STRUCTURE, FUNCTION AND APPLICATIONS OF METAL-REQUIRING ENZYMES: CARBONIC ANHYDRASE AND EPI-ISOZIZAEENE SYNTHASE

JULIE ANNE AARON

A Dissertation in

Chemistry

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

2010

____
Dr. David W. Christianson, Professor, Chemistry
Supervisor of Dissertation

____
Dr. Gary A. Molander, Professor, Chemistry
Department Chairman

Dissertation Committee

Dr. Ronen Marmorstein, Professor, Chemistry and Wistar Institute

Dr. Ivan J. Dmochowski, Associate Professor, Chemistry

Dr. Barry S. Cooperman, Professor, Chemistry

“One must still have chaos in oneself to be able to give birth to a dancing star.”
- Nietzsche

I dedicate this thesis to my family,
for supporting me,
and to Dr. Li and Mr. Doketis,
for inspiring me.

Acknowledgements

My graduate experience has been more rewarding then I ever imagined. I am deeply indebted to my mentors, colleagues, friends, and family who have supported, encouraged, and celebrated with me these past 5 years.

First and foremost, I would like to thank my dissertation advisor, Dr. David Christianson. David is an excellent mentor, both personally and professionally. I admire his dedication and enthusiasm for teaching, and am grateful for the opportunity to work in his lab. The balance of support and scientific independence with which he provided me was essential to my success. I would also like to extend my greatest appreciation to the members of my dissertation committee, Dr. Ronen Marmorstein, Dr. Ivan Dmochowski and Dr. Barry Cooperman for their advice and guidance. I would especially like to acknowledge Dr. Marmorstein, Director of the Chemistry-Biology Interface (CBI) Training Program, for his enthusiastic support of the CBI Program and Chemical Biophysics Mini-Symposia Series (CBMS).

I have had the privilege of working in a lab filled with incredibly supportive and engaging colleagues, with whom I have enjoyed friendships, both inside and outside the lab. I have especially enjoyed our camaraderie during countless synchrotron trips, where despite the lack of sleep, we managed to have a lot of fun. I would like to especially thank Dr. Luigi Di Costanzo, and Dr. Heather Gennadios who served as my mentors when I first joined the lab, and Dr. Dan Dowling, Dr. Kathryn Cole and Kate Thorn for

their friendship and great chats over coffee/tea over the years. I would also like to thank the other members of our lab, Dr. German Gomez, Dr. Hyunshun Shin, Dr. Kevin Jude, Dr. Sangeetha Vedula, Dr. Katya Shishova, Dr. Tatiana Zakharian, Dr. Monica Ilies, Dr. Mustafa Koksal Dr. Mo Chen, Patrick Lombardi and Cristina Virgilio for their friendship and advice over the years.

The work I present herein is the result of two excellent collaborations. First, I would like to thank Dr. Aru Hill and Jenny Chambers of Dr. Dmochowski's Lab at the University of Pennsylvania. Jenny Chambers is one of the most beautiful people I have ever met, inside and out, and Dr. Aru Hill, is the smartest graduate student I have had the privilege of knowing. I am very grateful to them for the success of the CAII-cryptophane project. Secondly, my thanks go to Dr. David Cane at Brown University. I am grateful for the opportunity to twice visit his labs in Providence, to perform experiments with Dr. Xin Lin and Dr. Wayne Chou on the *ep*i-isozizaene synthase project.

I am indebted to members of the Departments of Chemistry, and Biochemistry & Biophysics Departments for their financial and administrative support of my extracurricular endeavors. The success of the CBI and CBMS symposia would not have been possible without the gracious support of Matt Lane, Cheryl McFadden, Angie Young, Dr. Mike Brent, Candice Adams and Yvonne Kline. I also give thanks to other members of the Department of Chemistry; Andrea Carroll, Christine Zay and Mandy Swope and especially Judith Currano.

As a protein crystallographer, I had the unique opportunity to take numerous synchrotron data collection trips, where I not only collected data, but also learned extensively about new techniques and software. I thank the Cornell High Energy Synchrotron Source, the National Light Source at Brookhaven National Laboratories, and the Advanced Photon Source at Argonne National Laboratories, especially the beam line scientists at NE-CAT for access to X-ray facilities. This work was supported by NIH grants GM49758, GM56838 and a NIH CBI Training Grant.

I would like to thank the members of my graduate class, with whom I have had a lot of fun these past 5 years. Firstly I would like to thank Dr. Dan Himmelberger, my closest friend, for his love, support and patience. Thanks to Sara (Nichols) Hayik and Ian Farrell, whom together with Dan, have been my lunch group for the past 4 years – these are some of my best memories. My thanks go to Deidre Sandrock, Diana (Cabral) Challen, Maria (Gilleece) Bednar, Jenny (Muth) Chambers, Alice Chong, Andre Isaacs, Belgin Canturk and Ken Lassen – what a fantastic group of students to start with at Penn. I would also like to thank Ariane Perez-Gavilian, Liz Roesser, Brandon Kelley, Danielle Reifsnyder, Rob Hickey, Ben Dyme, Dave Perez, and especially Emily Berkeley for their friendship.

I would also like to thank my girlfriends from Canada who have been nothing but supportive, although they never really understood what I was up to; Dr. Sabrina Akhtar, Dr. Kristina Powles, Dr. Hinal Sheth and Dr. Monica Hau.

Finally, I would like to thank my family for their incredible support. First, to my New Jersey family-away-from-home, a special thanks to Aunt Judy and Uncle Mike Marinelli and Aunt Char and Uncle Phil McMorris who always treated me as one of their own. To my grandparents, Helen Freund (Gram), and Dorothy and Charlie Aaron (Nan & Pop) for their love, encouragement and good genes. Lastly, to my siblings, Jessie and Dave, and my parents Blanche and Dave, who always believed in me, often more than I did in my self, thank you for your love, I know I have made you proud.

ABSTRACT

STRUCTURE, FUNCTION, AND APPLICATIONS OF METAL-REQUIRING ENZYMES: CARBONIC ANHYDRASE AND EPI-ISOZIZAENE SYNTHASE

Julie Anne Aaron

David W. Christianson

Cryptophane Biosensors for Targeting Human Carbonic Anhydrase

Cryptophanes represent an exciting class of xenon-encapsulating molecules that can be exploited as probes for nuclear magnetic resonance imaging. A series of carbonic anhydrase-targeting, xenon-binding cryptophane biosensors were designed and synthesized. Isothermal titration calorimery and surface plasmon resonance measurements confirmed nanomolar affinity between human carbonic anhydrase II and the cryptophane biosensors. A 1.70 Å resolution crystal structure of a cryptophane-derivatized benezenesulfonamide human carbonic anhydrase II complex was determined, and shows how an encapsulated xenon atom can be directed to a specific biological target. Furthermore, this work illustrates the utility and promise of developing xenon biosensors to diagnose human diseases characterized by the upregulation of specific carbonic anhydrase biomarkers, specifically human carbonic anhydrase IX and XII.

Structural Studies of epi-Isozizaene Synthase from Streptomyces coelicolor

The X-ray crystal structure of recombinant *epi*-isozizaene synthase (EIZS), a sesquiterpene cyclase from *Streptomyces coelicolor* A3(2), has been determined at 1.60 Å resolution. Specifically, the structure of wild-type EIZS is that of its closed conformation in complex with three Mg^{2+} ions, inorganic pyrophosphate (PP_i), and the benzyltriethylammonium cation (BTAC). Additionally, the structure of D99N EIZS has been determined in an open, ligand-free conformation at 1.90 Å resolution. Comparison of these two structures provides the first view of conformational changes required for substrate binding and catalysis in a bacterial terpenoid cyclase, and enables a comparison of substrate recognition amongst terpenoid synthases from different domains of life. Mutagenesis of aromatic residues in the enzyme active site alters the cyclization template and results in the production of alternative sesquiterpene products. The structure and activity of several active site mutants have been explored. The 1.64 Å resolution crystal structure of F198A EIZS in a complex with three Mg^{2+} ions, PP_i, and BTAC reveals an alternative binding orientation of BTAC, whereas the crystal structures of L72V, A236G and V329A EIZS reveal an unchanged BTAC binding orientation. Alternative binding orientations of a carbocation intermediate could lead to the formation of alternative products.

Table of Contents

Dedication ii

Acknowledgements iii

Abstract vii

Table of Contents ix

List of Tables xiii

List of Figures xiv

Part I: Cryptophane Biosensors for Targeting Human Carbonic Anhydrase

Chapter 1: Introduction

1.1 Carbonic Anhydrase: A Model System 1

1.2 ^{129}Xe MRI 7

1.3 ^{129}Xe-Cryptophane Biosensors 8

Chapter 2: Binding Studies of ^{129}Xe-Cryptophane Biosensors and Carbonic Anhydrase

2.1 Design and Synthesis of ^{129}Xe-Cryptophane Biosensors 11

2.2 Isothermal Titration Calorimetry Measurements
  2.2.1 Introduction 14
  2.2.2 Experimental Methods 15
  2.2.3 Results and Discussion 16

2.3 Surface Plasmon Resonance Measurements
  2.3.1 Introduction 20
  2.3.2 Experimental Methods 21
  2.3.3 Results and Discussion 22

Chapter 3: Structural Studies of ^{129}Xe-Cryptophane-CAII

3.1 Experimental Methods 25

3.2 Results and Discussion 29

3.3 Future Applications of Cryptophane-based CA Biosensors 38

References, Part I 41

Part II: Structural and Functional Studies of the Sesquiterpene Cyclase *epi*-Isozizaene Synthase

Chapter 4: Introduction

4.1 Terpenes and Terpene Synthases 50

4.2 *Streptomyces* 55

4.3 *epi*-Isozizaene 56

Chapter 5: X-Ray Crystal Structure of *epi*-Isozizaene Synthase from *Streptomyces coelicolor*

5.1 Expression and Purification 65

5.2 Crystallization 68

5.3 Structure Determination with Heavy Atoms
    5.3.1 Introduction 70
    5.3.2 Results 72

5.4 Structure of EIZS-Mg^{2+}-PP_i-BTAC Complex 80

5.5 Structure of EIZS-$Hg^{2+}{}_4$ Complex 83

Chapter 6: X-ray Crystal Structure of D99N EIZS and Implications for Substrate Recognition

6.1 Introduction 88

6.2 Experimental Methods

6.2.1 Site-Directed Mutagenesis, Expression and Purification 89
6.2.2 Crystallization and Structure Determination 90

6.3 Structure of D99N EIZS and Implications for Substrate Recognition 93

Chapter 7: Structural and Biochemical Studies of the Active Site of EIZS

7.1 Introduction 101

7.2 Experimental Methods

7.2.1 Site-Directed Mutagenesis, Expression and Purification 103

7.2.2 Crystallization and Structure Determination of EIZS Active Site Mutants

7.2.2.1 F198A EIZS 106
7.2.2.2 L72V EIZS 106
7.2.2.3 A236G EIZS 107
7.2.2.4 V329A EIZS 107
7.2.3 Radioactive Substrate Kinetic Assay of EIZS Mutants 110
7.2.4 GC-MS Analysis of Product Arrays Generated by EIZS Mutants 111

7.3 Results

7.3.1 Radioactive Substrate Kinetic Assay 111
7.3.2 GC-MS Analysis 116

7.4 Crystal Structures of Mutant EIZS

7.4.1 F198A EIZS-Mg^{2+}_3-PP_i-BTAC complex 125
7.4.2 L72V, A236G and V329A EIZS- Mg^{2+}_3-PP_i-BTAC complexes 128

7.5 Discussion 130

Chapter 8: Trinuclear Metal Clusters in Catalysis by Terpenoid Synthases

8.1 Introduction 131

8.2 Isoprenoid Coupling Enzymes

8.2.1 Farnesyl Diphosphate Synthase 132
8.2.2 Geranylgeranyl Diphosphate Synthase 137
8.2.3 Nonspecific Prenyl Synthase 137

8.3 Isoprenoid Cyclization Enzymes

8.3.1 Fungal Cyclases 138
8.3.2 Bacterial Cyclases 143
8.3.3 Plant Cyclases 145

8.4 Discussion 151

Chapter 9: Future Directions 155

References, Part II 158

List of Tables

Table 1.1. Catalytic steady-state constants and protein data bank reference codes for human carbonic anhydrase isozymes. 3

Table 2.1. Summary of dissociation constants for biosensors-carbonic anhydrase complexation at 298 K. 17

Table 2.2. Dissociation constant determined by surface plasmon resonance for biosensors **7**-**9** with human CA II. 23

Table 3.1. Data collection and refinement statistics for CA II-**9**-Xe complex. 28

Table 5.1. Data collection statistics for native and mercury derviatized EIZS crystals. 75

Table 5.2. Refinement statistics for wild-type EIZS complexes. 79

Table 6.1. Refinement statistics for D99N EIZS. 92

Table 7.1. EIZS mutagenic primer sequences. 105

Table 7.2. EIZS active site mutant data collection and refinement statistics. 109

Table 7.3. Steady-state kinetic parameters for wild-type EIZS and site-specific mutants. 115

Table 7.4. Distribution of sesquiterpene products from wild-type EIZS and site-specific mutants. 122

List of Figures

Figure 1.1. A cartoon representation of the crystal structure of CA II. 4

Figure 1.2. Catalytic mechanism of carbonic anhydrase. 5

Figure 1.3. General structures of cryptophanes. 10

Figure 2.1. Cryptophane biosensor synthesis. 13

Figure 2.2. Isothermal calorimetric data for the interactions of **4** and **7** with CA I and CA II. 18

Figure 2.3. Isothermal calorimetric data for the interactions of **8** and **9** with CA I and CA II. 19

Figure 2.4. Surface plasmon resonance sensorgrams for the interaction between biosensors **7**-**9** and human CA II. 24

Figure 3.1. Crystals of Biosensor-**9**-CA II. 27

Figure 3.2. Anomalous Fourier map showing Xe location at opening of CA II active site. 31

Figure 3.3. Anomalous Fourier map showing Xe location in CA II hydrophobic pocket. 32

Figure 3.4. The M_oM_o and P_oP_o enantiomers of the cryptophane-A-derived CA biosensor. 35

Figure 3.5. The crystal structure of biosensor **9** bound to CA II. 36

Figure 3.6. The unit cell of CA II-**9**-Xe complex. 37

Figure 4.1. General scheme of terpenoid nomenclature and biosynthesis 52

Figure 4.2. Structural similarities among various terpenoid synthases. 54

Figure 4.3. The structures of zizaene sesquiterpenes. 59

Figure 4.4. Proposed mechanism for the formation of *epi*-Isozizane from FPP by EIZS. 60

Figure 4.5. Proposed *epi*-isozizaene cyclization scheme based on quantum chemical calculations 61

Figure 4.6. Albaflavenone biosynthetic pathway in *S. coelicolor*. 64

Figure 5.1. SDS-PAGE analysis of the purification of epi-isozizaene synthase. 67

Figure 5.2. Crystals of epi-isozizaene synthase. 69

Figure 5.3. Diffraction pattern of methylmercuric acetate derivatized EIZS crystals. 74

Figure 5.4. Ramachandran plot of the refined EIZS-$Mg^{2+}{}_3$-PP_i-BTAC structure. 78

Figure 5.5. Ribbon plot of the EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex, and the structure of BTAC. 81

Figure 5.6. Particle size distribution of EIZS sample, measured by dynamic light scattering. 82

Figure 5.7. Active site of EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex. 86

Figure 5.8. Structural changes between EIZS $Mg^{2+}{}_3$-PP_i-BTAC and $Hg^{2+}{}_4$ complexes. 87

Figure 6.1. Crystal packing of WT and D99N EIZS. 95

Figure 6.2. Structure of D99N EIZS. 97

Figure 6.3. Conservation of $Mg^{2+}{}_3$-PP_i and -diphosphate binding motifs among bacterial and fungal terpenoid cyclases. 100

Figure 7.1. Initial rate versus substrate concentration for the reaction of WT and aromatic mutant EIZS with FPP. 113

Figure 7.2. Initial rate versus substrate concentration for the reaction of aliphatic mutant EIZS with FPP. 114

Figure 7.3. Gas chromatographs of organic products of F96A, F198A, and W203F EIZS. 117

Figure 7.4. Mass spectra of identified sesquiterpene products of F96A, F198A, and W203F EIZS. 118

Figure 7.5. Mass spectra of unidentified sesquiterpene products of F96A, F198A, and W203F EIZS, and *epi*-Isozizaene. 121

Figure 7.6. Proposed cyclization cascade for observed products of WT and mutant EIZS. 123

Figure 7.7. Active site contour of WT and F198A EIZS. 126

Figure 7.8. Stereoview of the active site of F198A EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex. 127

Figure 7.9. Stereoview of an overlay of the active sites of WT, L72V, A236G and V329L EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complexes. 129

Figure 8.1. Conservation of $Mg^{2+}{}_3$-PP_i and -diphosphate binding motifs among isoprenoid coupling enzymes. 136

Figure 8.2. Conservation of $Mg^{2+}{}_3$-PP_i and -diphosphate binding motifs among terpenoid cyclases. 142

Figure 8.3. The diphosphate binding site of (+)-δ-cadinene synthase from *G. arboreum.* 150

Figure 8.4. Stereoview of the $Mg^{2+}{}_3$-PP_i cluster from *epi*-Isozizaene synthase. 154

Part I: Cryptophane Biosensors for Targeting Human Carbonic Anhydrase

Chapter 1: Introduction

1.1 Carbonic Anhydrase: A Model System

Carbonic anhydrases (CAs) are ubiquitous zinc-metalloenzymes that catalyze the hydration of carbon dioxide to form bicarbonate (Equation 1). Four distinct, unrelated CA gene families have been identified and characterized as follows; α-CAs found in vertebrates, bacteria, algae, and the cytoplasm of green plants, the β-CAs found in bacteria, algae, and chloroplasts of mono- and dicotyledons, γ-CAs found in archaea and some bacteria, and finally, the δ-CAs found in some marine diatoms (Supuran, 2008). There are 16 α-CA isozymes in mammals, with various sub-cellular localization including the cytosol, membrane-bound, mitochondrial, transmembrane, and secreted (Table 1.1) (Supuran, 2007). Not only are these enzymes important for pH homeostasis, they are also involved in a number of other biosynthetic pathways including gluconeogeneis, ureagenesis and lipogenesis. As such, CAs have been implicated in cellular proliferation, spermatozoan motility, and aqueous humor production, and are therefore targets to treat many diseases (Ekstedt, 2004; Pastorekova, 2006; Mincione, 2008).

$$CO_2 + H_2O \Leftrightarrow HCO_3^- + H^+ \qquad (1)$$

The X-ray crystal structures of several human CA isozymes have been solved to-date, including CA I, II, III, IV, VI, VIII, IX, XII and XIII (Table 1.1). The molecular structures of human CAs are highly conserved. The structure is characterized by a central anti-parallel β-sheet, and the active site contains a catalytically essential Zn^{2+} ion, coordinated by three histidine residues (Figure 1.1). The active site is located at the bottom of a roughly conical, 15 Å-deep cleft, which is predominantly hydrophobic one one side and predominately hydrophilic on the other (Liljas, 1972).

Kinetic experiments have helped to elucidate the catalytic mechanism of the enzyme (Figure 1.2). A zinc-bound water molecule makes a hydrogen bond to the hydroxyl moiety of T199. The pKa of the zinc-bound water is lowered to 7, and therefore can lose a proton and become a zinc-bound hydroxide. The hydroxide, a strong nucleophile, is well oriented to attack a CO_2 molecule bound in a nearby hydrophobic pocket (Figure 1.1), forming a zinc-bound bicarbonate. Next, the bicarbonate is displaced by a water molecule and diffuses into bulk solution. The active form of the enzyme is reformed by loss of proton from the new zinc-bound water via proton transfer to bulk solvent, assisted by H64.

The most effective CA inhibitors designed to date contain an aryl-sulfonamide moiety; the sulfonamide moiety coordinates to the Zn^{2+} ion as a sulfonamidate ion, and the aryl moiety binds in the hydrophobic cleft and acts as a scaffold from which the inhibitor can be tailored to target a specific isozyme (Krishnamurthy, 2008). Hundreds of CA inhibitors have been designed and tested as potential drugs for the treatment of a variety of diseases including glaucoma (CA II and CA XII), cancer (CA IX and CA XII), obesity (CA VA, CA VB), seizures (CA II, CA VII, CA XII, CA XIV), and bacterial infections (various CAs from pathogenic organisms) (Supuran, 2007).

Table 1.1. Catalytic steady-state constants[1] and protein data bank reference codes for human carbonic anhydrase isozymes.

Isozyme	**k_{cat} (s^{-1})+**	**k_{cat}/K_M (M^{-1} s^{-1})**	**PDB code**	**Reference**
CA I	2.0×10^5	5.0×10^7	2CAB	(Kannan, 1984)
CA II	1.4×10^6	1.5×10^8	1CA2	(Eriksson, 1988a)
CA III	1.0×10^4	3.0×10^5	1Z97	(Duda, 2005)
CA IV	1.1×10^6	5.1×10^7	3FW3	To be published. PDB released 12/01/2009
CA VA	2.9×10^5	2.9×10^7	-	-
CA VB	9.5×10^5	9.8×10^7	-	-
CA VI	3.4×10^5	4.9×10^7	3FE4	To be published. PDB released 12/16/2008
CA VII	9.5×10^5	8.3×10^7	-	-
CA VIII (CARP)[2]	-	-	2W2J	(Di Fiore, 2009)
CA IX	3.8×10^5	5.5×10^7	3IAI	(Alterio, 2009)
CA X	-	-	-	-
CA XI	-	-	-	-
CA XII	4.2×10^5	3.5×10^7	1JCZ	(Whittington, 2001)
CA XIII	1.5×10^5	1.1×10^7	3D0N	(Di Fiore, 2009)
CA XIV	3.1×10^5	3.9×10^7	1RJ5[2]	(Whittington, 2004)[3]
murine CA XV	4.7×10^5	3.3×10^7	-	-

[1] Kinetic parameters were taken from a recent review by C. T. Supuran (Supuran, 2008).
[2] CARP is a carbonic anhydrase-related protein.
[3] Crystal structure of murine carbonic anhydrase XIV

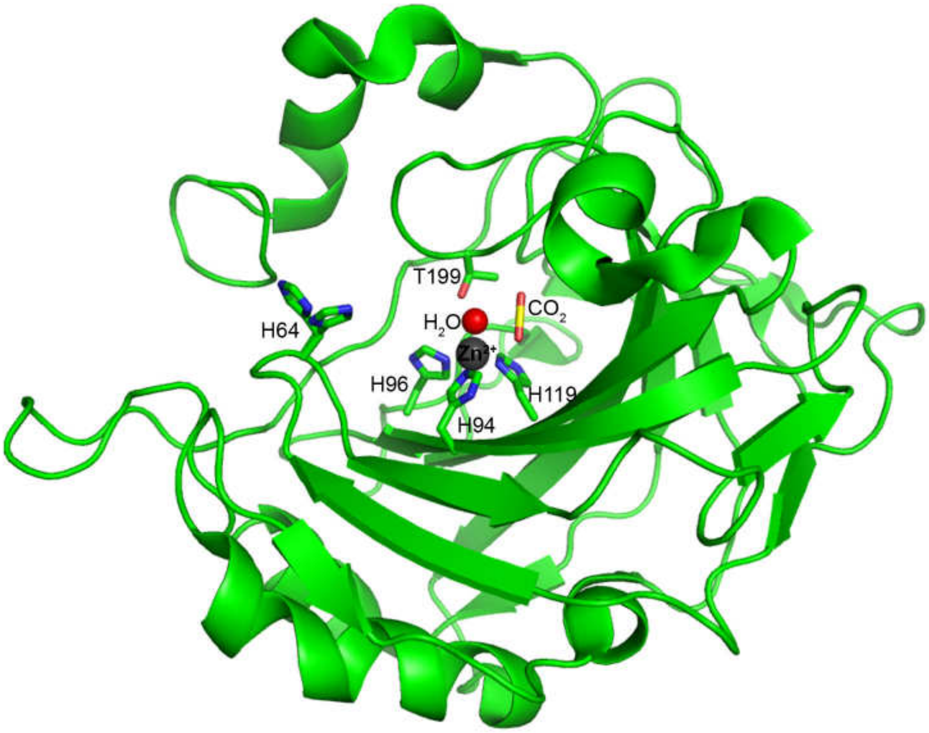


Figure 1.1. A cartoon representation of the crystal structure of human CA II, solved at 1.1 Å resolution (PDB 3D92) (Domsic, 2008) in complex with the substrate, CO_2. The catalytic Zn^{2+} ion is represented as a sphere (grey) coordinated by residues H94, H96 and H119, and is coordinated to water (red sphere), which makes a hydrogen bond to the side chain of T199. Carbon dioxide is bound in a hydrophobic pocket in the active site formed by V121, V143, and L198. A proton shuttle which transports a proton from the zinc-bound water molecule to bulk solvent is proposed to include H64, which is observed in two discrete conformations in the crystal structure.

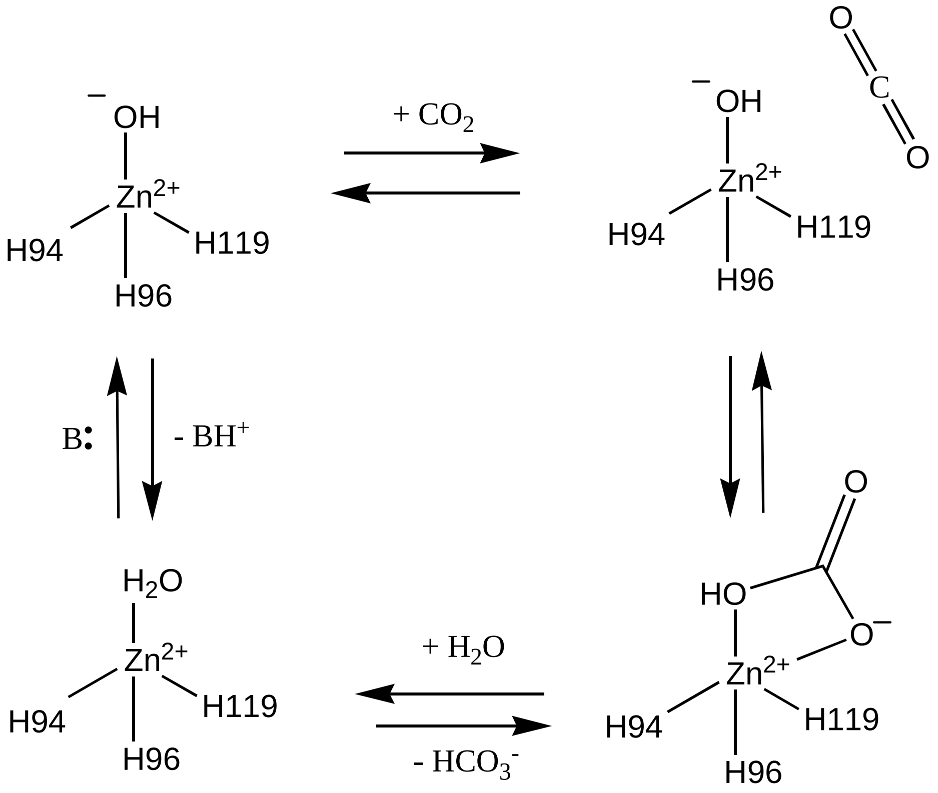


Figure 1.2. Catalytic mechanism of carbonic anhydrase. A schematic representation of the catalytic mechanism of the hydration of CO_2 by α-CAs. Carbon dioxide is bound in a hydrophobic pocket formed by V121, V143 and L198. Coordination to the zinc cation lowers the pKa of the bound water molecule from 15.7 to ~7, by stabilizing its conjugate base, hydroxide. Furthermore, the positively charged zinc stabilizes the negatively-charged transition state leading to bicarbonate formation.

Of the 12 catalytically active mammalian CAs, CA II is the most thoroughly studied isozyme and is regarded as a robust model system for systematically studying protein-ligand binding (Elbaum, 1996; Krishnamurthy, 2008). Human CA II is a particularly good model system for many practical reasons, namely it is a 30 kDa monomeric enzyme that is simple to overexpress and purify from *E. coli* due to its exceptional stability. The catalytic mechanism of CA, as well as the mechanism of inhibition by Zn^{2+}-binding ligands are well understood. Human CA II has the highest catalytic efficiency of the α-CAs, with $k_{cat}/K_M = 1.5 \times 10^8 \ M^{-1} \ s^{-1}$, approaching the limit of diffusion control.

The X-ray crystal structures of several CA isozymes have been determined at high-resolution. These structures have enabled a thorough study of ligand/inhibitor binding. CA II, a soluble cytosolic isozyme, is particularly amenable to crystallization, resulting in hundreds of protein data bank (PDB) submissions, many at high resolution (better than 1.2 Å). The use of CA as a model for biophysical and physical-organic studies of protein-ligand interactions has been recently extensively reviewed (Krishnamurthy, 2008). The use of CA II as a model system for studying protein-ligand interactions is extended in this dissertation; CA II is utilized for the structure-based design of a xenon (^{129}Xe) biosensor for use as a magnetic resonance imaging (MRI) contrast agent.

1.2 ^{129}Xe MRI

Proton (^{1}H) magnetic resonance imaging (MRI) is one of the most widely used and versatile techniques for scanning deep tissue, with important applications in disease diagnosis. MRI offers many advantages for *in vivo* imaging; it is non-invasive, uses non-ionizing radiation, and produces tomographic images. Although the intrinsic ^{1}H MRI signals from water and fat typically have low sensitivity, contrast agents have been developed that contain gadolinium or iron-oxide particles, which improve the diagnostic power of the resulting images (Degani, 1997; Foster-Gareau, 2003). More recently however, research has shifted focus to investigate the use of other nuclear magnetic resonance (NMR) active nuclei, including ^{3}He, ^{13}C, ^{83}Kr and ^{129}Xe, which can be hyperpolarized to significantly increase the NMR signal. Hyperpolarization is achieved through a process known as spin-exchange optical pumping; angular moment is transferred from circularly polarized light to an alkali metal, the newly polarized metal interacts with the nuclear spin of the isotope of interest through dipolar coupling to increase the population of unpaired spins (Kauczor, 1998). ^{129}Xe is particularly exciting for applications in imaging since it has a spin-½ nucleus, a >200-ppm chemical shift window in water, and a natural isotopic abundance of 26% (commercially available up to 86%). Moreover, ^{129}Xe hyperpolarization can enhance MRI signals ~10,000-fold (Cherubini, 2003).

Current *in vivo* ^{129}Xe MRI applications include functional lung imaging. Hyperpolarized ^{129}Xe MRI offers increased signal-to-noise ratios for lung tissue with respect to conventional ^{1}H MRI. Typically, a mixture of hyperpolarized ^{129}Xe gas and N_2 is inhaled by a patient, where it acts as a contract agent for visualizing the airways. Imaging the diffusion of xenon gas in the lungs has clinical applications in the diagnosis of asthma, chronic obstructive pulmonary disease, cystic fibrosis, and pediatric chronic lung disease (Fain, 2007). However, there are limitations to the application of hyperpolarized ^{129}Xe MRI imaging due to the reliance of the technique on the diffusion of xenon to the tissue of interest. This limitation can be overcome by the development of functional xenon biosensors, a strategy first proposed by Spence and colleagues in 2001 (Spence, 2001).

1.3 ^{129}Xe-Cryptophane Biosensors

Extending the application of hyperpolarized ^{129}Xe imaging beyond the lungs requires a biosensor that is able to bind xenon atoms, while simultaneously targeting the biological moiety of interest. An interesting class of organic supramolecular compounds known as cryptophanes can be used as the xenon cage. Cryptophane cages consist of two cup-shaped [1.1.1]orthocyclophane units connected by three bridging units (Figure 1.3). Cryptophanes of diverse shapes, sizes and chemical properties can be synthesized by varying the R1 and R2 substituents, the identity of the bridging units, Y, as well as the diastereomer (*syn* or *anti*). The type of bridging unit has a great effect on the size of the cage. The volume of a cryptophane-A cage [$Y=O(CH_2)_2O$, $R1=R2=OCH_3$] is 95 Å^3, and it can reversibly encapsulate xenon ($K_A \sim 3900$ M^{-1}). The highest affinity measured to date for a cryptophane-Xe interaction is $K_A \sim 3.3 \times 10^4$ M^{-1}, for tri-acetate cryptophane-A.

Furthermore, varying the R1 and R2 substituents facilitates the use of cryptophane cages as xenon carriers in a biosensor that can be targeted to specific proteins using an appropriate affinity tag (Lowery, 2006; Schroder, 2006). The affinity-tags and their respective targets that have been investigated include biotin-streptavidin, peptide-antigen and DNA-DNA hybrid (Spence, 2001; Roy, 2007; Schlundt, 2009). In this work, a series of racemic biosensors have been designed to target the active site of the CA isozymes, and structural and binding studies of the biosensor with human CA I and CA II follow.

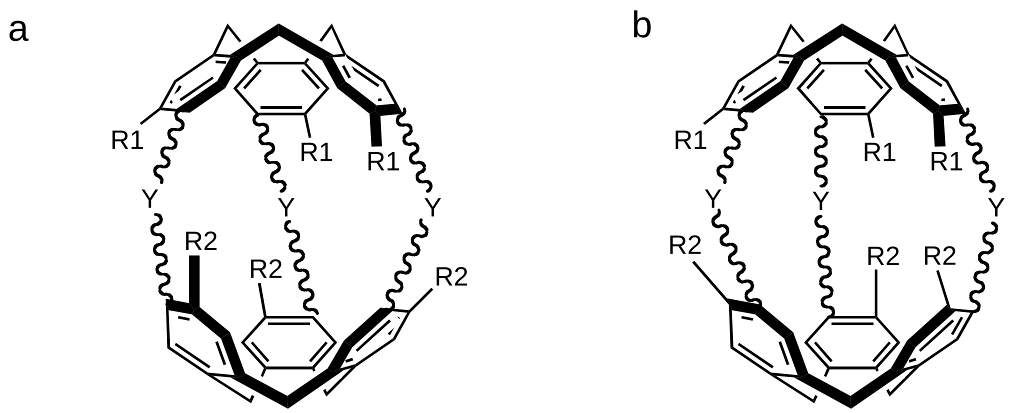

a
R1
R1
R1
Y
Y
Y
R2
R2
R2
b
R1
R1
R1
Y
Y
Y
R2
R2
R2

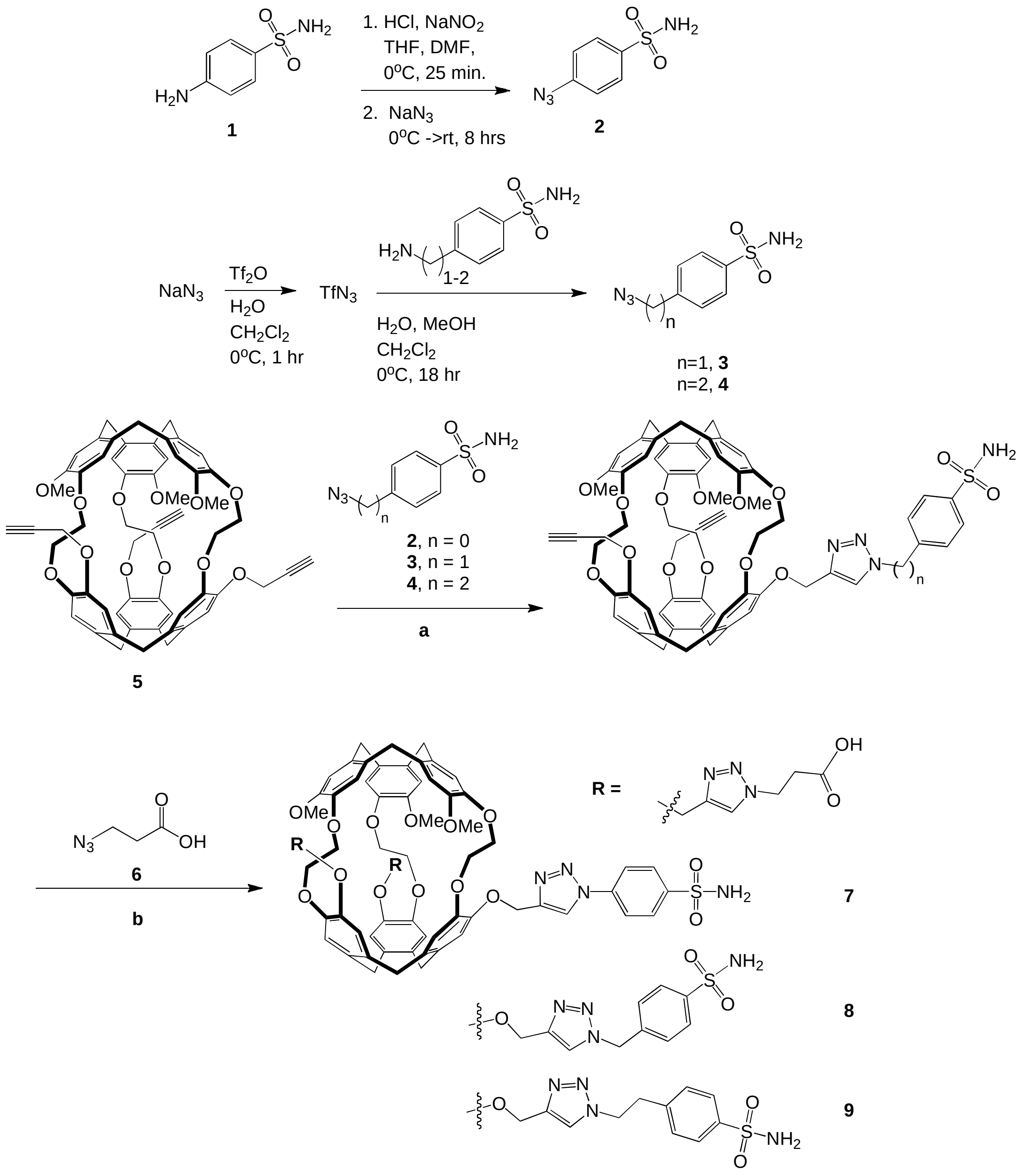
Figure 1.3. General structures of cryptophanes. (a) *anti* diasteromer. (b) *syn* diasteromer. Varying the bridging atoms (Y) and substituents (R1 and R2) give rise to cages of diverse size, shape and chemical properties.

Chapter 2: Binding Studies of ^{129}Xe-Cryptophane Biosensors and Carbonic Anhydrase

2.1 Design and Synthesis of ^{129}Xe-Cryptophane Biosensors

The design of a functional ^{129}Xe-cryptophane biosensor to target CA is quite simple. First, an appropriate cryptophane cage must chosen as the functional xenon binding moiety. Next, an appropriate CA affinity tag must be synthesized and attached at the R1 or R2 position, and finally, the remainder of R1 and R2 positions must be derivatized with appropriate side groups to impart sufficient biosensor solubility. As a starting point for the design, the cryptophane-A cage was chosen due to its superior xenon binding capabilities. To target the catalytically active Zn^{2+} ion in the active site of CA a benzene-sulfonamide functional group was chosen, and added at the R1 position. In order to increase the aqueous solubility of the biosensor, additional carboxylic acid functional groups were coupled to the two additional R1 positions, and $-OCH_3$ groups were added at the R2 positions. A series of three biosensors were synthesized with varying linker lengths between the sulfonamidate functional group and the cryptophane cage to investigate the optimal biosensor construct. A summary of the synthesis, performed by J. M. Chambers is summarized in Figure 2.1. The CA targeting portion of the biosensor was synthesized starting from **1**. The benzene-sulfonamide moiety was conjugated to an azide linker, varying the number of methylene groups between the benzene ring and the azide moiety from 0-2 to form **2-4**. Next, starting with tripropargyl cryptophane (R1= CH_2CCH and R2= OCH_3), **5**, the azide functionality on **2-4** was stoichiometrically coupled to the cryptophane via a copper catalyzed [3+2] cycloaddition.

Following silica column chromatography to remove unstoichiometric side products, the remaining R1 positions were coupled to 3-azidopropionic acid, **6**, to form a series of water-soluble CA targeting cryptophane biosensors **7-9**.



1. HCl, NaNO2
THF, DMF,
0oC, 25 min.
2. NaN3
0oC ->rt, 8 hrs
1
2
NaN3
Tf2O
H2O
CH2Cl2
0oC, 1 hr
TfN3
H2O, MeOH
CH2Cl2
0oC, 18 hr
1-2
n=1, 3
n=2, 4
5
2, n = 0
3, n = 1
4, n = 2
a
6
b
R =
7
8
9

Figure 2.1. Cryptophane biosensor synthesis was performed by J. M. Chambers, for a detailed procedure see (Chambers, 2009). Reaction conditions: a: $CuSO_4$, 2,6-Lutidine, Na Ascorbate, DMSO (for n=0, no light and dry dioxane), rt. b: $CuSO_4$, 2,6-Lutidine, Na Ascorbate, DMSO, rt.

2.2 Isothermal Titration Calorimetry Measurements

2.2.1 Introduction

Isothermal titration calorimetry (ITC) is a technique used to measure the thermodynamic parameters of interactions between molecules by measuring the change in heat upon mixing of the analytes of interest. Specifically, ITC directly measures the binding affinity (K_a), enthalpy change (ΔH), and binding stoichiometry (n) of an interaction between two molecules, in our case CA and a cryptophane biosensor. Then, using the Gibbs free energy equation (2), the entropy (ΔS) and Gibbs free energy (ΔG) of the interaction can be determined (Falconer, 2010).

$$\Delta G = -RT\ln K_a = \Delta H - T\Delta S \qquad (2)$$

An ITC experiment is performed in a calorimeter that contains two identical small cells surrounded by an adiabatic jacket. One cell serves as a control/reference cell while the second cell serves as a sample cell. To begin an experiment, one of the analytes of interest, in our case the protein CA, is placed in the sample cell. Next, the biosensor is slowly titrated into the protein solution in aliquots of a few microliters per addition. Binding of the cryptophane biosensor to CA is exothermic. The instrument detects temperature differences between the sample and reference cell and measures the time-dependent input of power required to maintain both cells at the same temperature. The raw data is integrated with respect to time to determine the total amount of heat released per analyte injection, and together with the molar ratio of biosensor to protein the thermodynamic parameters of the interaction can be determined. The advantages of ITC include the ability to measure free energy, enthalpy, entropy, association constant, and stoichiometry of an interaction simultaneously without incorporating unnatural labels, which could introduce bias to the measurement. The association constant for the interaction can be converted into a dissociation constant from the relationship $K_d = 1/K_a$ (Freyer, 2008).

2.2.2 Experimental Methods

All calorimetry experiments were conducted at 298 K on a VP-ITC titration microcalorimeter from MicroCal, Inc. (Northhampton MA), using standard protocols and data analysis (Wiseman, 1989; Fisher, 1995). Human CA I was purchased from Sigma and used without further purification, human CA II was overexpressed in *Escherichia coli* and purified as previously described (Alexander, 1993). Biosensors **7**, **8** and **9**, as well as the benzenesulfonamide-linker, **4**, were synthesized by J. M. Chambers (Chambers, 2009). CA I and CA II were diluted to ~ 20 μM and exhaustively dialyzed against 50 mM Tris-SO_4 (pH 8.0). Biosensors (~10 mM stock solutions in DMSO) were dissolved at a concentration of 135-300 μM in an aliquot of the same buffer, and an equivalent concentration of DMSO was added to the enzyme solution. Prior to the titration experiment, samples were degassed under vacuum for 5 min. The sample cell (effective volume = 1.4 mL) was overfilled with 1.8 mL of CA at a concentration of 14-26 μM, and the reference cell was filled with water. The contents of the sample cell were titrated with 30 aliquots (10 μL each) of inhibitor (two initial 2 μL injections were made, but not used in data analysis). After each injection, the heat change was measured and converted to the corresponding enthalpy value. The reaction mixture was continuously stirred at 300 rpm during titration. Control experiments were carried out by titrating the inhibitor into the buffer solution under identical experimental conditions. The calorimetric data are presented with the background titrations subtracted from the experimental data. The amount of heat produced per injection was calculated by integration of the area under each peak. Data were fit to the equation $q = V \Delta H[E]_t K[L]/(1 + K[L])$, where q is the heat evolved during the course of the reaction, V is the cell volume, ΔH is the binding enthalpy per mole of ligand, $[E]_t$ is the total enzyme concentration, K is the binding constant, and [L] is inhibitor concentration. Nonlinear regression fitting to the binding isotherm (ORIGIN 5.0 software, MicroCal) using a one-site model gave the equilibrium dissociation constant of the ligand, K_d, and estimates of the standard error. Representative isothermal calorimetric data and binding isotherms are shown in Figures 2.2 and 2.3 and a summary of dissociation constants for the CA-biosensor complexes are summarized Table 2.1. The error is $\sigma_i = \sqrt{(C_{ii}\chi 2)}$, where C_{ii} is the diagonal element of the variance-covariance matrix.

2.2.3 Results and Discussion

ITC binding measurements indicate that all three biosensors have nanomolar affinity for human CA I and CA II (Table 2.1). The K_d of the benzenesulfonamide affinity tag, **4**, was measured as a control to determine the contribution of the cryptophane on the binding constant. ITC binding studies indicate that the presence of cryptophane and length of the linker between the sulfonamide and the cage has little effect on the K_d. Overall, all three biosensors exhibited modestly higher affinity for CA I (20 – 80 nM) versus CA II (60-100 nM). Biosensor **9** exhibited the highest affinity for CA II (60 nM), while biosensor **7** exhibited the highest affinity for CA I (20 nM) (Chambers, 2009).

Table 2.1. Summary of dissociation constants for biosensors-carbonic anhydrase complexation determined by ITC at 298 K.

CA Isozyme	**Ligand**	**K_d (nM)**
I	4	30 ± 10
	7	20 ± 10
	8	80 ± 10
	9	30 ± 20
II	4	100 ± 10
	7	100 ± 20
	8	110 ± 30
	9	60 ± 20

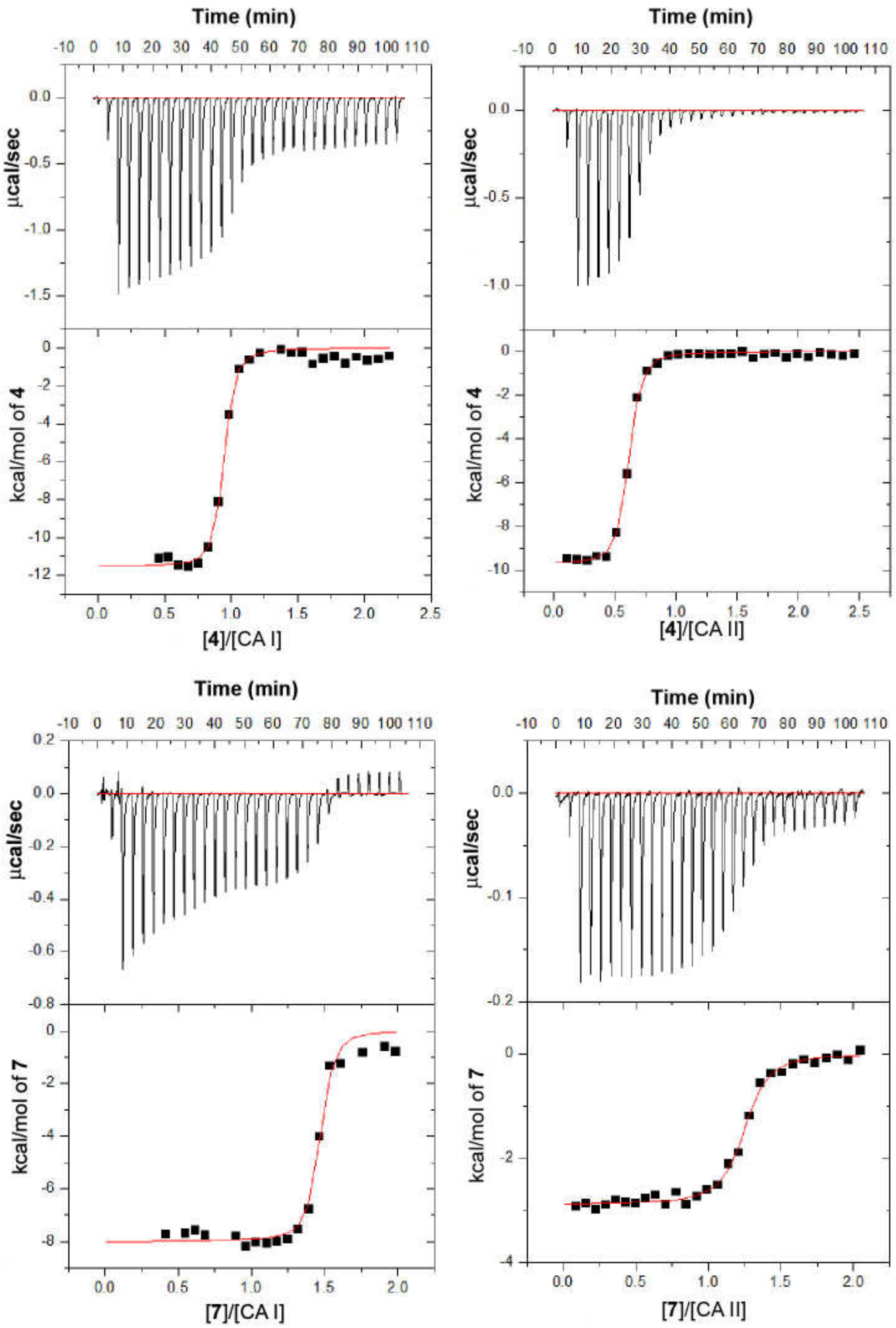


Figure 2.2. Isothermal calorimetric data for the interactions of **4** and **7** with CA I (left panel) and CA II (right panel). CA I (19.9 µM) titrated with **4** (200.0 µM); CA II (25.9 µM) titrated with **4** (292.3 µM); CA I (20.0 µM) titrated with **7** (183.0 µM); CA II (19.32 µM) titrated with **7** (182 µM).

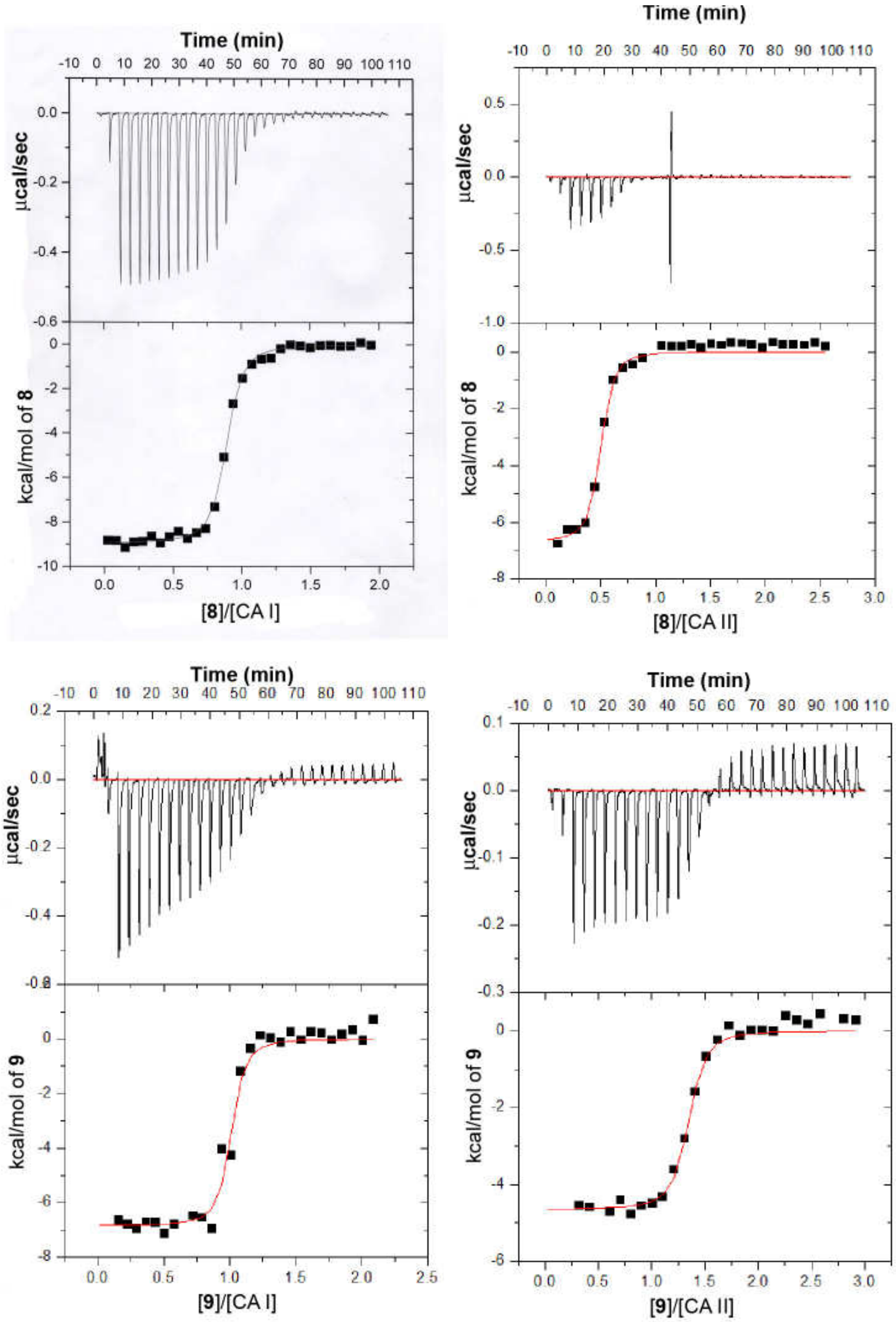


Figure 2.3. Isothermal calorimetric data for the interactions of **8** and **9** with CA I (left panel) and CA II (right panel). CA I (22.4 µM) titrated with **8** (200.0 µM); CA II (16.6 µM) titrated with **8** (194.9 µM); CA I (14.2 µM) titrated with **9** (135.7 µM); CA II (14.9 µM) titrated with **9** (200 µM).

2.3 Surface Plasmon Resonance Measurements

2.3.1 Introduction

Surface plasmon resonance (SPR) is a physical phenomenon that can be used to detect very small changes at a surface. Specifically, SPR experiments measure a change in the local index of refraction at a surface, resulting in a change in resonance conditions of surface plasmon waves (Pattnaik, 2005). Recently, a number of instruments based on SPR have become commercially available and are specifically designed to quantify macromolecular interactions (Jason-Moller, 2006). The experimental setup involves immobilization of one of the binding partners of interest (typically referred to as the “ligand”) to a sensor-chip. A typical sensor-chip is a glass slide coated with a thin layer of gold followed by a specific surface matrix, to which the ligand is attached (examples include carboxymethylated dextran, streptavidin, nickel cheltation, and hydrophobic monolayer). Surface plasmons are excited by incident light beam on the opposite side of the gold surface. The incident photons induce an evanescent light field into the gold film and at a certain incident angle are able to excite surface plasmons. When a plasmon is excited, the change in the reflected light is observed at that incident angle is measured by a charged couple device (CCD) chip (Pattnaik, 2005). The first step of an SPR experiment is to immobilize the “ligand” on the sensor chip surface. Next, the “analyte” flows over the chip, where it interacts with the “ligand” (the association phase), this interaction is correlated to the change in mass at the sensor surface, resulting in an observable change in the SPR angle, detected by a change in the intensity of the reflected light (measured in resonance units (RU)). Finally, the chip is regenerated by flowing buffer over the surface of the chip (dissociation phase). The time-dependant change in

SPR signal recorded during an experiment can be fit to calculate the binding constant, K_d, for the interaction of the specific ligand and analyte (Pattnaik, 2005).

SPR techniques offer many advantages for studying binding of small molecules and macromolecular targets; the technique can be label-free, interactions can be studied in real time, it requires very small volumes of protein, and one immobilized sample on a chip can be reused many times to study a variety of different analytes. However there are a number of challenges as well since the change in refractive index is relatively small, accurate results require optimization of experimental parameters and high-quality data for consistent, accurate results (Cannon, 2004).

2.3.2 Experimental Methods

Interaction analyses were performed using a Biacore 3000 SPR instrument (Biacore AB, Uppsala, Sweden) at The Protein Core Facility, Children's Hospital of Philadelphia. Recombinant human CA II was coupled to a carboxymethylated dextran (CM5) chip using amine coupling reagents *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide, *N*-hydroxysuccinimide (NHS), and ethanolamine HCl, using previously published procedures (Cannon, 2004). The CA II stock solution was prepared in 100 mM sodium acetate (pH 4.9). Samples of the analytes, biosensors **7-9**, were prepared in the running buffer (20 mM Na_2HPO_4-NaH_2PO_4, pH 7.4, 1.5 M NaCl, 3% DMSO) at the following analyte concentrations, 62.5 nM, 125 nM, 250 nM, 500 nM, 1000 nM, 2000 nM. Each concentration of analyte was tested in duplicate. Kinetic data were fit to a simple 1:1 interaction model (Langmuir binding) using the program BIAevaluation.

2.3.3 Results and Discussion

Dissociation constants for the interactions of human CA II and biosensors **7-9** were determined using SPR (sensorgrams are shown in Figure 2.4 and a summary of the results in Table 2.2). Comparison of the dissociation constants obtained from ITC and SPR confirm nanomolar-binding affinity of the cryptophane biosensors to CA II. Furthermore, SPR measurements indicate that biosensor **7** binds tightest to human CA II, followed by biosensor **9** then **8**. Interestingly, there is a greater than 10-fold discrepancy between the K_d determined by SPR and ITC for the biosensor **7**-CA II interaction. The origins of this discrepancy are unclear; as the fitting parameters for the SPR and ITC data are satisfactory. Overall, the SPR experiments illustrate the usefulness of this technique for studying cryptophane-biosensor-protein interactions, specifically in a high-throughput manner.

Table 2.2. Dissociation constant determined by surface plasmon resonance for biosensors **7-9** with human CA II.

Analyte	**K_d (nM)**	χ^2
7	7.33	0.853
8	207	0.442
9	40.9	1.1

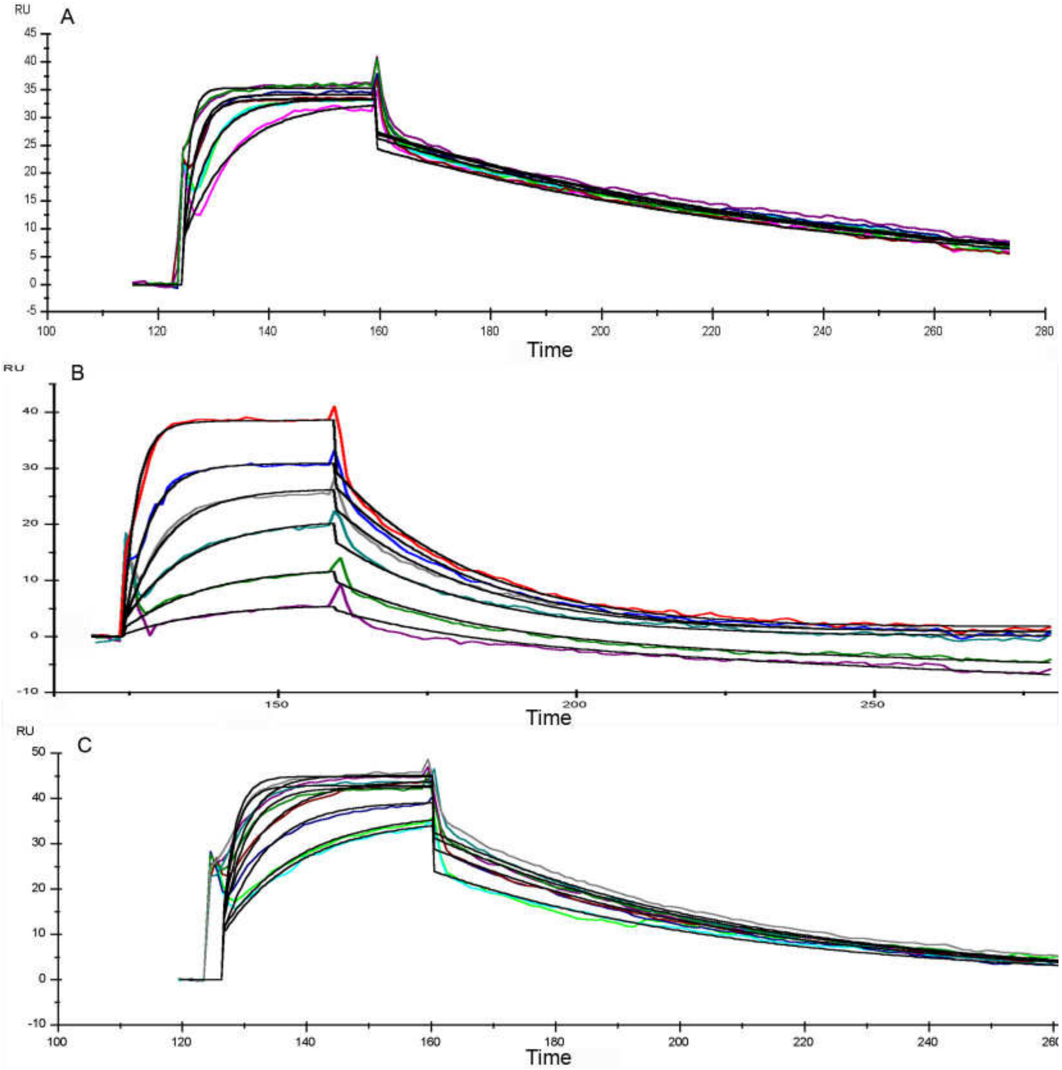


Figure 2.4. Surface plasmon resonance sensorgrams for the interaction between human CA II and (A) biosensor **7**, (B) biosensor **8**, and (C) biosensor **9**.

Chapter 3: Structural Studies of ^{129}Xe-Cryptophane-CA II

3.1 Materials and Methods

Human CA II was overexpressed in *Escherichia coli* and purified as previously described (Alexander, 1993). Crystals of the CA II-biosensor complex were formed by adding a two-fold excess of biosensor **9** (10 mM stock in DMSO) to 0.5 mg/mL CA II (50 mM Tris-SO_4, pH 7.5) and incubating at 4 ^oC for one hour. The mixture was concentrated using a YM-10 filter to a final CA II concentration of 10 mg/mL. Crystals were grown using the hanging drop method: a 5 μL drop of CA II solution was added to a 5 μL drop of precipitant solution (50 mM Tris-SO_4, 16% PEG 3350, 3.5 mM β-mercaptoethanol) and suspended over a reservoir containing 1 mL 50 mM Tris-SO_4, 27-32% PEG 3350 and 3.5 mM β-mercaptoethanol at 4 ^oC. Crystals formed within 1-2 weeks and were improved with seeding. Cubic crystals grew to typical dimensions 0.2 mm x 0.2 mm x 0.2 mm (Figure 3.1). Crystals were cryoprotected by augmentation of the mother liquor with 15% glycerol and then looped and pressurized with $Xe_{(g)}$ for 30 minutes at 20 atm using a Xenon Chamber (Hampton Research). Crystals were flash cooled 10 seconds after Xe pressurization. Crystals yielded diffraction data to 1.70 Å at the Cornell High Energy Synchrotron Source (CHESS) beamline F-2 (λ = 0.9795 Å, 100 K), using an ADSC Quantum 210 CCD detector (Szebenyi, 1997). Diffraction data were indexed, integrated and scaled using HKL2000 (Otwinowski, 1997). Crystals belonged to space group *C*2 (unit cell parameters a = 67.4 Å, b = 50.0 Å, c = 81.0 Å, β = 107.1^o) and were isomorphous with those of T199P CAII complexed with thiocyanate (PDB 1LG6) (Huang, 2002). Initial phases were obtained by molecular replacement using the program

Phaser (Storoni, 2004) with PDB 1LG6 (less water molecules and ligand) as a search probe for rotation and translation functions. The programs CNS (Brunger, 1998) and O (Jones, 1991) were used in refinement and rebuilding, respectively. Figures were generated using PyMOL. Molecular surface area was calculated with protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel, 2007). Data collection and refinement statistics are summarized in Table 3.1.

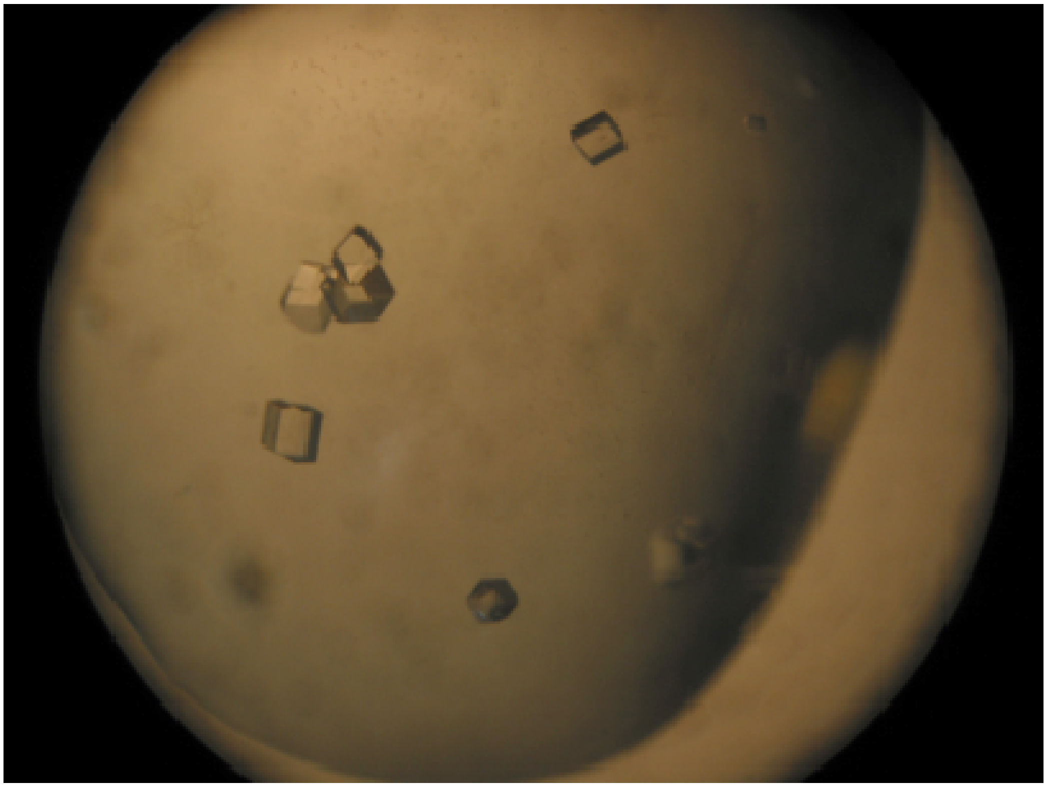


Figure 3.1. Crystals of Biosensor-**9**-CA II.

Table 3.1. Data Collection and Refinement Statistics for CA II-**9**-Xe complex.

	CA II-**9**-Xe complex
Data Collection	
PDB Code	3CYU
Resolution, Å	38.7 – 1.70
Total reflections measured[a]	52826 (4698)
Unique reflections measured[a]	27728 (2556)
R_{merge}[a,b]	0.078 (0.496)
$I/\sigma(I)$[a]	27.1 (2.3)
Completeness (%)[a]	97.0 (90.3)
Redundancy[a]	3.9 (3.7)
Refinement	
Reflections used in refinement/test set	24730/1139
R_{work}	0.226
R_{free}	0.249
Protein atoms[c]	2049
Water molecules[c]	185
Xe atoms[c]	2
Cryptophane-A-benzenesulfonamide atoms[c]	103
R.m.s deviations	
Bond lengths, Å	0.016
Bond angles, °	1.8
Dihedral angles, °	22.4
Improper dihedral angles, °	0.7
Average B-factors, Å^2	
Main chain	31
Side chain	35
Xe atoms	43
Zn atom	28
Cryptophane-A-benzenesulfonamide atoms	42
Solvent	40
Ramachandran Plot[d]	
Allowed (%)	86.6
Additionally allowed (%)	12.5
Generously allowed (%)	0.9
Disallowed (%)	0.0

[a] Number in parentheses refer to the outer 0.1 Å shell of data.

[b] $R_{merge} = \sum |I-\langle I\rangle| / \sum I$, where I is the observed intensity and ⟨I⟩ is the average intensity calculated for replicate data.

[c] Per asymmetric unit

[d] Ramachandran plot statistics calculated for non-proline and non-glycine residues using PROCHECK (Laskowski, 1993).

3.2 Results and Discussion

The X-ray crystal structure of the CA II-**9**-Xe complex was solved to 1.70 Å resolution, and refined to final R_{work} and R_{free} values of 0.23 and 0.25, respectively (Aaron, 2008). Two xenon sites were identified by inspection of the Bijvoet difference Fourier map calculated from anomalous data. The first site is near the opening of the active site cleft, 18 Å from Zn^{2+} and 8 Å from the protein chain, and corresponds to the Xe atom encapsulated by the cryptophane (Figure 3.2). The encapsulation of Xe within the cryptophane cage of **9** is confirmed by inspection of the Bijvoet difference Fourier map calculated from anomalous scattering data. X-ray diffraction data was collected at a wavelength $\lambda = 0.9795$ Å, which is far from the Xe L_I edge of 2.27 Å (Watanabe, 1965). Nevertheless, the anomalous scattering component *f*” is 3.4 e^- for Xe, so the anomalous signal is still prominent at the wavelength of data collection. The second Xe site is a hydrophobic pocket defined by A116, L148, V218, L157, V223 and F226, which is consistent with the known binding interactions of Xe in other systems (Figure 3.3) (Prange, 1998). The crystallographic occupancies of these Xe sites refine to 0.50 and 0.37, respectively. Anomalous scattering peaks are absent from crystals not subject to Xe pressurization.

The occupancy of the active site zinc ion was refined at 0.5, which was consistent with the occupancy of 0.5 determined for biosensor **9**. Because the crystallographic occupancy was thus 0.5 for Xe encapsulated within the cryptophane moiety, and the electron density map indicated the binding of both cryptophane enantiomers, each enantiomer was refined with an occupancy of 0.25 (average B-factor = 42 $Å^2$). A total of

185 water molecules were included in later cycles of refinement. Data reduction and refinement statistics are recorded in Table 2. The N-terminus (N1-H3) was disordered and is omitted from the final model.

Biosensor **9** coordinates to the active site Zn^{2+} ion as the sulfonamidate anion, displacing the zinc-bound hydroxide ion of the native enzyme as previously observed in other complexes of CA II with benzenesulfonamide derivatives (Eriksson, 1988b; Elbaum, 1996; Supuran, 2007; Krishnamurthy, 2008). The crystallographic occupancies of **9** and Zn^{2+} are refined at 0.5. It is unusual to observe diminished Zn^{2+} occupancy in a CA II-inhibitor complex, and the molecular origins of this effect are not clear. Notably, **9** contains a chiral axis and the electron density map reveals the binding of equal populations of both enantiomers (Figures 3.4 and 3.5), (each refined with an occupancy of 0.25) (Eliel, 1994; Collet, 1996; Ruiz, 2006). Overall, the binding of **9** does not cause any significant structural changes in the active site, and the root-mean-square deviation is 0.34 Å for 256 Cα atoms between the current structure and the unliganded enzyme (PDB 2CBA) (Hakansson, 1992). The total surface area of **9** is ~1500 $Å^2$, of which ~500 $Å^2$ becomes solvent inaccessible due to contacts of **9** within the active site cleft of CAII designated molecule I in Figure 3.6. Crystal contacts bury an additional 540 $Å^2$ of the surface of **9** as follows: 270 $Å^2$ with molecule III, and 240 $Å^2$ and 30 $Å^2$ with the front and back faces of molecule II, respectively. Molecule IV does not contact **9** bound to molecule I.

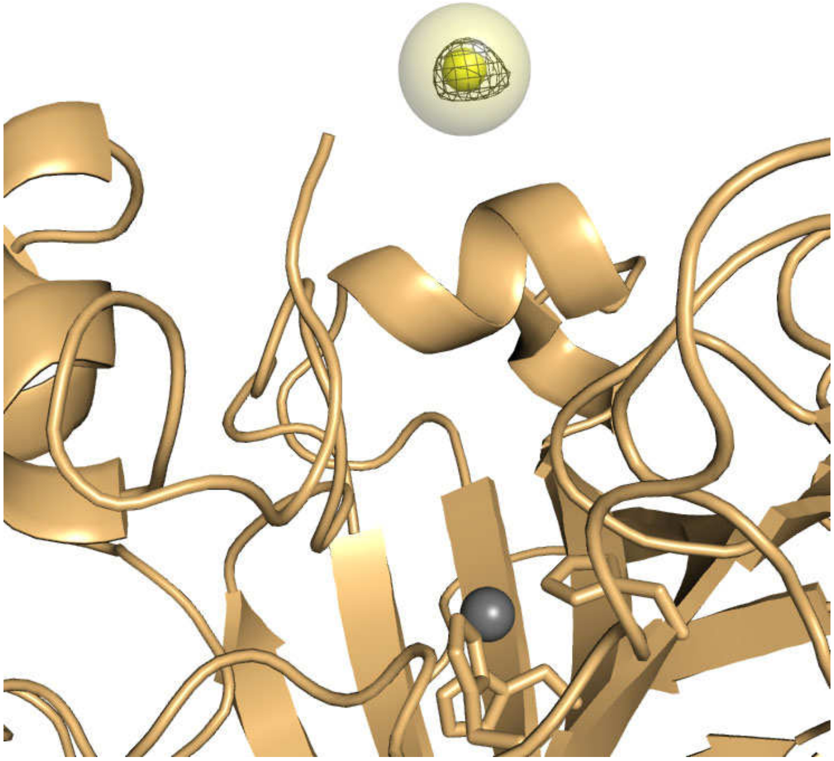


Figure 3.2. Anomalous Fourier map showing Xe location at opening of CA II active site. Xe (yellow sphere) was identified upon inspection of Bijvoet difference Fourier map (black) calculated from anomalous data to be 18 Å from Zn^{2+} (gray sphere) at opening of active site cleft. The Xe occupancy is refined at 0.50. The van der waals radius of Xe is shown as a translucent yellow sphere.

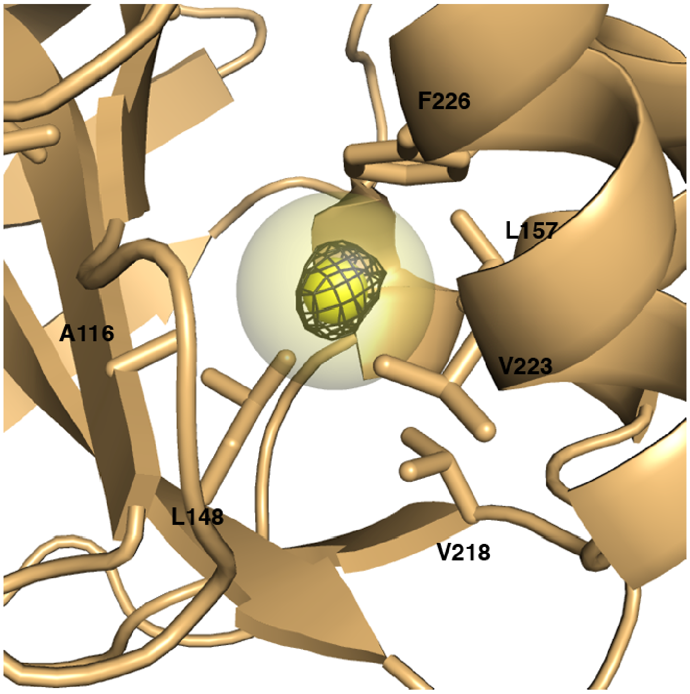
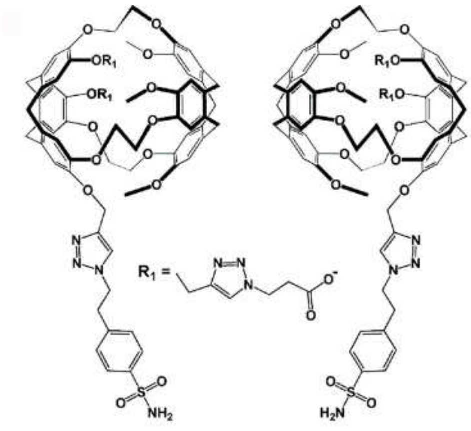


Figure 3.3. Anomalous Fourier map showing Xe location in CA II hydrophobic pocket. A second binding site is observed in a hydrophobic pocket defined by A116, L148, V218, L157, V223 and F226. Binding at this site is consistant with Xe binding in other proteins. Occupancy is 0.37. The van der waals radius of Xe is shown as a translucent yellow sphere.

Some structural changes are observed near the outer rim of the active site cleft where the cryptophane binds. The most notable change is observed for Q136, which rotates ~180° to make van der Waals contacts with the cryptophane and the symmetry-related cryptophane bound to molecule III in the crystal lattice. Other residues at the active site rim of molecule I that make close contacts with the cryptophane are G132 and P202. Additional structural changes in the crystal lattice result from the binding of **9** to molecule I: in molecule II, H36 rotates ~90° to make a van der Waals contact with the cage, and Q137 of molecule III rotates ~90° to donate a hydrogen bond to an ether oxygen atom of **9**.

Although the pendant propionates appear to be more disordered than the cryptophane and are characterized by correspondingly weaker electron density, a hydrogen bond between a propionate moiety and Q53 of molecule II is observed. The relative dearth of strong cryptophane-protein interactions may explain why the affinity of **9** measured by ITC is only slightly better than that measured for the parent triazole-benzenesulfonamide lacking the cryptophane (K_D = 100 ± 10 nM). Larger refined thermal (B) factors for **9** (<B> = 42 $Å^2$) compared with the overall CA II model (<B>main chain = 31 $Å^2$; <B>side chain = 35 $Å^2$) reflect the mobility and conformational heterogeneity of the bound biosensor.

Limited hydrogen bond interactions between CA II and the cryptophane moiety of **9** may be advantageous for the use of cryptophanes as ^{129}Xe biosensors. Translational and rotational freedom, the consequence of a flexible linker between the cryptophane and the benzensulfonamide, could allow the cage to reorient rapidly *in situ,* independently of the protein, to result in decreased correlation times and narrower line widths that increase the sensitivity of ^{129}Xe NMR measurements in solution. In conclusion, this work reveals the first experimentally determined structure showing how an encapsulated ^{129}Xe atom can be specifically directed to a biomedically relevant protein target. The possible implications for cancer diagnosis are profound, given that CA isozymes IX and XII are overexpressed on the surface of certain cancer cells (Pastorekova, 2006). Moreover, a search of the Protein Data Bank reveals that with its molecular mass of 1554, the **9**-Xe complex is one of the largest synthetic organic ligands ever cocrystallized with a protein. Thus, this work demonstrates the feasibility of preparing crystalline complexes between proteins and nonbiological, nanometer-scale ligands.



9-M_oM_o **9-P_oP_o**

Figure 3.4. The M_oM_o and P_oP_o enantiomers of the cryptophane-A-derived CA biosensor. The benzenesulfonamide moiety serves as an affinity tag that targets the Zn^{2+} ion, and the R_1 substituents contain triazole propionate moieties that enhance aqueous solubility.

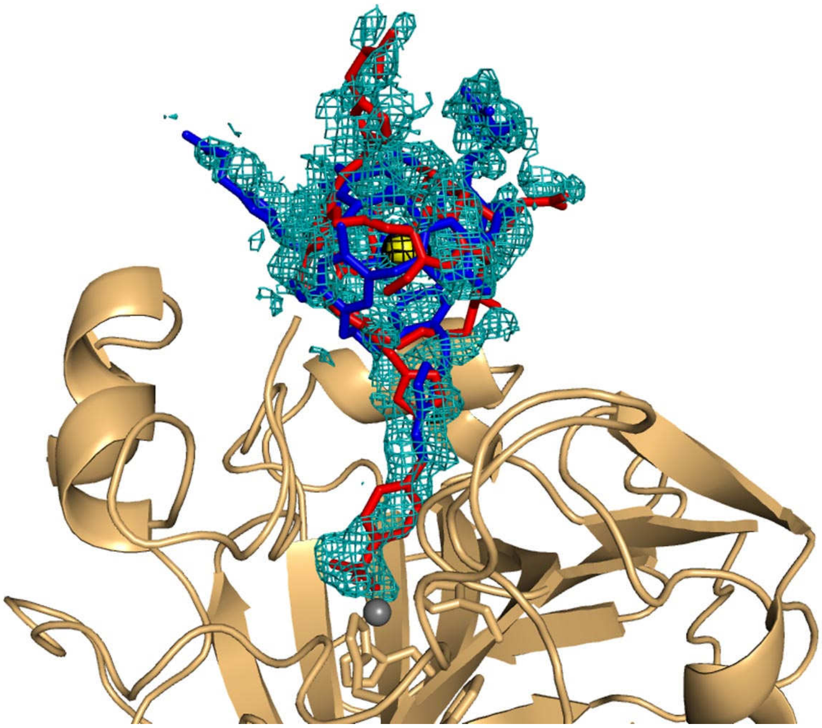


Figure 3.5. The crystal structure of biosensor **9** bound to CAII. Refinement revealed the binding of equal populations of both enantiomers of **9**. A simulated annealing omit map showing **9**-M_oM_o (blue) and **9**-P_oP_o (red) bound in the active site (1.9 σ contour, teal).

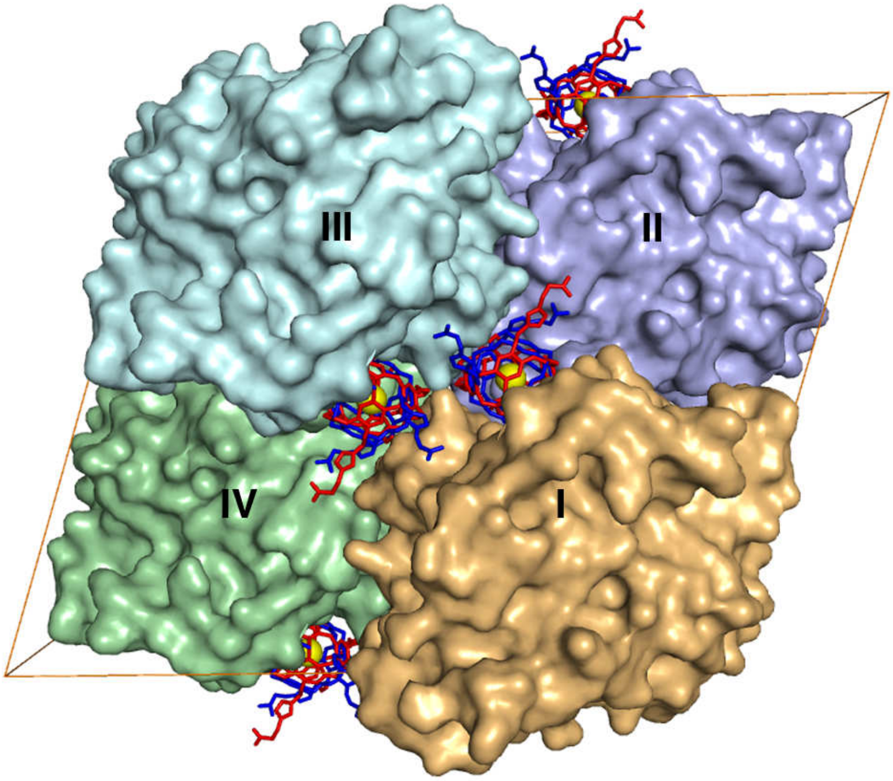


Figure 3.5. The unit cell of CAII-biosensor-**9** complex. The crystals in space group C2 contains four molecules: I (x,y,z), II (x+½,y+½,z), III (-x,y,-z) and IV (-x+½,y+½,-z).

3.3 Future Applications of Cryptophane-based CA Biosensors

Although a very valuable model system, molecular imaging of human CA II is not relevant for disease diagnosis since CA II is found in many tissues in the body, and is not upregulated in connection with a particular disease. In comparison, CA IX and CA XII overexpression is induced by hypoxia, a pathological condition caused by oxygen deprivation, and associated with many types of cancer (Supuran, 2008; De Simone, 2010). CA IX and CA XII are two of only three transmembrane human CAs (the third is CA XIV). In addition to tumors, CA IX is also localized in gastrointestinal mucosa, while CA XII is localized in renal, intestinal, reproductive epithelial, and eye tissues (Kivela, 2005; Supuran, 2008). Structural analyses of human CA IX and XII offer useful insight in the design of a selective biosensor for imaging cancerous tissue.

Human CA IX contains four domains in addition to an N-terminal 37 amino acid signal peptide. Attached to the signal peptide is a proteoglycan-like (PG) domain, followed by a 257 residue catalytic domain, a transmembrane segment that transverses the membrane once, and finally a short intracytoplasmic tail (De Simone, 2010). The crystal structure of the catalytic domain of CA IX was recently determined (Alterio, 2009), and revealed several interesting structural observations, which may aid in rational drug design of selective CA IX inhibitors. Human CA IX forms a dimer (Hilvo, 2008), mediated by an intermolecular disulfide bond between C41 of each monomer, in addition to ~1590 $Å^2$ buried surface area (Alterio, 2009). The structure of the dimer suggests that the proteoglycan domains are located at the border of the active site, and are proposed to be involved in assisting the catalytic domain in catalysis. pH dependent activity profiles indicate that the presence of the PG domains reduces the pKa of the zinc-bound water molecule from 7.01 to 6.49, making CA IX more active in solid and hypoxic tumors where the pH is typically slightly acidic, and coincidently where CA IX is typically overexpressed (Alterio, 2009). Overall, the structure of the catalytic region of CA IX is highly conserved with respect the other CA isozymes, the rmsd for the superposition of the backbone atoms of the catalytic domain of CA IX with human CA II is 1.4 Å. CA IX was co-crystallized with the acetazolamide, a sulfonamide inhibitor with a K_I of ~9-25 nM, depending upon the CA IX construct (Hilvo, 2008). However, some structural differences in the region of residues 125-137, namely residues 131, 132, 135, and 136 can be exploited for specific inhibitor design. Furthermore, the extracellular localization of CA IX allows for the design of positively charged, membrane-impermeable, inhibitors which can bind to membrane-associated CA isozymes without effecting cytosolic and mitochondrial CA isoforms (Alterio, 2009).

Since CA IX expression is highly tumor specific, and is primarily localized on the surface of solid tumors including gliomas/ependymomas, mesotheliomas, and many types of carcinomas (Potter, 2003), it is an ideal candidate for selective imaging for early cancer detection. Specifically, a functionalized cryptophane containing a CA IX targeting moiety coupled with positively-charged amine water-solubilizing groups are a starting point for design of a CA IX-specific ^{129}Xe MRI contrast agent (Taratula, 2009).

The crystal structure of the catalytic domain of human CA XII also indicated the formation of a biological isologous dimer (Whittington, 2001). The dimer interface contains 19 hydrogen-bond interactions and buries ~1,100 $Å^2$ of surface area per monomer, while leaving the enzyme active site accessible. The residues that define the core of the active site are highly conserved with respect to CA II, however the structures diverge in the “130’s segment” of the active site. Specifically, human CA XII has an alanine residue in place of F131 in human CA II, creating a larger cavity in the active site and exposing S135 (Whittington, 2001). This region of the human CA XII active site can be exploited for CA XII-specific biosensors.

Recently, bioreductive nitro-containing sulfonamides have been investigated as carbonic anhydrase inhibitors with selectivity for tumor associated CA IX and XII. Several compounds have been identified with selectivity ratios for the inhibition of CA IX and XII over CA II up to 17 times, and for the inhibition of CA IX and XII over CA I up to 1400 times (D'Ambrosio, 2008). These compounds serve as excellent lead compounds for the design of a selective CA IX or CA XII biosensor.

References

Aaron, J. A., Chambers, J. M., Jude, K. M., Di Costanzo, L., Dmochowski, I. J. and Christianson, D. W. (2008) Structure of a Xe-129-cryptophane biosensor complexed with human carbonic anhydrase II. *J. Am. Chem. Soc. 130*, 6942-6943.

Alexander, R. S., Kiefer, L. L., Fierke, C. A. and Christianson, D. W. (1993) Engineering the zinc binding site of human carbonic anhydrase II: Structure of the His94Cys apoenzyme in a new crystalline form. *Biochemistry 32*, 1510-1518.

Alterio, V., Hilvo, M., Di Fiore, A., Supuran, C. T., Pan, P., Parkkila, S., Scaloni, A., Pastorek, J., Pastorekova, S., Pedone, C., Scozzafava, A., Monti, S. M. and De Simone, G. (2009) Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX. *Proc. Natl. Acad. Sci. U. S. A. 106*, 16233-16238.

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. and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905-921.

Cannon, M. J., Papalia, G. A., Navratilova, I., Fisher, R. J., Roberts, L. R., Worthy, K. M., Stephen, A. G., Marchesini, G. R., Collins, E. J., Casper, D., Qiu, H., Satpaev, D., Liparoto, S. F., Rice, D. A., Gorshkova, I. I., Darling, R. J., Bennett, D. B., Sekar, M., Hommema, E., Liang, A. M., Day, E. S., Inman, J., Karlicek, S. M., Ullrich, S. J., Hodges, D., Chu, T., Sullivan, E., Simpson, J., Rafique, A., Luginbuhl, B., Nyholm Westin, S., Bynum, M., Cachia, P., Li, Y.-J., Kao, D., Neurauter, A., Wong, M., Swanson, M. and Myszka, D. G. (2004) Comparative analyses of a small molecule/enzyme interaction by multiple users of Biacore technology. *Anal. Biochem. 330*, 98-113.

Chambers, J. M., Hill, P. A., Aaron, J. A., Han, Z. H., Christianson, D. W., Kuzma, N. N. and Dmochowski, I. J. (2009) Cryptophane xenon-129 nuclear magnetic resonance biosensors targeting human carbonic anhydrase. *J. Am. Chem. Soc. 131*, 563-569.

Cherubini, A. and Bifone, A. (2003) Hyperpolarised xenon in biology. *Prog. Nucl. Magn. Reson. Spectrosc. 42*, 1-30.

Collet, A. (1996). Cryptophanes. In Comprehensive Supramolecular Chemistry, J. L. Atwood, J. E. D. Davis, D. D. MacNicol and F. Vogtle, eds. (New York, Pergamon), pp. 325-365.

D'Ambrosio, K., Vitale, R. M., Dogne, J. M., Masereel, B., Innocenti, A., Scozzafava, A., De Simone, G. and Supuran, C. T. (2008) Carbonic anhydrase inhibitors: Bioreductive nitro-containing sulfonamides with selectivity for targeting the tumor associated isoforms IX and XII. *J. Med. Chem. 51*, 3230-3237.

De Simone, G. and Supuran, C. T. (2010) Carbonic anhydrase IX: Biochemical and crystallographic characterization of a novel antitumor target. *Biochim. Biophys. Acta, Proteins Proteomics 1804*, 404-409.

Degani, H., Gusis, V., Weinstein, D., Fields, S. and Strano, S. (1997) Mapping pathophysiological features of breast tumors by MRI at high spatial resolution. *Nat. Med. 3*, 780-782.

Di Fiore, A., Monti, S. M., Hilvo, M., Parkkila, S., Romano, V., Scaloni, A., Pedone, C., Scozzafava, A., Supuran, C. T. and De Simone, G. (2009) Crystal structure of human carbonic anhydrase XIII and its complex with the inhibitor acetazolamide. *Proteins: Struct., Funct., Bioinf. 74*, 164-175.

Domsic, J. F., Avvaru, B. S., Kim, C. U., Gruner, S. M., Agbandje-McKenna, M., Silverman, D. N. and McKenna, R. (2008) Entrapment of carbon dioxide in the active site of carbonic anhydrase II. *J. Biol. Chem. 283*, 30766-30771.

Duda, D. M., Tu, C., Fisher, S. Z., An, H., Yoshioka, C., Govindasamy, L., Laipis, P. J., Agbandje-McKenna, M., Silverman, D. N. and McKenna, R. (2005) Human carbonic anhydrase III: Structural and kinetic study of catalysis and proton transfer. *Biochemistry 44*, 10046-10053.

Ekstedt, E., Holm, L. and Ridderstrale, Y. (2004) Carbonic anhydrase in mouse testis and epididymis; transfer of isozyme IV to spermatozoa during passage. *J. Mol. Histol. 35*, 167-173.

Elbaum, D., Nair, S. K., Patchan, M. W., Thompson, R. B. and Christianson, D. W. (1996) Structure-based design of a sulfonamide probe for fluorescence anisotropy detection of zinc with a carbonic anhydrase-based biosensor. *J. Am. Chem. Soc. 118*, 8381-8387.

Eliel, E. L. and Wilen, S. H. (1994). Stereochemistry of Organic Compounds (New York, John Wiley & Sons, Inc.).

Eriksson, A. E., Jones, T. A. and Liljas, A. (1988a) Refined structure of human carbonic anhydrase II at 2.0 A resolution. *Proteins 4*, 274-282.

Eriksson, A. E., Kylsten, P. M., Jones, T. A. and Liljas, A. (1988b) Crystallographic studies of inhibitor binding-sites in human carbonic anhydrase-II - A penta-coordinated binding of the SCN- ion to the zinc at high pH. *Proteins 4*, 283-293.

Fain, S. B., Korosec, F. R., Holmes, J. H., O'Halloran, R., Sorkness, R. L. and Grist, T. M. (2007) Functional lung imaging using hyperpolarized gas MRI. *J. Magn Reson. Imaging 25*, 910-923.

Falconer, R. J., Penkova, A., Jelesarov, I. and Collins, B. M. (2010) Survey of the year 2008: applications of isothermal titration calorimetry. *J. Mol. Recognit.*

Fisher, H. F. and Singh, N. (1995). Calorimetric methods for interpreting protein-ligand interactions. In Energetics of Biological Macromolecules (San Diego, Academic Press Inc), pp. 194-221.

Foster-Gareau, P., Heyn, C., Alejski, A. and Rutt, B. K. (2003) Imaging single mammalian cells with a 1.5 T clinical MRI scanner. *Magn. Reson. Med. 49*, 968-971.

Freyer, M. W. and Lewis, E. A. (2008). Isothermal titration calorimetry: Experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. In Biophysical Tools for Biologists: Vol 1 in Vitro Techniques, pp. 79-113.

Hakansson, K., Carlsson, M., Svensson, L. A. and Liljas, A. (1992) Structure of native and apo carbonic anhydrase-II and structure of some of its anion ligand complexes. *J. Mol. Biol. 227*, 1192-1204.

Hilvo, M., Baranauskiene, L., Salzano, A. M., Scaloni, A., Matulis, D., Innocenti, A., Scozzafava, A., Monti, S. M., Di Fiore, A., De Simone, G., Lindfors, M., Janis, J., Valjakka, J., Pastorekova, S., Pastorek, J., Kulomaa, M. S., Nordlund, H. R., Supuran, C. T. and Parkkila, S. (2008) Biochemical characterization of CA IX, one of the most active carbonic anhydrase isozymes. *J. Biol. Chem. 283*, 27799-27809.

Huang, S., Sjoblom, B., Sauer-Eriksson, A. E. and Jonsson, B. H. (2002) Organization of an efficient carbonic anhydrase: Implications for the mechanism based on structure-function studies of a T199P/C206S mutant. *Biochemistry 41*, 7628-7635.

Jason-Moller, L., Murphy, M. and Bruno, J. (2006). Overview of Biacore systems and their applications. In Current Protocols in Protein Science (John Wiley & Sons, Inc.).

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr., Sect. A: Found. Crystallogr. 47*, 110-119.

Kannan, K. K., Ramanadham, M. and Jones, T. A. (1984) Structure, refinement, and function of carbonic anhydrase isozymes: refinement of human carbonic anhydrase I. *Ann. N. Y. Acad. Sci. 429*, 49-60.

Kauczor, H. U., Surkau, R. and Roberts, T. (1998) MRI using hyperpolarized noble gases. *Eur. Radiol. 8*, 820-827.

Kivela, K. J., Kivela, J., Saarnio, J. and Parkkila, S. (2005) Carbonic anhydrases in normal gastrointestinal tract and gastrointestinal tumours. *World J. Gastroenterol. 11*, 155-163.

Krishnamurthy, V. M., Kaufman, G. K., Urbach, A. R., Gitlin, I., Gudiksen, K. L., Weibel, D. B. and Whitesides, G. M. (2008) Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev. 108*, 946-1051.

Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774-797.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK - A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283-291.

Liljas, A., Lovgren, S., Bergsten, P. C., Carlbom, U., Petef, M., Waara, I., Strandbe.B, Fridborg, K., Jarup, L. and Kannan, K. K. (1972) Crystal structure of human carbonic anhydrase-C. *Nat. New Biol. 235*, 131-137.

Lowery, T. J., Garcia, S., Chavez, L., Ruiz, E. J., Wu, T., Brotin, T., Dutasta, J. P., King, D. S., Schultz, P. G., Pines, A. and Wemmer, D. E. (2006) Optimization of xenon biosensors for detection of protein interactions. *Chembiochem 7*, 65-73.

Mincione, F., Scozzafava, A. and Supuran, C. T. (2008) The development of topically acting carbonic anhydrase inhibitors as antiglaucoma agents. *Current Curr. Pharm. Des. 14*, 649-654.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307-326.

Pastorekova, S., Parkkila, S. and Zavada, J. (2006). Tumor-associated carbonic anhydrases and their clinical significance. In Advances in Clinical Chemistry, Vol 42 (San Diego, Elsevier Academic Press Inc), pp. 167-216.

Pattnaik, P. (2005) Surface plasmon resonance: applications in understanding receptor-ligand interaction. *Appl. Biochem and Biotech. 126*, 79-92.

Potter, C. P. S. and Harris, A. L. (2003) Diagnostic, prognostic and therapeutic implications of carbonic anhydrases in cancer. *Br. J. Cancer 89*, 2-7.

Prange, T., Schiltz, M., Pernot, L., Colloc'h, N., Longhi, S., Bourguet, W. and Fourme, R. (1998) Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins: Struct., Funct., Bioinf. 30*, 61-73.

Roy, V., Brotin, T., Dutasta, J. P., Charles, M. H., Delair, T., Mallet, F., Huber, G., Desvaux, H., Boulard, Y. and Berthault, P. (2007) A cryptophane biosensor for the detection of specific nucleotide targets through xenon NMR spectroscopy. *Chemphyschem 8*, 2082-2085.

Ruiz, E. J., Sears, D. N., Pines, A. and Jameson, C. J. (2006) Diastereomeric Xe chemical shifts in tethered cryptophane cages. *J. Am. Chem. Soc. 128*, 16980-16988.

Schlundt, A., Kilian, W., Beyermann, M., Sticht, J., Gunther, S., Hopner, S., Falk, K., Roetzschke, O., Mitschang, L. and Freund, C. (2009) A xenon-129 biosensor for monitoring MHC-peptide interactions. *Angew. Chem., Int. Ed. 48*, 4142-4145.

Schroder, L., Lowery, T. J., Hilty, C., Wemmer, D. E. and Pines, A. (2006) Molecular imaging using a targeted magnetic resonance hyperpolarized biosensor. *Science 314*, 446-449.

Spence, M. M., Rubin, S. M., Simitrov, I. E., Ruiz, E. J., Wemmer, D. E., Pines, A., Yao, S. Q., Tian, F. and Schultz, P. G. (2001) Functionalized xenon as a biosensor. *Proc. Natl. Acad. Sci. U. S. A. 98*, 10654-10657.

Storoni, L. C., McCoy, A. J. and Read, R. J. (2004) Likelihood-enhanced fast rotation functions *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 432-438.

Supuran, C. T. (2008) Carbonic anhydrases - An overview. *Curr. Pharm. Des. 14*, 603-614.

Supuran, C. T. and Scozzafava, A. (2007) Carbonic anhydrases as targets for medicinal chemistry. *Bioorg. Med. Chem. 15*, 4336-4350.

Szebenyi, D. M. E., Arvai, A., Ealick, S., LaIuppa, J. M. and Nielson, C. (1997) A system for integrated collection and analysis of crystallographic diffraction data. *J. Synch. Rad. 4*, 128-135.

Taratula, O. and Dmochowski, I. J. (2009) Functionalized 129Xe contrast agents for magnetic resonance imaging. *Curr. Opin. Chem. Biol. 14*, 1-8.

Watanabe, T. (1965) Measurement of L absorption spectra of xenon. *Physical Review 137*, 1380-1382.

Whittington, D. A., Grubb, J. H., Waheed, A., Shad, G. N., Sly, W. S. and Christianson, D. W. (2004) Expression, assay, and structure of the extracellular domain of murine carbonic anhydrase XIV: Implications for selective inhibition of membrane-associated isozymes. *J. Biol. Chem. 279*, 7223-7228.

Whittington, D. A., Waheed, A., Ulmasov, B., Shah, G. N., H., G. J., Sly, W. S. and Christianson, D. W. (2001) Crystal structure of the dimeric extracellular domain of human carbonic anhydrase XII, a bitopic membrane protein overexpressed in certain cancer tumor cells. *Proc. Natl. Acad. Sci. U. S. A. 98*, 9545-9550.

Wiseman, T., Williston, S., Brandts, J. F. and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem. 179*, 131-137.

Part II: Structural and Functional Studies of the Sesquiterpene Cyclase *epi*-Isozizaene Synthase

Chapter 4: Introduction

4.1 Terpenes and Terpene Synthases

The terpenome is comprised of a family of more than 55,000 structurally and stereochemically diverse natural products, all of which ultimately derive from the universal 5 carbon precursors dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) (Figure 4.1) (Christianson, 2007). Increasingly longer polyisoprenoids are formed by the coupling of DMAPP and IPP, in a head to tail fashion to form geranyldiphosphate, farnesyl diphosphate, gernarylgeranyl diphosphate, and geranylfarnesyl diphosphate, the linear precursors of the mono-, sesqui-, di-, and sester-terpenes, respectively (Figure 4.1) (Tholl, 2006). Terpenoids are ubiquitous throughout nature and serve a multitude of specific functions in plants, animals, insects, bacteria and fungi. For example, terpenoids are critical for plant survival and account for a large number of primary metabolites, including molecules involved in photosynthesis, respiration, and membrane structure. Terpenoids also account for a wide range of secondary metabolites in plants, where they bestow unique flavors and fragrances, provide chemical defense against pests, and facilitate interactions between plants and other organisms (Aharoni, 2005; Pichersky, 2006). From a medicinal perspective, terpenoids are of great interest because many of these natural products exhibit anti-cancer, anti-malarial, and anti-microbial activities (Aharoni, 2005). Furthermore, two molecules of farnesyldiphosphate can be coupled together via a head-to-head condensation to form squalene, the triterpene linear precursor to the steroids (Abe, 1993).

A family of enzymes known as terpenoid synthases, are responsible for the tremendous structural diversity of the terpenoids. Terpenoid synthases can be divided into two categories: class I enzymes adopt the FPP synthase α-helical fold and initiate catalysis by metal triggered ionization of the substrate diphosphate group, and class II enzymes adopt an unrelated double α-barrel fold and initiate catalysis by protonation of an epoxide ring or carbon-carbon double bond (Figure. 4.2). Class I terpenoid synthases can be further subdivided into three categories: coupling enzymes that catalyze chain elongation reactions to form increasingly longer polyisoprenoid diphosphates, coupling enzymes that catalyze irregular (i.e., non-head-to-tail) isoprenoid condensation reactions such as cyclopropanation, cyclobutanation, or branching reactions (Thulasiram, 2007), and cyclization enzymes that catalyze the conversion of linear isoprenoid substrates into single and multi-ringed hydrocarbon products (Croteau, 1985; Cane, 1990; Lesburg, 1998; Wendt, 1998). The potential diversity of carbon-carbon bond formation afforded by the flexible linear isoprenoid substrate, and the chemical potential for subsequent biosynthetic functionalization of cyclic terpenoids by cytochrome P450, monooxygenases, etc., make terpenoid biosynthesis an attractive system for engineering novel compounds (Aharoni, 2005; Yoshikuni, 2006; Austin, 2008).

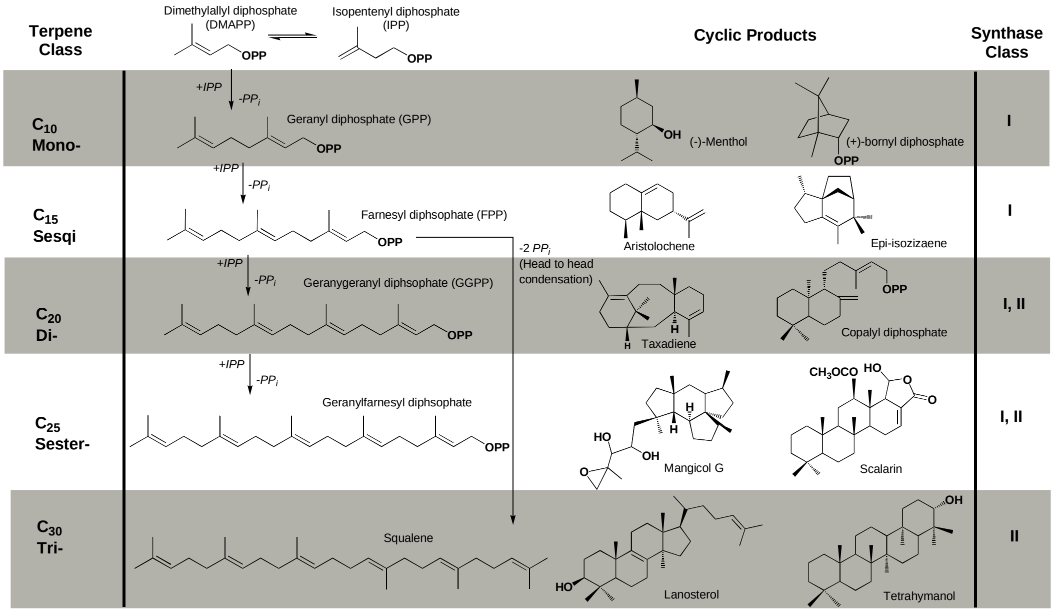

Terpene Class
Dimethylallyl diphosphate (DMAPP)
OPP
Isopentenyl diphosphate (IPP)
OPP
Cyclic Products
Synthase Class
+IPP
-PP$_i$
C$_{10}$ Mono-
Geranyl diphosphate (GPP)
OPP
OH
(-)-Menthol
(+)-bornyl diphosphate
OPP
I
+IPP
-PP$_i$
C$_{15}$ Sesqi
Farnesyl diphsophate (FPP)
OPP
Aristolochene
Epi-isozizaene
I
-2 PP$_i$
(Head to head condensation)
+IPP
-PP$_i$
C$_{20}$ Di-
Geranygeranyl diphsophate (GGPP)
OPP
H
H
Taxadiene
OPP
Copalyl diphosphate
I, II
+IPP
-PP$_i$
C$_{25}$ Sester-
Geranylfarnesyl diphsophate
OPP
H
H
HO
OH
O
Mangicol G
CH$_3$OCO
HO
O
O
Scalarin
I, II
C$_{30}$ Tri-
Squalene
HO
Lanosterol
OH
Tetrahymanol
II

Figure. 4.1. General scheme of terpenoid nomenclature and biosynthesis (OPP = diphosphate, PP$_i$ = inorganic diphosphate).

The crystal structures of several class I terpenoid coupling and cyclization enzymes have been solved, revealing a conserved α-helical terpenoid synthase fold across the domains of life. Structures of enzyme complexes with substrates, inhibitors, and/or products have also revealed the universal conservation of a trinuclear metal cluster implicated in the molecular recognition of the substrate diphosphate group as well as the initiation of catalysis. Metal ions are coordinated by metal binding motifs on opposing helices near the mouth of the active site. The metal binding motifs are generally described as either "aspartate-rich" [**D**DXX(XX)**D/E**] or as a secondary metal binding motif "NSE/DTE" [(**N,D**)D(L,I,V)X(**S,T**)XXX**E**], in which boldface residues typically coordinate to catalytically obligatory Mg^{2+} or Mn^{2+} ions (where metal ligands are indicated in boldface) (Christianson, 2006). X-ray crystal structures have been instrumental in understanding the catalytic mechanisms of terpenoid synthases: the active site of each synthase provides a template that binds the flexible substrate(s) in the proper orientation and conformation so that, upon the departure of the diphosphate leaving group and resultant generation of a reactive carbocation, the active site template ensures a specific trajectory of intermolecular and intramolecular carbon-carbon bond formation in the ensuing cyclization cascade (Christianson, 2008).

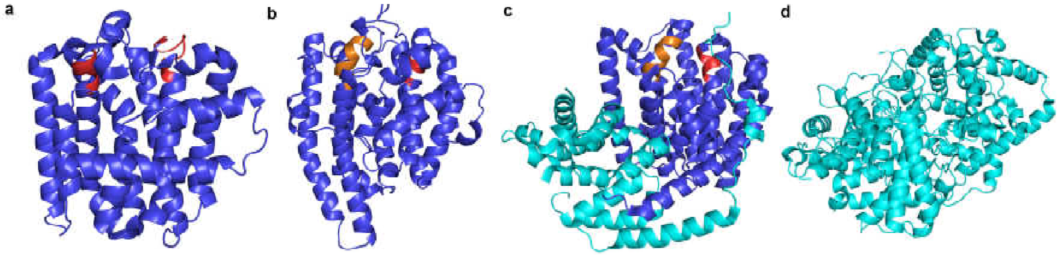


Figure 4.2. Structural similarities among various terpenoid synthases define the core class I terpenoid cyclase fold (blue). Conserved metal binding motifs are the aspartate-rich motifs (red) and “NSE/DTE” motifs (orange) highlighted in (a) *E. coli* FPP synthase (PDB code 1RQI), (b) *epi*-isozizaene synthase (PDB code 3KB9), and (c) (+)-bornyl diphosphate synthase (PDB code 1N22), which contains an additional N-terminal domain (cyan). This α-helical domain is topologically similar to the α-barrel fold of the class II terpenoid cyclases, which occurs in a double domain architecture in the triterpene cyclase (d) oxidosqualene cyclase (PDB code 1W6K).

4.2 Streptomyces

Streptomyces are gram-positive, filamentous, saprophytic, soil-dwelling bacteria. To-date, more than 500 species of *Streptomyces* have been identified. In addition to their central role in carbon recycling, *Streptomyces* are also characterized by their complex secondary metabolism (Challis, 2003). *Streptomyces* are a very abundant source of antibiotics, amazingly, over two-thirds of naturally derived antibiotics currently in use are produced by *Streptomyces* (Bentley, 2002). Of the thousands of secondary metabolites that are isolated from *Streptomyces* many are polyketides (ie. tetracycline) (Pickens, 2009) and aminoglycosides (ie. neomycin) (Kudo, 2009), however very few are cyclic terpenoids. One such example of a cyclic terpenoid is pentalenonelactone, an antibiotic derived from the sesquiterpene pentalenene, produced by a number of *Streptomyces* strains (Cane, 1994). Prior to this work, the X-ray crystal structure of pentalenene synthase (Lesburg, 1997) was the only known structure of a bacterial sesquiterpene cyclase. However, the structure of pentalenene synthase, determined at 2.6 Å resolution, was is in an open, unliganded conformation, and therefore did not provide evidence of a conserved trinuclear metal cluster among bacterial sesquiterpene cyclases. Additional X-ray crystal structures of bacterial terpenoid cyclases are necessary to draw conclusions about the evolutionary relationships among these enzymes. Furthermore, comparative studies of the harmless *Streptomyces* genus with other members of the Actinomycetales order, for example disease causing *Mycobacterium tuberculosis* (tuberculosis) and *Mycobacterium leprae* (leprosy), may offer insight into the treatment of these pathogens (Bentley, 2002).

The complete genome of *Streptomyces coelicolor* A3(2) was sequenced in 2002, and contained a surprisingly high 7,825 predicted genes, compared with 4,289 genes in the Gram-negative bacterium *Escherichia coli*; and a predicted 31,780 in humans (Bentley, 2002). The genome contains many genes involved in secondary metabolism including polyketide synthases, chalcone synthases, non-ribosomal peptide synthetases as well as several gene clusters coding for terpene synthesis including geosmin, hopanoid, and albaflavenone biosynthesis (Bentley, 2002).

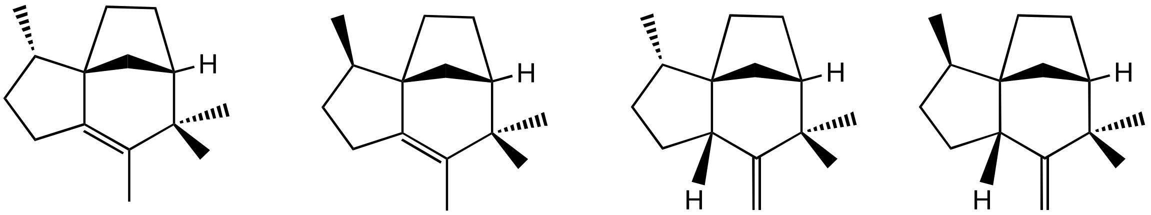
4.3 epi-Isozizaene

epi-Isozizaene is a member of a unique family of tricyclic sesquiterpene, like its parent hydrocarbon, zizaene (Coates, 1972), it has a highly strained ring system including a quaternary center (Figure 4.3). *epi*-Isozizaene was first observed as a natural product from the bacteria *Streptomyces coelicolor* A3(2), where it is formed by a novel terpene cyclase, *epi*-isozizaene synthase (EIZS). EIZS was first characterized by Lin and Cane (Lin, 2006), due to its 23.8% sequence identity with pentalenene synthase, a sesquiterpene cyclase isolated from *Streptomyces exfoliates* UC5319. Furthermore, the EIZS protein contained the two conserved Mg^{2+}-binding motifs, an aspartate-rich motif (**D**^{99}**D**RH**D**) and the secondary metal binding motif (**N**^{240}DLC**S**LPK**E**). EIZS catalyzes the Mg^{2+}-dependant cyclization of the FPP (the linear precursor of the sesquiterpenes).

The proposed mechanism for the cyclization of farnesyl diphosphate to *epi*-isozizaene was elucidated by 1-D and 2-D NMR analysis of the products isolated from the incubation of EIZS with several isotopically labeled substrates, namely [1,1-$^2H_2$]-

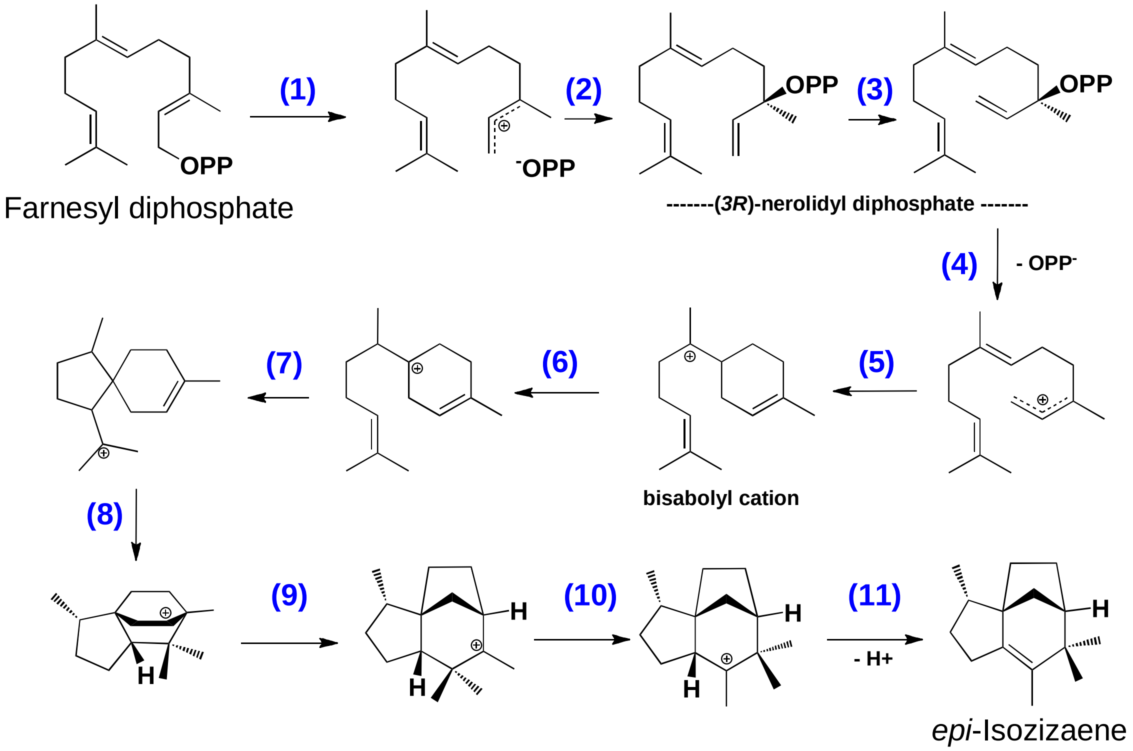
FPP, (1*R*)-[1-^{2}H]-FPP and (1S)-[1-^{2}H]-FPP (Figure 4.4) (Lin, 2006). Recently, new mechanistic insights into the *epi*-isozizaene folding pathway have been provided by computational quantum chemistry (Hong, 2009). Using the computational program GAUSSIAN03 (Frisch, 2003) to conduct a thorough analysis of carbocation intermediates and transition states from several sesquiterpene cyclization pathways, Hong and Tantillo suggest that the conformation of the bisabolyl cation attainable in the enzyme active site is a primary determinate of the structure and stereochemistry of the resultant sesquiterpenes. They report four unique bisabolyl cation conformers; each proposed to be involved in formation of a specific set of sesquiterpene products. Surprisingly, outwardly related products, for example *epi*-isozizaene and isozizaene, are proposed to be formed by different conformers of the bisabolyl cation, which vary in the orientation of the acyclic hydrocarbon chain. However, some folding pathways appear to be somewhat permissive, certain products, including *epi*-isozizaene, can be formed via more than one bisabolyl cation conformer.

Using an arbitrarily defined zero-energy bisabolyl conformer, **A0**, as a basis for their calculations, Hong and Tantillo present a detailed theoretical cyclization scheme for the formation of *epi*-isozizaene, consisting of 6 cationic intermediates, beginning from (3*R*)-nerolidyl diphosphate in the cisoid conformation (Figure 4.5). The first intermediate, **A1**, is bisabolyl cation conformer formed by 1,6-cyclization of cisoid (3*R*)-nerolidyl diphosphate. The **A1** conformer is in a productive conformation to undergo a [1,2]-hydride shift between carbons 6 and 7 via the exterior face of the acyclic chain, to form cationic intermediate **B1**. Intermediate **B1** is in an appropriate conformation to then undergo a subsequent 6,10-cyclization via attack of the C10=C11 π-bond on the pro-*R* face of the C6 cationic center for form the acorenyl cation, **C1**. Next, a thermodynamically favorable rearrangement to the **C2** acorenyl cation confomer facilitates the direct conversion of **C2** to **E1** via a concerted cyclization and alkyl shift, circumventing the formation of a discrete secondary cation. The final intermediate, **F1**, is the result of a [1,2]-methyl shift of either C12 or C13, via a bridged nonclassical carbocation transition state. The final product, *epi*-isozizaene is formed by direct deprotonation of **F1** at C10. Although the total calculated energy change for the pathway depends on the density functional theory method chosen, the overall rearrangement beginning from the bisabolyl cation is considerably exothermic, and involves a minimal number of conformational changes between steps. Overall, the theoretical study of the *epi*-isozizaene folding pathway is consistent with that presented by Lin and Cane, and offers additional insight into several steps in the pathway, namely the concerted 3,11-cyclization and C4 alkyl shift going from **C2** to **E1**, and the approximately equal energy for the [1,2]-methyl shift of C12 or C13 to form **F1** (Hong, 2009). This study illustrates the utility of quantum calculations for studying terpenoid cyclase reaction coordinates. By expanding theoretical calculations to include contributions from an enzyme active site, we can continue to develop an understanding of terpenoid cyclase structure-function relationships and work towards engineering novel, efficient terpenoid cyclases.



epi-Isozizaene Isozizaene *epi*-Zizaene *Z*izaene

Figure 4.3. The structures of zizaene sesquiterpenes.



(1)
(2)
(3)
(4)
- OPP-
(5)
(6)
(7)
(8)
(9)
(10)
(11)
- H+
OPP
-OPP
OPP
OPP
Farnesyl diphosphate
-------(3R)-nerolidyl diphosphate -------
bisabolyl cation
H
H
H
H
H
H
epi-Isozizaene

Figure 4.4. Proposed mechanism for the formation of *epi*-isozizane from FPP by EIZS. **(1)** Ionization of FPP. **(2)** Isomerization will give (3*R*)-nerolidyl diphosphate. **(3)** Rotation about the newly generated C-2/C-3 bond generates the corresponding *cisoid* (3*R*)-nerolidyl diphosphate conformer. **(4)** Ionization of (3*R*)-nerolidyl diphosphate. **(5)** Cyclization to form bisabolyl cation. **(6)** A 1,2-hydride shift. **(7)** Spirocyclization. **(8)** Cyclization. **(9)** Ring contraction. **(10)** Methyl migration. **(11)** Deprotonation to yield (+)-*epi*-isozizaene.

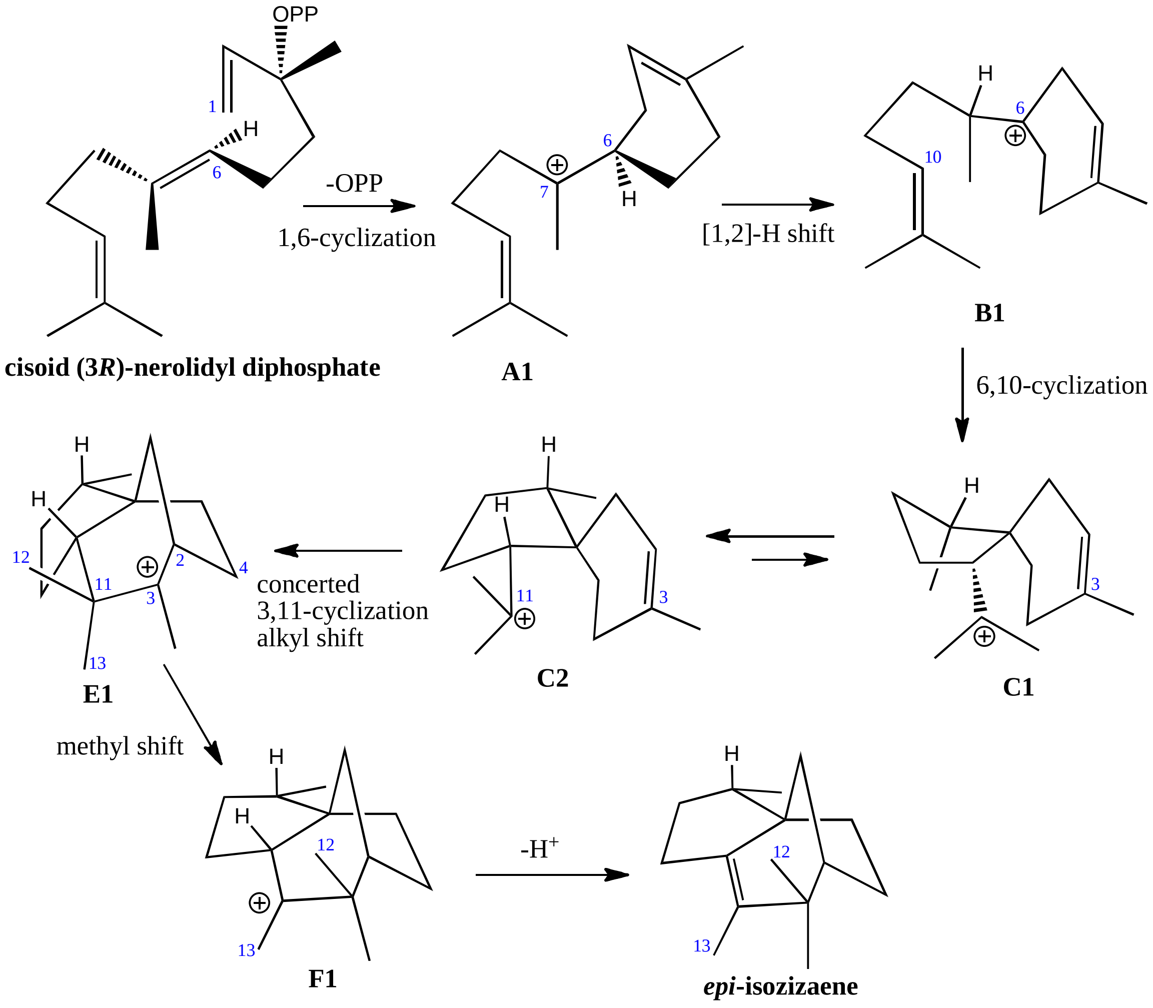

OPP
1
H
6
cisoid (3*R*)-nerolidyl diphosphate
-OPP
1,6-cyclization
7
6
H
A1
[1,2]-H shift
H
6
10
B1
6,10-cyclization
H
3
C1
H
H
11
3
C2
concerted
3,11-cyclization
alkyl shift
H
H
12
11
3
2
4
13
E1
methyl shift
H
H
12
13
F1
-H+
H
12
13
epi-isozizaene

Figure 4.5. Proposed *epi*-isozizaene cyclization scheme based on quantum chemical calculations (Hong, 2009). The C12 1,2-methyl shift (**E1** to **F1** step) is preferred over a C13 1,2-methyl shift by ~2kcal/mol, however both are energetically accessible and could be affected by the active site of the enzyme.

In the *Streptomyces colicolor* A3(2) genome, EIZS is transcriptionally coupled to to cytochrome P450 170A1 (CYPA170 A1) (Zhao, 2008). Cytochrome P450 monooxygenases belong to a superfamily of heme-containing proteins that catalyze redox reactions. Specifically, P450 monooxygenases catalyze the oxidation of organic products using atmospheric dioxygen as the oxygen source and electrons from NAD(P)H, and producing a molecule of water as a side product (Bernhardt, 2006). CYP170A1 from *S. coelicolor* carries out a two-step allylic oxidation to convert *epi*-isozizaene to an epimeric mixture of albaflavenols and ultimately to the sesquiterpene antibiotic albaflavenone (Figure 4.6). Therefore, *epi*-isozizaene has been identified as an intermediate in albaflavenone synthesis, and is only detected in bacterial extracts in CYP170A1 knockout strains of *S. coelicolor.* The final product of the two-gene cluster, albaflavenone, has also been isolated from *S. albidoflavus*, and exhibits modest antibacterial activity against *Bacillus subtilis* (Zhao, 2008).

EIZS was chosen as a target for structure determination via protein X-ray crystallography in order to investigate the structural changes that occur upon the binding of the three Mg^{2+} ions and the substrate (FPP) or substrate analogues, triggering active site closure and substrate ionization. Previous work has indicated that the details of these structural changes generally differ between plant (Starks, 1997; Gennadios, 2009) and fungal (Rynkiewicz, 2001; Shishova, 2007) sesquiterpene cyclases. Until now, the mechanism of active site closure of a bacterial terpenoid cyclase has remained unknown since the only available crystal structure of a bacterial cyclase has been that of *S.*

exfoliatus UC5319 pentalenene synthase, which was determined only in an open, ligand-free conformation (Lesburg, 1997).

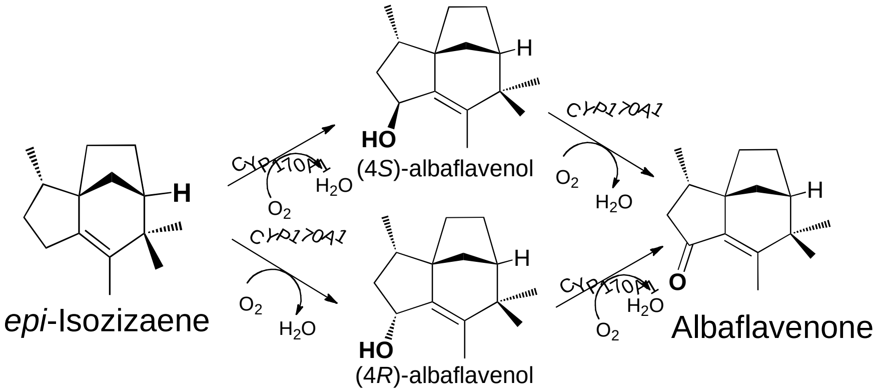

epi-Isozizaene
H
CYP170A1
O2
H2O
HO
H
(4S)-albaflavenol
CYP170A1
O2
H2O
HO
H
(4R)-albaflavenol
CYP170A1
O2
H2O
CYP170A1
O2
H2O
O
H
Albaflavenone

Figure 4.6. Albaflavenone biosynthetic pathway in *S. coelicolor*. EIZS catalyzes the Mg^{2+} dependent cyclization of FPP to form *epi*-isozizaene, then CYP170A1 catalyzes a two-step allylic oxidation to albaflavenone.

Chapter 5: X-Ray Crystal Structure of *epi*-Isozizaene Synthase from *Streptomyces coelicolor*

5.1 Expression and Purification

Recombinant EIZS from *Streptomyces coelicolor* A3(2) was expressed at high levels in *Escherichia coli* BL21(DE3) and purified as previously described (Lin, 2006) with minor modifications. Briefly, *E. coli* BL21(DE3) carrying pET28a(+)/SCO5222 was inoculated into Luria-Bertani (LB) medium containing kanamycin and grown overnight at 37 °C. A total of 4 L of LB/kanamycin medium was inoculated with 5 mL of the overnight seed culture, and *E. coli* was grown at 37 °C until the OD_{600} reached 0.5. The temperature was reduced to 20 ^{o}C and the cells were induced with 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 18 h. Cells were harvested, resuspended in 50 mL of Talon buffer A [50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 20 % glycerol and 5 mM β-mercaptoethanol (BME)], supplemented with phenylmethylsulfonyl fluoride and DNase, and sonicated for 6 min using a 40 % duty cycle and power range 30 %. After three cycles of sonication, the cell lysate was clarified by centrifugation at 16000*g* and 4 °C for 75 min. The clarified lysate was loaded on a Talon (Clontech) Co^{2+} metal affinity resin (5 mL), and a step gradient from 0 to 200 mM imidazole in Talon buffer was applied to elute the enzyme. The fractions were analyzed using SDS-PAGE, and the most concentrated fractions were pooled and applied to a Superdex gel filtration column (HiLoad 26/60 Superdex, GE Healthcare) equilibrated in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10 % glycerol, 10 mM $MgCl_2$ and 2 mM Tris(2-carboxyethyl)phosphine (TCEP). The fractions were analyzed by SDS-PAGE and the fractions containing the enzyme were pooled and concentrated to 8 mg/mL enzyme using a YM-10 centricon. The resulting protein preparation was >99 % pure on the basis of SDS-PAGE.

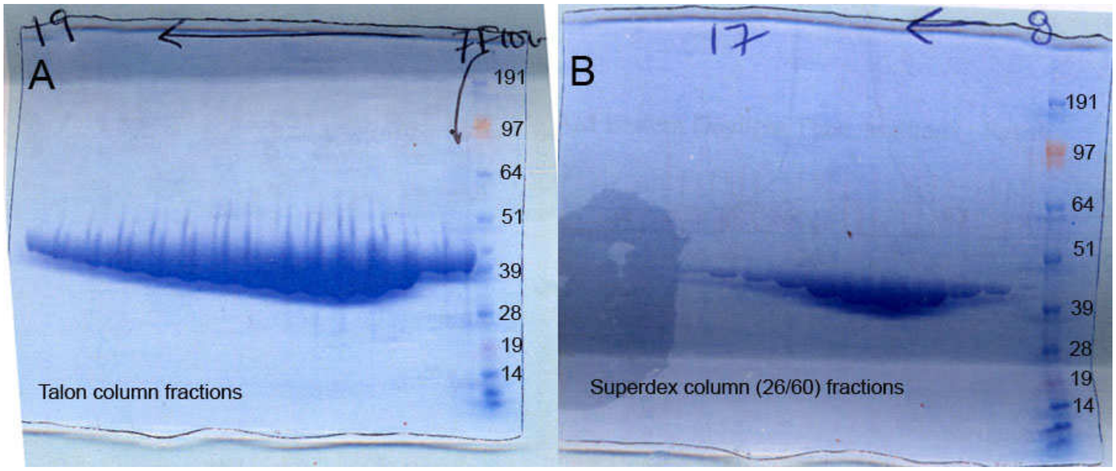

19
Flow
A
191
97
64
51
39
28
19
14
Talon column fractions
17
8
B
191
97
64
51
39
28
19
14
Superdex column (26/60) fractions

Figure 5.1. SDS-PAGE analysis of the purification of *epi*-isozizaene synthase. (A) Fractions from the Talon column: Molecular weight markers are shown on in the right lane, and the column flow-through is in the second lane from the right. The strong band is due to *epi*-isozizaene synthase (44 kDa). (B) Fractions from the Superdex (2660) column. Molecular weight markers are shown on in the right lane.

5.2 Crystallization

EIZS was crystallized by the sitting drop, and hanging drop vapor diffusion methods. Protein was freshly filtered using a 0.22 μm filter prior to setting up drops. Typically, a 4 μL drop of protein solution [8-10 mg/mL EIZS, 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10 mM $MgCl_2$, 10 % glycerol, 2 mM TCEP, 2 mM sodium pyrophosphate, 2 mM benzyltriethylammonium chloride] was added to 4 μL of precipitant solution [100 mM Bis-Tris (pH 5.5), 25-28 % polyethylene glycol 3350, 0.2 M $(NH_4)_2SO_4$] and equilibrated against a 1 mL well reservoir of precipitant solution. Initial crystallization conditions were identified using the commercially available Index screen from Hampton Research (condition #66). Crystals appeared within 2-3 days and grew to maximal dimensions of 100 μm × 10 μm × 10 μm. Initial crystals diffracted to 2.15 Å at the home source, where the space group and unit cell parameters were first determined. Higher resolution data was then collected at the synchrotron. Crystals diffracted X-rays to 1.60 Å resolution at the Advanced Photon Source, beamline NE-CAT 24-ID-C (Argonne, IL), and belonged to space group $P2_1$ with unit cell parameters a = 53.185 Å, b = 47.374 Å, c = 75.376 Å and $\beta = 95.53^o$; with one monomer in the asymmetric unit, the Matthews coefficient V_M = 2.1 $Å^3$/Da, corresponding to a solvent content of 43 %.

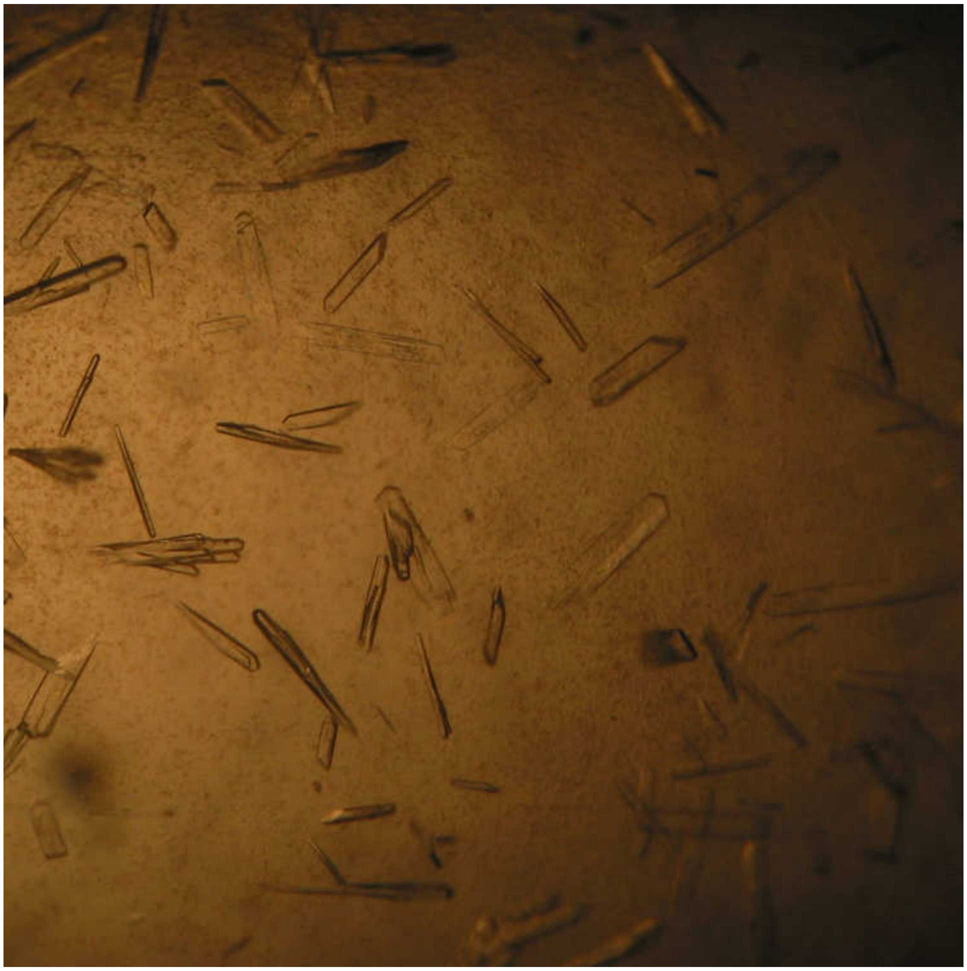


Figure 5.2. Crystals of *epi*-isozizaene synthase.

5.3 Structure Determination with Heavy Atoms

5.3.1 Introduction

A crystal structure can be described as the best-fit model to a contour map of the electron density throughout the unit cell. The electron density, $\rho(x,y,z)$, can be represented by equation **5.1**, a periodic function represented by a Fourier series using the

$$\rho(x,y,z) = 1/\mathrm{V} \sum_{h}\sum_{k}\sum_{l} \boldsymbol{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \qquad \mathbf{5.1}$$

structure factors, $\boldsymbol{F}_{hkl}$. A structure factor is a complete description of a diffracted X-ray. Since an X-ray can be described as a wave, it has a frequency, amplitude, and phase. To calculate the electron density according to equation **5.1**, and thus build a model of the protein structure, the frequency, amplitude and phase of each reflection, *hkl*, must be known. When an X-ray diffracts off a protein crystal, the frequency of the X-ray does not change, therefore the frequency is equal to that of the X-ray source. The amplitude of the diffracted X-ray is proportional to $(\boldsymbol{I}_{hkl})$, the square root of the measured intensity of reflection *hkl* (Rhodes, 2000). However, the phase of the diffracted X-ray is unknown, and is the missing piece of the crystallography puzzle that must be found to complete a crystal structure determination.

Protein crystallographers have developed several methods to obtain the “lost” phase information from the diffraction data. A very commonly used method is known as molecular replacement (MR), when the phases of a similar, known structure are used as the initial phase estimates for the new structure. This method works best when a homologous structure is known, typically the phasing model has >25 % sequence identity and an r.m.s. deviation of <2.0 Å between the Cα atoms of the model and the new structures (Taylor, 2003). When MR is not successful, other methods that rely upon the incorporation of heavy atoms to the crystal lattice are used. When heavy atoms are incorporated into a crystal lattice, measurable intensity changes are observed in the diffraction pattern, with respect to the native diffraction pattern. Upon incorporation of the heavy atoms, if the crystal lattices remain isomorphous, except for the presence of the heavy atom, the isomorphous difference between the reflection amplitudes of native and derivative crystals can be used as an estimate of the heavy-atom structure-factor amplitude, which can be used to determine the positions of the heavy-atoms. Once the phases of the heavy atoms are known, a first approximation of the phase of the native protein can be made. When a single heavy atom derivative is used, this method is known as single isomorphous replacement (SIR). This method can be extended, when necessary, to include information from more than one heavy atom derivative, and is known as multiple isomorphous replacement (MIR) (Taylor, 2003). SIR and MIR are typically very effective methods for structure determination, however in some cases the incorporation of heavy atoms into native crystals causes unexpected protein conformational changes. These conformational changes are sometime evident from changes in the unit cell parameters, however in other cases non-isomorphism is not apparent, yet critically hampers structure solution.

In addition to their mere presence, heavy atoms can additionally aid in structure solution due to their ability to anomalously scatter X-rays. Heavy atoms absorb X-rays at specific wavelengths, known as the element's absorption edge, resulting in unequal intensities between the *hkl* and –*h* –*k* –*l* reflections. The anomalous difference can be used to determine the location of the anomalous scattering atom in the unit cell, thus providing phase information. Single-wavelength anomalous diffraction (SAD), can be advantageous since it does not require an isomorphous native data set. Also, anomalous scattering methods can be extended include multiwavelength anomalous diffraction (MAD), where several data sets are collected from the same crystal at different wavelengths. Furthermore, isomorphous and anomalous heavy atom methods can be combined, for example in SIRAS (single isomorphous replacement with anomalous scattering), and MIRAS (multiple isomorphous heavy-atom replacment using anomalous scattering) (Taylor, 2003). Some of the most common heavy atoms used for phasing are mercury, platinum, gold, and uranium. These metals are typically incorporated into the crystal lattice by soaking preformed crystals in a precipitant solution supplemented with a millimolar concentration of the metal ion, where they typically interact with surface cysteine, histidine, and methionine residues (Rhodes, 2000). Selenium can also be used as a heavy atom for phasing, and can be incorporated into the protein itself by expressing recombinant protein in media supplemented with selenomethionine (Hendrickson, 1990). Recently, xenon has also been explored as a heavy atom for phasing. Protein crystals are briefly pressurized with xenon gas prior to flash cooling, causing xenon atoms to be trapped in hydrophobic pockets within the protein, which can be used for phasing (Evans, 2003).

5.3.2 Results

A number of heavy metal compounds were tested to identify a good heavy metal derivative for phasing. These heavy metal compounds included thimerosal

($C_9H_9HgNaO_2S$), mercuric chloride ($HgCl_2$), mersalyl acid ($C_{13}H_{18}HgNO_6$), ethylmercuryl chloride (C_2H_5HgCl), methylmercuric acetate ($C_3H_6HgO_2$), ytterbium chloride (YbCl3), PIP (di-μ-iodobis(ethylenediamine)diplatinum (II) nitrate) ($C_4H_{16}I_2N_6O_6Pt_2$), dipotassium platinum hexachloride (K_2PtCl_6), samarium chloride ($SmCl_2$), manganese chloride ($MnCl_2$), and sodium bromide (NaBr). Additionally, selenomethionine containing EIZS protein was expressed and crystallized.

Of the many heavy metal derivatives that were investigated, two mercury compounds were identified which were used for successful phasing. Mercury derivatives were prepared by soaking crystals in precipitant buffer supplemented with 1 mM ethyl mercury chloride or methylmercuric acetate for 1-2 days, followed by flash cooling. Diffraction data were collected at the absorption edge of mercury (1.0548 Å) on beamline X29 at the National Synchrotron Light Source (NSLS). A sample diffraction image of a methylmercuric acetate dervatized EIZS crystal is shown in Figure 5.3. Complete data sets of ethylmercury chloride, and methylmercuric acetate derivatized EIZS crystals were collected. Crystals diffracted X-rays to 1.90 Å and 2.10 Å, respectively and belonged to space group $P2_1$ with unit cell parameters in Table 5.1; data reduction was achieved with Denzo and Scalepack (Otwinowski, 1997), derivative data sets were scaled anomalously.

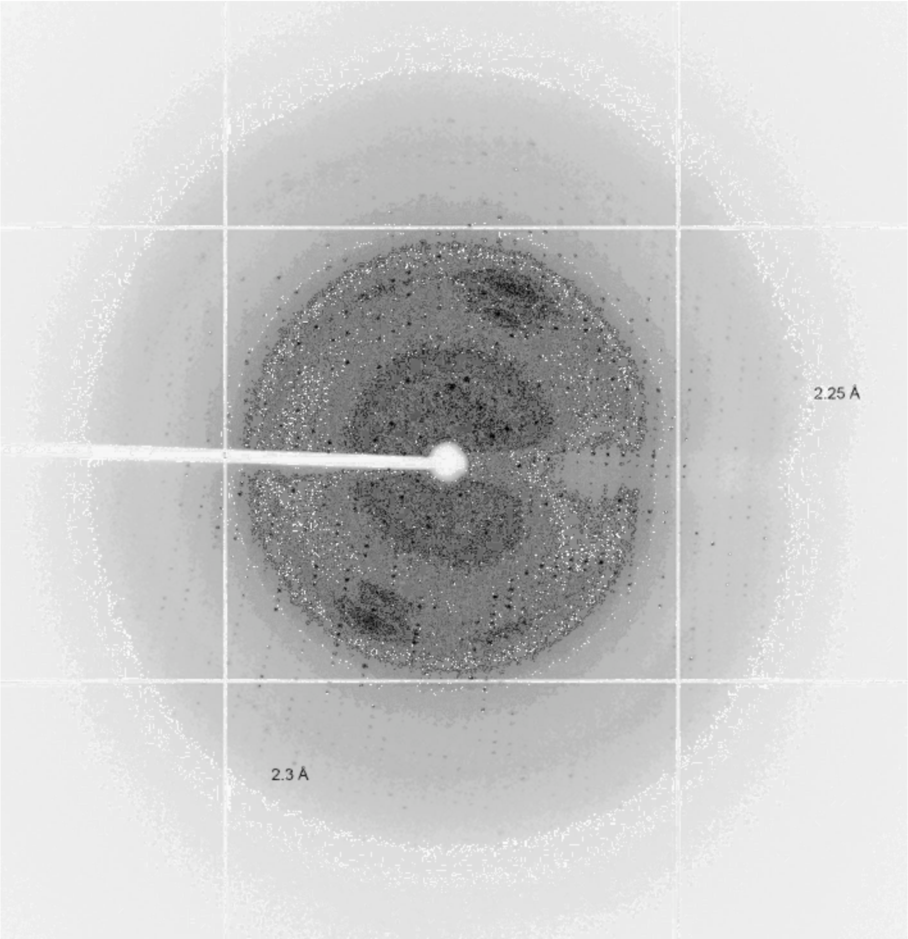


Figure 5.3. Diffraction pattern of methylmercuric acetate derivatized EIZS crystals, collected at beamline X29, National Synchrotron Light Source. The crystal diffracted to 2.1 Å resolution.

Table 5.1. Data collection statistics for native and mercury derviatized EIZS crystals. The space group of all crystals is $P2_1$.

EIZS Structure	Native	Ethylmercury chloride	Methylmercuric acetate
Data			
Wavelength, Å	1.0080	1.0548	1.075
Resolution, Å	1.60	1.90	2.10
Completeness*, %	92.9 (100)	99.4 (98.6)	99.6 (97.8)
Redundancy*	3.6 (3.6)	5.8 (5.1)	3.5 (3.0)
R_{sym}*†	0.058 (0.208)	0.075 (0.325)	0.108 (0.390)
I/σ	34 (8.3)	18 (4.4)	10 (2.7)
Unit Cell Parameters			
a, Å	53.185	51.693	52.931
b, Å	47.374	46.549	46.462
c, Å	75.376	75.551	75.689
β, (°)	95.53	92.338	94.992

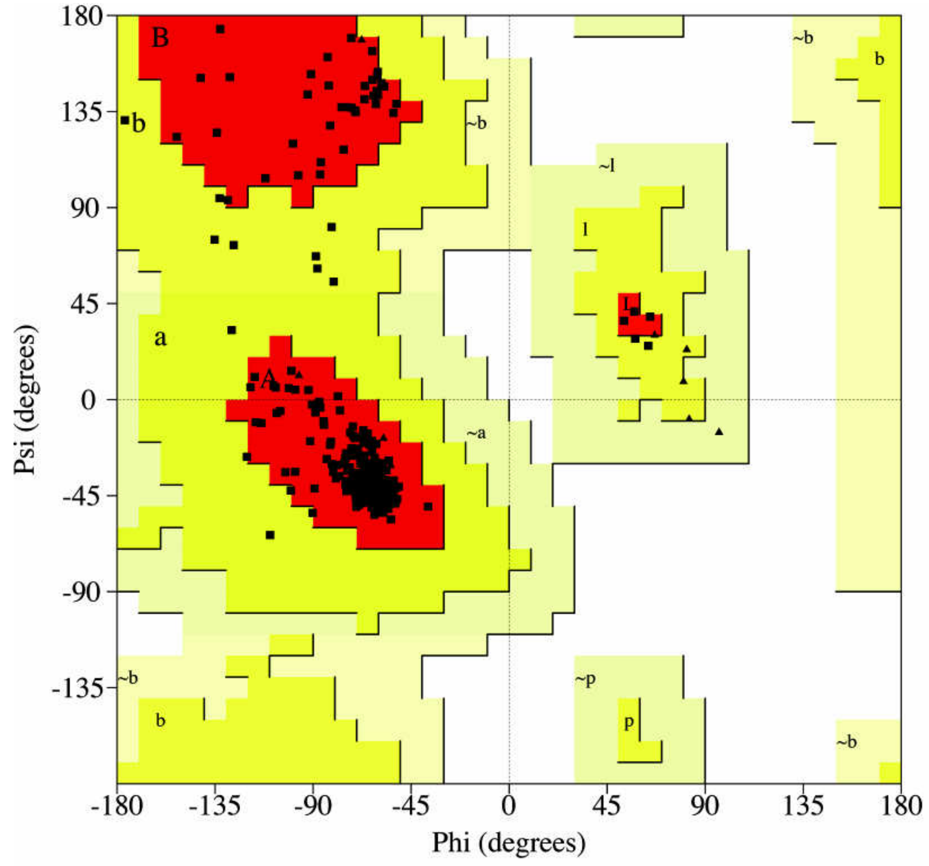
*Values in parentheses refer to the highest shell

† $R_{sym} = \sum|I_h - \langle I_h \rangle|/\sum I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent refections.

Difference Patterson maps generated with anomalous scattering data revealed large peaks corresponding to mercury. Initial attempts to solve the crystal structure using single or multiple isomorphous replacement were not successful due to apparent non-isomorphism between the native and mercury-derivatized crystals. The program autoSHARP (Vonrhein, 2007) was ultimately used for structure determination. AutoSHARP is an automated structure solution program built around the phasing program SHARP (Bricogne, 2003), the density modification program SOLOMON (Abrahams, 1996), and the ARP/wARP package (Langer, 2008) for automated model building and refinement using REFMAC (Skubak, 2004). The following data is required to begin an autoSHARP run, the type of heavy atom present in the crystal, the number of expected heavy atom sites, the space group and wavelength of data collection. The file format of the reflection data input can be either MTZ or SCALEPACK. Additionally, the contents of the asymmetric can be input as a molecular weight, number of residues or the amino acid sequence, the later enables automatic building of the structure.

The initial electron density map was phased by single-wavelength anomalous dispersion (SAD) of with the 1.90 Å data set of the ethylmercury chloride derivative using autoSHARP. Four mercury sites were identified, and the phase correlation coefficient calculated by autoSHARP for this data set was 0.466, with anomalous phasing power 0.447. Following density modification and automatic building, approximately 50% of the protein residues were built into the electron density map. At this point, the molecular model of EIZS was refined against the 1.60 Å resolution data set collected from the native crystal instead of the mercury derivative. Molecular replacement using the partially built model was achieved with PHASER (Storoni, 2004). Iterative cycles of refinement and manual model building using CNS (Brunger, 2007), O (Jones, 1991), COOT (Emsely, 2004), and PHENIX (Adams, 2002) allowed for the assembly of the complete protein model. Individual atomic B-factors were utilized during refinement. Buffer molecules, ions, water, glycerol, the benzyl triethylammonium cation (BTAC), PP_i, and a sulfate ion (hydrogen bonded to R163, H164, R220 and R226) were included in later cycles of refinement. Data reduction and refinement statistics are recorded in Table 5.2. A total of 340 of 381 residues (A16-N355) are present in the final model, as the N- and C- termini are disordered. According to the program PROCHECK (Laskowski, 1993), 95 % of the residues were located in the most favorable regions of the Ramachandran plot, and no residues are in disallowed regions (Figure 5.4).

In order to investigate the mercury binding sites of the ethyl mercury chloride derivatized crystals used for structure determination, refinement of the 1.90 Å-resolution structure was completed. Four mercury binding sites were identified by strong peaks in a Bijvoet difference Fourier map adjacent to C68, C213, C243 and C283. Iterative cycles of refinement and manual model building were achieved with PHENIX and COOT, respectively. Ions and water molecules were included in later cycles of refinement. Individual atomic B-factors were utilized, and the mercury, chloride and sulfur atoms were refined anisotropically. Data reduction and refinement statistics are recorded in Table 5.2. A total of 323 of 381 residues are present in the final model, as there was no interpretable density for the A251-L267 loop as well as the N- and C- termini.



Plot statistics

```
Residues in most favoured regions  [A,B,L]                   287     95.0%
Residues in additional allowed regions  [a,b,l,p]             15      5.0%
Residues in generously allowed regions  [~a,~b,~l,~p]          0      0.0%
Residues in disallowed regions                                 0      0.0%
                                                            ----    ------
Number of non-glycine and non-proline residues               302    100.0%

Number of end-residues (excl. Gly and Pro)                     2

Number of glycine residues (shown as triangles)               19
Number of proline residues                                    17
                                                            ----
Total number of residues                                     340
```

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 5.4. Ramachandran plot of the refined EIZS-$Mg^{2+}{}_3$-PP_i-BTAC structure.

Table 5.2. Refinement Statistics for wild-type EIZS complexes

EIZS Structure:	Mg^{2+}_3-PP_i-BTAC complex	Hg^{2+}_4 complex
Data		
Wavelength, Å	1.0080	1.0548
Resolution, Å	40 - 1.60	50 - 1.90
Unique reflections	46,113	28,446
Completeness*, %	92.9 (100)	99.4 (98.6)
Redundancy*	3.6 (3.6)	5.1 (5.1)
R_{sym}*†	0.058 (0.208)	0.075 (0.325)
Refinement		
R_{cryst}/R_{free}‡	0.158 / 0.189	0.169 / 0.202
r.m.s.d. bonds, Å	0.012	0.005
r.m.s.d. angles, °	1.4	0.8
r.m.s.d. dihedral angles, °	21	15
No. of atoms		
Protein atoms	2858	2512
Solvent atoms	431	335
Ligand atoms	55	13
Ramachandran plot		
Allowed, %	94.7	95.2
Additionally allowed, %	5.3	4.8

*Values in parentheses refer to the highest shell

† $R_{sym} = \sum|I_h - \langle I_h\rangle|/\sum I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent refections.

‡ $R_{cryst} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$, where summation is over the data used for refinement. R_{free} was calculated as for R_{cryst} by using 5% of the data that was excluded from refinement.

5.4 Structure of EIZS-Mg^{2+}-PP_i-BTAC Complex

At 1.60 Å resolution, the crystal structure of this complex is the highest resolution structure of any terpenoid cyclase determined to date (Aaron, 2010). EIZS adopts the class I terpenoid synthase α-helical fold and consists of a bundle of 10 α-helices, designated A-J (Figure 5.5), in which the 20 Å-deep active site is defined mainly by helices C, D, G, H, and J. Among the terpenoid cyclases of known structure, EIZS is structurally most similar to pentalenene synthase (r.m.s. deviation = 3.3 Å for 304 Cα atoms). EIZS crystallizes as a monomer, consistent with dynamic light scattering measurements indicating that the protein is a monomer in solution (Figure 5.6).

The electron density map clearly reveals 3 Mg^{2+} ions, PP_i, and a BTAC molecule bound in the active site (Figure 5.7 (a)). The side chain of D99 in the aspartate-rich motif $\mathbf{D}^{99}DRHD^{103}$ coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,syn*-bidentate geometry, while $Mg^{2+}{}_B$ is chelated by the “NSE” motif $\mathbf{N^{240}}DLC\mathbf{S}LPK\mathbf{E}^{248}$ (boldface indicates Mg^{2+} ligands). Each Mg^{2+} ion is coordinated with octahedral geometry, with nonprotein coordination sites occupied by the oxygen atoms of PP_i and by water molecules. The PP_i anion also accepts hydrogen bonds from the side chains of R194, K247, R338, and Y339 (Figure 5.7 (b)). It is likely that similar metal coordination and hydrogen bond interactions are normally formed with the diphosphate group of the substrate FPP in the precatalytic Michaelis complex. These interactions stabilize the closed active site conformation that sequesters FPP from bulk solvent and triggers ionization to initiate the electrophilic cyclization cascade.

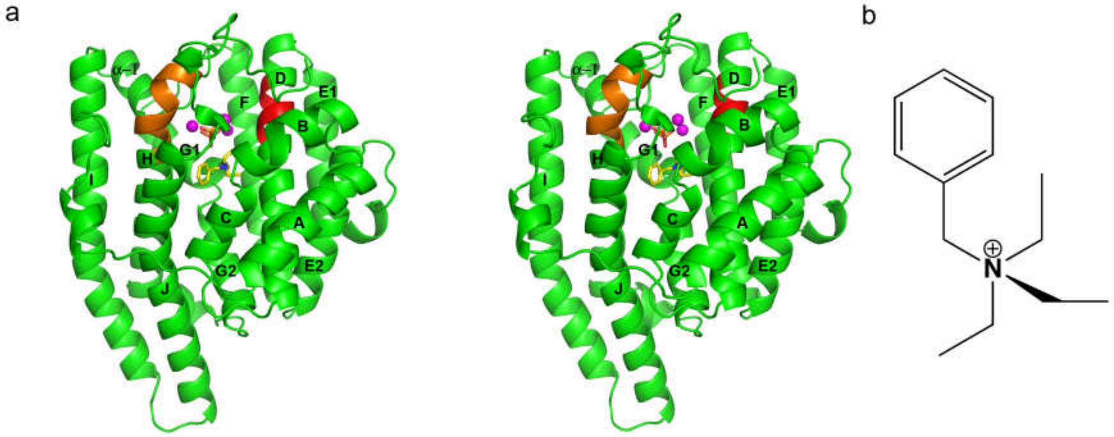


Figure 5.5. (a) Ribbon plot of the EIZS-Mg^{2+}_3-PP_i-BTAC complex showing the aspartate-rich motif (red) and the NSE motif (orange) flanking the mouth of the active site. The Mg^{2+} ions are shown as magenta spheres, PP_i and BTAC molecules are color coded by atom [carbon (yellow), nitrogen (blue), oxygen (red), and phosphate (orange)]. Helices are labeled according to the convention first established for FPP synthase (Tarshis, 1994). (b) The quaternary ammonium group of the benzyltriethylammonium cation (BTAC) serves as a mimic of a carbocation intermediate in catalysis.

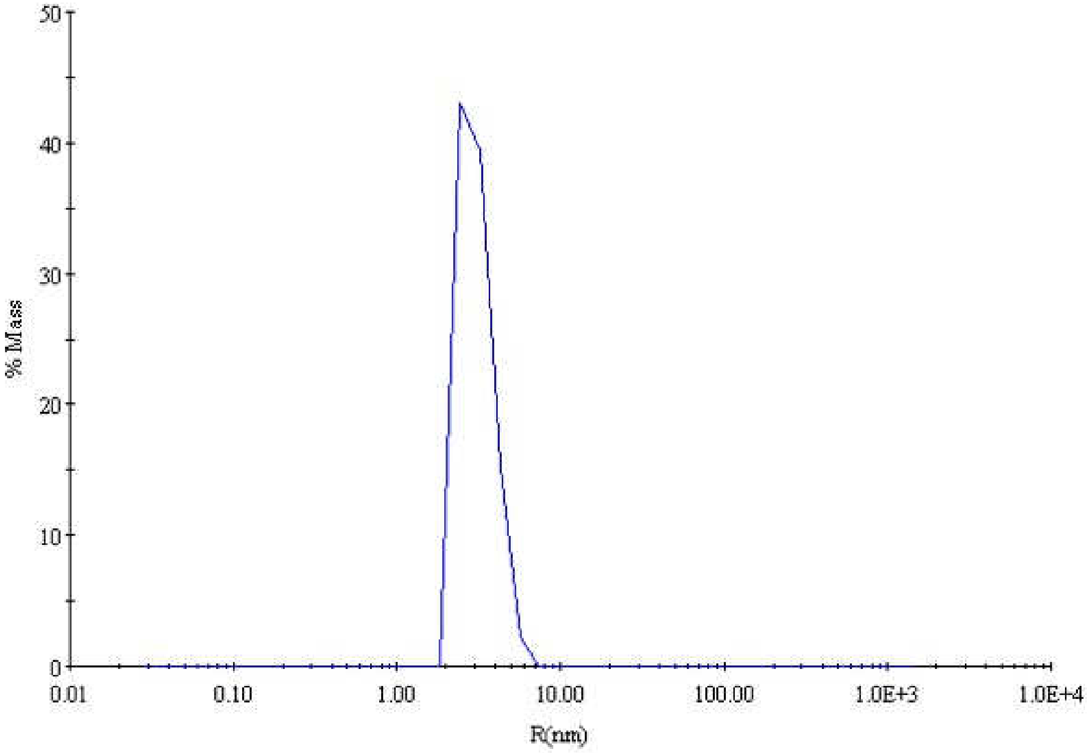


Figure 5.6. Particle size distribution of EIZS sample, measured by dynamic light scattering. Results indicated a monomeric sample, with an average particle radius of 3 nm.

Crystals of EIZS form upon addition of BTAC (Figure 5.5 (b)), commonly used in synthetic organic chemistry as a phase transfer catalyst. As a hydrophobic cation, BTAC mimics the bisabolyl carbocation intermediate proposed for the EIZS-catalyzed cyclization mechanism (Lin, 2006). The positively charged quaternary ammonium group of BTAC appears to be stabilized by long-range electrostatic interactions with the PP_i anion (the shortest N---O separation is 4.1 Å), as well as cation-π interactions with the aromatic side chains of F95, F96 and F198 (N---ring centroid separations range 4.7 – 5.4 Å) (Figure 5.6 (a)). Such cation-π interactions are proposed to play a critical role in stabilizing the highly reactive carbocation intermediates found in all enzyme-catalyzed terpenoid cyclizations (Jenson, 1997; Lesburg, 1997; Ma, 1997). The binding of BTAC in the active site of EIZS demonstrates that the bacterial terpenoid cylase active site can accommodate and stabilize a positively charged ligand resembling the positively charged carbocation intermediates of the normal cyclization cascade. Although determined at a much lower resolution of 2.85 Å, the recently reported structure of the trichodiene synthase-Mg^{2+}_3-PP_i-BTAC complex provides an example of the stabilization of positively charged ligands in the active site of a fungal terpenoid cyclase (Vedula, 2007).

5.5 Structure of EIZS-Hg^{2+}_4 Complex

To date, we have been unable to prepare crystals of wild-type EIZS in a completely ligand-free state, since the additives BTAC and PP_i required for crystallization are not readily dialyzed out of the crystals. However, soaking crystals of the EIZS-Mg^{2+}_3-PP_i-BTAC complex with ethyl mercury chloride displaces 3 Mg^{2+} ions,

PP_i, and BTAC to yield the 1.90 Å resolution structure of the EIZS-Hg^{2+}_4 complex in which all Hg^{2+} ions bind remotely from the active site (Figure 5.8 (a)). In this complex, Hg^{2+}_A is coordinated by C68, Hg^{2+}_B is coordinated by C283 (the electron density is best interpreted as 2 alternate positions), Hg^{2+}_C is coordinated by C243, and Hg^{2+}_D is coordinated by C213 (2 alternate positions). Surprisingly, however, the now ligand-free active site is not open and empty, as it is in all other ligand-free terpenoid cyclase structures. Instead, helix G, which forms one side of the active site (and contains R194, a residue that donates a hydrogen bond to PP_i in the ligand-bound structure), bends by ~110° to occupy the location formerly occupied by Mg^{2+}_3-PP_i (Figure 5.8 (c)). Additionally, helix D moves ~1.5 Å outward, helix J moves ~4 Å outward, while helix H, which was bent in the ligand-bound structure to enable Mg^{2+}_B chelation by the NSE motif, becomes straight and also moves outward (Figure 5.8 (c)). The positions and conformations of helices C, E, and F remain essentially unchanged, and there is no interpretable electron density for helix I or the loop (A251-L267) connecting it to helices H and J. Salt bridges between helix G and helices D and F stabilize the closed ligand-free structure: R194 makes a salt-bridge with E175 (helix F) and R195 makes a salt bridge with D103 (helix D) (Figure 5.8 (c)). Although these residues are found in pentalenene synthase as R173, F174, D84 and E152, a comparable closed ligand-free conformation was not observed by Lesburg and colleagues (Lesburg, 1997). Although the EIZS-Hg^{2+}_4 structure illustrates a unique terpenoid cyclase structure, the role of the Hg^{2+} ions in dislodging the Mg^{2+}_3-PP_i-BTAC complex to result in the closed ligand-free conformation of EIZS is unclear. However, this structure does not appear to be biologically relevant, as this structural transition would result in an inactive enzyme and does not likely reflect the state of the wild type enzyme in the absence of substrate.

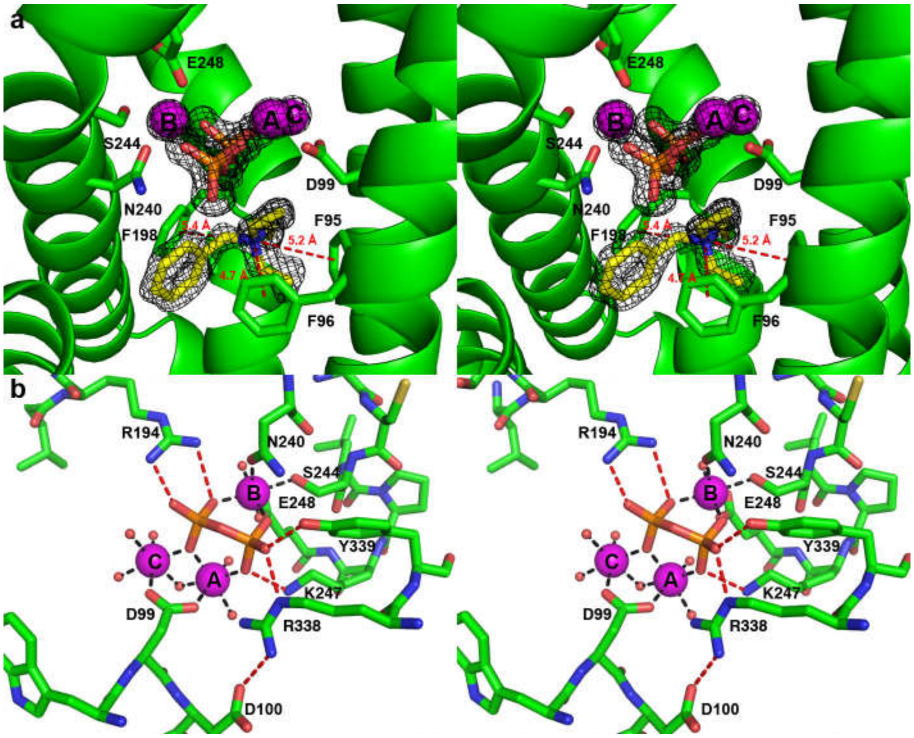

a
E248
B
A
C
S244
D99
N240
F198
5.4 Å
5.2 Å
F95
4.7 Å
F96
b
R194
N240
S244
B
E248
Y339
C
A
K247
D99
R338
D100

Figure 5.7. Active site of EIZS-Mg^{2+}_3-PP_i-BTAC complex. (a) Simulated annealing omits maps (black) of the PP_i anion, Mg^{2+} ions, and BTAC, contoured at 5σ. Note the cation-π interactions between the positively charged quaternary ammonium group of BTAC and the aromatic rings of F95, F96 and F198 (red dashed lines). (b) Metal coordination interactions (black dashed lines) and hydrogen bond interactions (red dashed lines) in the EIZS-Mg^{2+}_3-PP_i complex.

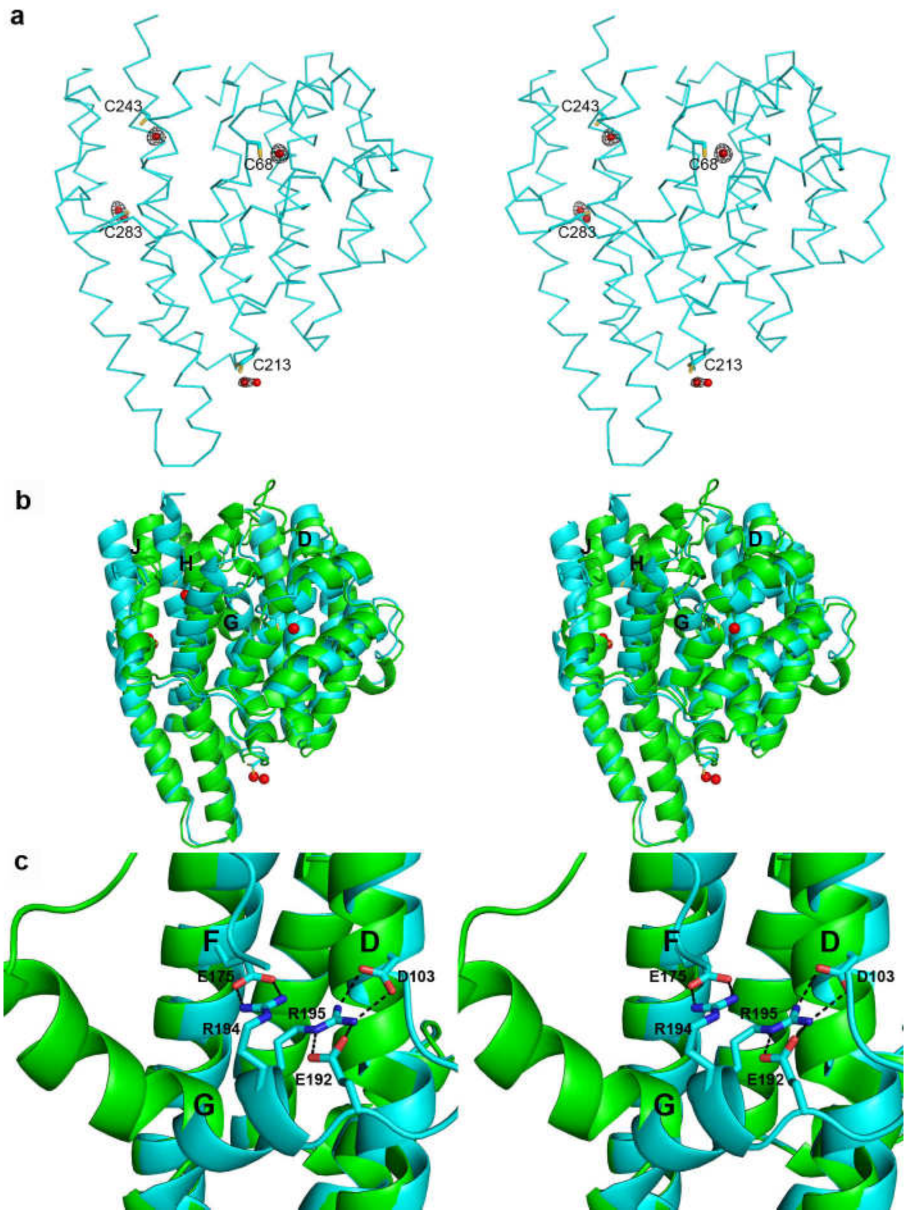

a
C243
C68
C283
C213
C243
C68
C283
C213
b
J
H
D
G
J
H
D
G
c
F
D
E175
D103
R195
R194
E192
G
F
D
E175
D103
R195
R194
E192
G

Figure 5.8. Structural changes between EIZS $Mg^{2+}{}_3$-PP_i-BTAC and $Hg^{2+}{}_4$ Complexes. (a) Ribbon diagram of EIZS-$Hg^{2+}{}_4$ complex. Anomalous Fourier map (black) contoured to 6 σ, indicates the location of Hg atoms (red spheres). (b) Superposition of ribbon plots of EIZS-$Mg^{2+}{}_3$-PP_i-BTAC (green) and EIZS-$Hg^{2+}{}_4$ (cyan; Hg^{2+} ions appear as red spheres). Helices D, G, H and J, which undergo the largest changes, are labeled. (c)

Salt bridges (black dashes) between helix G and helices D and F stabilize the closed ligand-free structure.

Chapter 6: X-ray Crystal Structure of D99N *epi*-Isozizaene Synthase and Implications for Substrate Recognition

6.1 Introduction

Interestingly, the detailed structural changes that occur between open and closed active site conformations differ between fungal and plant terpenoid cyclases. For example, the r.m.s. deviation between ligand-free and $Mg^{2+}{}_3$-PP_i–complexed trichodiene synthase from *F. sporotrichioides* is 1.4 Å (Rynkiewicz, 2001), whereas that of bornyl diphosphate synthase from *S. officinalis* (culinary sage) is only 0.6 Å for the catalytic domain (Whittington, 2002). Until now, it had not been possible to study the structural changes between the open and closed active site conformations of bacterial terpenoid cyclases because pentalenene synthase had been determined only in the ligand-free state (Lesburg, 1997). Since the residues that interact with Mg^{2+} ions or PP_i in *epi*-isozizaene synthase (EIZS) are conserved in pentalenene synthase, superposition of the closed, ligand-bound structure of EIZS with ligand-free pentalenene synthase enables an approximated comparison of the open-to-closed structural transition in bacterial terpenoid cyclases. Alignment of the two structures reveals a very similar alignment of the metal binding motifs, with pentalenene synthase helices D ("aspartate-rich" motif) and H ("NSE" motif) being 1.5 Å further apart than in EIZS. We hypothesize that upon binding of $Mg^{2+}{}_3$ and substrate or PP_i, the active site of pentalenene synthase undergoes a change to a closed conformation comparable to that observed for the EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex. To further explore structural changes resulting from the dissociation of $Mg^{2+}{}_3$-PP_i-BTAC from the active site of EIZS, we attempted to determine the structure of EIZS in the absence of Mg^{2+} and PP_i, however we were unable to obtain crystals of the WT enzyme without Mg^{2+}, PP_i and BTAC. Furthermore, we were not able to remove, via soaking, Mg^{2+}-PP_i-BTAC from crystals of the EIZS-Mg^{2+}-PP_i-BTAC complex.

Therefore, an alternative approach was taken to investigate the open structure of EIZS. The D99N amino acid substitution severely decreases catalytic efficiency (Lin, 2009) by compromising the *syn,syn*-bidentate coordination of $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ (Figure 5.7). Thus this single-site mutant enables the structural study of an unliganded conformation of EIZS.

6.2. Experimental Methods

6.2.1 Site-Directed Mutagenesis, Expression and Purification

The D99N single site-specific mutation was introduced into the EIZS wild-type plasmid using primers 1 and 2 as follows (lower case letters represent the mutant codon introduced): primer 1, 5’GGT TCT TCG TCT GGa acG ACC GTC ACG AC-3’, primer 2, 5’-GTC GTG ACG GTC gtt CCA GAC GAA GAA CC-3’. The optimal reaction mixture for PCR amplification of the insert was 100 ng of each forward and reverse primer, 3 μL of 10 mM dNTP mix, 100 ng plasmid, 5 μL of *Pfu* turbo polymerase buffer, and 1 unit of *Pfu* turbo polymerase diluted with water to a final volume of 50 μL. Optimal PCR conditions required initial denaturation of the reaction mixture at 95 °C for 5 min, addition of polymerase followed by thirty cycles (1 min denaturation at 95 °C, 1 min annealing at 60 °C, 8 min extension at 72 °C), and a final 10 min extension at 72 °C followed by a final hold at 4 °C. One μL of Dpn1 was added to the PCR mixture and incubated at 37 °C for 1 h to digest the template. PCR products were transformed into XL1-Blue cells for DNA isolation and sequencing. DNA was purified (Qiagen mini-prep kit) from cultures from single colonies, and DNA sequencing (DNA Sequencing Facility, University of Pennsylvania) confirmed incorporation of the mutation. Mutant proteins were expressed and purified using the same procedures as described for the wild-type enzyme (Section 5.1).

6.2.2 Crystallization and Structure Determination

The D99N EIZS mutant was crystallized by the hanging drop vapor diffusion method. Crystals formed with a precipitant solution at a slightly higher pH than the wild type crystals [100 mM Bis-Tris (pH 6.0), 25-28% polyethylene glycol 3350, 0.2 M $(NH_4)_2SO_4$], and were improved by successive rounds of micro-streak seeding using D99N crystals as the seed stock. Crystals diffracted to 1.9 Å resolution at APS beamline NE-CAT 24-ID-C and belonged to space group $P2_12_12_1$ with unit cell parameters a = 41.144 Å, b = 81.952 Å, c = 106.693 Å. The Matthew's coefficient (V_M) of 2.04 $Å^3$/Da, and solvent content of 40 % suggest the presence of one monomer per asymmetric unit. Molecular replacement calculations were performed with PHASER (Storoni, 2004) using the atomic coordinates of native EIZS (less ligand and solvent atoms) as a search probe. Iterative cycles of refinement and manual model building were achieved with PHENIX (Adams, 2002) and COOT (Emsely, 2004), respectively. Sulfate and water molecules were included in later cycles of refinement. Individual atomic B-factors were utilized.

Data reduction and refinement statistics are recorded in Table 6.1. A total of 316 of 381 residues (P18-E335) are present in the final model, as the N- and C- termini are disordered.

Table 6.1. Data Collection and Refinement Statistics for D99N EIZS.

EIZS Structure:	D99N (ligand-free)
Data	
Wavelength, Å	0.9795
Resolution, Å	50 – 1.90
Unique reflections	28,694
Completeness*, %	97.6 (95.6)
Redundancy*	3.3 (3.0)
R_{sym}*†	0.089 (0.634)
Refinement	
R_{cryst}/R_{free}‡	0.162 / 0.207
r.m.s.d. bonds, Å	0.009
r.m.s.d. angles, °	1.1
r.m.s.d. dihedral angles, °	16
No. of atoms	
Protein atoms	2638
Solvent atoms	316
Ligand atoms	5
Ramachandran plot	
Allowed, %	94.4
Additionally allowed, %	5.6

*Values in parentheses refer to the highest shell

† $R_{sym} = \sum|I_h - \langle I_h\rangle|/\sum I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent refections.

‡$R_{cryst} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$, where summation is over the data used for refinement. R_{free} was calculated as for R_{cryst} by using 5% of the data that was excluded from refinement.

6.3. Structure of D99N EIZS and Implications for Substrate Recognition

Unexpectedly, D99N EIZS crystallizes in a unique packing arrangement with respect to wild-type EIZS, resulting in a different space group and unit cell parameters. Both crystal forms contain one monomer per asymmetric unit, however wild-type EIZS crystallized in space group $P2_1$, with two monomers in the unit cell (Figure 6.1 (a)), whereas D99N EIZS crystallized in space group $P2_12_12_1$, with four equivalent monomer positions in the unit cell (Figure 6.1. (b)).

In the closed, ligand-bound conformation of EIZS only the first aspartate of the D^{99}DRHD motif coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$; substitution of a neutral asparagine residue for the negatively-charged aspartate residue at position 99 is sufficient to disrupt the assembly and stability of the trinuclear metal cluster required for substrate recognition and activation. The crystal structure of D99N EIZS reveals the complete absence of $Mg^{2+}{}_3$-PP_i-BTAC in the active site. Moreover, the active site of D99N EIZS adopts an open conformation (Figure 6.2) without any conformational changes of helix G. Therefore, we conclude that the conformational change of helix G shown in Figure 5.8 is somehow caused by Hg^{2+} binding and is not biologically relevant.

Comparison of EIZS-$Mg^{2+}{}_3$-PP_i-BTAC and D99N EIZS structures enables a structural understanding of the molecular recognition of substrate by EIZS, and the ensuing structural changes that result in closing of the active site upon binding of substrate and Mg^{2+} ions. The crystal structures are particularly informative in that they reveal two biologically relevant active site conformations: a closed, ligand-bound conformation and an open, ligand-free conformation, respectively. Alignment of the two structures reveals ligand-induced conformational changes of helix H and loop H-α-1, as well as the J-K loop (which is completely disordered in D99N EIZS). The overall r.m.s. deviation between wild-type and D99N EIZS structures is 1.6 Å for 318 Cα atoms. The D99N EIZS structure suggests that in the absence of PP_i, or a PP_i-containing substrate, the C-terminal residues, beginning at E335, become disordered, which results in an open active site. Upon Mg^{2+} and PP_i (or PP_i-containing substrate) binding, the C-terminus becomes ordered and closes the active site and blocks the solvent from entering the active site. Corresponding structural changes generally accompany active site closure in fungal (Shishova, 2008) and plant (Whittington, 2002) terpenoid cyclases. The template for FPP cyclization is fully formed in the closed active site conformation of EIZS as well as all other class I terpenoid cyclases.

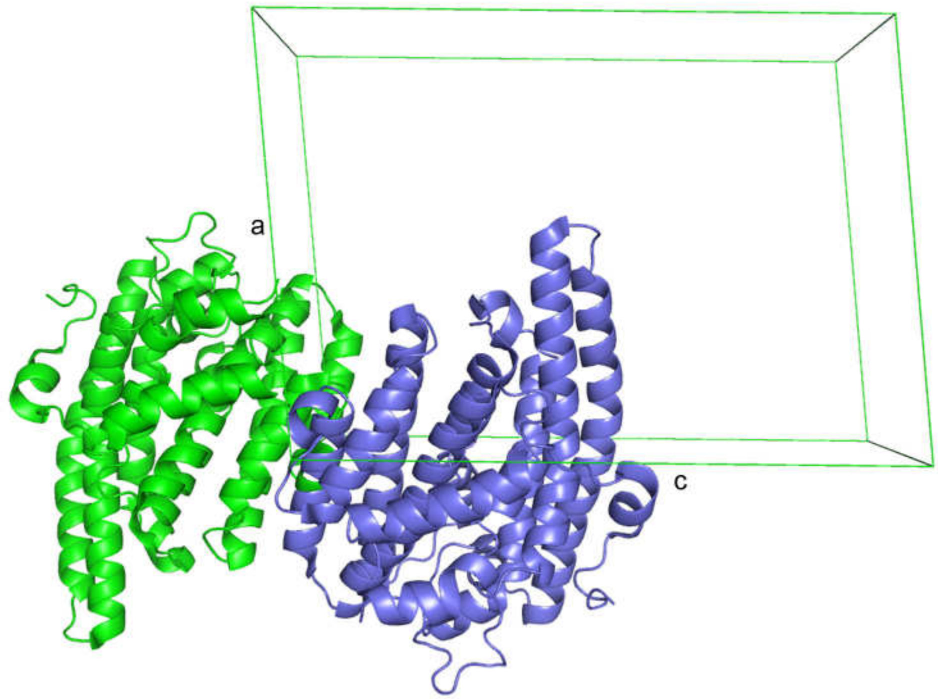


Figure 6.1. (a) The unit cell of EIZS-Mg^{2+}_3-PP_i-BTAC crystals viewed along the b axis. The crystals belong to space group $P2_1$, the asymmetric unit is a monomer, and the two positions are related in the unit cell by a screw axis along b.

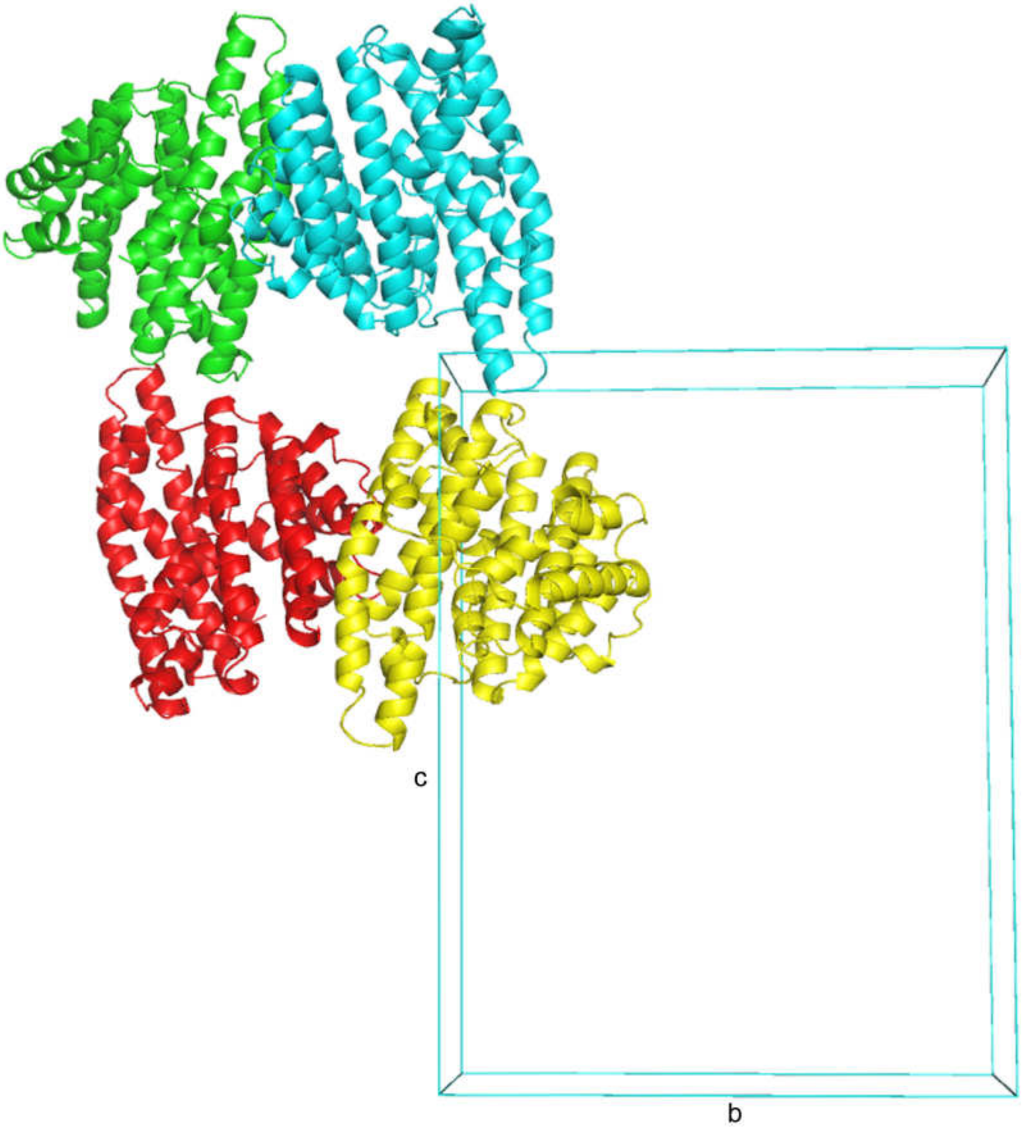


Figure 6.1. (b) The unit cell of D99N EIZS crystals viewed along the a axis. The crystals belong to space group $P2_12_12_1$, the asymmetric unit is a monomer. There are four equivalent positions in the unit cell, which possesses three perpendicular twofold screw axes.

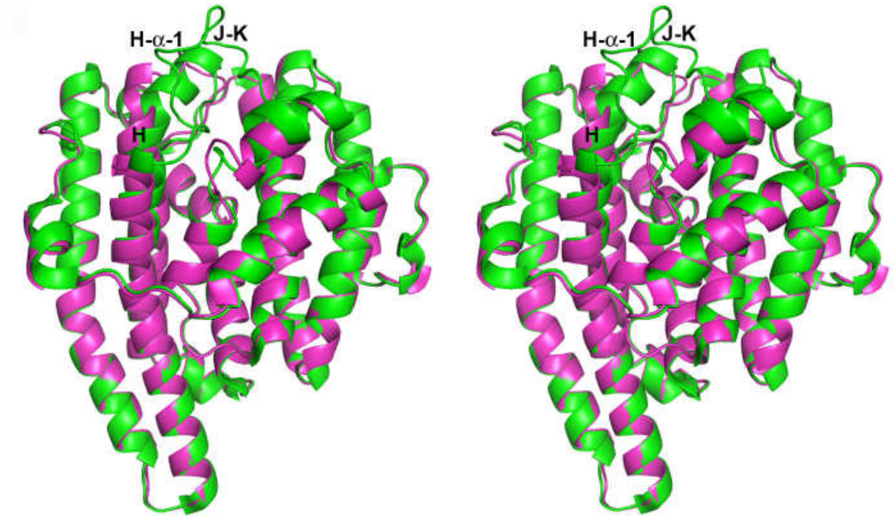


Figure 6.2. A stereoview of a superposition of the structures of EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex (green) and D99N EIZS (purple), illustrating the structural changes in helix H and the H-α-1 and J-K loops that accompany active site closure.

The newly determined open and closed structures of EIZS enable a comparison of substrate recognition and active site closure amongst bacterial, fungal and plant terpenoid cyclases. Although the first and third aspartate residues in the aspartate-rich metal binding motif coordinate to Mg^{2+}_A and Mg^{2+}_C in the plant terpenoid cyclases containing a complete trinuclear metal cluster such as (+)-bornyl diphosphate synthase (Figure 6.3 (c)) (Whittington, 2002), limonene synthase (Hyatt, 2007), and (+)-δ-cadinene synthase (Gennadios, 2009), only the first aspartate residue in the aspartate-rich motif coordinates to Mg^{2+}_A and Mg^{2+}_C in Mg^{2+}_3-PP_i complexes with the fungal cyclases trichodiene synthase (Rynkiewicz, 2001) and aristolochene synthase (Figure 6.3 (b)) (Shishova, 2007). EIZS is similar to the fungal cyclases in that only D99 coordinates to Mg^{2+}_A and Mg^{2+}_C. A critical role for D99 in metal complexation is reflected in the dramatic losses of catalytic activity measured for the D99N and D99E mutants (Lin, 2009). Although D100 of EIZS does not directly interact with the Mg^{2+} ions or PP_i, it does accept a hydrogen bond from R338, which also donates a hydrogen bond to PP_i (Figure 6.3 (a)). Site-directed mutagenesis reveals that the D100N mutant has lost >95 % activity compared to the native enzyme (Lin, 2009), suggesting that the D100N mutation disrupts the D100-R338-PP_i hydrogen bond network. In aristolochene synthase, the second aspartate in the aspartate-rich motif, D91, similarly stabilizes a hydrogen bond network with R314 and PP_i (Shishova, 2007). Surprisingly, in (+)-bornyl diphosphate synthase, the second aspartate, D352, is involved in a hydrogen bond network with R314 and PP_i (Whittington, 2002), illustrating the importance of the arginine residue in stabilizing the closed conformation. Furthermore, a second conserved arginine residue makes a hydrogen bond to the opposite end of the diphosphate moiety, R194 in EIZS, R175 in aristolochene synthase and R493 in (+)-bornyl disphosphate synthase, suggesting that the bacterial, fungal, and plant cyclases share the same molecular strategy for linking the molecular recognition of the substrate diphosphate group with the mechanism of active site closure. Two additional PP_i coordinating interactions are conserved amongst EIZS and aristolochene synthase; K247 and Y339 in EIZS donate H-bonds to diphosphate oxygen atoms and are conserved as K226 and Y315 in aristolochene synthase. A higher degree of conservation amongst active site residues suggests bacterial and fungal terpenoid cyclases derive from a more recent common ancestor than plant terpenoid cyclases.

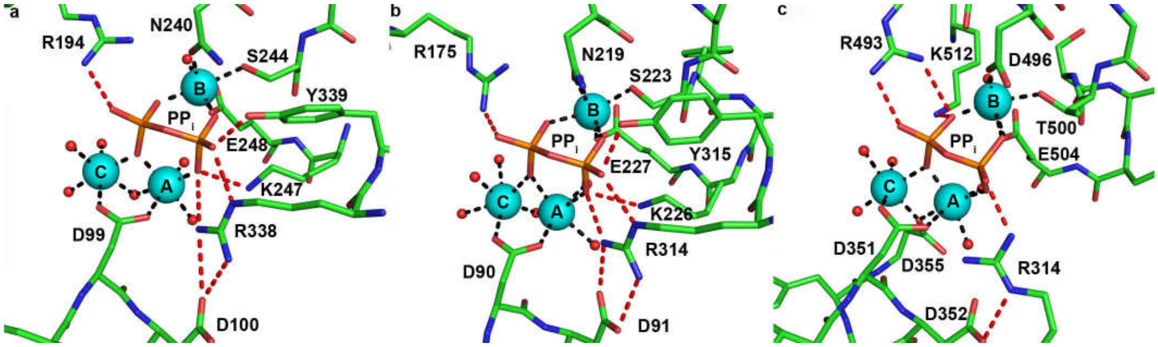

a
R194
N240
S244
B
PP$_i$
Y339
E248
C
A
K247
D99
R338
D100
b
R175
N219
S223
B
PP$_i$
E227
Y315
C
A
K226
R314
D90
D91
c
R493
K512
D496
B
PP$_i$
T500
E504
C
A
D351
D355
R314
D352

Figure 6.3. Conservation of $Mg^{2+}{}_3$-PP_i and -diphosphate binding motifs among bacterial and fungal terpenoid cyclases. Metal coordination (black) and hydrogen bond (red) interactions with phosphate(s) are indicated. (a) Bacterial sesquiterpene cyclase *S. coelicolor epi*-isozizaene synthase-$Mg^{2+}{}_3$-PP_i complex (PDB code 3KB9); (b) Fungal sesquiterpene cyclase *A. terreus* aristolochene synthase-$Mg^{2+}{}_3$-PP_i complex (PDB code 2OA6); (c) Plant monoterpene cyclase *S. officinalis* (+)-bornyl diphosphate synthase-$Mg^{2+}{}_3$-PP_i complex (PDB code 1N22; metal ions are labelled according to the convention first established for trichodiene synthase).

Chapter 7: Structural and Biochemical Studies of the Active Site of EIZS

7.1 Introduction

In general, the permissiveness and promiscuity of terpenoid cyclases vary, both in terms of the substrates they accept and the product(s) they generate. These properties are dictated by the three-dimensional contour of the fully formed template in the closed active site conformation. Many terpenoid cyclases, such as *A. terreus* aristolochene synthase (Felicetti, 2004), are high-fidelity cyclases that generate one product exclusively. However other cyclases, such as EIZS, generate one major product and minor quantities of one or more alternative products. Detailed gas chromatography-mass spectrometry (GC-MS) analysis of the organic products that result from incubation of WT EIZS with FPP reveal the promiscuity of the EIZS template. Specifically, 79 % of the total sesquiterpene product mixture is *epi*-isozizaene, and the remaining 21 % is identified as a mixture of β-farnesene (5 %), sesquisabinene-A (3 %), zizaene (9 %), α-cedrene (2 %), sesquiphellandrene (1 %), and 2 % of an unidentified sesquiterpene (Lin, 2009). The structural basis for such mechanistic promiscuity is presumably rooted in how well the active site contour enforces the correct regiochemistry and stereochemistry for cyclization and eventual quenching of the carbocation intermediates by chaperoning the conformations of reactive intermediates. Intriguingly, the conformations and orientations of such intermediates may not reflect the original conformation and orientation of the substrate if the template is somewhat permissive (Hong, 2009). A more permissive template allows alternative premature quenching of on-pathway intermediates or off-pathway conformations that lead to the formation of aberrant products.

Manipulation of the cyclization template by site-directed mutagenesis can redirect the biosynthetic trajectory of a terpenoid cyclase. This result can be achieved by modification of active site contour residues (Yoshikuni, 2006) or of residues that are more distant from the active site (O'Maille, 2008). With EIZS, two different strategies have been employed to manipulate the cyclization template: mutagenesis of metal-binding motif residues, and residues that directly contribute to forming the unique active-site contour.

Mutagenesis of the conserved Mg^{2+}-binding domains severely compromises catalytic efficiency. Single-site metal-binding motif mutants D99E, D100N, N240D, S244A, and E248D all retain less than 5% of WT activity, however GC-MS analysis of their respective organic products indicates these mutations have only a modest effect on the cyclization template, resulting in slightly altered relative proportions of *epi*-isozizaene and alternate sesquiterpene side-products. Specifically, *epi*-isozizaene accounts for 62 to 91 % of the relative sesquiterpene products of these mutants, and only one additional sesquiterpene side product is identified, α-neocallitropsene (3 % D99E, 3 % N240 D, <1 % E248D), thus the fidelity of *epi*-isozizaene biosynthesis is not significantly compromised. Indeed, certain amino acid substitutions involving the Mg^{2+}-binding residues, such as D100N, N240D, S244A, and E248D, actually lead to increased proportions of *epi*-isozizaene and lower levels of the alternative sesquiterpene products, although with significantly decreased overall catalytic efficiency (Lin, 2009).

It is hypothesized that the structural and stereochemical diversity achieved by the terpenoid cyclases is the result of an enzyme active site that merely chaperones a multi-step interamolecular cyclization cascade. As such, a terpenoid cyclase active site is primarily lined with hydrophobic and aromatic residues, which serve to guide the cyclization cascade by assisting in the stabilization of carbocation intermediates (Christianson, 2008). Aromatic residues are able to stabilize carbocation intermediates via charge-quadrupole (cation-π) interactions. The active site of EIZS has several aromatic residues, namely F95, F96, F198, W203, F332, H333, and W325. As a first step in exploring the importance of active site aromatic residues for catalysis and product diversity in EIZS, we have prepared the F96A, F198A, and W203F mutants. These residues were selected for mutation based on their proximity to the BTAC cation that was observed to coordinate in the active site of the WT EIZS-Mg^{2+}-PP_i-BTAC complex presented in Chapter 5 (Figure 5.6). Furthermore, the effect of mutating aliphatic residues in the active site was also explored. Specifically L72, A236 and V329 were chosen for investigation, and each site was selectively mutated to investigate the following single-site EIZS mutants: L72V, L72I, A236G, A236V, V329A and V329L to explore the effects of subtle changes to the active site contour on enzyme activity and product distribution.

7.2 Experimental Methods

7.2.1 Site-Directed Mutagenesis, Expression and Purification

The following EIZS single-site mutants were investigated: F96A, F198A, W203A, W203F, L72V, L72I, A236G, A236V, V329A and V329L. Single site-specific mutations were introduced into the EIZS wild-type plasmid using forward and reverse primers for each respective mutant (Table 7.1). The optimal reaction mixture for PCR amplification of the insert was 100 ng of each forward and reverse primer, 3 μL of 10 mM dNTP mix, 100 ng plasmid, 5 μL of *Pfu* turbo polymerase buffer, and 1 unit of *Pfu* turbo polymerase diluted with water to a final volume of 50 μL. Optimal PCR conditions required initial denaturation of the reaction mixture at 95 °C for 5 min, addition of polymerase followed by thirty cycles (1 min denaturation at 95 °C, 1 min annealing at 60 °C, 8 min extension at 72 °C), and a final 10 min extension at 72 °C followed by a final hold at 4 °C. One μL of Dpn1 was added to the PCR mixture and incubated at 37 °C for 1 h to digest the template. PCR products were transformed into XL1-Blue cells for DNA isolation and sequencing. DNA was purified (Qiagen mini-prep kit) from cultures from single colonies, and DNA sequencing (DNA Sequencing Facility, University of Pennsylvania) confirmed incorporation of the mutations. Mutant proteins were expressed and purified using the same procedures as described for the wild-type enzyme (Section 5.1). The W203A mutant expressed poorly, and could not be successfully purified, therefore a more conservative W203F mutant was investigated. The L72I, A236V and V329L EIZS mutants did not express as well as their respective counterparts L72V, A236G and V329A, and were not investigated further.

Table 7.1. EIZS mutagenic primer sequences. Lower case letters represent the mutant codon introduced.

Protein		**Primers**
F96A	Forward	5′-CTA CAG CGC GTG GTT Cgc aGT CTG GGA CGA CCG TC-3′
	Reverse	5′-GAC GGT CGT CCC AGA Ctg cGA ACC ACG CGC TGT AG-3′
F198A	Forward	5′-GAA CTG CGC CGG CTC ACG gca GCG CAC TGG ATC TGG AC-3′
	Reverse	5′-GTC CAG ATC CAG TGC GCt gcC GTG AGC CGG CGC AGT TC-3′
W203A	Forward	5’-GTT CGC GCA CTG CAT Cgc aAC CGA CCT GCT GGA G-3’
	Reverse	5’-CTC CAG CAG GTC GGT tgc CAT CCA GTG CGC GAA C-3’
W203F	Forward	5′-GTT CGC GCA CTG GAT Ctt tAC CGA CCT GCT GG-3′
	Reverse	5′-GCT CCA GCA GGT CGG Taa aGA TCC AGT GCG CG-3′
L72I	Forward	5’-CTG TGC TAC ACG GAC att ATG GCG GGC TAC TAC C-3’
	Reverse	5’-GGT AGT AGC CCG CCA Taa tGT CCG TGT AGC ACA G-3’
L72V	Forward	5’-CTG TGC TAC ACG GAC gtg ATG GCG GGC TAC TAC-3’
	Reverse	5’-GTA GTA GCC CGC CAT cac GTC CGT GTA GCA CAG-3’
A236G	Forward	5’-GTC AGG AAT TCg gcG CCT GGT ACA AC-3’
	Reverse	5’-GTT GTA CCA GGC gcc GAA TTC CTG AC-3’
A236V	Forward	5’-GAG TCA GGA ATT Cgt gGC CTG GTA CAA CGA C-3’
	Reverse	5’-GTC GTT GTA CCA GGC cac GAA TTC CTG ACT C-3’
V329A	Forward	5’-CTG GTT CAG TTC Cgc gTA CTG GTT CCA CC-3’
	Reverse	5’-GGT GGA ACC AGT Acg cGG AAC TGA ACC AG-3’
V329L	Forward	5’-GAA CTG GTT CAG TTC Cct gTA CTG GTT CCA CCA CG-3’
	Reverse	5’-CGT GGT GGA ACC AGT Aca gGG AAC TGA ACC AGT TC-3’

7.2.2 Crystallization and Structure Determination of EIZS active site mutants

7.2.2.1 F198A EIZS-Mg^{2+}_3-PP_i-BTAC complex.

The F198A EIZS mutant was crystallized by the hanging drop vapor diffusion method with the same conditions used to crystallize the wild-type enzyme, but with successive rounds of micro-streak seeding using native crystals as the seed stock. Crystals diffracted to 1.64 Å resolution at NSLS beamline X29 and belonged to space group $P2_1$ with unit cell parameters a = 53.241 Å, b = 47.179 Å, c = 75.568 Å and β = 95.57°. Molecular replacement calculations were performed with PHASER (Storoni, 2004) using the atomic coordinates of native EIZS (less ligands and solvent atoms) as a search probe. The electron density clearly revealed the F198A mutation. Iterative cycles of refinement and manual model building were achived with PHENIX and COOT, respectively. Ions, PP_i, BTAC, and water molecules were included in later cycles of refinement. Individual atomic B-factors were utilized. Data collection and refinement statistics for the F198A EIZS-Mg^{2+}_3-PP_i-BTAC complex are listed in Table 7.2.

7.2.2.2 L72V EIZS-Mg^{2+}_3-PP_i-BTAC complex.

The L72V EIZS mutant was crystallized by the hanging drop vapor diffusion method with the same conditions used to crystallize the wild-type enzyme. Crystals diffracted to 2.10 Å resolution at NSLS beamline X29 and belonged to space group $P2_1$ with unit cell parameters a = 52.977 Å, b = 47.236 Å, c = 75.084 Å and β = 95.66°. Molecular replacement calculations were performed with PHASER (Storoni, 2004) using the atomic coordinates of native EIZS (less ligands and solvent atoms) as a search probe. The electron density clearly revealed the L72V mutation. Iterative cycles of refinement and manual model building were achived with PHENIX and COOT, respectively. Ions, PP_i, BTAC, and water molecules were included in later cycles of refinement. Individual atomic B-factors were utilized. Data collection and refinement statistics for the L72V EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex are listed in Table 7.2.

7.2.2.3 A236G EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex.

The A236G EIZS mutant was crystallized by the hanging drop vapor diffusion method with the same conditions used to crystallize the wild-type enzyme. Crystals diffracted to 1.76 Å resolution at APS beamline NE-CAT-ID-C and belonged to space group $P2_1$ with unit cell parameters a = 53.137 Å, b = 47.220 Å, c = 75.211 Å and β = 95.51°. Molecular replacement calculations were performed with PHASER (Storoni, 2004) using the atomic coordinates of native EIZS (less ligands and solvent atoms) as a search probe. The electron density clearly revealed the A236G mutation. Iterative cycles of refinement and manual model building were achived with PHENIX and COOT, respectively. Ions, PP_i, BTAC, and water molecules were included in later cycles of refinement. Individual atomic B-factors were utilized. Data collection and refinement statistics for the A236G EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex are listed in Table 7.2.

7.2.2.4 V329A EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex.

The V329A EIZS mutant was crystallized by the hanging drop vapor diffusion method with the same conditions used to crystallize the wild-type enzyme. Crystals diffracted to 1.95 Å resolution at NSLS beamline X29 and belonged to space group $P2_1$ with unit cell parameters a = 53.214 Å, b = 47.485 Å, c = 75.283 Å and β = 95.50°.

Molecular replacement calculations were performed with PHASER (Storoni, 2004) using the atomic coordinates of native EIZS (less ligands and solvent atoms) as a search probe. The electron density clearly revealed the V329A mutation. Iterative cycles of refinement and manual model building were achieved with PHENIX and COOT, respectively. Ions, PP_i, BTAC, and water molecules were included in later cycles of refinement. Individual atomic B-factors were utilized. Data collection and refinement statistics for the V329A EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex are listed in Table 7.2.

Table 7.2. EIZS active site mutant data collection and refinement statistics.

EIZS Structure:	F198A $Mg^{2+}{}_3$-PP_i-BTAC complex	L72V $Mg^{2+}{}_3$-PP_i-BTAC complex	A236G $Mg^{2+}{}_3$-PP_i-BTAC complex	V329A $Mg^{2+}{}_3$-PP_i-BTAC complex
Data				
Wavelength, Å	1.075	0.9795	0.9795	0.9795
Resolution, Å	50 – 1.64	50 – 2.10	50 – 1.76	50 – 1.95
Unique reflections	45,831	22, 011	37, 050	26,839
Completeness*, %	99.8 (100)	99.9 (99.2)	99.2 (98.2)	97.3 (97.4)
Redundancy*	3.6 (3.5)	3.6 (3.5)	3.3 (3.1)	4.2 (3.7)
R_{sym}*†	0.062 (0.238)	0.104 (0.301)	0.071 (0.339)	0.116 (0.432)
Refinement				
R_{cryst}/R_{free}‡	0.156/0.190	0.156/0.206	0.159/0.203	0.152/0.201
r.m.s.d. bonds, Å	0.015	0.007	0.009	0.007
r.m.s.d. angles, °	1.6	1.050	1.167	1.014
r.m.s.d. dihedral angles, °	18	15	17	17
No. of atoms				
Protein atoms	2858	2788	2801	2812
Solvent atoms	467	272	357	376
Ligand atoms	31	31	31	42
Ramachandran plot				
Allowed, %	95.0	94.0	94.7	94.0
Additionally allowed, %	5.0	6.0	5.3	6.0

*Values in parentheses refer to the highest shell

† $R_{sym} = \sum|I_h - \langle I_h \rangle|/\sum I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent refections.

‡$R_{cryst} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$, where summation is over the data used for refinement. R_{free} was calculated as for R_{cryst} by using 5% of the data that was excluded from refinement.

7.2.3 Radioactive substrate kinetic assay of EIZS mutants

EIZS mutants were assayed as previously described (Lin, 2006) in 50 mM piperazine-N,N′ -bis(2-ethanesulfonic acid) (PIPES) (pH 6.5), 20 % glycerol, 100 mM NaCl, 10 mM $MgCl_2$, and 5 mM BME. Each series of assays was performed 2-3 times using concentrations of [1-^{3}H]FPP (100 mCi/mmol) ranging from 0.025 to 50 μM. The optimal enzyme concentration for each mutant was determined where the dependence of product formation on enzyme concentration was linear and less than 10% of the substrate was turned over: wild-type (1 nM), F96A (20 nM) , F198A (20 nM), W203F (20 nM), L72V (2.5 nM), A236G (2.5 nM), and V329L (2.5 nM). A 1-mL reaction mixture in a test-tube was overlaid with 1 mL hexane immediately after addition of substrate, covered with aluminum foil, and incubated for 15 min at 30 °C. The reaction was quenched by addition of 75 μL of 500 mM EDTA (pH 8.0) and vortexed for 20 s. The hexane extract was passed through a silica gel column directly into a scintillation vial containing 5 mL of scintillation fluid. The aqueous phase was extracted with an additional 2 x 1 mL hexane and passed through the same silica gel column. Finally, the column was washed with an additional 1 mL hexane. A Beckman scintillation counter was used to measure product formation, and the substrate concentration versus rate of product formation data was fit by nonlinear regression using the program Prism to determine k_{cat} based on the known total enzyme concentration. For the L72V, A236G and V329A EIZS mutants, the method was adapted to a 0.5 mL reaction volume format, however the volume of hexane used for the extraction was not changed.

7.2.4 GC-MS Analysis of Product Arrays Generated by EIZS Mutants.

The substrate, farnesyl diphosphate (60 μM), was incubated with 40 μM mutant EIZS (F96A, F198A or W203F) in 6 mL buffer (50 mM PIPES, 15 mM $MgCl_2$, 100 mM NaCl, 20% glycerol, 5 mM BME) and overlaid with 3 mL HPLC-grade *n*-pentane in a glass test tube at 30 °C for 18 h. Reaction products were extracted with *n*-pentane 3 times, dried with anhydrous $MgSO_4$, and concentrated on an ice-water mixture under reduced pressure until the volume was reduced to 100 μL. The products were analyzed using an Agilent 6890 GC/JEOL JMS-600H mass spectrometer, using a 30 m x 0.25 mm HP5MS capillary column (Department of Chemistry, Brown University) in EI (positive) mode mode using a temperature program of 60-280 °C, with a gradient of 20 °C /min and a solvent delay of 3.5 min. Analysis of the organic extracts resulting from the incubation of FPP with the mutant cyclases by GC-MS reveals the formation of mixtures of sesquiterpene hydrocarbons with $m/z = 204$. Compounds were identified by comparison of their individual mass spectra and chromatographic retention indices with those of authentic compounds in the MassFinder 3.0 Database (Harangi, 2003).

7.3 Results

7.3.1 Radioactive Substrate Kinetic Assay

The steady-state kinetic parameters, K_M and k_{cat}, of the WT and mutant enzymes were measured by a radioactive substrate (^{3}H-FPP) assay. Scintillation counting was used to monitor the amount of tritium labeled organic products generated by the enzyme during a fixed time. For WT and mutant EIZS, a plot of initial velocity versus substrate concentration (Figures 7.1 and 7.2) was used to determine the steady-state kinetic parameters using non-linear regression. The resulting kinetic parameters, summarized in Table 7.3, indicate that the mutations only modestly affect K_M, with respect to WT EIZS. This result is expected since the residues required for recognition of the substrate diphosphate (R194, K247, R338, and Y339) are not affected by the mutations. However, the rate of the reaction is affected by some of the mutations. Specifically, the three aromatic mutants (F98A, F198A, and W203F), cause a dramatic decrease in k_{cat} of two orders of magnitude, resulting in overall catalytic efficiencies (k_{cat}/K_M) decreased 205- to 275-fold. The aliphatic mutations (L72V, A236G, and V329A), however, do not result in decreased reaction rates, and thus retain the catalytic efficiency of the WT protein. It is important to note however that this assay simply measures tritium-labeled products extracted in hexanes, therefore the total product yield of the WT or mutant EIZS contains a mixture of sesquiterpene products. Furthermore, the activity of the enzymes may be underestimated, since hydroxylated products may not be extracted with this procedure. This is particularly important in the case of the aromatic mutants, which may be able to accommodate solvent molecules in their active sites (Section 7.4.1), which may increase the proportion of hydroxylated products.

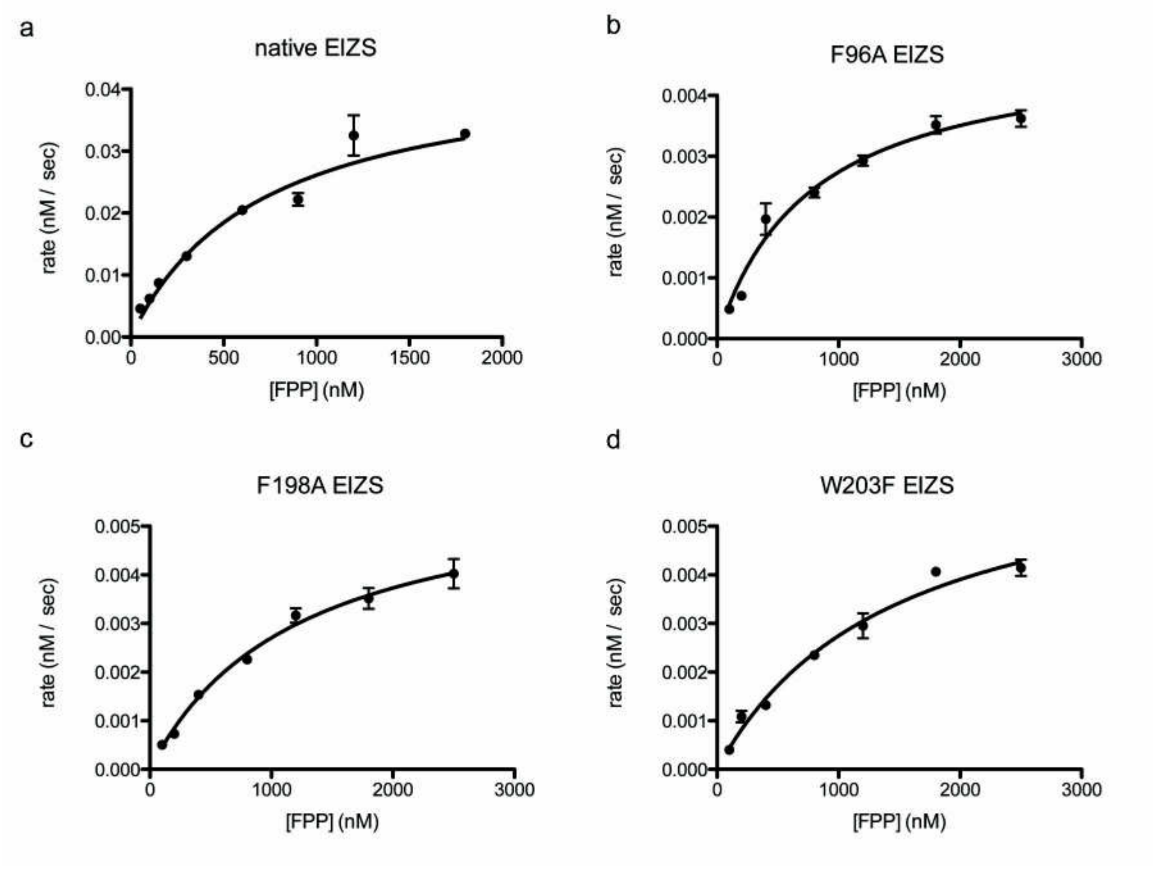


Figure 7.1. Initial rate versus substrate concentration for the reaction of WT and aromatic mutant EIZS with FPP. (a) WT EIZS. (b) F96A EIZS. (c) F198A EIZS. (d) W203F EIZS.

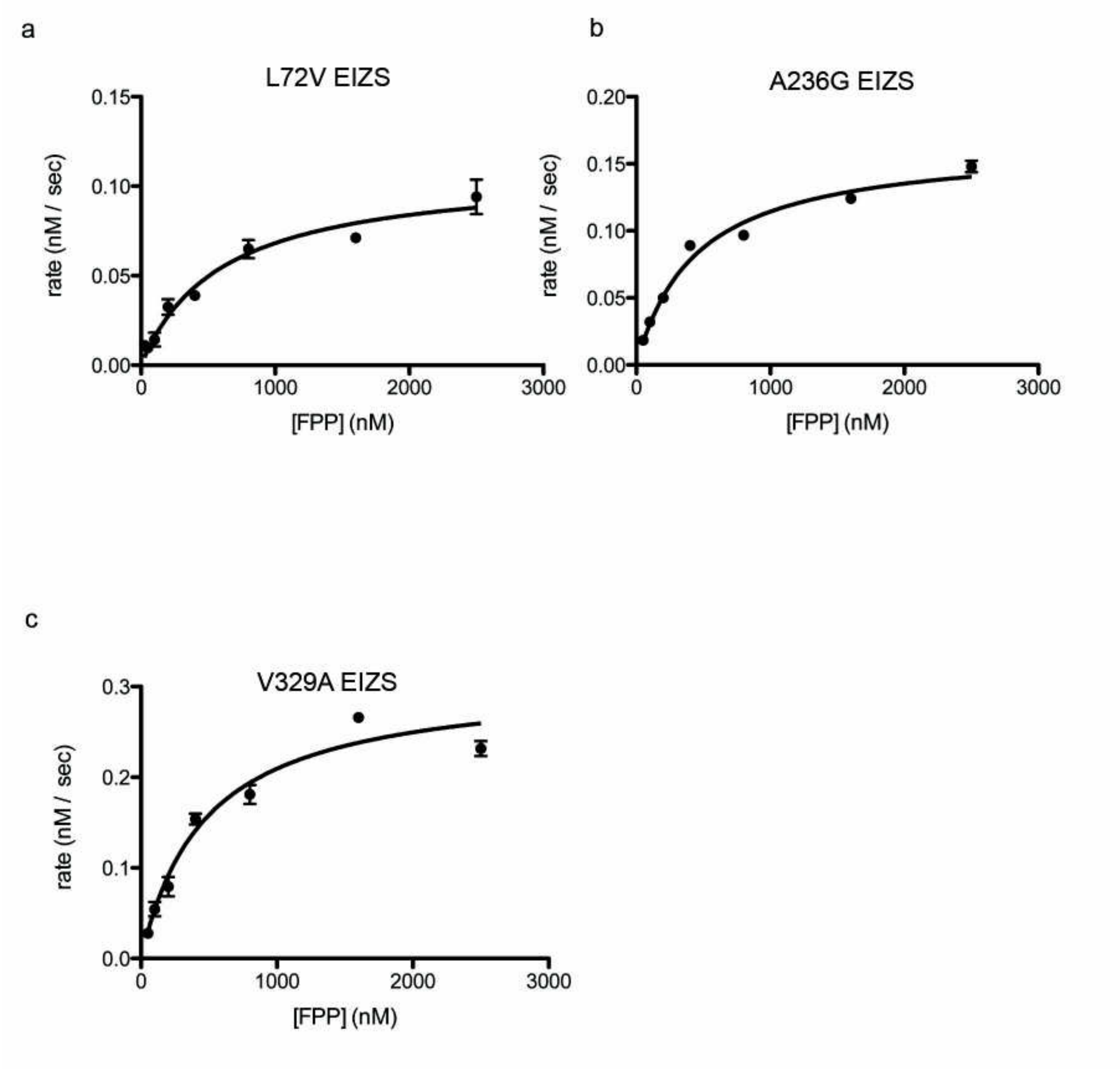


Figure 7.2. Initial rate versus substrate concentration for the reaction of aliphatic mutant EIZS with FPP. (a) L72V EIZS. (b) A236G EIZS. (c) V329A EIZS.

Table 7.3. Steady-state kinetic parameters for wild-type EIZS and site-specific mutants.

Protein	k_{cat} (s^{-1})	K_M (nM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT	0.045 ± 0.003	710 ± 100	$6.3 \pm 0.1 \times 10^4$
F96A	0.00024 ± 0.00002	770 ± 130	310 ± 60
F198A	0.00030 ± 0.00002	1200 ± 200	250 ± 45
W203F	0.00034 ± 0.00003	1450 ± 200	250 ± 45
L72V	0.044 ± 0.003	600 ± 120	$7.3 \pm 1.5 \times 10^4$
A236G	0.066 ± 0.002	450 ± 50	$1.5 \pm 0.2 \times 10^5$
V329A	0.126 ± 0.007	470 ± 80	$2.7 \pm 0.5 \times 10^5$

7.3.2 GC-MS Analysis

The sesquiterpene products of the F98A, F198A, and W203F mutants were separated and identified by gas chromatography-mass spectrometry (GC-MS) as GC peaks with corresponding MS parent ions having m/z of 204 ($C_{15}H_{24}$) in the MS. The gas chromatograms of the products of the three aromatic mutants, shown in Figure 7.3, illustrate very unique product arrays for each mutant. The identity of each peak was confirmed by comparison of the observed mass spectra from the MassFinder 3.0 database (Figure 7.4). Several products could not be identified from the available MS data, and are recorded as unknown sesquiterpene products (Figure 7.5). Results are summarized in Table 7.4.

As expected, the product array is severely altered in the aromatic site-specific EIZS mutants. However, the formation of each sesquiterpene product identified can be reasoned to be a result of derailing the proposed EIZS cyclization pathway (Figure 7.6). Notably, none of the three active site aromatic mutants generate *epi*-isozizaene as a major product; indeed, no *epi*-isozizaene whatsoever is generated by the F198A mutant, and the active site contour of this variant as observed in the crystal structure of the F198A EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex is more complementary in shape to bisabolene-derived cyclization products such as β-acoradiene, as illustrated in Figure 7.7. Notably, β-acoradiene is not generated by wild-type EIZS, so the appearance of this spiroterpenoid represents a new catalytic activity introduced by a single amino acid substitution.

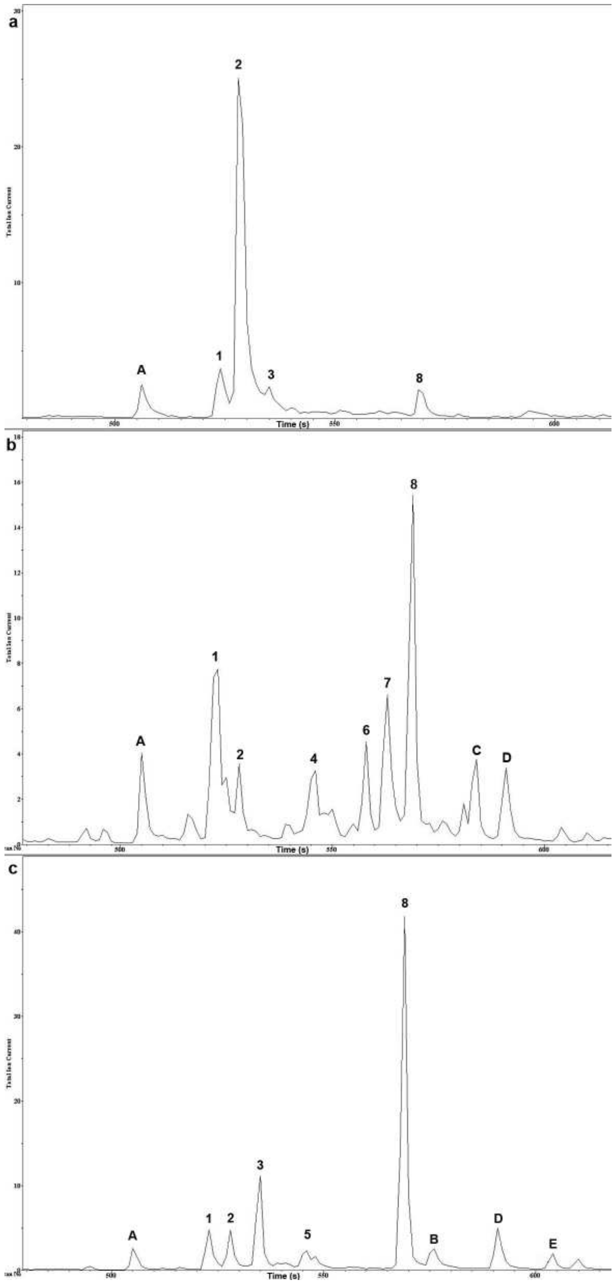
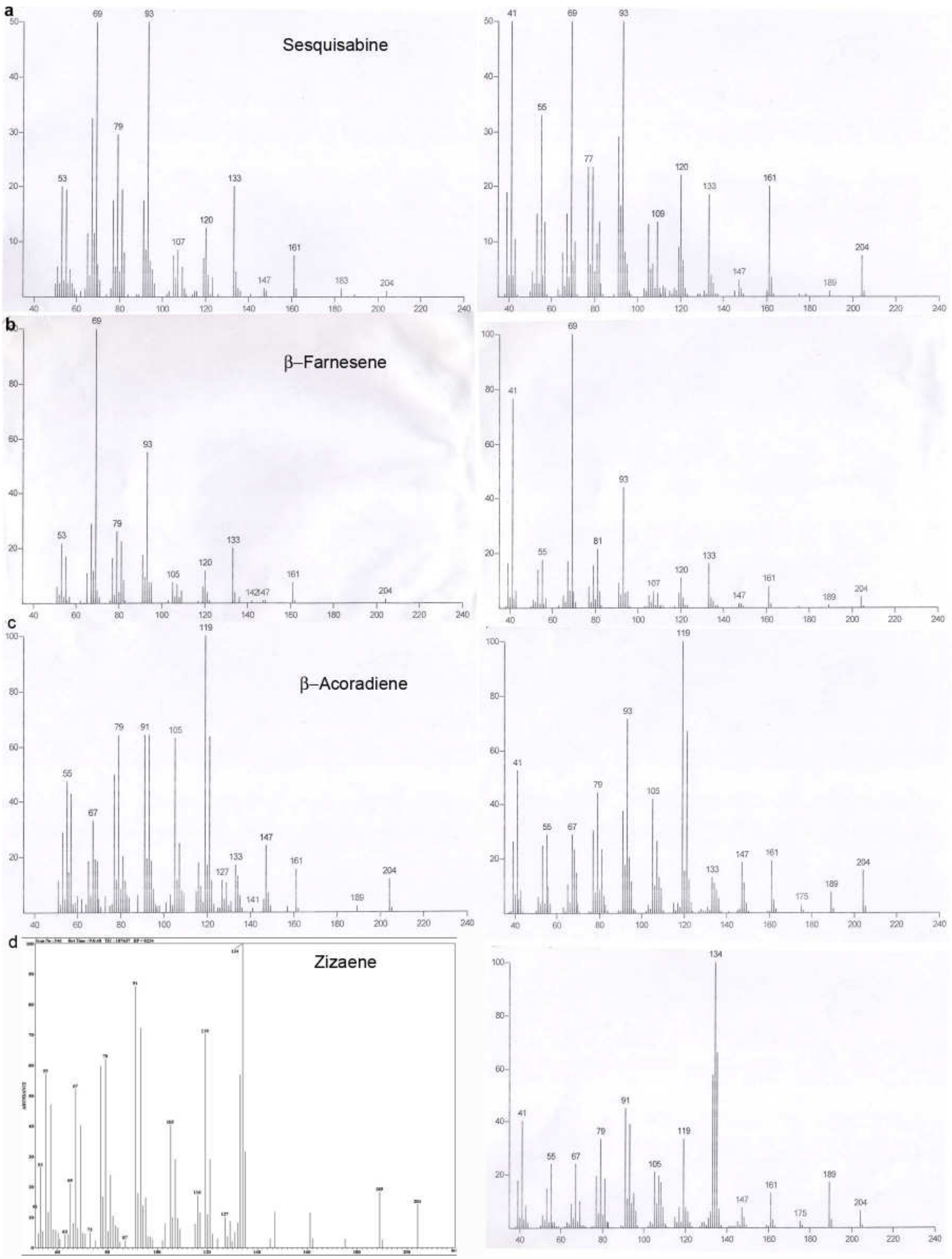
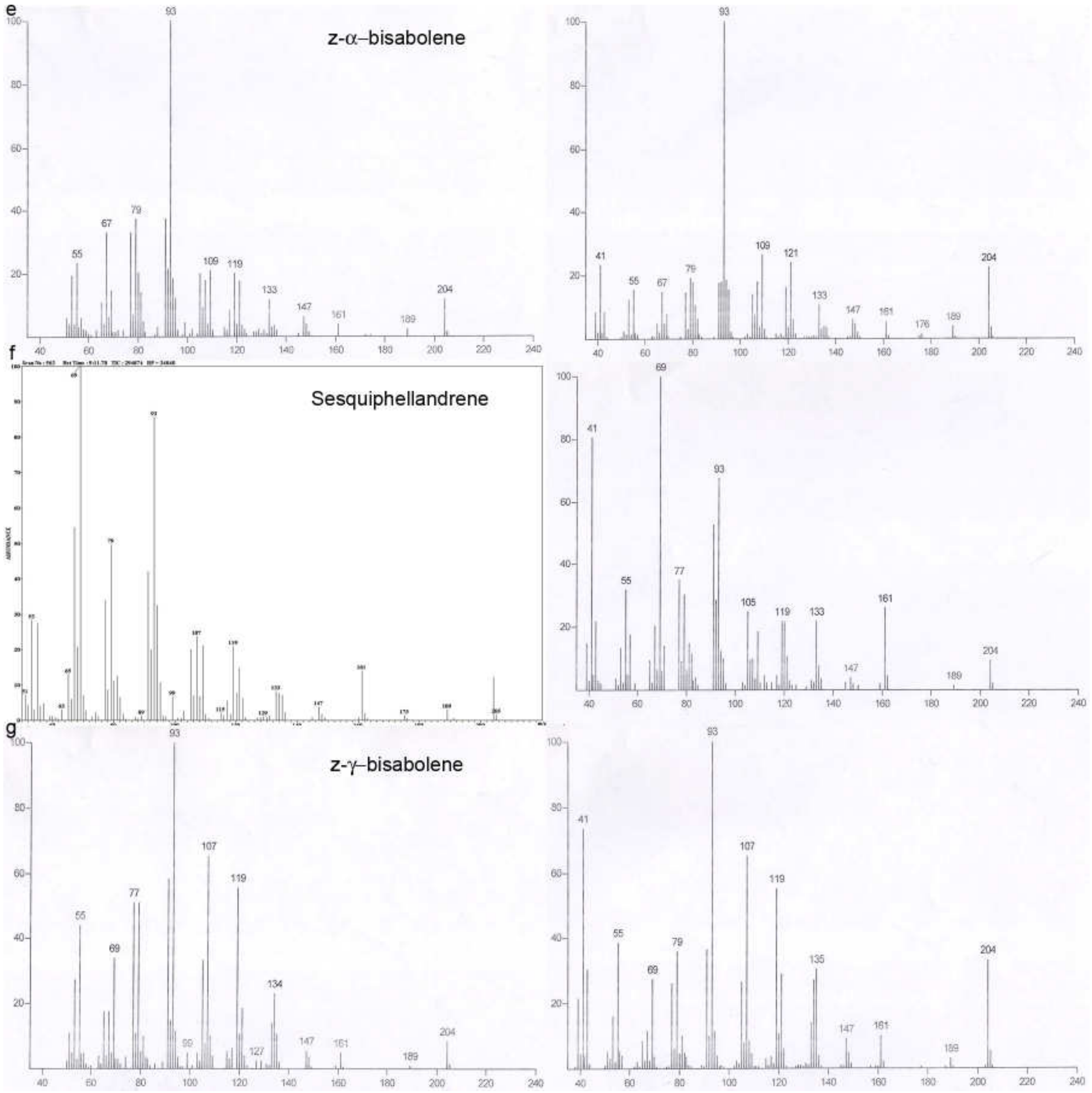


Figure 7.3. Gas chromatographs of hexane extracts of reaction of mutant EIZS with FPP. Unidentified products are labeled as **A**-**E**, and identified products are labeled as follows: Sesquisabine **(1)**, β-farnesene **(2)**, *epi*-Isozizaene **(3)**, Zizaene **(4)**, β-acoradiene **(5)**, Z-α-bisabolene **(6)**, Sesquiphellandrene **(7)**, Z-γ-bisabolene **(8)**. (a) F96A EIZS. (b) F198A EIZS. (c) W203F EIZS.

Figure 7.4 (Following pages). Mass spectra of identified sesquiterpene products of F96A, F198A, and W203F EIZS. For each product the experimental MS is shown on the left and the reference MS from the MassFinder 3.0 database is shown on the right. (a) Sesquisabine. (b) β-farnesene. (c) β-acoradiene. (d) Zizaene. (e) Z-α-bisabolene. (f) Sesquiphellandrene. (g) Z-γ-bisabolene.


a
Sesquisabine
69
93
79
53
133
120
107
161
147
183
204
41
69
93
55
77
120
133
161
109
147
189
204
b
β–Farnesene
69
93
79
53
133
120
105
161
142
147
204
69
41
93
81
55
133
120
107
161
147
189
204
c
β–Acoradiene
119
79
91
105
55
67
147
133
127
161
141
189
204
119
93
41
79
105
55
67
147
161
133
204
189
175
d
Zizaene
134
91
119
79
55
67
105
189
204
134
41
91
79
119
55
67
105
189
161
147
204
175


e
z-α–bisabolene
93
79
67
55
109
119
133
147
161
189
204
93
41
55
67
79
109
121
133
147
161
176
189
204
f
Sesquiphellandrene
69
93
79
41
55
77
105
119
133
161
147
189
204
g
z-γ–bisabolene
93
107
119
77
55
69
134
99
127
147
161
189
204
93
41
107
119
55
79
69
135
147
161
189
204

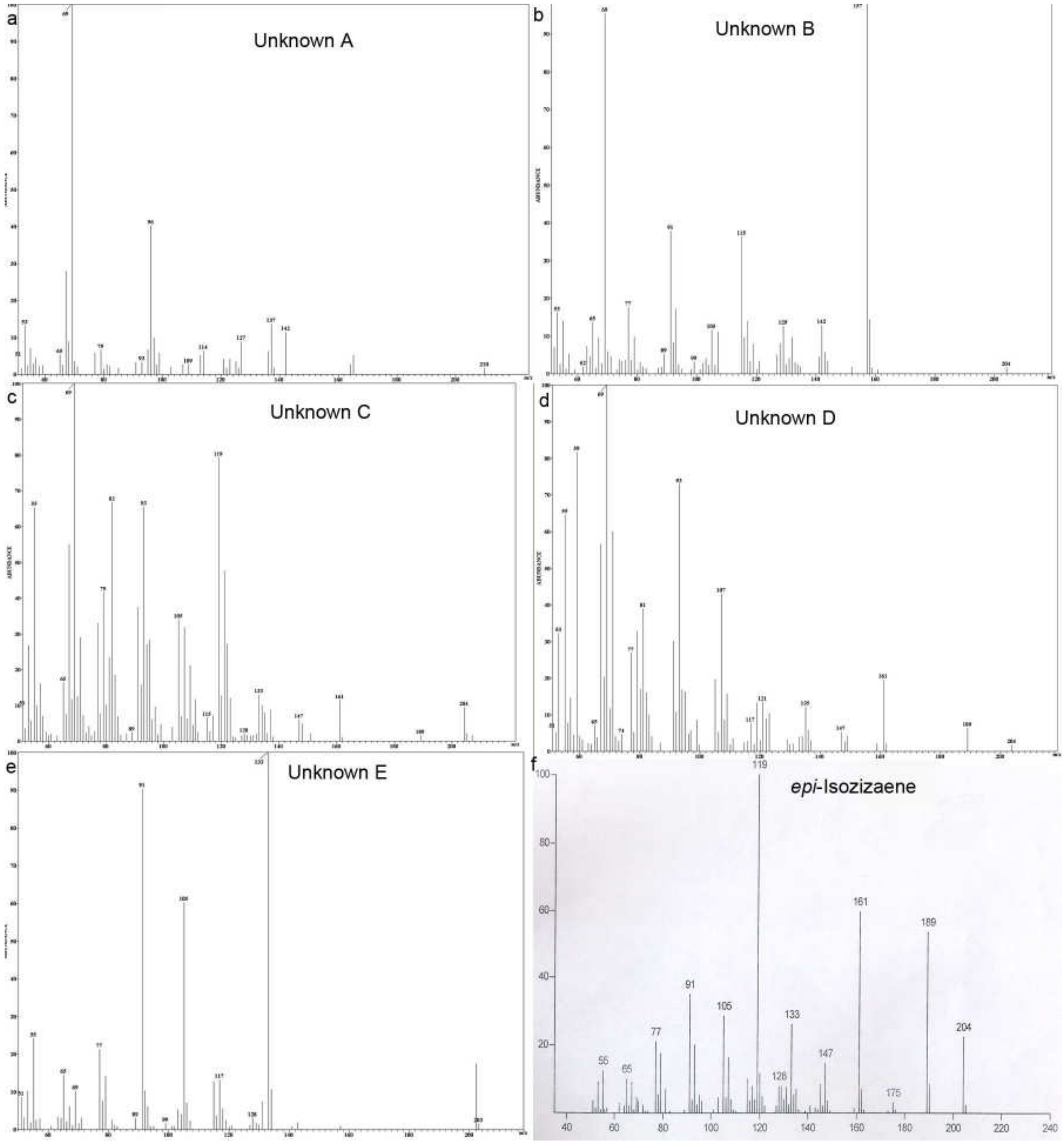


Figure 7.5. Mass spectra of unidentified sesquiterpene products (a-e) of F96A, F198A, and W203F EIZS, and (f) *epi*-Isozizaene.

Table 7.4. Distribution of sesquiterpene products from wild-type EIZS and site specific mutants. Relative product percentages were calculated from the relative peak areas of the sesquiterpene products. The relative product proportions are based on the assumptions that the relative areas of the peaks in the gas chromatogram represent the relative proportion of each sesquiterpene analyte, and the concentration of each sesquiterpene analyte is within the linear range for detection by the instrument.

Protein	Relative product percentage (%)													
	Unknown A	Sesquisabine	β–Farnesene	*epi*-Isozizaene	Zizaene	β-acoradiene	Z-α-bisabolene	Sesquiphellandrene	Z-γ-bisabolene	Unknown B	Unknown C	Unknown D	Unknown E	α-cedrene
WT	-	2	5	79	9		-	1	-	-	-	-	-	2
F96A	7	9	70	8	-	-	-	-	7	-	-	-	-	-
F198A	6	20	5	-	-	12	6	13	24	-	7	7	-	-
W203F	4	7	6	14	7	-	-	-	47	6	-	8	3	-

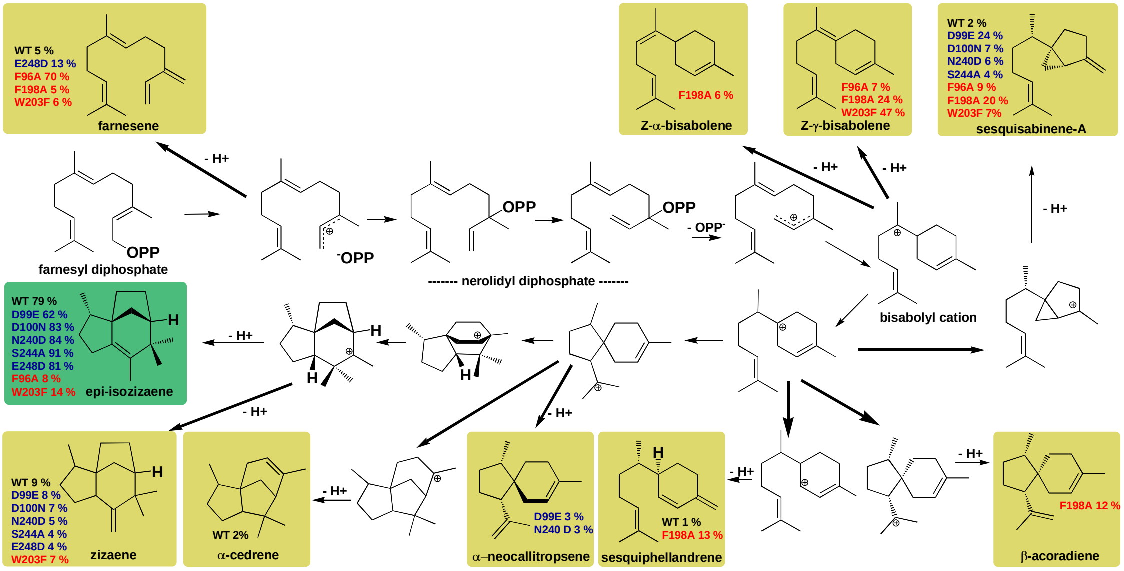

WT 5 %
E248D 13 %
F96A 70 %
F198A 5 %
W203F 6 %
farnesene
F198A 6 %
Z-α-bisabolene
F96A 7 %
F198A 24 %
W203F 47 %
Z-γ-bisabolene
WT 2 %
D99E 24 %
D100N 7 %
N240D 6 %
S244A 4 %
F96A 9 %
F198A 20 %
W203F 7%
sesquisabinene-A
- H+
OPP
farnesyl diphosphate
⁻OPP
OPP
OPP
- OPP⁻
------- nerolidyl diphosphate -------
bisabolyl cation
WT 79 %
D99E 62 %
D100N 83 %
N240D 84 %
S244A 91 %
E248D 81 %
F96A 8 %
W203F 14 %
epi-isozizaene
H
WT 9 %
D99E 8 %
D100N 7 %
N240D 5 %
S244A 4 %
E248D 4 %
W203F 7 %
zizaene
WT 2%
α-cedrene
D99E 3 %
N240 D 3 %
α–neocallitropsene
WT 1 %
F198A 13 %
sesquiphellandrene
F198A 12 %
β-acoradiene

Figure 7.6. Proposed cyclization cascade for observed products of WT and mutant EIZS. Biosynthetic versatility of EIZS can be manipulated by site-directed mutagenesis, as illustrated for sesquiterpene products identified for wild-type (WT) and mutant cyclases. In general, more diverse sesquiterpene product arrays result from the substitution of aromatic residues defining the active site contour (red labels) than from substitution of residues that coordinate the Mg^{2+} ions required for catalysis (blue labels (Lin, 2009)). For example, F198A EIZS does not generate *epi*-isozizaene at all, but instead generates a mixture of sesquisabinene-A, *Z*-α- and *Z*-γ-bisabolenes, sesquiphellandrene, and β-acoradiene as its major cyclization products. Remolding the active site contour permits the generation of alternative products as long as they can be accommodated within the remolded template, as illustrated for β-acoradiene in Figure 7.7.

Altering the active site contour of EIZS alters the conformations and cyclization trajectories of FPP and reactive carbocation intermediates, and the formation of alternative products appears to depend on how well a particular alternative product fits the remolded active site contour in a mutant cyclase. Additionally, the F96A and F198A substitutions could compromise the potential stabilization of carbocation intermediates by cation-π interactions. Some of the sesquiterpenes generated by these aromatic mutants have previously been observed as side products generated by both wild-type and mutant EIZS enzymes (Lin, 2009), while three sesquiterpene products have not previously been observed with this cyclase.

(*E*)-β-Farnesene, the major product (70%) formed by F96A EIZS, results from deprotonation at C-3^1 of the allylic cation that results from the initial ionization of FPP; other reaction products include sesquisabinene-A, *Z*-γ-bisabolene, and an unidentified hydrocarbon presumed to be a sesquiterpene based on mass spectrometry, with m/z = 204. F198A EIZS generates sesquisabinene-A, (*E*)-β-farnesene, zizaene, β-acoradiene, *Z*-α-bisabolene, sesquiphellandrene, *Z*-γ-bisabolene and 3 unidentified sesquiterpenes. Interestingly, F198A EIZS generates no *epi*-isozizaene. The predominant product of W203F EIZS is *Z*-γ-bisabolene, generated by the abstraction of a proton from the intermediate bisabolyl cation (Figure 7.6). *epi*-Isozizaene accounts for 14% of the products of the W203F mutant; the remaining products include sesquisabinene, (*E*)-β-farnesene, zizaene, and 4 unidentified sesquiterpenes. It is notable that *epi*-isozizaene biosynthetic activity is preserved, if only partially so, in the EIZS mutant with the most conservative aromatic-aromatic substitution, which presumably preserves more of the general contour and electrostatic profile of the active site.

7.4 Crystal Structures of Mutant EIZS

7.4.1 F198A EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex

In order to investigate structural changes in the active site resulting from mutagenesis of aromatic residues, the X-ray crystal structure of F198A EIZS was determined at 1.64 Å resolution. This mutant was selected for X-ray crystallographic study because it retains some catalytic activity (Table 7.3) and exhibits a remarkably altered product array (Table 7.4). Overall, the F198A mutation results in minimal structural perturbations, and the r.m.s. deviation between the structure of wild-type EIZS and F198A EIZS is 0.10 Å for 340 Cα atoms. In the active site, the largest structural changes resulting from the F198A substitution are a ~30° rotation of the side chain of F95 and an alternative rotamer of M73. The binding mode of $Mg^{2+}{}_3$-PP_i is identical to that observed in the wild-type enzyme; however, the BTAC molecule occupies an alternative position such that its benzyl ring makes quadrupole-quadrupole interactions with the aromatic rings of F95, F96, W203, and W325 (Figure 7.8). Surprisingly, 4 solvent molecules are observed in the active site, forming a hydrogen bonding network with N233 and the backbone carbonyl of A236. Comparison of the contours of WT and F198A EIZS helps to explain the production of side products such as β-acoradiene by F198A EIZS, due to the higher complementary in shape to bisabolene-derived cyclization products (Figure 7.7).

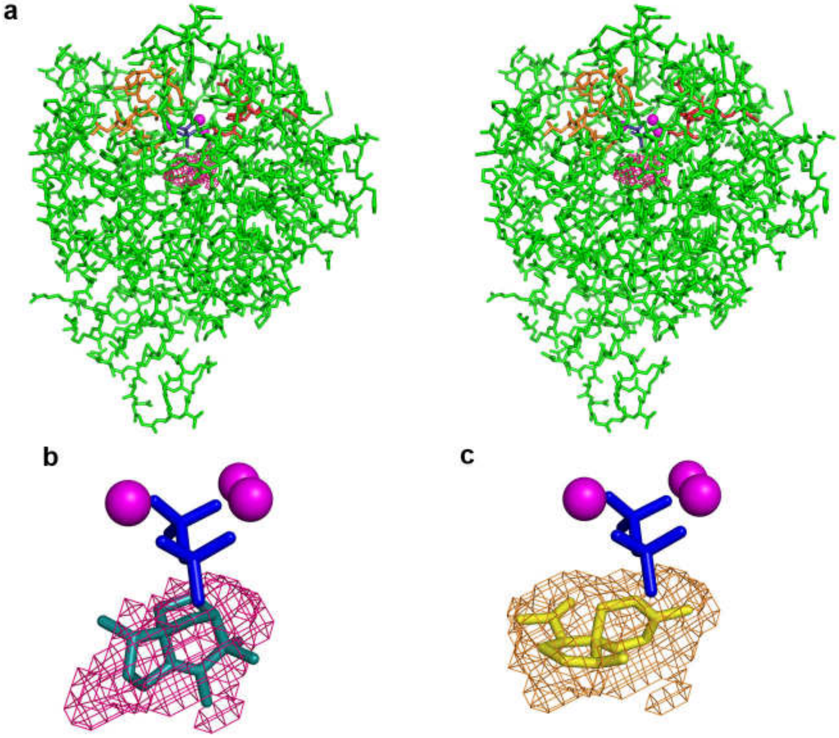


Figure 7.7. (a) A stereoview of the active site surface contour encapsulated by the closed conformation of EIZS is shown as magenta meshwork. The aspartate-rich motif (red) and the NSE motif (orange) are oriented as in Figure 1. (b) The cyclization product, *epi*-isozizaene, is modeled into the enclosed active site contour of EIZS (magenta meshwork), and the location of the $Mg^{2+}{}_3$-PP_i cluster is shown as a visual reference. (c) The enclosed active site contour of F198A EIZS (light brown meshwork) into which the new cyclization product β-acoradiene is modeled. The remolded active site contour in this mutant prevents *epi*-isozizaene formation but permits the formation of new or alternative sesquiterpene products predominantly derived from the bisabolyl carbocation intermediate.

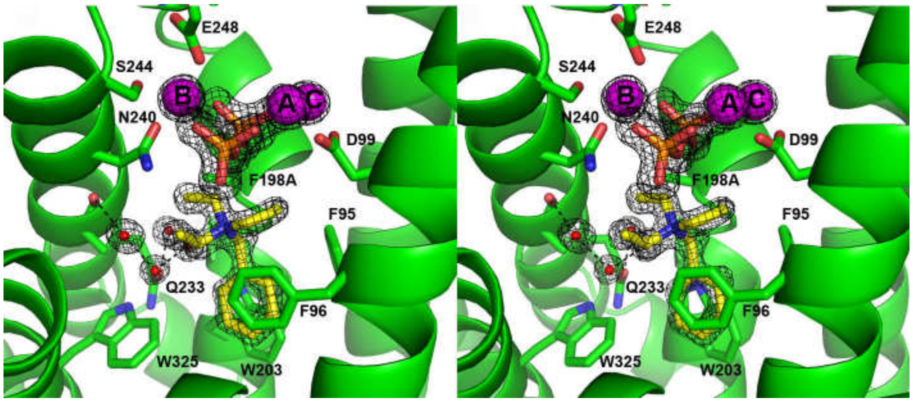

E248
S244
B
A
C
N240
D99
F198A
F95
Q233
F96
W325
W203
E248
S244
B
A
C
N240
D99
F198A
F95
Q233
F96
W325
W203

Figure 7.8. Stereoview of the active site of F198A EIZS-$Mg^{2+}{}_3$-PP_i-BTAC complex. Simulated annealing omit maps (black) of the PP_i anion, Mg^{2+} ions, and BTAC in the active site of F198A EIZS, contoured at 5σ. Note the alternative position of BTAC resulting from the F198A mutation.

7.4.2 L72V, A236G and V329A EIZS-Mg^{2+}_3*-*PP_i*-BTAC complexes*

Crystal structures of the L72V, A236G and V329A EIZS- Mg^{2+}_3-PP_i-BTAC complexes were determined at 2.10 Å, 1.76 Å, and 1.95 Å respectively. These mutants were chosen for crystallography because they expressed well and, once purified, were stable and well behaved in solution. Notably, each of these three mutations replaced a larger residue with a smaller residue, resulting in slight changes to the active site cavity contour without altering the polarity of the cavity. All three mutants crystallized under the same conditions as the WT enzyme, including the addition of pyrophosphate and BTAC, resulting in the respective mutant EIZS-Mg^{2+}_3-PP_i-BTAC complexes. Overall, each single mutation caused minimal structural perturbations; the rmsd between the structure of WT EIZS and the L72V, A236G and V329A mutants is 0.13 Å, 0.096 Å, and 0.125 Å respectively for 340 Cα atoms. The binding mode of the trinuclear Mg^{2+} cluster and pyrophosphate was unchanged in the three mutants, and furthermore the position and orientation of BTAC was also unchanged, except for the L72V mutant in which BTAC moved ~0.5 Å deeper into the active site and rotated by ~20° (Figure 7.9). Minimal changes in the active site are confirmed by active site volume calculations using the program VOIDOO (Kleywegt, 1994). After removing BTAC's coordinates from the active site, the calculated volume of the WT cavity is 62.22 Å^3, compared with 62.71 Å^3 for L72V, 62.86 Å^3 for A236G and 61.71 Å^3 for V329A. To determine whether these seemingly trivial volume changes result in altered products arrays, future experiments include GC-MS analysis of the products of these mutants.

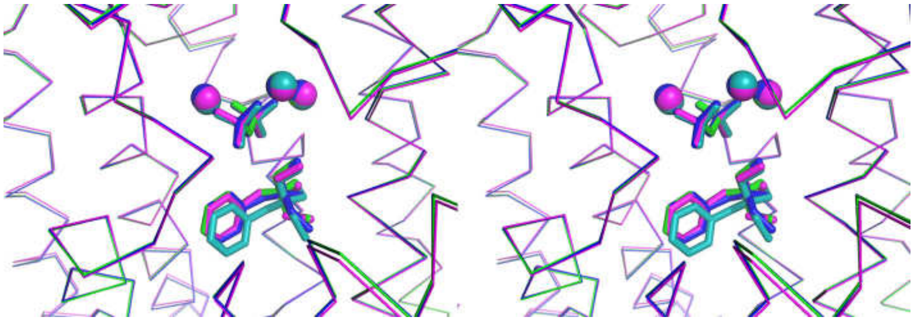


Figure 7.9. Stereoview of an overlay of the active sites of WT (green), L72V (cyan), A236G (blue) and V329L (purple) EIZS-Mg^{2+}_3-PP_i-BTAC complexes.

7.5 Discussion

Mutagenesis of aromatic residues that contribute directly to the active site contour significantly compromises the fidelity of *epi*-isozizaene biosynthesis (Figure 7.6), with the F198A substitution completely suppressing *epi*-isozizaene formation and redirecting the cyclization cascade toward the generation of alternative acyclic, monocyclic, and bicyclic sesquiterpenes. Thus, remolding the active site contour by mutagenesis opens up new cyclization trajectories while closing off old ones.

The appearance of low levels of new or alternative cyclization products resulting from mutagenesis of the active site contour in a terpenoid cyclase may reflect past or future evolutionary potential, i.e., catalytic promiscuity in enzyme function may provide a "toehold of evolution" (Petsko, 1993). The evolution of biosynthetic diversity in this family of enzymes is achieved by simply remolding the active site contour to promote one cyclization pathway while suppressing hundreds of others, and it is notable that this is readily achieved by only a handful of amino acid substitutions. The current work represents the first step in deciphering the relationship between the structure of the EIZS active site and its biosynthetic specificity as a product-like template for terpenoid cyclization reactions: the three-dimensional contour of the active site can be remolded to better fit another product and disfavor others, even to the point of excluding *epi*-isozizaene formation. That the biosynthetic specificity of a terpenoid cyclase is so sensitive to and so readily manipulated by minimal mutagenesis in nature or in the laboratory will likely contribute to the growing structural and stereochemical diversity of the terpenome.

Chapter 8: Trinuclear Metal Clusters in Catalysis by Terpenoid Synthases

8.1 Introduction

Terpenoid synthases are ubiquitous enzymes that catalyze the formation of structurally and stereochemically diverse isoprenoid natural products. Many isoprenoid coupling enzymes and terpenoid cyclases from bacteria, fungi, protists, plants, and animals share the class I terpenoid synthase fold. This family of enzymes, which is responsible for such a diverse range of products, is of great interest medicinally since many terpenoid natural products exhibit anti-cancer, anti-malarial, and anti-microbial activities (Aharoni, 2005). Furthermore, in humans, the 15 carbon linear isoprenoid farnesyl diphosphate (FPP) is a precursor in the biosynthesis of steroids and is also utilized for posttranslational prenylation of Ras in GTPase signaling (Rondeau, 2006; Agrawal, 2009). Recently, human farnesyl diphosphate synthase (FPP synthase) has been identified as the target of nitrogen-containing bisphosphonate drugs used for the treatment of bone diseases such as osteoporosis, hypercalcemia, and metastatic bone disease (Ebetino, 2005; Licata, 2005). Moreover, protozoan FPP synthase homologues have recently been identified as targets for the treatment of parasitic infections, including Chagas disease and African sleeping sickness (Ferella, 2008).

Despite generally low amino acid sequence identity class I terpenoid synthases, which adopt the FPP synthase α-helical fold, contain conserved metal binding motifs that coordinate to a trinuclear metal cluster. This cluster not only serves to bind and orient the flexible isoprenoid substrate in the precatalytic Michaelis complex, but it also triggers the departure of the diphosphate leaving group to generate a carbocation that initiates catalysis. Review of the available crystal structures of class I terpenoid synthases complexed with trinuclear metal clusters and either isoprenoid diphosphates or inorganic pyrophosphate highlights the conserved structural aspects of the trinuclear metal cluster and the additional conserved hydrogen bond donors that are required for catalysis.

8.2 Isoprenoid Coupling Enzymes

8.2.1 Farnesyl diphosphate synthase

Farnesyl disphosphate synthase, the archetypical prenyltransferase, catalyzes the formation of farnesyl diphosphate (FPP), the linear isoprenoid precursor of sesquiterpene natural products. Chain elongation to form FPP proceeds in two distinct steps (Figure 4.1): first, isopentenyl disphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are coupled to form geranyl diphosphate (GPP), and then a second molecule of IPP is coupled to GPP to form FPP. The first crystal structure of FPP synthase was that of the avian enzyme (Tarshis, 1994), which revealed a novel α-helical fold. The structure revealed two conserved aspartate-rich (DDXXD) sequences (Ashby, 1990) on helices D and H, which flank the mouth of the active site cavity. Additionally, a single Sm^{3+} ion, used for heavy metal derivatization for MIR phasing, was bound by each DDXXD motif.

A decade later, the crystal structure of *E. coli* FPP synthase was the first to reveal the binding of a trinuclear magnesium cluster in the active site of an isoprenoid coupling enzyme (Hosfield, 2004), similar to the trinuclear magnesium clusters previously observed in fungal and plant terpenoid cyclases (Rynkiewicz, 2001; Whittington, 2002).

The structure of *E. coli* FPP synthase was solved as the enzyme-substrate ternary complex with the noncleavable DMAPP analogue, dimethylallyl S-thiolodiphosphate (DMSPP), and a molecule of IPP. Applying the $Mg^{2+}{}_A$, $Mg^{2+}{}_B$, and $Mg^{2+}{}_C$ nomenclature first established for the trinuclear magnesium cluster of trichodiene synthase (Rynkiewicz, 2001), the crystal structure of the *E. coli* FPP synthase-$Mg^{2+}{}_3$-DMSPP-IPP complex reveals octahedral coordination of all three metal ions (Figure 8.1 (a)): $Mg^{2+}{}_A$ is coordinated by D105 and D111 of the first aspartate-rich motif on helix D, two diphosphate oxygen atoms, and two water molecules; $Mg^{2+}{}_C$ is coordinated by the side chains of D105, and D111, as well as one diphosphate oxygen and three water molecules; and $Mg^{2+}{}_B$ is coordinated by D244 of the second aspartate-rich motif, two diphosphate oxygen atoms, and three water molecules. The diphosphate group of DMSPP also accepts hydrogen bonds from R116, K202, and K258.

More recently, the structure of the human FPP synthase-$Mg^{2+}{}_3$-zoledronate-IPP complex (Rondeau, 2006) reveals complete conservation of $Mg^{2+}{}_3$-diphosphate recognition between *E. coli* and human FPP synthases (Figure 8.1 (b)). In human FPP synthase, two DDXXD motifs coordinate to the $Mg^{2+}{}_3$ cluster: the first aspartate of the **D^{103}**DXX**D^{107}** (hereafter boldface residues indicate Mg^{2+} coordinating residues) motif coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,syn*-bidentate geometry, and one oxygen atom of D107 bridges $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,anti*-coordination stereochemistry; the first aspartate of the second **D^{243}**DXXD motif coordinates to $Mg^{2+}{}_B$. The diphosphate moiety additionally accepts hydrogen bonds from R112, K200, and K257. Interestingly, the closed active site conformation is also stabilized by newly formed hydrogen bonds between K266 and D107 and D174. The r.m.s. deviation between the unliganded enzyme and the closed conformation of the $Mg^{2+}{}_3$-zoledronate-IPP complex is 1.3 Å (341 Cα atoms). Analysis of X-ray crystal structures of several human FPP synthase-$Mg^{2+}{}_3$-bisphosphonate complexes suggests a two-step mechanism for substrate binding (Rondeau, 2006). First, the binding of DMAPP and 3 Mg^{2+} ions brings together the two DDXXD motifs, and loops D-E and H-I come together to form a hydrogen bond between T111 and the backbone of I258. These structural changes close the entrance to the allylic binding site and complete the formation of the IPP binding site. Secondly, as IPP binds, and as the basic C-terminal tail of the enzyme becomes ordered and closes the IPP binding site, IPP and DMAPP are properly oriented for catalysis.

The binding of a trinuclear magnesium cluster is similarly conserved in FPP synthases from parasitic organisms. The flagellated protozoan *T. cruzi* causes Chagas disease, primarily in Latin America (Tanowitz, 2009). Bisphosphonates have emerged as a potential treatment for Chagas disease by inhibiting *T. cruzi* FPP synthase (Garzoni, 2004). The crystal structures of *T. cruzi* FPP synthase-$Mg^{2+}{}_3$-inhibitor complexes (Gabelli, 2006) suggest conservation of the trinuclear magnesium cluster for substrate binding and catalysis. In the complex with risedronate (Figure 8.1 (c)), the first carboxylate of the **D**$^{\mathbf{98}}$DXX**D**$^{\mathbf{102}}$ motif on helix D coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,syn*-bidentate geometry, and one oxygen atom of D102 bridges $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,anti*-coordination stereochemistry. The first carboxylate of the **D**$^{\mathbf{250}}$DXXD motif on helix H is the only residue that directly coordinates to $Mg^{2+}{}_B$; however, D251 and D254 indirectly interact with $Mg^{2+}{}_B$ via bridging water molecules. Each Mg^{2+} ion is coordinated with octahedral geometry, with non-protein coordination sites occupied by oxygen atoms of the inhibitor phosphonate groups and water molecules. Oxygen atoms of the two risedronate phosphonate groups accept hydrogen bonds from the side chains of R107, K207, and K264.

Trypanosoma brucei is an African parasitic protist, and its FPP synthase is related to that of *T. cruzi* by 70 % amino acid sequence identity. The crystal structure of *T. brucei* FPP synthase complexed with 3 Mg^{2+} ions and the bisphosphonate inhibitor BPH-721 (Zhang, 2009) reveals conservation of the trinuclear magnesium cluster for substrate binding and catalysis (Figure 8.1 (d)). The first carboxylate of the $\mathbf{D^{103}}$DXX$\mathbf{D^{107}}$ motif on helix D coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn*,*syn*-bidentate geometry, and one oxygen atom of D107 bridges $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn*,*anti*-coordination stereochemistry. The first aspartate in the $\mathbf{D^{255}}$DXXD motif on helix H is the only residue that directly coordinates to $Mg^{2+}{}_B$; however, D256 and D259 indirectly interact with $Mg^{2+}{}_B$ via bridging water molecules. Oxygen atoms of the two phosphonate groups of the inhibitor BPH-721 also accept hydrogen bonds from the side chains of R112, K212, and K269 (Mao, 2006).

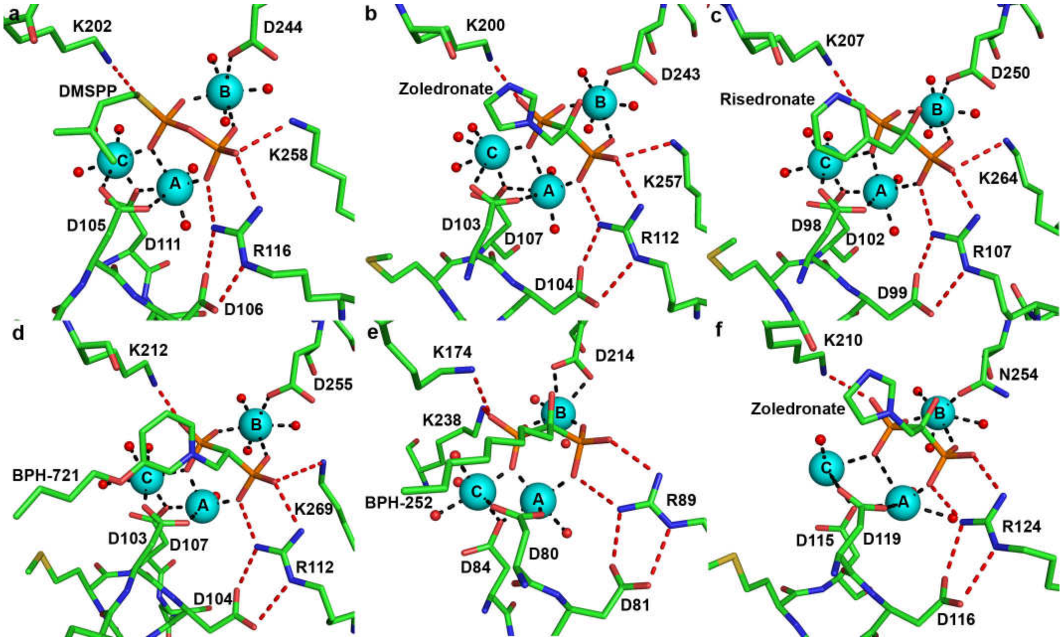

a
K202
D244
DMSPP
B
C
A
K258
D105
D111
R116
D106
b
K200
D243
Zoledronate
B
C
A
K257
D103
D107
R112
D104
c
K207
D250
Risedronate
B
C
A
K264
D98
D102
R107
D99
d
K212
D255
B
BPH-721
C
A
K269
D103
D107
R112
D104
e
K174
D214
K238
B
BPH-252
C
A
R89
D84
D80
D81
f
K210
N254
Zoledronate
B
C
A
R124
D115
D119
D116

Figure 8.1. Conservation of $Mg^{2+}{}_3$-PP_i and -diphosphate binding motifs among isoprenoid coupling enzymes. Metal coordination (black) and hydrogen bond (red) interactions with phosphate(s) are indicated. (a) *E. coli* FPP synthase-$Mg^{2+}{}_3$-DMSPP-IPP complex (PDB code 2EGW); (b) human FPP synthase-$Mg^{2+}{}_3$-zoledronate complex (PDB code 2F8Z); (c) *T. cruzi* FPP synthase-$Mg^{2+}{}_3$-risedronate complex (PDB code 1YHL); (d) *T. brucei* FPP synthase-$Mg^{2+}{}_3$-BPH-721 complex (PDB code 3DYH); (e) *S. cerevisae* GGPP synthase-$Mg^{2+}{}_2$-BPH-252 complex (PDB code 2Z4X); (f) *C. parvum* nonspecific prenyl synthase-$Mg^{2+}{}_3$-zoledronate complex (PDB code 2Q58).

8.2.2 Geranylgeranyl diphosphate synthase

Geranylgeranyl disphosphate synthase (GGPP synthase) catalyses the condensation of IPP and FPP to form GGPP (Figure 4.1). Recently, GGPP synthase has emerged as a pharmaceutical target for the treatment of cancer since geranylgeranylation is involved in Rac, Rap and Rho signaling pathways (Russell, 2006). The crystal structures of GGPP synthases from *Thermus thermophilus* (Nishio, 2004), *Sinapis alba* (Kloer, 2006), and *Saccharomyces cerevisiae* (Chang, 2006) have been determined in addition to that of human GGPP synthase (Kavanagh, 2006). However, a crystal structure containing a complete trinuclear magnesium cluster has only been observed in the active site of monomer B of the *S. cerevisiae* GGPP synthase-Mg^{2+}_3-BPH-252 complex (Figure 8.1 (e)) (Chen, 2008). The first aspartate of the $\mathbf{D^{80}}$DIE$\mathbf{D^{84}}$ motif coordinates to Mg^{2+}_A and Mg^{2+}_C with *syn,syn*-bidentate geometry, and one oxygen atom of D84 bridges Mg^{2+}_A and Mg^{2+}_C with *syn,anti*-coordination stereochemistry; the first aspartate of the second $\mathbf{D^{214}}$DYLN motif coordinates to Mg^{2+}_B. Each Mg^{2+} ion is coordinated with octahedral geometry, with non-protein coordination sites occupied by oxygen atoms of the inhibitor phosphonate groups and water molecules. Oxygen atoms of the two phosphonate groups of the inhibitor BPH-252 also accept hydrogen bonds from R89, K174 and K238.

8.2.3 Nonspecific prenyl synthase

Recently, the crystal structure of a nonspecific prenyl synthase from *Cryptosporidium parvum* has been determined (Artz, 2008). *C. parvum* causes livestock infections and is classified as a bioterrorism threat by the Centers for Disease Control and Prevention (Hashsham, 2004). The enzyme has a unique ability to catalyze chain elongation reactions with isoprenoid substrates of various lengths to generate C_{20}-C_{45} linear isoprenoids products. The crystal structure of the enzyme reveals conservation of the classic α-helical terpenoid synthase fold, and its complex with the inhibitor risedronate reveals that a complete trinuclear magnesium cluster is coordinated by DDXXD and NDXXD motifs (Figure 8.1 (f)) (Artz, 2008). The first carboxylate of the $\mathbf{D^{115}}$DXX$\mathbf{D^{119}}$ motif on helix D is oriented for coordination to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,syn*-bidentate geometry; however, the distance between D115 and $Mg^{2+}{}_C$ is 3.14 Å, thus too long to be considered an inner-sphere metal coordination interaction. One oxygen atom of the third aspartate, D119, bridges $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,anti*-coordination stereochemistry. N254 of the $\mathbf{N^{254}}$DXXD motif on helix H coordinates to $Mg^{2+}{}_B$, and D255 and D258 indirectly interact with $Mg^{2+}{}_B$ via bridging waters. The diphosphate moiety additionally accepts hydrogen bonds from K210 and G251, and the closed active site conformation is stabilized by hydrogen bonds between D116 of the DDXXD motif and R124, and D255 of the NDXXD motif and Q251.

8.3 Isoprenoid Cyclization Enzymes

8.3.1 Fungal cyclases

The sesquiterpene cyclase trichodiene synthase from *Fusarium sporotrichioides* catalyzes the first committed step in the biosynthesis of nearly 100 different trichothecene mycotoxins. Trichodiene synthase is one of the most thoroughly studied terpenoid cyclases, and enzymological and crystallographic studies have illuminated important features in the cyclization mechanism (recently reviewed in (Christianson, 2006)). Recent computational studies have also provided new insight on the catalytic mechanism (Hong, 2009). The structures of unliganded trichodiene synthase and the trichodiene synthase-$Mg^{2+}{}_3$-PP_i complex were the first to reveal the binding of a trinuclear magnesium cluster in the active site of a terpenoid synthase (Figure 8.2 (a)) (Rynkiewicz, 2001). The first aspartate of the $\mathbf{D}^{\mathbf{100}}$DXXD motif on helix D coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,syn*-bidentate geometry. The second metal binding motif $\mathbf{N}^{\mathbf{225}}DLM\mathbf{S}^{\mathbf{229}}FYK\mathbf{E}^{\mathbf{333}}$ is located on helix H and coordinates to $Mg^{2+}{}_B$. All three metal ions are additionally coordinated by PP_i and solvent molecules to complete octahedral coordination polyhedra.

Superposition of the unliganded and $Mg^{2+}{}_3$-PP_i complexed trichodiene synthase structures reveals conformational changes that cap the active site upon ligand binding. Overall, the r.m.s. deviation between the native and liganded structures is 1.4 Å for 349 $C\alpha$ atoms. Interestingly, upon ligand binding, D101 in the aspartate-rich motif forms a salt bridge with R304, which donates a hydrogen bond to PP_i. In addition to Mg^{2+} coordination interactions, the PP_i anion also accepts hydrogen bonds from R182, K232, and Y305. The D101-R304-PP_i hydrogen bond network appears to link substrate binding with the transition between the open and closed active site conformations. Assuming that the diphosphate group of FPP triggers the same structural changes as observed for PP_i, the substrate is sequestered from bulk solvent and the complete trinuclear magnesium cluster triggers departure of the diphosphate leaving group to generate the carbocation that initiates the cyclization cascade. The seemingly conservative D100E mutation results in a 22-fold loss in catalytic activity (measured by k_{cat}/K_M) and structural studies indicate that the additional methylene group of E100 perturbs the $Mg^{2+}{}_3$-PP_i complex such that $Mg^{2+}{}_A$ binding is weakened, E233 breaks its coordination interaction with $Mg^{2+}{}_B$, and

Mg^{2+}_C is dissociated; additionally, hydrogen bond interactions between PP_i and R182 and R304 are broken (Rynkiewicz, 2002; Vedula, 2005b).

The role of the D101-R304 salt bridge in closing the trichodiene synthase active site has been explored in mutagenesis studies. The D101E mutation results in a moderate 5-fold decrease in catalytic activity; however, there are no crystal structures of this mutant available for study (Cane, 1996). In contrast, the R304K mutant results in a 5000 fold decrease in catalytic activity, and the crystal structure of R304K trichodiene synthase complexed with $Mg^{2+}{}_3$-PP_i-*(R)*-azabisabolene reveals the loss of the expected hydrogen bond between K304 and D101 (Vedula, 2005a). Although the PP_i binding motif remains intact in the R304K mutant, it is evident that the R304-D101 hydrogen bond is critical for properly activating the substrate diphosphate group. In contrast, while Y305 donates a hydrogen bond to PP_i in the wild-type enzyme complex with $Mg^{2+}{}_3$-PP_i, catalytic activity and PP_i binding are not significantly affected in the Y305F mutant (Cane, 1995; Vedula, 2005b).

Another fungal cyclase that has been the subject of extensive structural and functional study is aristolochene synthase, which is a sesquiterpene cyclase that catalyzes the cyclization of FPP to form (+)-aristolochene. Structures of aristolochene synthases from *Penicillium roqueforti* (Caruthers, 2000) and *Aspergillus terreus* (Shishova, 2007) have been solved; these enzymes are related by 61 % amino acid sequence identity. Although there is no crystal structure of *P. roqueforti* aristolochene synthase complexed with $Mg^{2+}{}_3$-PP_i, the structure of the *A. terreus* aristolochene synthase $Mg^{2+}{}_3$-PP_i complex

(Shishova, 2007) (Figure 8.2 (b)) indicates conservation of the $Mg^{2+}{}_3$-PP_i binding motif first observed in trichodiene synthase (Rynkiewicz, 2001).

The aspartate-rich motif of aristolochene synthase on helix D appears as $\mathbf{D}^{\mathbf{90}}$DXXE. The carboxylate side chain of D90 coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_B$ with *syn*,*syn*-bidentate geometry, and is the only residue in the aspartate motif that coordinates to the Mg^{2+} ions. The carboxylate group of D91 makes a salt bridge with R304, and the final carboxylate in the motif is E119, which accepts a hydrogen bond from a water molecule coordinated to $Mg^{2+}{}_A$. The second metal binding motif $\mathbf{N}^{\mathbf{219}}DIY\mathbf{S}^{\mathbf{223}}YEK\mathbf{E}^{\mathbf{227}}$ is located on helix H and chelates $Mg^{2+}{}_B$ (Shishova, 2007), consistent with the structures of terpenoid cyclases from plants, bacteria, and fungi (Starks, 1997; Rynkiewicz, 2001; Whittington, 2002; Hyatt, 2007; Aaron, 2010). As found in the active sites of trichodiene synthase (Rynkiewicz, 2001) and *epi*-isozizaene synthase (Aaron, 2010), PP_i binding in aristolochene synthase is similarly accommodated by hydrogen bonds donated from two arginines (R175 and F314), one lysine (K226), and one tyrosine (Y315) (Figure 8.2 (b)).

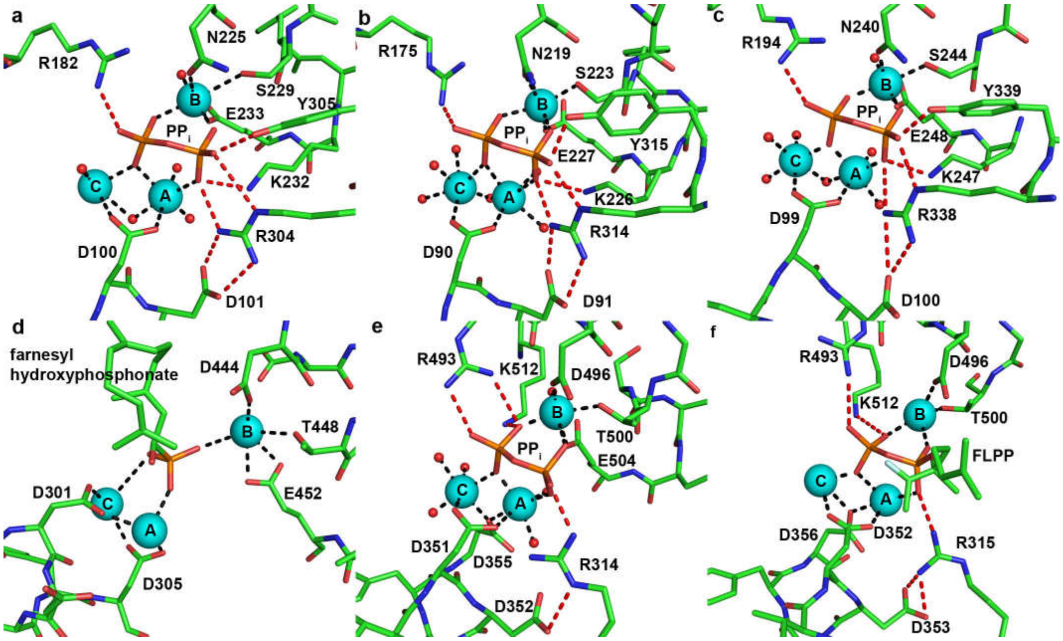

a
R182
N225
B
S229
Y305
E233
PPi
C
A
K232
R304
D100
D101
b
R175
N219
S223
B
PPi
E227
Y315
C
A
K226
R314
D90
D91
c
R194
N240
S244
B
Y339
PPi
E248
C
A
K247
D99
R338
D100
d
farnesyl
hydroxyphosphonate
D444
B
T448
E452
D301
C
A
D305
e
R493
K512
D496
B
T500
PPi
E504
C
A
D351
D355
R314
D352
f
R493
D496
K512
B
T500
FLPP
C
A
D356
D352
R315
D353

Figure 8.2. Conservation of Mg$^{2+}_3$-PP$_i$ and -diphosphate binding motifs among terpenoid cyclases. Metal coordination (black) and hydrogen bond (red) interactions with phosphate(s) are indicated. (a) *F. sporotrichioides* trichodiene synthase-Mg$^{2+}_3$-PP$_i$ complex (PDB code 1JFG); (b) *A. terreus* aristolochene synthase-Mg$^{2+}_3$-PP$_i$ complex (PDB code 2OA6); (c) *S. coelicolor epi*-isozizaene synthase-Mg$^{2+}_3$-PP$_i$ complex (PDB code 3KB9); (d) *N. tabacum* 5-*epi*-aristolochene synthase-Mg$^{2+}_3$-farnesyl hydroxyphosphonate complex (PDB code 5EAT; note that many of the metal-phosphate interactions indicated are too long to be considered inner-sphere metal coordination interactions); (e) *S. officinalis* (+)-bornyl diphosphate synthase-Mg$^{2+}_3$-PP$_i$ complex (PDB code 1N22; metal ions are labeled according to the convention first established for trichodiene synthase); (f) *M. spicata* limonene synthase-Mn$^{2+}_3$-FLPP complex (PDB code 2ONG; conserved hydrogen bonding is indicated between D353 and R315 despite poor geometry).

8.3.2 Bacterial cyclases

In recent years, prokaryotes have emerged as sources of diverse isoprenoids. Specifically, a large number of novel isoprenoids have been isolated from organisms belonging to the taxonomical order *Actinomycetales* (Daum, 2009). The crystal structures of two bacterial sesquiterpene cyclases have been solved, and both derive from *Actinomycetales*: pentalenene synthase from *Streptomyces* UC5319 (Lesburg, 1997) and *epi*-isozizaene synthase from *Streptomyces coelicolor* (Aaron, 2010). Pentalenene synthase catalyzes the cyclization of FPP to form the tricyclic sesquiterpene pentalenene in the first committed step in the biosynthesis of the pentalenolactone family of antibiotics (Seto, 1980). Although the structure of pentalenene synthase was the first to be reported of a terpenoid cyclase and demonstrated that the terpenoid cyclase shared the FPP synthase fold first observed for avian FPP synthase (Tarshis, 1994), no structure of this bacterial terpenoid cyclase complexed with metal ions is available.

However, the crystal structure of *epi*-isozizaene synthase complexed with $Mg^{2+}{}_3$-PP_i-BTAC (BTAC is the benzyltriethylammonium cation, a crystallization additive) reveals that $Mg^{2+}{}_3$-PP_i binding motifs are conserved between fungal and bacterial terpenoid cyclases (Figure 8.2 (c)) (Aaron, 2010). The first aspartate of the aspartate-rich motif $\mathbf{D^{99}}$DRHD coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn*,*syn*-bidentate geometry, and $Mg^{2+}{}_B$ is chelated by $\mathbf{N^{240}}$DLC$\mathbf{S^{244}}$LPK$\mathbf{E^{248}}$. Each Mg^{2+} ion is coordinated with octahedral geometry and nonprotein coordination sites are occupied by oxygen atoms of PP_i and water molecules. The PP_i anion also accepts hydrogen bonds from the side chains of R194, K247, R338, and Y339 which correspond to R182, K232, R304 and Y305 of trichodiene synthase (Rynkiewicz, 2001) and R175, K226, R314 and Y315 of aristolochene synthase (Shishova, 2007). Pentalenene synthase and *epi*-isozizaene synthase share 24 % amino acid sequence identity, and residues that interact with Mg^{2+} ions or PP_i in *epi*-isozizaene synthase are conserved in pentalenene synthase. Superposition of the liganded closed structure of *epi*-isozizaene synthase with unliganded pentalenene synthase reveals a very similar alignment of the metal binding motifs, with pentalenene synthase helices D and H 1.5 Å further apart than in *epi*-isozizaene synthase. Accordingly, upon substrate or PP_i binding the active site of pentalenene synthase presumably undergoes a conformational change to a closed conformation comparable to that observed for the *epi*-isozizaene synthase $Mg^{2+}{}_3$-PP_i-BTAC complex.

Although the second aspartate of the *epi*-isozizaene synthase aspartate-rich motif, D100, does not directly interact with the Mg^{2+} ions or PP_i, it does accept a hydrogen bond from R338, which also donates a hydrogen bond to PP_i. Site-directed mutagenesis reveals that the D100N mutation causes a >95 % loss of activity with respect to the native enzyme (Lin, 2009), suggesting that the D100N mutation disrupts the D100-R338-PP_i hydrogen bond network presumed to be important for substrate recognition. As previously discussed, the second aspartate in the aspartate-rich motifs of trichodiene synthase and aristolochene synthase similarly stabilizes a hydrogen bond network with R304 and PP_i (Rynkiewicz, 2001; Shishova, 2007), so it appears that the bacterial and fungal cyclases share the same molecular strategy for linking the molecular recognition of the substrate diphosphate group with the active site closure mechanism (Aaron, 2010). The third aspartate in the aspartate rich motif of *epi*-isozizaene synthase, D103, points away from the active site and makes no hydrogen bond interactions that are involved in substrate binding, as also observed in $Mg^{2+}{}_3$-PP_i complexes of trichodiene synthase. The absence of a structural or catalytic role for the terminal aspartate in the aspartate-rich motif of bacterial terpenoid cyclases is supported by mutagenesis of the corresponding residue in pentalenene synthase: the D84E mutation results in a mere 3-fold loss of catalytic activity (as measured by k_{cat}/K_{M}), whereas the D80E and D81E mutations yield 3500- and 400-fold reductions in activity, respectively (Seemann, 2002).

8.3.3 Plant Cyclases

5-*epi*-Aristolochene synthase from *Nicotiana tabacum* catalyzes the cyclization of FPP to form 5-*epi*-aristolochene in the first committed step in the biosynthesis of the antifungal phytoalexin capsidiol (Starks, 1997). As the first crystal structure determined of a plant terpenoid cyclase and the second terpenoid cyclase structure to be reported, the structure of 5-*epi*-aristolochene synthase reveals the presence of 2 domains (Starks, 1997): a catalytically active C-terminal domain that adopts the α-helical class I terpenoid synthase fold, and an N-terminal domain of unknown function that exhibits an α-helical fold similar to that of a class II terpenoid synthase (Wendt, 1998). Two metal-binding motifs are identified: an aspartate-rich motif $\mathbf{D}^{\mathbf{301}}DXX\mathbf{D}^{\mathbf{305}}$, and a $\mathbf{D}^{\mathbf{444}}DTA\mathbf{T}^{\mathbf{448}}YEV\mathbf{E}^{\mathbf{452}}$ motif. While the binding of a trinuclear magnesium cluster was identified in the 5-*epi*-aristolochene synthase farnesyl hydroxyphosphonate complex (Figure 8.2 (d)), analysis of the structure reveals that many of the coordination interactions with Mg^{2+} ions range 2.2 Å – 3.7 Å, longer than expected for ideal Mg^{2+} coordination (Zheng, 2008). This could suggest that the structure is that of a partially closed conformation. Nonetheless, D301 and D305 of the aspartate-rich motif coordinate to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$, while the "DTE" motif chelates $Mg^{2+}{}_B$.

Interestingly, metal binding motifs are shared between sesquiterpene cyclases and monoterpene cyclases from plants. The monoterpene cyclase (+)-bornyl diphosphate synthase catalyzes the cyclization of geranyl diphosphate (GPP) to form (+)-bornyl diphosphate. This cyclization is unusual in that the substrate diphosphate group is reincorporated into the product. The structure of (+)-bornyl diphosphate synthase from *Salvia officinalis* was the first of a monoterpene cyclase (Whittington, 2002), and remains the only monoterpene cyclase for which structures have been solved in unliganded and liganded states. The crystal structure of (+)-bornyl diphosphate synthase reveals the two-domain α-helical architecture first observed for the plant sesquiterpene synthase 5-*epi*-aristolochene synthase: a catalytically active C-terminal domain adopting the class I terpenoid synthase fold, and an N-terminal domain adopting the class II terpenoid synthase fold (however, the N-terminal polypeptide caps the active site of the C-terminal domain in ligand complexes) (Whittington, 2002). The (+)-bornyl diphosphate synthase-$Mg^{2+}{}_3$-PP_i complex reveals that conserved metal-binding motifs and the PP_i anion (or the diphosphate group of the product itself, (+)-bornyl diphosphate) coordinate to 3 Mg^{2+} ions (Figure 8.2 (e)). The first carboxylate of the $\mathbf{D}^{\mathbf{351}}DXX\mathbf{D}^{\mathbf{355}}$ motif coordinates to $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,syn*-bidentate geometry, and D355 bridges $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ with *syn,anti*-coordination stereochemistry. Interestingly, unlike metal binding in the active sites of bacterial and fungal cyclases, both the first *and* third aspartates in the DDXXD motif of plant terpenoid cyclases coordinate to the catalytic metal ions. The second metal binding motif, **D**$^{\mathbf{496}}$DKG**T**$^{\mathbf{500}}$SYF**E**$^{\mathbf{504}}$, chelates $Mg^{2+}{}_B$ (Whittington, 2002). In addition to metal ion coordination interactions, the PP_i anion accepts hydrogen bonds from R314, R493, and K512. Comparison of the structures of unliganded and $Mg^{2+}{}_3$-PP_i complexed (+)-bornyl diphosphate synthase reveals several $Mg^{2+}{}_3$-PP_i induced conformational changes; however, the r.m.s. deviation of 306 Cα atoms in the catalytic C-terminal domain is only 0.6 Å (Whittington, 2002), significantly lower than observed for ligand-induced conformational changes in trichodiene synthase (1.4 Å) (Rynkiewicz, 2001) and aristolochene synthase (1.8 Å) (Shishova, 2007).

The recent structure determination of another plant monoterpene cyclase, limonene synthase from *Mentha spicata* (Hyatt, 2007), similarly reveals conservation of a trinuclear metal cluster in a cyclization reaction that generates 94% (–)-(4*S*)-limonene, and ~2% myrcene, (-)-α-pinene, and (-)-β-pinene (Williams, 1998). Limonene synthase shares the 2-domain α-helical fold common to plant terpenoid cyclases. Limonene synthase displays similar activity with Mg^{2+} or Mn^{2+} (a common feature of some terpenoid cyclases), and the structure of the enzyme has been determined in complex with 3 Mn^{2+} ions and the intermediate analogue 2-fluorolinalyl diphosphate (FLPP) (Figure 8.2 (f)) (Hyatt, 2007). Metal coordination interactions are similar to those observed in (+)-bornyl diphosphate synthase (Whittington, 2002). In limonene synthase, the first carboxylate of the **D**$^{\mathbf{352}}$DXX**D**$^{\mathbf{356}}$ motif coordinates to $Mn^{2+}{}_A$ and $Mn^{2+}{}_C$ with *syn,syn*-bidentate geometry, and one oxygen atom of D356 bridges $Mn^{2+}{}_A$ and $Mn^{2+}{}_C$ with *syn,anti*-coordination stereochemistry. Two out of three residues in the second metal binding motif, **D**$^{\mathbf{496}}$DLG**T**$^{\mathbf{500}}$SVEE^{504}, chelate $Mg^{2+}{}_B$; the position of the side chain of

E504 is not indicated and is presumably disordered. Additionally, the γ-hydroxyl of T500 is 3.2 Å away from Mg^{2+}_B, which is too long to be considered an inner sphere coordination interaction. The diphosphate group of the bound intermediate analogue FLPP accepts hydrogen bonds from R315, R493, and K512 (Hyatt, 2007).

Finally, the sesquiterpene cyclase (+)-δ-cadinene synthase from *Gossypium arboreum* (tree cotton) catalyzes the first committed step in the biosynthesis of the triterpene phytoalexin gossypol, a major defense metabolite synthesized by cotton plants (Chen, 1995). The recently determined structure of the unliganded enzyme and its complex with 2-fluorofarnesyl diphosphate (2F-FPP) reveals that minimal structural deviations result from ligand binding (the r.m.s. deviations are 0.28 Å and 0.50 Å between unliganded and liganded monomers A (514 Cα atoms) and B (494 Cα atoms), respectively) (Gennadios, 2009). In contrast with the plant terpenoid cyclases previously discussed (Starks, 1997; Whittington, 2002; Hyatt, 2007), (+)-δ-cadinene synthase contains a second aspartate-rich motif in place of the DTE motif on helix H. As previously discussed, this motif on helix H is common to chain elongation enzymes such as farnesyl diphosphate synthase, and (+)-δ-cadinene synthase is unique among known class I terpenoid cyclases in that it contains two aspartate-rich motifs for metal coordination. The structure of the liganded enzyme reveals a putative Mg^{2+}_3 cluster (weak electron density characterizes the three Mg^{2+} ions); Mg^{2+}_A and Mg^{2+}_C are coordinated by D307 and D311 of the first $\mathbf{D^{307}}$DXX$\mathbf{D^{311}}$ motif, and Mg^{2+}_B is coordinated by D451 and E455 of the second aspartate-rich motif, $\mathbf{D^{451}}$DVA$\mathbf{E^{455}}$ (Figure 8.3). However, many of the carboxylate-Mg^{2+} distances observed are too long for inner sphere metal coordination interactions; therefore, the structure may reflect an incomplete transition between the “open” and “closed” active site conformations. The diphosphate moiety of 2F-FPP accepts one hydrogen bond from a nearby basic residue, R448.

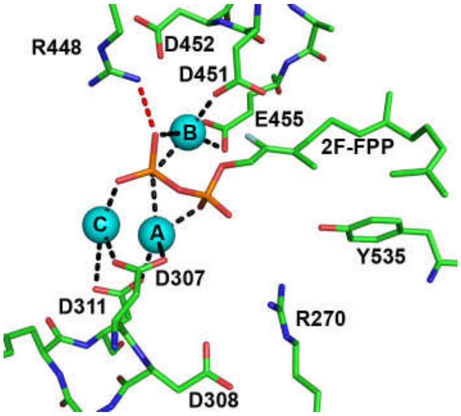

R448
D452
D451
B
E455
2F-FPP
C
A
Y535
D307
D311
R270
D308

Figure 8.3. The diphosphate binding site of (+)-δ-cadinene synthase from *G. arboreum* (PDB code 3G4F) with a putative $Mg^{2+}{}_3$ cluster and 2F-FPP bound. Metal ions are labeled according to convention established for trichodiene synthase. Some metal-phosphate interactions are too long to be considered inner-sphere metal coordination interactions, which could be a consequence of the nonproductive binding mode observed for 2F-FPP.

8.4 Discussion

Although the metal-dependence of catalysis by class I terpenoid synthases has been known for decades (Robinson, 1970), it was not until 2001 that the crystal structure of a terpenoid cyclase-Mg^{2+}_3-PP_i complex (trichodiene synthase) revealed that a trinuclear metal cluster accommodates PP_i binding; this trinuclear metal cluster is similarly implicated in binding and activating substrate farnesyl diphosphate for catalysis (Rynkiewicz, 2001). Since then, many X-ray crystal structures of isoprenoid coupling enzymes and terpenoid cyclases have been determined containing Mg^{2+}_3 (or Mn^{2+}_3) clusters. Comparisons of these structures reveal significant conservation in the constellation of metal ions and the residues that coordinate to these metal ions (Figures 8.1 and 8.2) despite generally insignificant amino acid sequence identity among these enzymes.

Trinuclear metal cluster coordination in FPP synthases is conserved among humans, bacteria and protozoans. Two aspartate-rich DDXXD binding motifs coordinate to 3 Mg^{2+} ions, which are also coordinated by the substrate diphosphate group. The first and last aspartate in the first DDXXD motif coordinate to Mg^{2+}_A and Mg^{2+}_C, and the first aspartate of the second DDXXD motif coordinates to Mg^{2+}_B. Also conserved are one arginine and two lysine residues that donate hydrogen bonds to diphosphate oxygens; the conserved arginine residue also donates hydrogen bond(s) to the second aspartate in the first DDXXD motif (Figure 8.1 (a)-(d)). The crystal structures of other isoprenoid coupling enzymes, GGPP synthase and nonspecific prenyl synthase, similarly reveal conservation of Mg^{2+}_3 binding motifs. Hydrogen bond interactions with PP_i are also conserved (Figure 8.1 (e)-(f)).

It is notable that the constellation of three metal ions and hydrogen bond donors is also conserved, with minor variations, in terpenoid cyclases from plants, bacteria, and fungi (Figure 8.2). First, $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ are coordinated by the first DDXXD motif: bacterial and fungal cyclases utilize only the first aspartate of this motif, whereas plant cyclases utilize the first and third aspartates of this motif (analogous to isoprenoid coupling enzymes). Second, the second aspartate-rich motif is usually replaced by an **N**DXX**S**XXX**E** motif in bacterial and fungal terpenoid cyclases and a **D**XXX**T**XXX**E** motif in plant terpenoid cyclases, in which boldface residues chelate $Mg^{2+}{}_B$ (although there can be some variations in this sequence, e.g., see (Zhou, 2009)). One exception, however, is (+)-δ-cadinene synthase, in which two aspartate-rich motifs coordinate to the trinuclear metal cluster. Third, residues that donate hydrogen bonds to PP_i are conserved in terpenoid cyclases across different domains of life. Specifically, two arginines donate hydrogen bonds to diphosphate oxygens: one appears to replace a conserved lysine serving this function in the isoprenoid coupling enzymes, and the other also donates a hydrogen bond to the second aspartate of the first DDXXD motif (as observed in the isoprenoid coupling enzymes). In bacterial and fungal terpenoid cyclases, conserved lysine and tyrosine residues additionally donate hydrogen bonds to PP_i.

In all cases in which a complete $Mg^{2+}{}_3$-PP_i cluster is bound, two 6-membered ring chelates are formed with $Mg^{2+}{}_A$ and $Mg^{2+}{}_B$ (Figure 8.4). The conformations of these 6-membered rings can vary, e.g., sofa, half-chair, etc. Such 6-membered ring chelates are occasionally observed in metal-diphosphate binding interactions, e.g., in the binding of the substrate analogue imidodiphosphate to inorganic pyrophosphatase (Fabrichniy, 2007).

In summary, conservation of a trinuclear metal cluster is critical for catalysis by class I terpenoid synthases. This cluster not only serves to bind and orient the flexible isoprenoid substrate in the precatalytic Michaelis complex, but it also triggers leaving group departure and initial carbocation formation. Conserved hydrogen bond donors in the terpenoid synthase active site assist the metal cluster in this function. That the trinuclear metal cluster is conserved for catalysis by terpenoid synthases from many domains of life suggests a common ancestry for this family of enzymes in the evolution of terpenoid biosynthesis.

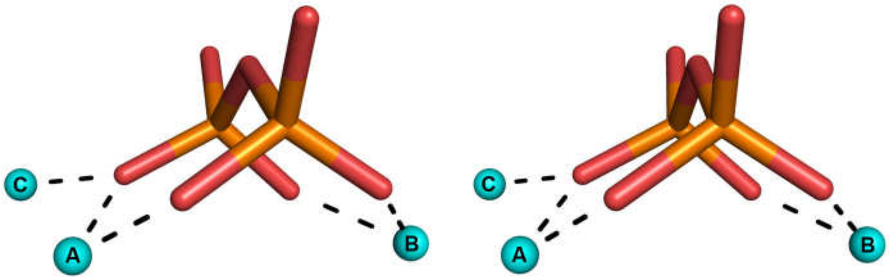


Figure 8.4. Stereoview of the $Mg^{2+}{}_3$-PP_i cluster from *epi*-isozizaene synthase. Dashed lines (black) represent metal-coordination interactions. The PP_i anion forms 6-membered ring chelates with $Mg^{2+}{}_A$ and $Mg^{2+}{}_B$, both of which adopt distorted sofa conformations.

Chapter 9: Future Directions

Terpenoid cyclases initiate and chaperone cyclization reactions to generate a multitude of structurally complex terpenoid products with precise regio- and stereo-specificity. The striking diversity of the terpenome is a direct result of the plasticity of the terpenoid synthases (Segura, 2003). It has been shown that the active site of a terpenoid synthase is predominantly lined with inert amino acids, which play a minimal role in the chemistry of catalysis beyond serving as a template and chaperone for the reaction (Christianson, 2008). The plasticity of the terpenoid cyclase active site has been studied in many systems (Greenhagen, 2006; Yoshikuni, 2006; Aaron, 2010).

A study of the sesquiterpene cyclase γ-humulene synthase provides an excellent example of the engineering potential of terpenoid cyclases. *Abies grandis* γ-humulene synthase, a promiscuous sesquiterpene cyclase, produces a mixture of 52 different terpenoid products. In the absence of a crystal structure of γ-humulene synthase, a homology model based on the known 5-*epi*-arisolochene synthase structure was used to identify “plasticity” residues in the γ-humulene synthase active site. The altered product profiles of a library of single-site mutants of the “plasticity” residues were determined and used to develop an algorithm to rationally design mutants using a combination of single-site mutations, based on the hypothesis that each plasticity residue is independent, meaning that the effect of a single mutation on the reaction mechanism is the same for the WT or any mutant form of the enzyme. Using their rational design algorithm, two to five mutations were combined to create mutant enzymes with up to 13 times greater relative yields of the preferentially desired sesquiterpene product (Yoshikuni, 2006).

In this work, *epi*-isozizaene synthase (EIZS) has been identified as an excellent model system for studying the structure-function relationships of sesquiterpene cyclases. EIZS is a stable, monomeric enzyme that readily forms crystals which diffract to ~1.6 Å resolution, and accommodates single amino acids mutations to active site residues, facilitating its potential use as a template for the rational design of novel terpenoid cyclases. Proposed experiments to continue this work include completing a GC-MS analysis of the products of the aliphatic active site mutants (L72V, A236G, and V329A) discussed in Chapter 7, to determine whether these small modifications to the contour of the active site result in perturbed product ratios. In addition, determining the kinetic activity, product-arrays by GC-MS, and crystal structures of several additional EIZS active site mutants, namely F95A, F332A, H333A, and W325F, would provide a thorough understanding of which residues directly affect the chaperoned cyclization cascade. Furthermore, crystal structures of the active site mutants provide an accurate picture of the enzymatic template, which can be used for modeling and quantum chemical calculations. To test the hypothesis that the product of a terpenoid cyclase can be predicted by how well the contour of the active site complements the shape of the product, modeling software, such as AutoDock (Morris, 2009), will be tested to determine a matching score for the respective enzymatic products of each EIZS mutant. These scores will be used to predict and test EIZS double and triple mutants, and to facilitate engineering new terpenoid cyclization templates.

It is important to remember that protein crystal structures provide a static picture of a dynamic system. Therefore, the orientation of the residues that form the active site cyclization template may occupy alternative conformations when the enzyme binds FPP in the closed conformation, with respect to the observed positions of the side chains in the $Mg^{2+}{}_3$-PP_i-BTAC complexes determined. Additional proposed crystallography experiments include determining crystal structures of WT and mutant EIZS with substrate, or intermediate, analogues in order to observe the position of a partially folded substrate in the active site. A crystal structure containing a partially folded intermediate would offer insight into the role of the active site aromatic residues in stabilizing the cationic intermediates via cation-π interactions.

The ultimate goal of the structure-function studies of the terpenoid cyclases is to increase our understanding of these enzymes to the point where it is possible to systematically alter the function of a terpene cyclase using a rational design strategy. The potential terpenoid rational design has also recently led to the launch of Allylix, a start-up company aiming to exploit the versatility and plasticity of these enzymes to cost effectively produces useful commercial quantities of useful and novel terpenoids (Allylix, 2010). Exploiting the specificity and efficiency of these enzymes may have profound effects on the large-scale production of terpenoid products useful in the food, cosmetics and pharmaceutical industries.

References, Part II

Aaron, J. A., Lin, X., Cane, D. E. and Christianson, D. W. (2010) Structure of *epi*-isozizaene synthase from *Streptomyces coelicolor* A3(2), a platform for new terpenoid cyclization templates. *Biochemistry 49*, 1787-1797.

Abe, I., Rohmer, M. and Prestwich, G. D. (1993) Enzymatic cyclization of squalene and oxidosqualene to triterpenes. *Chem. Rev. 93*, 2189-2206.

Abrahams, J. P. and Leslie, A. G. W. (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 52*, 30-42.

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1948-1954.

Agrawal, A. G. and Somani, R. R. (2009) Farnesyltransferase inhibitor as anticancer agent. *Mini Rev. Med. Chem. 9*, 638 - 652.

Aharoni, A., Jongsma, M. A. and Bouwmeester, H. J. (2005) Volatile science? Metabolic engineering of terpenoids in plants. *Trends Plant Sci. 10*, 594-602.

Allylix. (2010). "http://www.allylix.com."

Artz, J. D., Dunford, J. E., Arrowood, M. J., Dong, A., Chruszcz, M., Kavanagh, K. L., Minor, W., Russell, R. G. G., Ebetino, F. H., Oppermann, U. and Hui, R. (2008) Targeting a uniquely nonspecific prenyl synthase with bisphosphonates to combat cryptosporidiosis. *Chem. Biol. 15*, 1296-1306.

Ashby, M. N. and Edwards, P. A. (1990) Elucidation of the deficiency in two yeast coenzyme Q mutants. Characterization of the structural gene encoding hexaprenyl pyrophosphate synthetase. *J. Biol. Chem. 265*, 13157-13164.

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

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J. and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature 417*, 141-147.

Bernhardt, R. (2006) Cytochromes P450 as versatile biocatalysts. *J. Biotechnol. 124*, 128-145.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. and Paciorek, W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 59*, 2023-2030.

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

Cane, D. E. (1990) Enzymatic formation of sesquiterpenes. *Chem. Rev. 90*, 1089-1103.

Cane, D. E., Shim, J. H., Xue, Q., Fitzsimons, B. C. and Hohn, T. M. (1995) Trichodiene Synthase. Identification of Active Site Residues by Site-Directed Mutagenesis. *Biochemistry 34*, 2480-2488.

Cane, D. E., Sohng, J. K., Lamberson, C. R., Rudnicki, S. M., Wu, Z., Lloyd, M. D., Oliver, J. S. and Hubbard, B. R. (1994) Pentalenene synthase - Purification, molecular-cloning, sequencing, and high-level expression in *Escherichia coli* of a terpenoid cyclase from *Streptomyces* UC5319. *Biochemistry 33*, 5846-5857.

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

Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2000) Crystal structure determination of aristolochene synthase from the blue cheese mold, Penicillium roqueforti. *J. Biol. Chem. 275*, 25533-25539.

Challis, G. L. and Hopwood, D. A. (2003) Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc. Natl. Acad. Sci. U. S. A. 100*, 14555-14561.

Chang, T.-H., Guo, R.-T., Ko, T.-P., Wang, A. H.-J. and Liang, P.-H. (2006) Crystal structure of type-III geranylgeranyl pyrophosphate synthase from saccharomyces cerevisiae and the mechanism of product chain length determination. *J. Biol. Chem. 281*, 14991-15000.

Chen, C. K.-M., Hudock, M. P., Zhang, Y., Guo, R.-T., Cao, R., No, J. H., Liang, P.-H., Ko, T.-P., Chang, T.-H., Chang, S.-C., Song, Y., Axelson, J., Kumar, A., Wang, A. H. and Oldfield, E. (2008) Inhibition of geranylgeranyl diphosphate synthase by bisphosphonates: A crystallographic and computational investigation. *J. Med. Chem. 51*, 5594-5607.

Chen, X.-Y., Chen, Y., Heinstein, P. and Davisson, V. J. (1995) Cloning, expression, and characterization of (+)-δ-cadinene synthase: A catalyst for cotton phytoalexin biosynthesis. *Arch. Biochem. Biophys. 324*, 255-266.

Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev. 106*, 3412-3442.

Christianson, D. W. (2007) Roots of biosynthetic diversity. *Science 316,* 60-61.

Christianson, D. W. (2008) Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol. 12*, 141-150.

Coates, R. M. and Sowerby, R. L. (1972) Stereoselective total synthesis of (+/-)-zizaene. *J. Am. Chem. Soc. 94*, 5386-5396.

Croteau, R. and Cane, D. E. (1985). Monoterpene and sesquiterpene cyclases. In Methods in Enzymology, J. H. Law and H. C. Rilling, eds. (Academic Press), pp. 383-405.

Daum, M., Herrmann, S., Wilkinson, B. and Bechthold, A. (2009) Genes and enzymes involved in bacterial isoprenoid biosynthesis. *Curr. Opin. Chem. Biol. 13*, 180-188.

Ebetino, F. H., Roze, C. N., McKenna, C. E., Barnett, B. L., Dunford, J. E., Russell, R. G. G., Mieling, G. E. and Rogers, M. J. (2005) Molecular interactions of nitrogen-containing bisphosphonates within farnesyl diphosphate synthase. *J. Organomet. Chem. 690*, 2679-2687.

Emsely, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126-2132.

Evans, P. (2003) Phasing the AP2 core complex with Xe, Hg and Se. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 59*, 2039-2043.

Fabrichniy, I. P., Lehtio, L., Tammenkoski, M., Zyryanov, A. B., Oksanen, E., Baykov, A. A., Lahti, R. and Goldman, A. (2007) A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J. Biol. Chem. 282*, 1422-1431.

Felicetti, B. and Cane, D. E. (2004) Aristolochene synthase: mechanistic analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc. 126*, 7212-7221.

Ferella, M., Li, Z. H., Andersson, B. and Docampo, R. (2008) Farnesyl diphosphate synthase localizes to the cytoplasm of Trypanosoma cruzi and T. bruci. *Exp. Parasitol. 119*, 308-312.

Frisch, M. J. (2003). et al. Gaussian03. Pittsburg, PA, Gaussian Inc.

Gabelli, S. B., McLellan, J. S., Montalvetti, A., Oldfield, E., Docampo, R. and Amzel, L. M. (2006) Structure and mechanism of the farnesyl diphosphate synthase from *Trypanosoma cruzi*: Implications for drug design. *Proteins: Struct. Funct. Bioinf. 62*, 80-88.

Garzoni, L. R., Caldera, A., Meirelles, M. d. N., de Castro, S. L., Docampo, R., Meints, G. A., Oldfield, E. and Urbina, J. A. (2004) Selective in vitro effects of the farnesyl pyrophosphate synthase inhibitor risedronate on Trypanosoma cruzi. *Int. J. Antimicrob. Agents. 23*, 273-285.

Gennadios, H. A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., Allemann, R. K. and Christianson, D. W. (2009) Crystal structure of (+)-δ-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry 48*, 6175-6183.

Greenhagen, B. T., O'Maille, P. E., Noel, J. P. and Chappell, J. (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc. Natl. Acad. Sci. U. S. A. 103*, 9826-9831.

Harangi, J. (2003) Retention index calculation without n-alkanes--the virtual carbon number. *J. Chromatogr. A. 993*, 187-195.

Hashsham, S. A., Wick, L. M., Rouillard, J.-M., Gulari, E. and Tiedje, J. M. (2004) Potential of DNA microarrays for developing parallel detection tools (PDTs) for microorganisms relevant to biodefense and related research needs. *Biosens. Bioelectron. 20*, 668-683.

Hendrickson, W. A., Horton, J. R. and Lemaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD) - A vehicle for direct determination of 3-dimensional structure. *EMBO J. 9*, 1665-1672.

Hong, Y. J. and Tantillo, D. J. (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.

Hosfield, D. J., Zhang, Y., Dougan, D. R., Broun, A., Tari, L. W., Swanson, R. V. and Finn, J. (2004) Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J. Biol. Chem. 279*, 8526-8529.

Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B. 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.

Jenson, C. and Jorgensen, W. L. (1997) Computational investigations of carbenium ion reactions relevant to sterol biosynthesis. *J. Am. Chem. Soc. 119*, 10846-10854.

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr., Sect. A: Found. Crystallogr. 47*, 110-119.

Kavanagh, K. L., Dunford, J. E., Bunkoczi, G., Russell, R. G. G. and Oppermann, U. (2006) The crystal structure of human geranylgeranyl pyrophosphate synthase reveals a novel hexameric arrangement and inhibitory product binding. *J. Biol. Chem. 281*, 22004-22012.

Kleywegt, G. J. and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr 50*, 178-185.

Kloer, D. P., Welsch, R., Beyer, P. and Schulz, G. E. (2006) Structure and reaction geometry of geranylgeranyl diphosphate synthase from *Sinapis alba*. *Biochemistry 45*, 15197-15204.

Kudo, F. and Eguchi, T. (2009). Biosynthetic enzymes for the aminoglycosides butirosin and neomycin In Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates (San Diego, Elsevier Academic Press Inc), pp. 493-519.

Langer, G., Cohen, S. X., Lamzin, V. S. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc. 3*, 1171-1179.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK - A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283-291.

Lesburg, C. A., Caruthers, J. M., Paschall, C. M. and Christianson, D. W. (1998) Managing and manipulating carbocations in biology: Terpenoid cyclase structure and mechanism. *Curr. Opin. Struct. Biol. 8*, 695-703.

Lesburg, C. A., Zhai, G., Cane, D. E. and Christianson, D. W. (1997) Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology. *Science 277*, 1820-1824.

Licata, A. A. (2005) Discovery, clinical development, and therapeutic uses of bisphosphonates. *Annals of Pharmacotherapy 39*, 668-677.

Lin, X. and Cane, D. E. (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.

Lin, X., Hopson, R. and Cane, D. E. (2006) Genome mining in *Streptomyces coelicolor*: Molecular cloning and characterization of a new sesquiterpene synthase. *J. Am. Chem. Soc. 128*, 6022-6023.

Ma, J. C. and Dougherty, D. A. (1997) The cation-π interaction. *Chem. Rev. 97*, 1303-1324.

Mao, J., Mukherjee, S., Zhang, Y., Cao, R., Sanders, J. M., Song, Y., Zhang, Y., Meints, G. A., Gao, Y. G., Mukkamala, D., Hudock, M. P. and Oldfield, E. (2006) Solid-State NMR, crystallographic, and computational investigation of bisphosphonates and farnesyl diphosphate synthase-bisphosphonate complexes. *J. Am. Chem. Soc. 128*, 14485-14497.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem. 30*, 2785-2791.

Nishio, K., Nodake, Y., Hamada, K., Suto, K., Nakagawa, N., Kuramitsu, S. and Miura, K. (2004) Expression, purification, crystallization and preliminary X-ray studies of geranylgeranyl diphosphate synthase from Thermus thermophilus HB8. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 178-180.

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

Otwinowski, Z. and Minor, M. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. Part A. 276*, 307-326.

Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D. and Kozarich, J. W. (1993) On the origin of enzymatic species. *Trends Biochem. Sci. 18*, 372-376.

Pichersky, E., Noel, J. P. and Dudareva, N. (2006) Biosynthesis of plant volatiles: Nature's diversity and ingenuity. *Science 311*, 808-811.

Pickens, L. B. and Tang, Y. (2009) Decoding and engineering tetracycline biosynthesis. *Metab. Eng. 11*, 69-75.

Rhodes, G. (2000). Crystallography made crystal clear (New York, Elsevier).

Robinson, D. R. and West, C. A. (1970) Biosynthesis of cyclic diterpenes in extracts from seedlings of Ricinus communis L. II. Conversion of geranylgeranyl pyrophosphate into diterpene hydrocarbons and partial purification of the cyclization enzymes. *Biochemistry 9*, 70-79.

Rondeau, J.-M., Bitsch, F., Bourgier, E., Geiser, M., Hemmig, R., Kroemer, M., Lehmann, S., Ramage, P., Rieffel, S., Strauss, A., Green, J. R. and Jahnke, W. (2006) Structural basis for the exceptional in vivo efficacy of bisphosphonate drugs. *ChemMedChem 1*, 267-273.

Russell, R. G. G. (2006) Bisphosphonates. *Ann. N.Y. Acad. Sci. 1068*, 367-401.

Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2001) Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. U. S. A. 98*, 13543-13548.

Rynkiewicz, M. J., Cane, D. E. and Christianson, D. W. (2002) X-ray crystal structures of D100E trichodiene synthase and its pyrophosphate complex reveal the basis for terpene product diversity. *Biochemistry 41*, 1732-1741.

Seemann, M., Zhai, G., de Kraker, J.-W., Paschall, C. M., Christianson, D. W. and Cane, D. E. (2002) Pentalenene Synthase. Analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc. 124*, 7681-7689.

Segura, M. J. R., Jackson, B. E. and Matsuda, S. P. T. (2003) Mutagenesis approaches to deduce structure-function relationships in terpene synthases. *Nat. Prod. Rep. 20*, Seto, H. and Yonehara, H. (1980) Studies on the biosynthesis of pentalenolactone. III. Isolation of a biosynthetic intermediate hydrocarbon, pentalenene. *J. Antibiot. (Tokyo) 33*, 92-93.

Shishova, E. Y., Di Costanzo, L., Cane, D. E. and Christianson, D. W. (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.

Shishova, E. Y., Yu, F., Miller, D. J., Faraldos, J. A., Zhao, Y., Coates, R. M., Allemann, R. K., Cane, D. E. and Christianson, D. W. (2008) X-ray crystallographic studies of substrate binding to Aristolochene Synthase suggest a metal ion binding sequence for catalysis. *J. Biol. Chem. 283*, 15431-15439.

Skubak, P., Murshudov, G. N. and Pannu, N. S. (2004) Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr 60*, 2196-2201.

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

Storoni, L. C., McCoy, A. J. and Read, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr., Sect. D: Biol. Crystallogr 60*, 432-438.

Tanowitz, H. B., Machado, F. S., Jelicks, L. A., Shirani, J., de Carvalho, A. C., Spray, D. C., Factor, S. M., Kirchhoff, L. V. and Weiss, L. M. (2009) Perspectives on Trypanosoma cruzi-Induced Heart Disease (Chagas Disease). *Prog. Cardiovasc. Dis. 51*, 524-539.

Tarshis, L. C., Yan, M., Poulter, C. D. and Sacchettini, J. C. (1994) Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-Å resolution. *Biochemistry 33*, 10871-10877.

Taylor, G. (2003) The phase problem. *Acta Crystallogr., Sect. D: Biol. Crystallogr 59*, 1881-1890.

Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol. 9*, 297-304.

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

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

Vedula, L. S., Rynkiewicz, M. J., Pyun, H. J., Coates, R. M., Cane, D. E. and Christianson, D. W. (2005b) Molecular recognition of the substrate diphosphate group governs product diversity in trichodiene synthase mutants. *Biochemistry 44*, 6153-6163.

Vedula, L. S., Zhao, Y., Coates, R. M., Koyama, T., Cane, D. E. and Christianson, D. W. (2007) Exploring biosynthetic diversity with trichodiene synthase. *Arch Biochem. Biophys. 466*, 260-266.

Vonrhein, C., Blanc, E., Roversi, P. and Bricogne, G. (2007) Automated structure solutoin with autoSHARP. *Methods Mol. Biol. 364*, 215-230.

Wendt, K. U. and Schulz, G. E. (1998) Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure 6*, 127-133.

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

Williams, D. C., McGarvey, D. J., Katahira, E. J. and Croteau, R. (1998) 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 37*, 12213-12220.

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

Zhang, Y., Cao, R., Yin, F., Hudock, M. P., Guo, R.-T., Krysiak, K., Mukherjee, S., Gao, Y.-G., Robinson, H., Song, Y., No, J. H., Bergan, K., Leon, A., Cass, L., Goddard, A., Chang, T.-K., Lin, F.-Y., Van Beek, E., Papapoulos, S., Wang, A. H., Kubo, T., Ochi, M., Mukkamala, D. and Oldfield, E. (2009) Lipophilic bisphosphonates as dual farnesyl/geranylgeranyl diphosphate synthase inhibitors: An X-ray and NMR investigation. *J. Am. Chem. Soc. 131*, 5153-5162.

Zhao, B., Lin, X., Lei, L., Lamb, D. C., Kelly, S. L., Waterman, M. R. and Cane, D. E. (2008) Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *J. Biol. Chem. 283*, 8183-8189.

Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L. and Minor, W. (2008) Data mining of metal ion environments present in protein structures. *J. Inorg. Biochem. 102*, 1765-1776.

Zhou, K. and Peters, R. J. (2009) Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. *Phytochemistry 70*, 366-369.