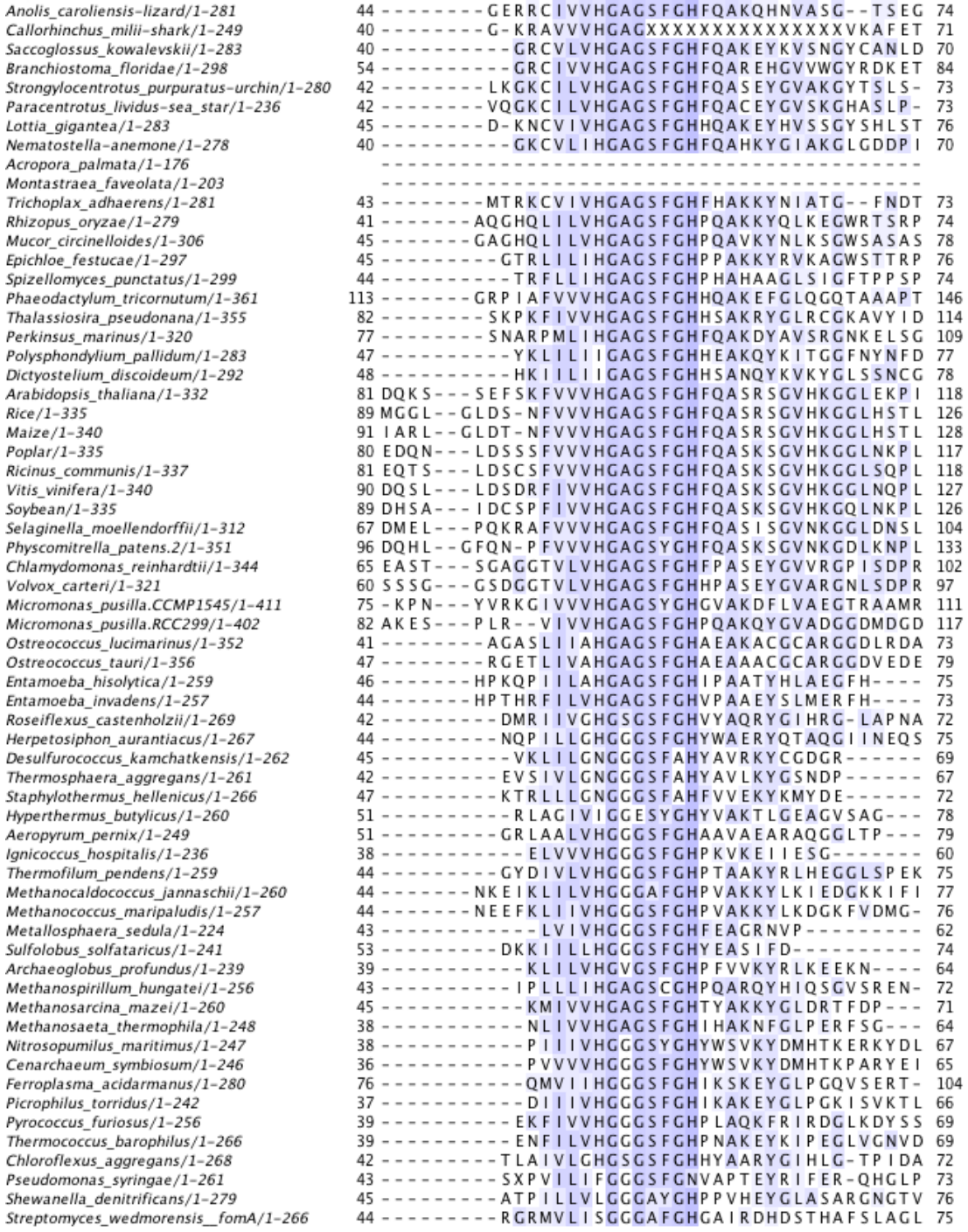


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

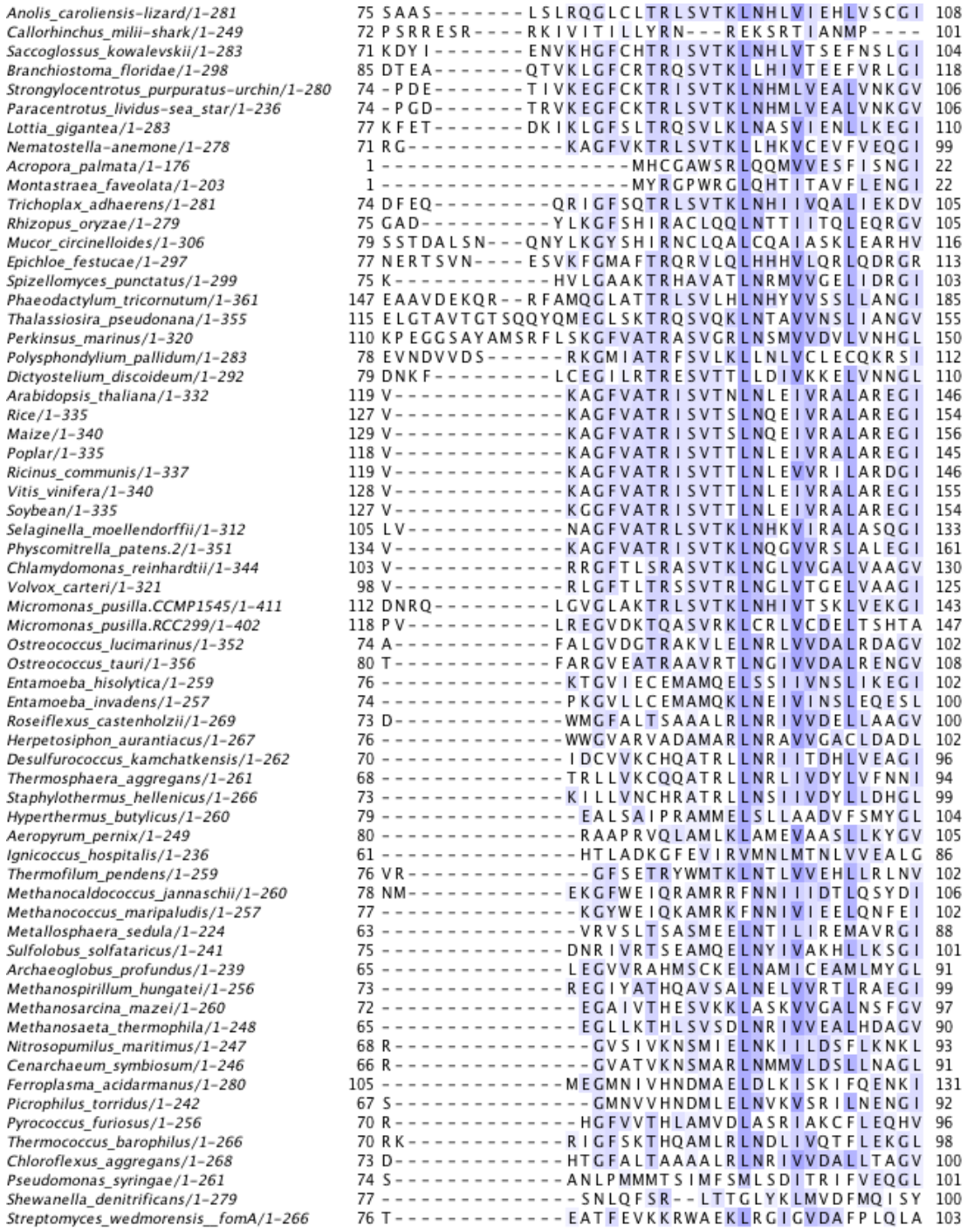


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

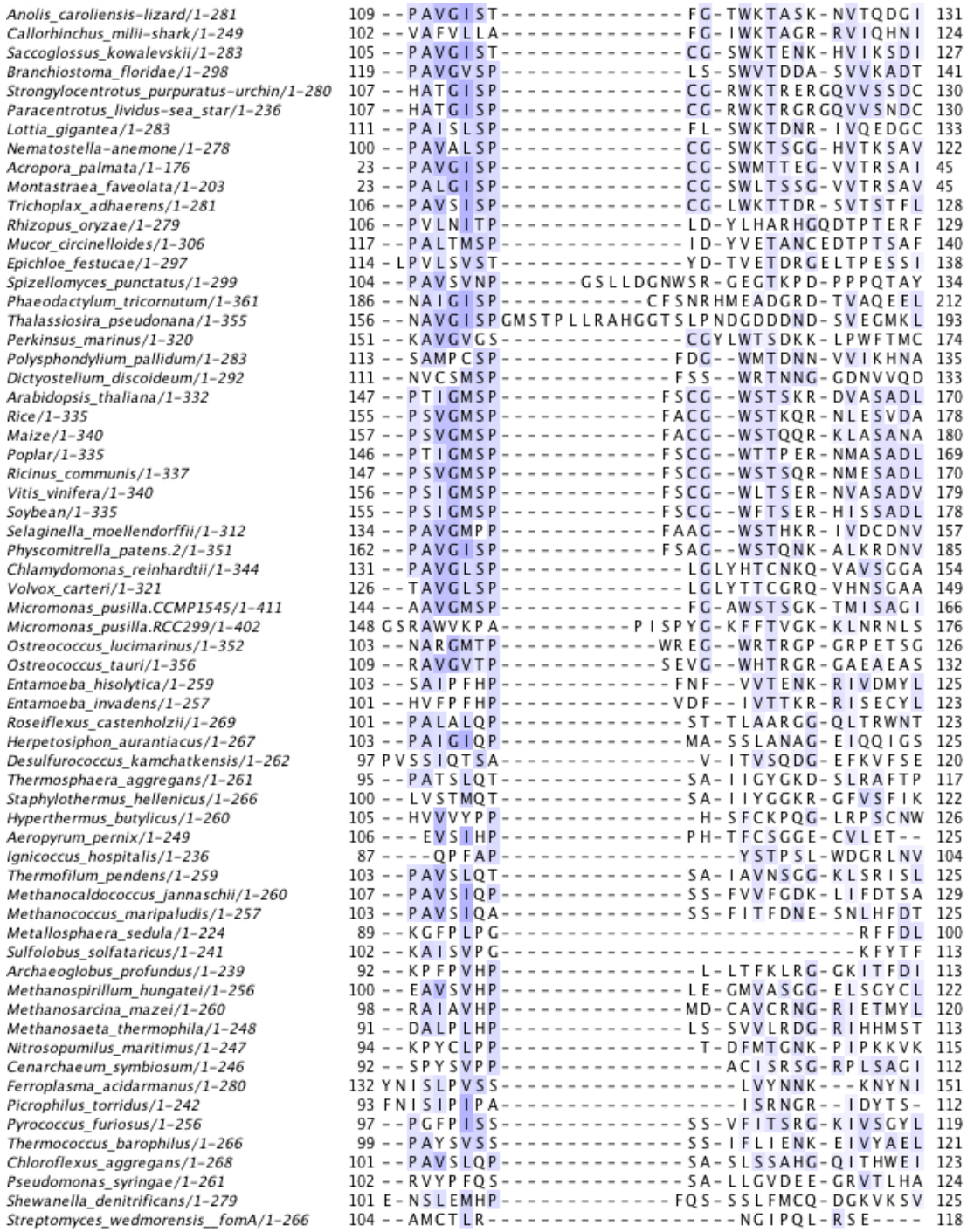


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

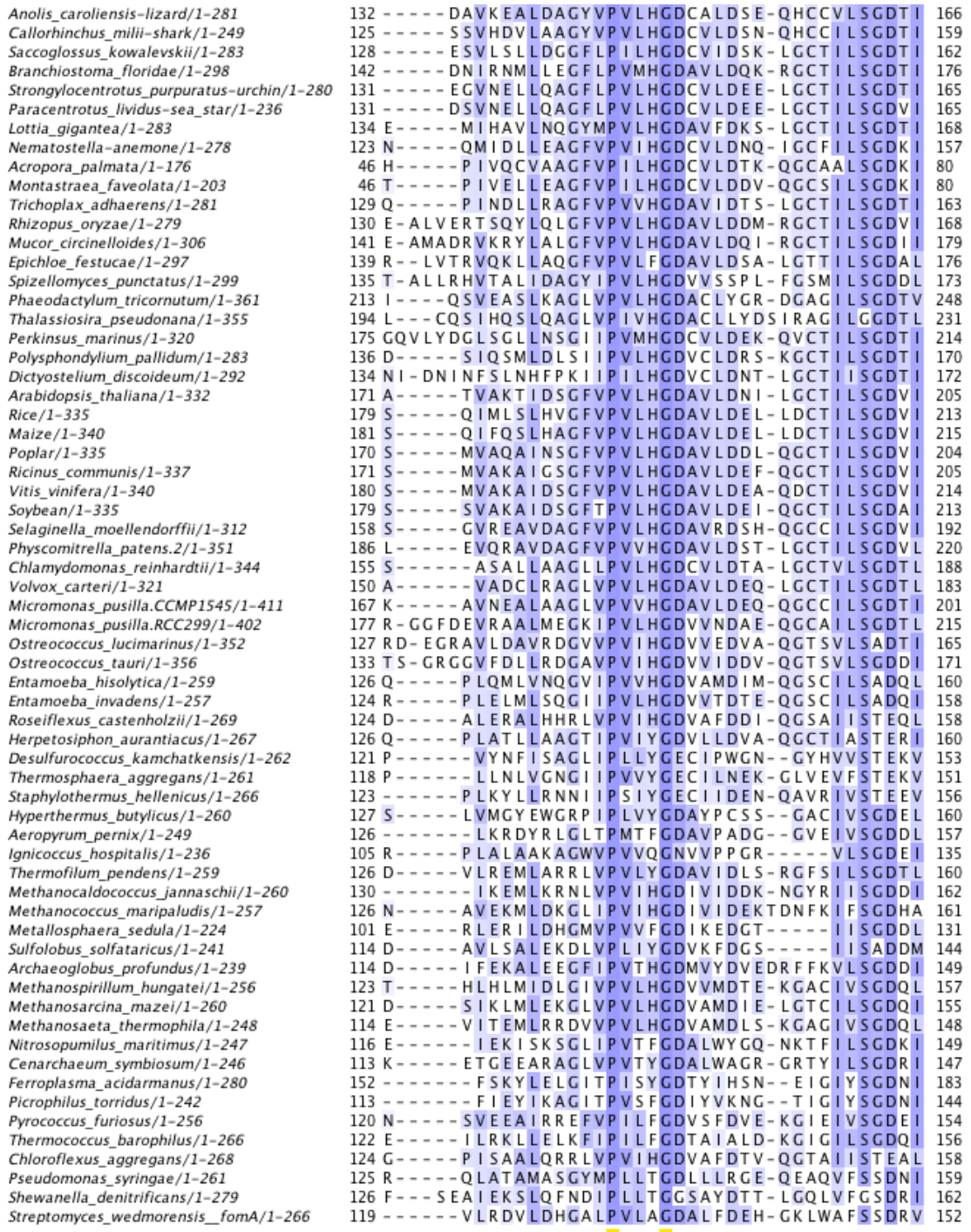


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

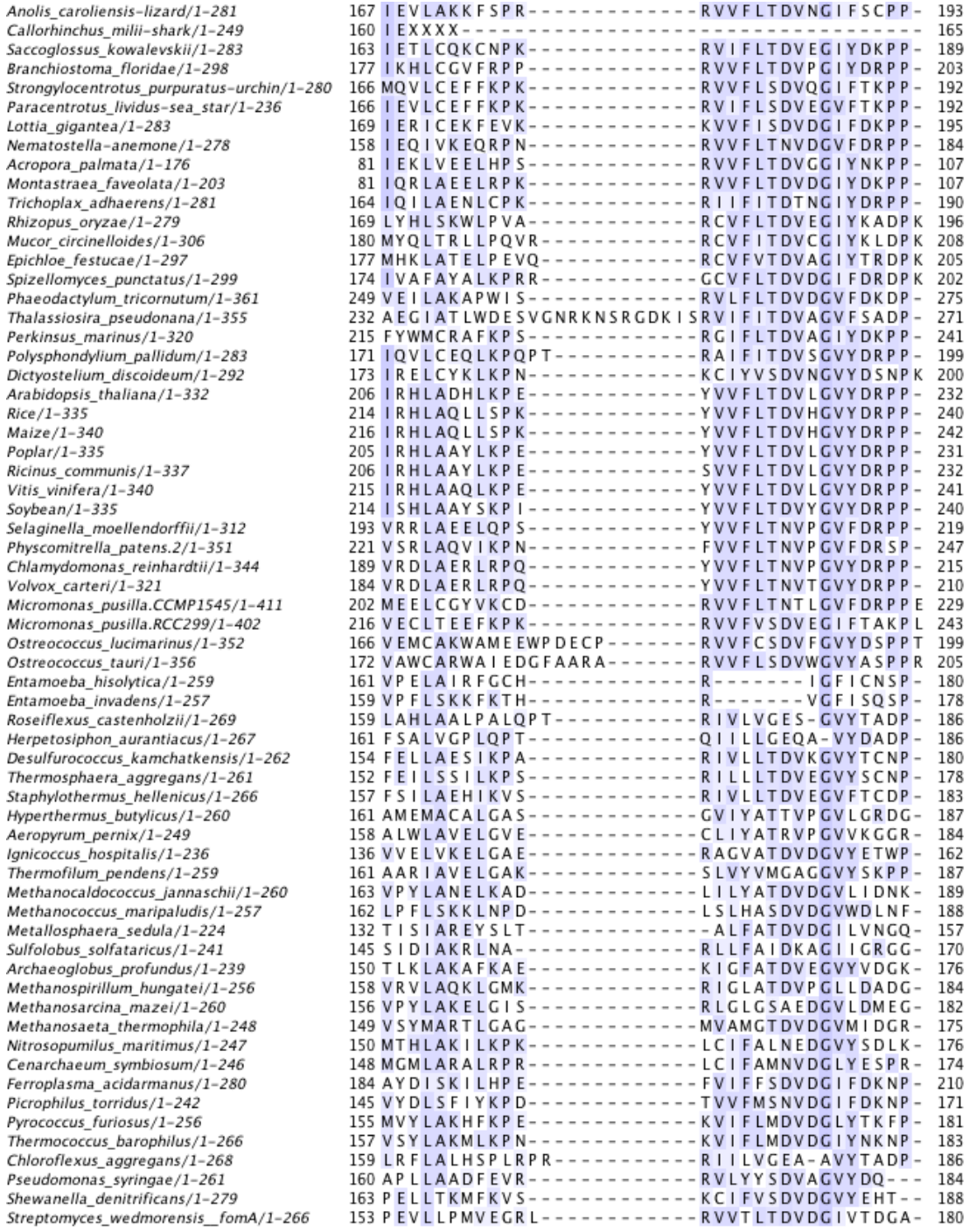


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

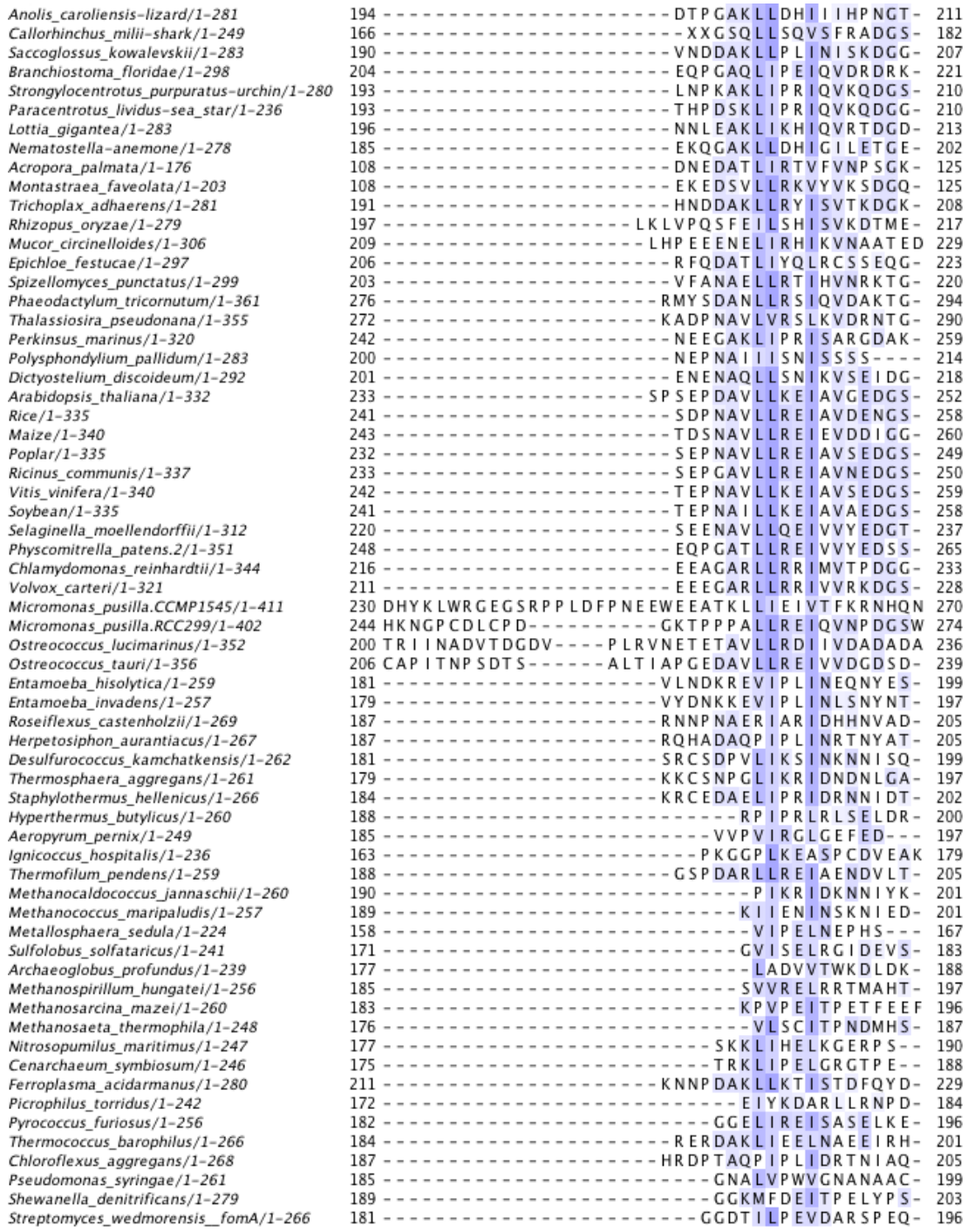


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

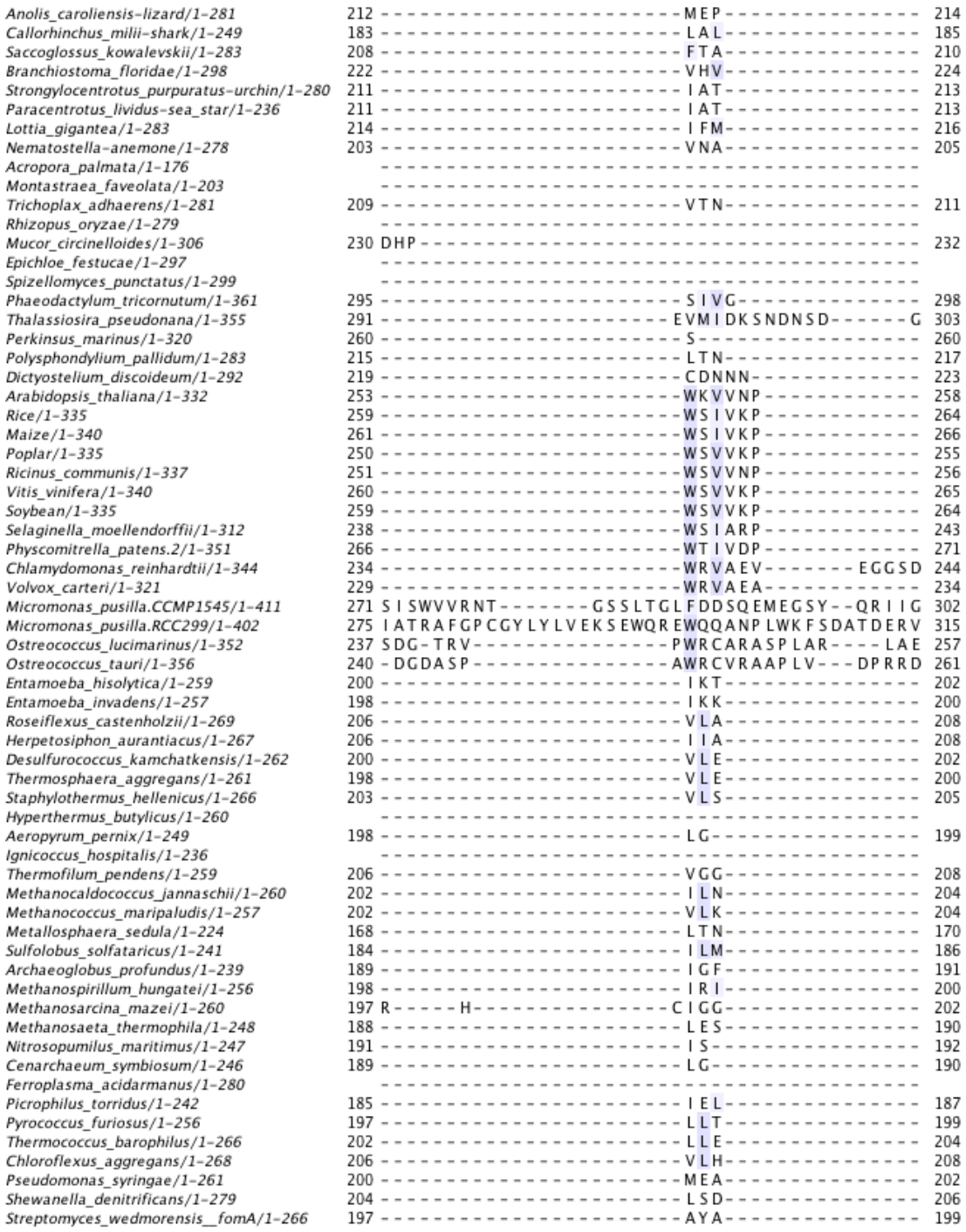


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

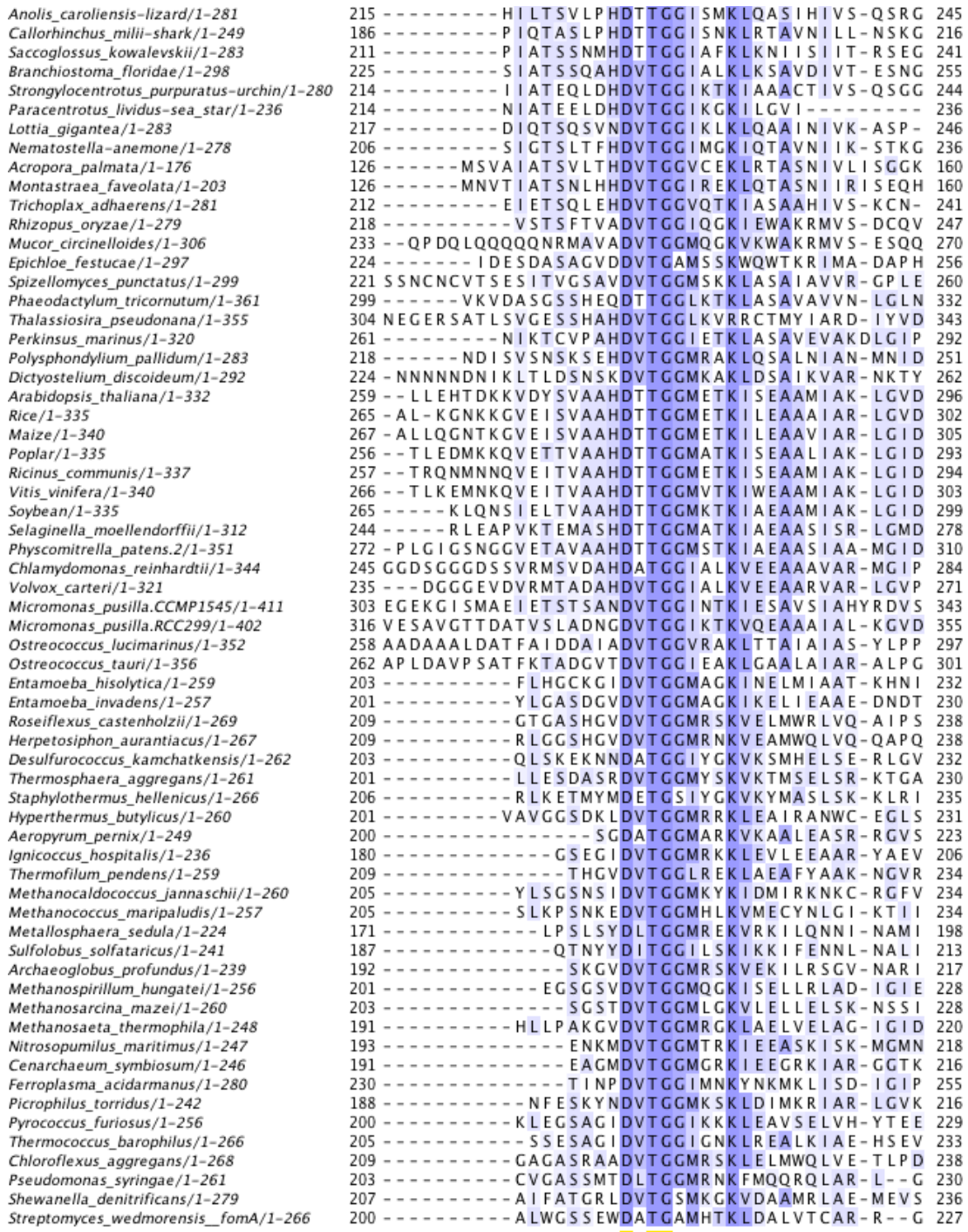


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

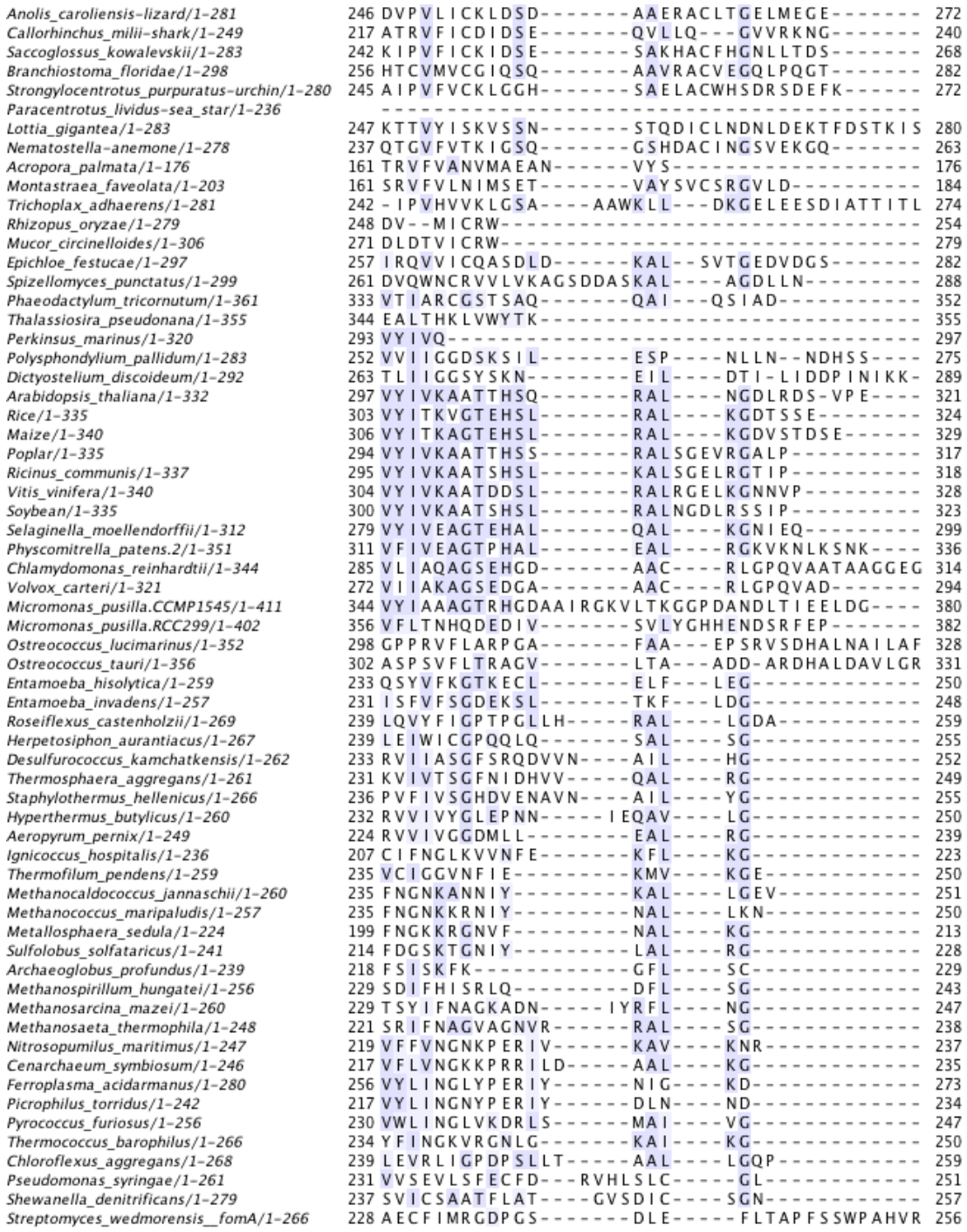


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

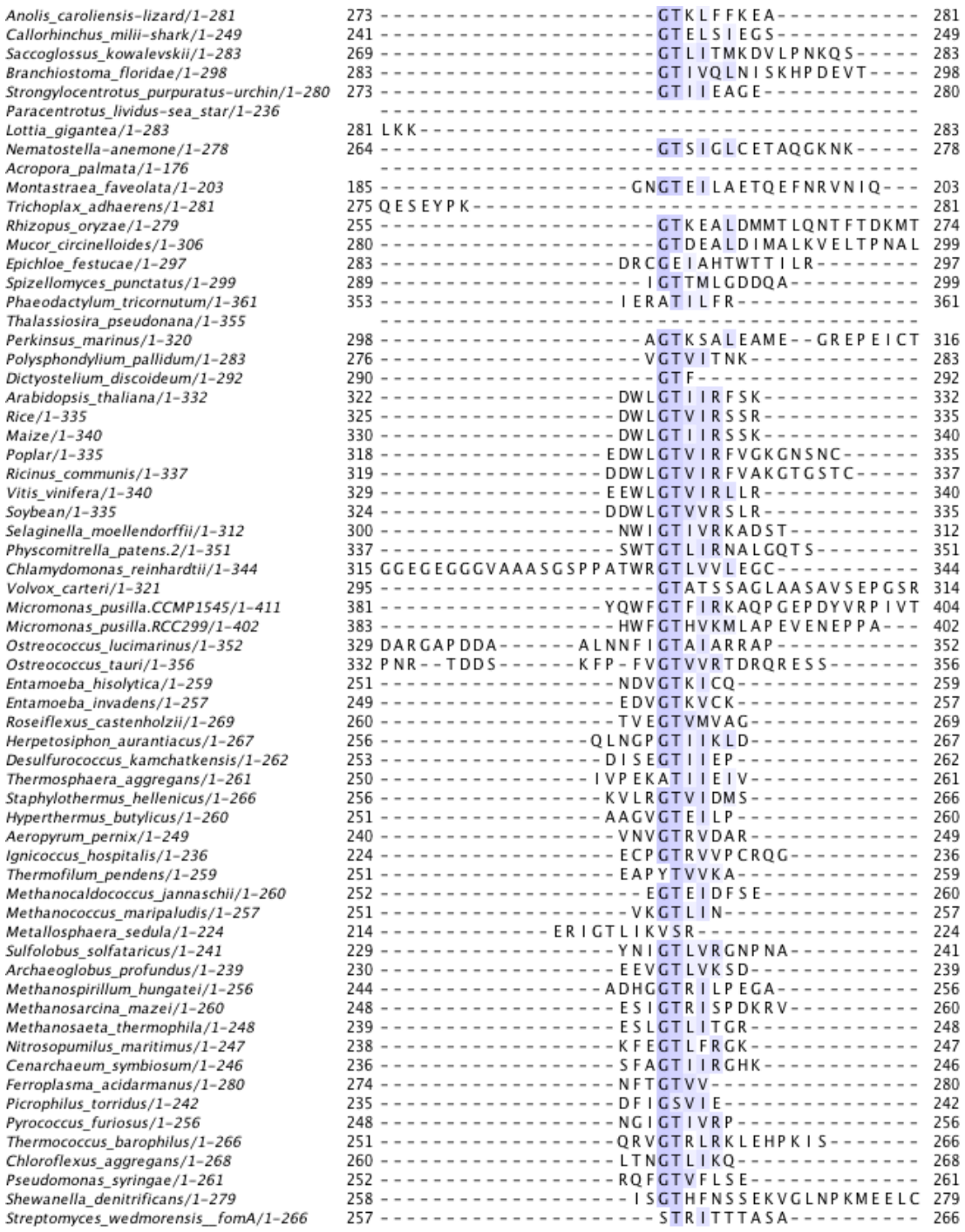


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

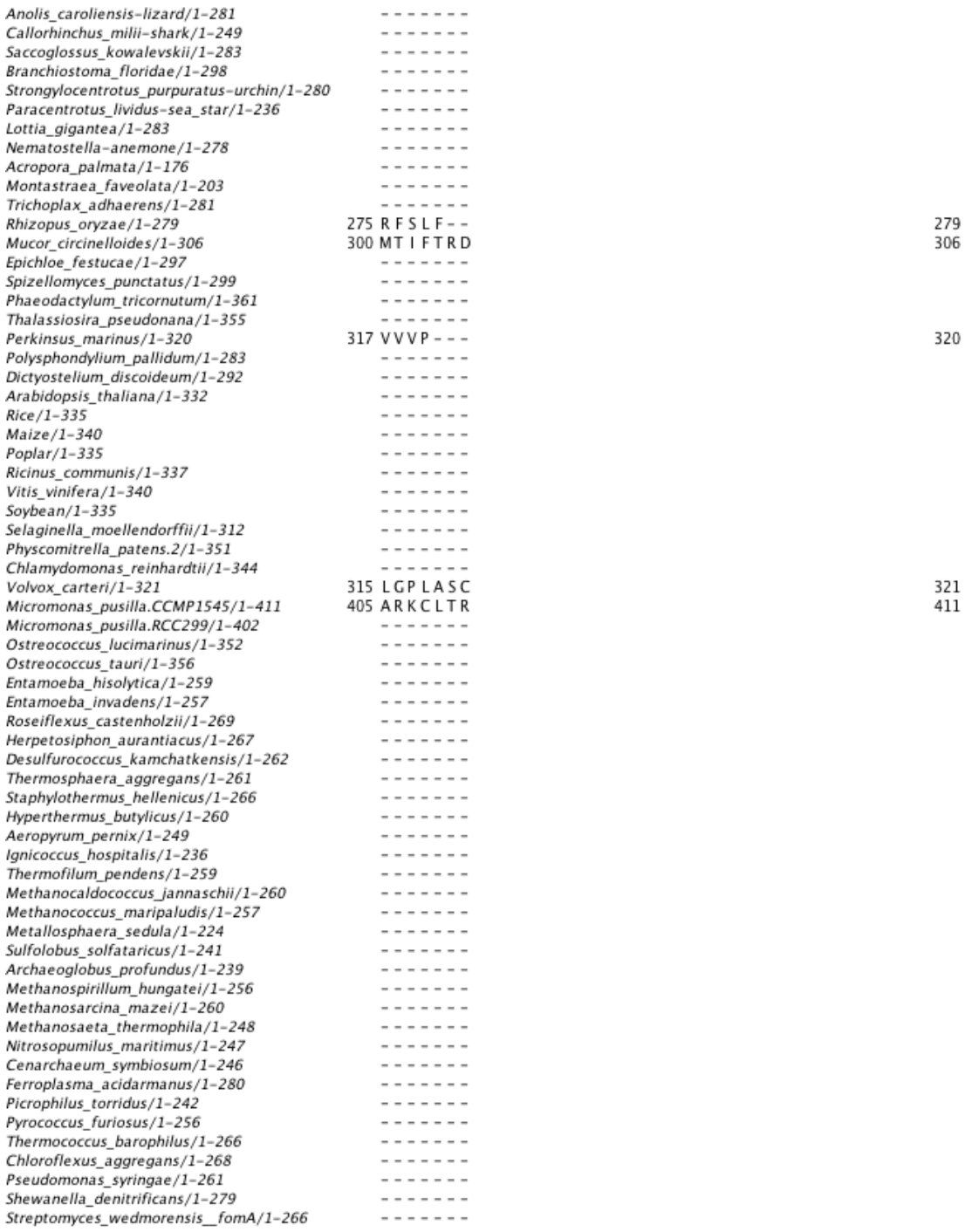


Figure 6.4. Alignment of IPKs from the three domains of life (cont).

ACKNOWLEDGEMENTS

The text of chapter 6, in part, has been submitted for publication of the material as it may appear in Chemical Communications, 2010, Dellas, Nikki; Manning, Gerard, Noel, Joseph P. I am the first author of this paper. Gerard Manning and Joseph P. Noel are the corresponding authors. I was responsible for all gene cloning, enzyme expression, purification, and kinetic characterization of IPK and its homologs. Gerard Manning was responsible for the bioinformatic and phylogenetic analysis of IPK and its homologs. All experiments were performed under the supervision of Joseph P. Noel.

REFERENCES

1. Smit, A.; Mushegian, A., Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome research* **2000,** *10* (10), 1468-1484.

2. Grochowski, L. L.; Xu, H.; White, R. H., Methanocaldococcus jannaschii uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *Journal of Bacteriology* **2006,** *188* (9), 3192-3198.

3. Lombard, J.; Moreira, D., Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol*.

4. Dellas, N.; Noel, J. P., Mutation of archaeal isopentenyl phosphate kinase highlights mechanism and guides phosphorylation of additional isoprenoid monophosphates. *ACS Chem Biol* **2010,** *5* (6), 589-601.

5. Pakhomova, S.; Bartlett, S. G.; Augustus, A.; Kuzuyama, T.; Newcomer, M. E., Crystal structure of fosfomycin resistance kinase FomA from Streptomyces wedmorensis. *The Journal of biological chemistry* **2008,** *283* (42), 28518-28526.

6. Nagegowda, D. A., Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Lett* **2010,** *584* (14), 2965-73.

7. Chen, M.; Poulter, C. D., Characterization of thermophilic archaeal isopentenyl phosphate kinases. *Biochemistry* **2010,** *49* (1), 207-17.

8. Birney, E.; Clamp, M.; Durbin, R., GeneWise and Genomewise. *Genome Res* **2004,** *14* (5), 988-95.

9. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **2004,** *32* (5), 1792-7.

10. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23* (21), 2947-8.

11. Waterhouse, A. M.; Procter, J. B.; Martin, D. M.; Clamp, M.; Barton, G. J., Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009,** *25* (9), 1189-91.

12. Boguski, M. S.; Lowe, T. M.; Tolstoshev, C. M., dbEST--database for "expressed sequence tags". *Nat Genet* **1993,** *4* (4), 332-3.

13. Birney, E.; Andrews, D.; Bevan, P.; Caccamo, M.; Cameron, G.; Chen, Y.; Clarke, L.; Coates, G.; Cox, T.; Cuff, J.; Curwen, V.; Cutts, T.; Down, T.; Durbin, R.; Eyras, E.; Fernandez-Suarez, X. M.; Gane, P.; Gibbins, B.; Gilbert, J.; Hammond, M.; Hotz, H.; Iyer, V.; Kahari, A.; Jekosch, K.; Kasprzyk, A.; Keefe, D.; Keenan, S.; Lehvaslaiho, H.; McVicker, G.; Melsopp, C.; Meidl, P.; Mongin, E.; Pettett, R.; Potter, S.; Proctor, G.; Rae, M.; Searle, S.; Slater, G.; Smedley, D.; Smith, J.; Spooner, W.; Stabenau, A.; Stalker, J.; Storey, R.; Ureta-Vidal, A.; Woodwark, C.; Clamp, M.; Hubbard, T., Ensembl 2004. *Nucleic Acids Res* **2004,** *32* (Database issue), D468-70.

Chapter 7

Conclusions

7.1. Overview

Isoprenoid biosynthesis constitutes an immensely diverse, highly branched network of pathways that spans both primary and secondary (specialized) metabolism in all organisms. The scope of this work includes two types of enzymes: terpene cyclases of secondary metabolism and isopentenyl phosphate kinase of primary metabolism.

Terpene cyclases are a fascinating class of enzymes that, based on structure, function, and some sequence motif conservation, are thought to have evolved from short-chain prenyl diphosphates of primary metabolism, which are responsible for the biosynthesis of GPP, FPP, or GGPP molecules that monoterpene, sesquiterpene, or diterpene synthases utilize as substrates for their electrophilic cyclization reactions, respectively. From a global perspective, this work demonstrates the adaptability of sesquiterpene cyclases to mutation without significant loss of function, but instead a gain of product promiscuity.[1] This research also shows how both substrate and product promiscuity are important from a structural perspective in terms of both substrate orientation and dynamics of the isoprenoid tail within the active site.[2]

Isopentenyl monophosphate kinase was originally thought to be solely important for archaeal isoprenoid biosynthesis.[3, 4] However structural and functional studies described here have allowed for the identification of a uniquely important residue within the enzyme active site.[5] This residue behaves as a marker to locate IPK homologs from other kingdoms, and successful characterization of these homologs proves that they are indeed true IPKs. These results imply the potential existence of a branched mevalonate pathway in Archaea and Eukarya.

7.2. Terpene synthases of specialized metabolism

Terpene cyclases constitute a class of enzymes that biosynthesize a chemically diverse profile of compounds known as terpenes. In this work, the study of several sesquiterpene synthases, including TEAS, HPS, and PAS, describes experimental findings associated with the structural, functional, and chemical properties of these enzymes and their small molecule products, which are all derived from the common substrate, FPP.

TEAS, whose major product is 5-epi aristolochene (5EA), can be converted to an HPS-like enzyme (termed "TEAS M9") that produces premnaspirodiene (PSD) as its major product by mutation of nine amino acids that are located in and around the TEAS active site.[6] On the initial mutational pathway towards M9, TEAS mutants display significant upregulation of a minor product, 4-epi eremophilene (4EE); the mechanism for its formation represents a hybrid of the 5EA mechanism and the PSD mechanism.[6] In order to fully characterize the catalytic landscape of the enzymes spanning the sequence space between TEAS and M9, a mutant library including all possible combinations of these nine mutations was created, as described in Chapter 2.[1] Although the catalytic landscape shows, on average, that the pathway towards the upregulation of another major product requires navigation through a promiscuous terrain, certain mutants bypass this terrain, demonstrating "jumps" to other products upon mutation of only one or two residues. From an evolutionary standpoint, these results cannot address the question of ancestry, that is, which enzyme (TEAS or HPS) came before the other. No individual mutations are found to control product specificity in one direction or the other; for example, no single mutation always upregulates a specific product, regardless of context. Instead, these mutations are very context dependent; that is, one amino acid mutation can contribute differently to phenotype in the context of its local environment. However, these results do, in part, support the theory that terpene cyclases were derived from a promiscuous

ancestor. In a selection of mutants, the observation of drastic product shifts accompanying single amino acid changes indicates that these cyclases have the ability to rapidly evolve a significantly different chemical profile with only a small change in sequence space. This ability could be a reflection of a sessile organism's approach toward environmental adaptability.

More recent discoveries suggest that not only product promiscuity, but also substrate promiscuity may play a role in controlling these interesting and often chemically complex terpene cyclase product profiles.[2] For example, the observation that FPP synthases can produce a certain percentage of *cis*-FPP in addition to its major product, *trans*-FPP, indicates the availability of an additional substrate for sesquiterpene cyclases.[7] Chapter 3 addresses this complex question in both a structural and a functional sense. Surprisingly, both *trans*-FPP and *cis*-FPP are substrates for TEAS, and each generates a *trans*-derived or *cis*-derived product spectrum including unique major products, 5EA and (+)-2-epi-prezizaene, respectively.[2] Functional comparisons between TEAS wt and TEAS M4 (a product-promiscuous mutant from the M9 library that produces equal amounts of 5EA, 4EE, and PSD) reveal that TEAS M4 is also more promiscuous than wild type when using *cis*-FPP as a substrate; this indicates that the level of product promiscuity, at least in this case, is independent of a *cis*- or *trans*-derived substrate. Comparisons of crystal structures of TEAS wt and TEAS M4 in complex with non-hydrolyzable substrate analogs 2F-FPP and cis-2F-FPP demonstrate that product promiscuity is directly related to dynamics of the isoprenoid chain in the active site. For example, the structures of TEAS M4 in complex with either 2F-FPP or *cis*-2F-FPP show significantly less electron density for the isoprenyl tail of the ligand compared with both TEAS wt structures.[2]

The product promiscuity of another terpene cyclase, patchoulol synthase (PAS) can also be altered in a rather surprisingly way: through mutations in an amino terminal region. The amino-terminal region is not thought to have a direct role in the terpene cyclase catalytic reaction. It is, however, thought to aid in active site capping to prevent premature release of a carbocation intermediates during the course of the reaction.[8-11] Mutation of both promiscuous and nonpromiscuous sesquiterpene cyclases reveals that certain cyclases, such as PAS, exhibit drastic product profile changes upon mutation at the RP motif in its amino terminal region, while other less promiscuous sesquiterpene cyclases exhibit little to no change. Although previous work speculates a general role for the amino-terminal region of these proteins,[8-11] this work articulates a direct role for this region, involving the RP motif. The series of mutations performed at both Arg and Pro of this motif in PAS suggest that while the Arg provides an anchor for the N-terminal tail through a salt bridge interaction with a C-terminal residue, the Pro provides the structural rigidity necessary to complete this task.

7.3. IPK of primary metabolism

7.3.1. Overview

Isopentenyl phosphate kinase (IPK) is an enzyme initially characterized from the thermophilic archaeon *M. jannaschii* that phosphorylates isopentenyl monophosphate (IP) to isopentenyl diphosphate (IPP).[4] IPP is one of two building blocks for all downstream isoprenoids, and it is therefore essential that its mechanism(s) of formation are understood among all three domains of life. IPK in particular was originally thought to be an enzyme exclusive to archaea, representing one of two enzymes required to complete the missing steps of the MVA pathway in this domain of life.[3, 4] In the classic mevalonate pathway, the last two steps leading to the production of IPP include genes encoding PMK and DPM-DC that

perform phosphorylation and decarboxylation of phosphomevalonate and diphosphomevalonate, respectively. In archaea, based on lack of evidence for these two orthologs and also the partial identification of an alternative route for the production of IPP, the reaction is thought to proceed in the reverse order, involving a decarboxylase followed by a kinase.[4] The kinase step is performed by IPK, and its structural and functional characterization, as discussed in chapter 5, allows for: 1) engineering of a deeper active site cavity for successful turnover of longer chained isoprenoid phosphates, and 2) the identification of an active site histidine residue that is unique to this member of the family and can therefore be used as a marker to identify IPK homologs from other kingdoms of life.[5]

7.3.2. Applications for IPK chain-length mutants

The goals behind engineering IPK to accept longer chained isoprenoid phosphates are two-fold: 1) to design a synthetic metabolic pathway, and 2) to synthesize isoprenoid diphosphate analogs. Over the past decade, there have been immense efforts on the front of MVA pathway upregulation, which has been accomplished through heterologous expression of MVA pathway enzymes in *E. coli* or upregulation in *S. cerevisiae.*[12-15] These efforts are geared towards the production of large quantities of terpenes, carotenoids, and other secondary metabolites that have, or can easily be derivatized into, compounds that have biological activity and medicinal value. Upregulation or overexpression of this pathway causes problems associated with metabolic flux,[14] production of unwanted byproducts (such as farnesol),[12] and feedback inhibition (such FPP inhibition of mevalonate kinase).[16] As discussed in chapter 5, IPK has been engineered to bind and turn over FP to FPP, which is extremely valuable towards the design of a much simpler synthetic metabolic pathway that would not suffer from any of the problems discussed above. For example, the overproduction of any given

sesquiterpene includes only three enzymatic steps: 1) phosphorylation of an inexpensive substrate such as farnesol (or an ester of farnesol) to FP, 2) phosphorylation of FP to FPP performed by the IPK chain-length mutant, and 3) cyclization by a sesquiterpene cyclase to the final sesquiterpene product. Although steps two and three are characterized, future work will address the design of a kinase that can phosphorylate isoprenoid alcohols of varying lengths.

Examples of chemoenzymatic synthesis of isoprenoid diphosphate analogs include the synthesis of fluorescent derivatives[17] and radiolabeled derivatives (through reaction of the substrate with radiolabeled ATP or ATP-γS), which would be useful for following *in vivo* prenylation or any other primary or secondary metabolic process involving isoprenoids.

7.3.3. Implications for active eukaryotic IPKs

The identification of His60 as a critical residue for binding and catalysis in IPK has been monumental in the location of IPK homologs in other kingdoms of life, as discussed in chapter 6.[5] This identification and successful characterization of eukaryotic IPKs has implications for the presence of an alternative MVA pathway in organisms that already have a fully functioning classic MVA pathway. Although IPK has a spotty distribution throughout the animal kingdom, the presence of IPK homologs in all green plants is an indication of its marked importance within this kingdom of life. In contrast to other kingdoms of life, isoprenoid biosynthesis within the plant kingdom is already very complex and includes both the DXP pathway, which operates in plant plastids, and the MVA pathway, which operates in the cytosol and/or other organelles.[18-20] Since compartmentalization is already a feature of isoprenoid biosynthesis in plants, it is tempting to speculate that a branched MVA pathway may allow for even further compartmentalization of certain enzymes within this pathway.

Since the DXP pathway and MVA pathway play different roles in plant isoprenoid biosynthesis (for example, GPP and GGPP are synthesized from the DXP pathway while FPP is synthesized from the MVA pathway), it is also possible that the branching mevalonate pathway directs the biosynthesis of specific primary or secondary metabolites, exerting yet another dimension of control over isoprenoid biosynthesis in the plant kingdom.

Ultimately, these hypotheses remain speculative until the missing piece to the alternative mevalonate pathway has been identified and characterized. This missing piece is the decarboxylase that catalyzes the first step after bifurcation from the classical pathway: the step that converts phosphomevalonate to isopentenyl monophosphate. One gene candidate that has been proposed to catalyze this reaction is the gene MJ0403 from *M. jannaschii,*[4] which is a putative dioxygenase that has sequence homology to LigAB and MEMO (mediator of ErbB2-driven cell motility). LigAB is a ring-cleaving extradiol dioxygenase that binds non-heme Fe^{2+} and plays a role in lignin degradation,[21] while MEMO is a human protein with homology to dioxygenases but no known catalytic function, and no experimental evidence demonstrating ability to bind a metal ion.[22] Although we have cloned and purified the MJ0403 putative decarboxylase, we have not been able to establish assay conditions that demonstrate successful turnover of phosphomevalonate to IP. Although efforts on assay optimization (including variation of metal ion type, metal ion concentration, and presence or absence of potential co-factors) are ongoing, the search for other decarboxylase candidates is under way.

REFERENCES

1. O'Maille, P. E.; Malone, A.; Dellas, N.; Andes Hess, B., Jr.; Smentek, L.; Sheehan, I.; Greenhagen, B. T.; Chappell, J.; Manning, G.; Noel, J. P., Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nature chemical biology* **2008,** *4* (10), 617-623.

2. Noel, J. P.; Dellas, N.; Faraldos, J. A.; Zhao, M.; Hess, B. A., Jr.; Smentek, L.; Coates, R. M.; O'Maille, P. E., Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. *ACS chemical biology* **2010,** *5* (4), 377-392.

3. Smit, A.; Mushegian, A., Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome research* **2000,** *10* (10), 1468-1484.

4. Grochowski, L. L.; Xu, H.; White, R. H., Methanocaldococcus jannaschii uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *Journal of Bacteriology* **2006,** *188* (9), 3192-3198.

5. Dellas, N.; Noel, J. P., Mutation of archaeal isopentenyl phosphate kinase highlights mechanism and guides phosphorylation of additional isoprenoid monophosphates. *ACS Chem Biol* **2010,** *5* (6), 589-601.

6. Greenhagen, B. T.; O'Maille, P. E.; Noel, J. P.; Chappell, J., Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (26), 9826-9831.

7. Thulasiram, H. V.; Poulter, C. D., Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity. *J Am Chem Soc* **2006,** *128* (49), 15819-23.

8. Whittington, D. A.; Wise, M. L.; Urbansky, M.; Coates, R. M.; Croteau, R. B.; Christianson, D. W., Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (24), 15375-15380.

9. Starks, C. M.; Back, K.; Chappell, J.; Noel, J. P., Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **1997,** *277* (5333), 1815-1820.

10. Hyatt, D. C.; Youn, B.; Zhao, Y.; Santhamma, B.; Coates, R. M.; Croteau, R. B.; Kang, C., Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (13), 5360-5365.

11. Little, D. B.; Croteau, R. B., Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases delta-selinene synthase and gamma-humulene synthase. *Archives of Biochemistry and Biophysics* **2002,** *402* (1), 120-135.

12. Asadollahi, M. A.; Maury, J.; Moller, K.; Nielsen, K. F.; Schalk, M.; Clark, A.; Nielsen, J., Production of plant sesquiterpenes in Saccharomyces cerevisiae: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol Bioeng* **2008,** *99* (3), 666-77.

13. Ohto, C.; Muramatsu, M.; Obata, S.; Sakuradani, E.; Shimizu, S., Overexpression of the gene encoding HMG-CoA reductase in Saccharomyces cerevisiae for production of prenyl alcohols. *Appl Microbiol Biotechnol* **2009,** *82* (5), 837-45.

14. Martin, V. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D., Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature biotechnology* **2003,** *21* (7), 796-802.

15. Pitera, D. J.; Paddon, C. J.; Newman, J. D.; Keasling, J. D., Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metabolic engineering* **2007,** *9* (2), 193-207.

16. Fu, Z.; Voynova, N. E.; Herdendorf, T. J.; Miziorko, H. M.; Kim, J. J., Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism. *Biochemistry* **2008,** *47* (12), 3715-24.

17. Hovlid, M. L.; Edelstein, R. L.; Henry, O.; Ochocki, J.; DeGraw, A.; Lenevich, S.; Talbot, T.; Young, V. G.; Hruza, A. W.; Lopez-Gallego, F.; Labello, N. P.; Strickland, C. L.; Schmidt-Dannert, C.; Distefano, M. D., Synthesis, properties, and applications of diazotrifluropropanoyl-containing photoactive analogs of farnesyl diphosphate containing modified linkages for enhanced stability. *Chemical biology & drug design* **2010,** *75* (1), 51-67.

18. Sapir-Mir, M.; Mett, A.; Belausov, E.; Tal-Meshulam, S.; Frydman, A.; Gidoni, D.; Eyal, Y., Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. *Plant Physiol* **2008,** *148* (3), 1219-28.

19. Carrero-Lerida, J.; Perez-Moreno, G.; Castillo-Acosta, V. M.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D., Intracellular location of the early steps of the isoprenoid biosynthetic pathway in the trypanosomatids Leishmania major and Trypanosoma brucei. *Int J Parasitol* **2009,** *39* (3), 307-14.

20. Hartman, I. Z.; Liu, P.; Zehmer, J. K.; Luby-Phelps, K.; Jo, Y.; Anderson, R. G.; DeBose-Boyd, R. A., Sterol-induced dislocation of 3-hydroxy-3-methylglutaryl coenzyme A reductase from endoplasmic reticulum membranes into the cytosol through a subcellular compartment resembling lipid droplets. *J Biol Chem* **2010,** *285* (25), 19288-98.

21. Sugimoto, K.; Senda, T.; Aoshima, H.; Masai, E.; Fukuda, M.; Mitsui, Y., Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. *Structure* **1999,** *7* (8), 953-65.

22. Qiu, C.; Lienhard, S.; Hynes, N. E.; Badache, A.; Leahy, D. J., Memo is homologous to nonheme iron dioxygenases and binds an ErbB2-derived phosphopeptide in its vestigial active site. *J Biol Chem* **2008,** *283* (5), 2734-40.