

Exploring Novel Enzyme Activity and Genetic Diversity Through Biocatalytic Engineering and Microbiome Mapping

by

Rosa M. C. Vásquez Espinoza

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2021

Doctoral Committee:

Professor John Montgomery, Co-Chair
Professor David H. Sherman, Co-Chair
Professor Roland Kersten
Professor Pavel Nagorny
Professor Alison Narayan
Professor Emily Scott

Rosa M. C. Vásquez Espinoza

rosave@umich.edu

ORCID ID: 0000-0003-0007-7959

© Rosa M. C. Vásquez Espinoza 2021

Dedication

This work is dedicated to my loving grandmother, Eugenia Matos Valerio, who created a natural pharmacy in our backyard and dreamed of becoming a doctor.

Acknowledgements

My time and experiences in graduate school have been essential to my development as a scientist and educator and I could not envision being where I am today without my family, friends and the stellar scientists I have been fortunate to work with and learn from. Words are not enough to express my gratitude to everyone who has contributed to my scientific and professional journey, whether that meant providing kind words and advice when most needed, taking the time to mentor me, celebrating any small or big success, providing a safe space to share and grow, or sharing laughs and joy in the day-to-day.

Looking back to when my journey in science began, I remember multiple instances where my curiosity was sparked or where I was encouraged to explore my creativity. When I was little, I remember my grandmother patiently showing me how she grows and uses medicinal plants such as *muña* for stomachaches or my dad telling me stories on genetic engineering of potatoes and grapes for agricultural development. They both dreamed of being scientists and their joy when talking about nature or scientific discoveries was contagious. I also remember my mom encouraging me to pursue art in any form even if that meant coming back to a house with nonsense doodles in the walls, or buying me multidisciplinary books that described space travel from a poetic and artistic perspective since I was obsessed with the idea of being an astronaut for many years.

In high school, our professors took us to a one-week trip in Madre de Dios located in the Peruvian Amazon, where we completed field experiments on biology and geography. I remember professor Dr. David Bruggers cutting through the dense jungle with a machete and jumping into the river in the middle of the night to catch a *caimán* so we could appreciate it from a closer view before safely releasing it back to the water. This was the first time I realized one could combine field work and adventure with science education and exploration.

Inspired by all these experiences, I moved to the US from Peru when I was 18 years old to pursue my undergraduate studies in Biochemistry and Molecular Biology at Tennessee Technological University, where I worked under the mentorship of Dr. David Beck for four years.

Dr. Beck not only took me under his wings and taught me how to do science in a careful and methodical manner, but he also supported my interest in completing a summer internship in a Traditional Chinese Medicine hospital in Beijing, China, which ultimately helped solidify my scientific interests leading me to the University of Michigan (U-M) and the world of natural products and biocatalysis.

I feel beyond fortunate and grateful to have Dr. David Sherman and Dr. John Montgomery as my co-advisors in graduate school. Not only have they guided me and supported me to become the scientist I am today, but they have also encouraged me to ask interesting questions, pursue independent research ideas and take on invaluable opportunities that helped advance my career. I recall meeting David prior joining U-M to talk about science, and learning about his long-standing collaboration with Peruvian institutions to advance scientific research in our country. From the start, he supported my long-term goals of doing research work in the Peruvian Amazon, which eventually led us to apply for a National Geographic grant to study the microbiome of an unexplored area of the jungle. His trust and commitment enabled us to become the first laboratory in the United States to have governmental permits for the collection and study of Peruvian microbes. I also have to thank Dr. Roger Cone and J.D. Anna Schork, for their continued and invaluable support in starting and completing the permit process, a journey that would have not been possible without them.

I will always remember John giving me an opportunity to join his lab despite my limited experience in synthetic chemistry research. His commitment to education and training is unwavering and continues to inspire me to become a better mentor myself. He recognized my interest and hard work in the collaborative project with the Sherman group, and assigned me great mentors in his lab to develop my skills in synthesis and methods development. I will always cherish the summer and Christmas parties at his house too, especially the one where I first learned what Pink Elephant is.

I am also very grateful for the wonderful collaborators I've had the pleasure to work with throughout my graduate studies including Dr. Ken Houk from the University of California Los Angeles, whose exemplary work in computational methods to study enzymes served as a critical foundation for many of the biochemical and kinetic experiments I completed. I also deeply thank Dr. Gretty Villena Chávez from Universidad Nacional Agraria La Molina in Lima Peru, whose expertise in fungal systems and support in initiating the Peruvian microbiome project was essential

for the success of our work, and Andrés Ruza from the Boiling River Project, whose friendship, mentorship and expertise have contributed to my scientific development and professional and personal growth. I'm also deeply grateful to the generous LSI donors whose support was essential in helping the Peru project move forward. Without any of them, the Boiling River microbiome work would have not been possible. Moreover, I am very thankful to my committee members, Dr. Alison Narayan, Dr. Emily Scott, Dr. Pavel Nagorny and Dr. Roland Kersten. Their insightful questions, thoughtful suggestions and guidance were instrumental in helping me see things from different perspectives and in my development as an independent scientist. I also would like to acknowledge Felipe Huanachín Carahuanco, Kersti Haatveit, Aneta Turlík and Mark Maskeri for thought-provoking discussions and for their invaluable contributions to the research work I present in this dissertation. Felipe is a friend I deeply admire and a talented Peruvian Ph.D. candidate that learned how to code to advance the bioinformatics section of our collaboration and take our work to the next level. I thank the undergraduate students that have worked with me throughout graduate school for their hard work, dedication, passion and patience as I fine-tuned my mentoring abilities. I especially thank Sari Grossman who completed an outstanding undergraduate thesis and became a dear friend over the late hours we stayed in lab talking about science and life, and Jeanie Tan whose passion for conservation and natural products inspired me. I'd also like to thank the Program in Chemical Biology and Laura Howe and Traci Swan who worked tirelessly to ensure we had a smooth graduate school process. I'm also forever happy and honored to be part of the National Geographic family and for all the undivided support they continue providing me with, and for giving me the confidence to embrace all parts of me and trust my voice. Special thanks to Gael Almeida, Kate Freshwater, Katia Andreassi, David Levy, Stacey McClain, Arslan Ahmad, Peg Keiner, and my educator collaborator Sharee Barton.

My talented friends and colleagues in the Sherman and Montgomery Labs also made this experience a joyful, fun, and enriching one. Not only were they always available to provide great scientific insight but were also willing to engage in interesting discussions that helped me explore new ideas and kept me inspired. I especially thank Dr. Sean Newmister for his great mentorship in all enzymes related work, Dr. Yogan Khatri for his invaluable teachings in all P450 research, Dr. Jennifer Schmidt for her important support in NMR experiments and mentorship in all things related to graduate school and lab, Dr. Andrew Robertson for his NMR expertise, Dr. Ashu Tripathi and Pam Schultz for their expert advice on growing difficult bacterial strains, Dr. Fengan

Yu for teaching me bioinformatic methods and *Streptomyces* engineering, Dr. Michael Gilbert for helping me set up my first organic reaction, Dr. Jessica Stachowski for her continuous help in synthesis and Dr. Amy Fraley, Dr. Matt DeMars and Dr. Vik Shende, my previous bay-mates, for their support throughout the first years of graduate school. I also thank Dr. Annabel Ansel, Dr. Wesley Pein, Dr. Alex Rand, Maria Adrover, Nicole Rivera, Maribel Okiye, Robert Hohlman, Sam Kelly, Dr. Alanna Condren, and all Sherman and Montgomery lab members who always brought a smile to my face and made my experience in graduate school an irreplaceable one.

I would be remiss if I didn't also thank my incredible outside-of-lab loved ones, friends and family who have been with me through thick and thin, never judging and always supporting me, and have brought happiness to my life especially during my years in graduate school and in Ann Arbor including my sisters Paige Vega and Alexandra Arriaga who are always there for a laugh or a cry and would take a flight to come see me whenever I needed them, Dr. Christopher Perry for being a constant source of light, joy and inspiration, Dr. Abhi Sharma, Dr. Natacha Bohin, Alyse Waldhorn, Renzo Milón, Gerardo Sztrancman, Andrés Ruzo, Felipe Huanachín Carahuanco and Kelly Chunga. I also thank Mathieu Trimbur, Erick Iciarte and Hoang Le for their support for various years especially at the beginning of graduate school and for encouraging me to look at schools in the north of the US despite of my love for warmer climates.

Lastly, I wouldn't be who I am today without my family. *Mi mami, papi, mama Rosa, abuelita Leonor, tía Julia, tío Carlos, tío Alejandro, Lauri, Kitty, Rodrigo, Alicito y tío Melín, ustedes siempre están conmigo a pesar de la distancia, siempre apoyándome en mis locuras aún así no sepan qué es lo que estoy haciendo. Ustedes nunca necesitaron entenderme, sólo han querido apoyarme y quererme y eso me ha dado las alas para llegar a donde estoy hoy. Gracias por hacerme creer que sí puedo, que no tengo límites aún así mi acento no sea perfecto o no sea de alguna religión o raza en particular. Soy el producto de su amor, apoyo, coraje, y sueños, y cada éxito es tanto suyo como mío.*

Table of Contents

Dedication	ii
Acknowledgments	iii
List of Figures	xii
List of Tables	xvii
List of Schemes	xix
List of Appendices	xxi
Abstract	xxiii
Chapter 1: Introduction and Background	1
1.1 Natural Products in Medicine	1
1.1.1 Medicinal Natural Products: History and Sources	1
1.1.2 The Present and Future of Natural Products in Drug Discovery	2
1.2 Biocatalysis and C-H Functionalization	1
1.2.1 General Biosynthesis of Secondary Metabolites	3
1.2.2 Chemical and Enzymatic C-H Functionalization	5
1.3 Cytochrome P450 Enzymes as Biocatalysts	7
1.3.1 History and Structure of P450s	7
1.3.2 P450 Mechanism	9
1.3.3 P450 Classification and Redox Partner Systems	10
1.3.4 P450s in Late-Stage Natural Product Diversification	12
1.3.4.1 Iterative Oxidation Cascades in Microbial P450s	12
1.3.4.2 Selective C-H Functionalization in Plant P450s	15
1.4 Thesis Outline	18
1.5 References	19
Chapter 2: Investigating and Engineering P450 TamI as an Iterative Biocatalyst for Late-Stage C-H Oxidation of Complex Natural Products	23
2.1 Introduction to the Tirandamycin Family of Tetramic Acid Antibiotics	23

2.1.1	Discovery of Tirandamycin Natural Products	23
2.1.2	Biological Applications and Mode of Action of Tirandamycin	23
2.1.3	Elucidation of the Tirandamycin Biosynthetic Pathway	24
2.2	Molecular and Structural Basis of Iterative C-H Oxidation by TamI	25
2.2.1	Background	25
2.2.2	Structure of P450 TamI in Complex with Tirandamycin C and Identification of Key Residues	26
2.2.3	Mutagenesis and Biochemical Characterization	26
2.2.3.1	Probing a Tetramic Acid-Based Anchoring Mechanism	27
2.2.3.2	Probing a Hydrophobic Pocket in the Active Site of TamI	27
2.2.4	Computational Investigations of Iterative Oxidations	29
2.2.4.1	Density Functional Theory on Intrinsic Reactivity	29
2.2.4.2	Molecular Dynamics with Tirandamycin Congeners	30
2.2.5	Discussion	31
2.3	Engineering P450 TamI for Selective C-H Functionalization and Epoxidation of Tirandamycin Antibiotics	32
2.3.1	Introduction and Background	32
2.3.2	Design and Engineering of a Toolbox of P450 TamI Biocatalysts	33
2.3.2.1	Identifying Key Active Site Residues for Mutagenesis	33
2.3.2.2	TamI Mutants with Divergent Selectivity from Wild-Type	34
2.3.2.3	TamI Mutants with Enhanced Iterative Capabilities	35
2.3.2.4	Catalytically Self-Sufficient TamI Mutant Biocatalysts	37
2.3.3	Elucidation of New Tirandamycin Congeners and Mechanisms	37
2.3.3.1	Optimization of Large-Scale Fermentation Conditions	37
2.3.3.2	TamI Mutants L101A_L295I and L295A Interrupt the Stepwise Oxidative Cascade Generating Tirandamycin L	38
2.3.3.3	Dual-Function TamI L295V Produces Tirandamycin M	40
2.3.3.4	TamI L101A_L295I Bypasses Oxidation at C10 Forming Tirandamycin N	41
2.3.3.5	Multifunctional TamI L295A Catalyzes an Unexpected Oxidative Cascade Generating Tirandamycin O and O'	42

2.3.3.6 Iterative TamI L244A_L295V Enables a Continuous 8 e ⁻ Oxidation	43
2.3.4 Biochemical and Kinetic Characterization of TamI Mutants	43
2.3.4.1 Substrate Scope, Total Turnover Capacity and Substrate Binding Properties of TamI Mutants	43
2.3.4.2 Michaelis-Menten Kinetic Parameters of TamI Variants	45
2.3.5 Antimicrobial Activities of New Tirandamycin	47
2.3.6 Discussion	48
2.4 References	52
Chapter 3: Multiple Activated Oxygen Species in P450 TamI Iterative Catalysis	55
3.1 Background	55
3.1.1 Involvement of Various Oxidizing Species in P450 Catalysis	55
3.1.2 Proposed Mechanisms in P450 TamI Oxidative Cascade Supporting Distinct Functional Oxidants	57
3.2 Mechanism of Epoxidation via Cpd 0 to Form Tirandamycin L (6)	60
3.2.1 Structure-Derived Mechanism	60
3.2.2 Catalytic Evidence and Mutagenesis Experiments	61
3.2.3 Use of Oxygen Surrogates in TamI-Catalyzed Reactions	63
3.3 Iterative Oxidation Cascade to Produce Tirandamycin O and O'	64
3.3.1 Evidence for a Reactive, Nucleophilic Peroxo-Iron Oxidant	64
3.3.2 Assessing Cpd 0 Reactivity	68
3.3.3 Evidence for a Cpd I-Mediated Mechanism	69
3.4 Discussion	70
3.5 References	73
Chapter 4: Substrate Engineering of P450 TamI to Expand Substrate Scope	76
4.1 Background and Introduction	76
4.1.1 Previous Syntheses of Tirandamycin Natural Products	76
4.1.2 Biological Relevance of Derivatizing Tirandamycin	77
4.1.3 Substrate Engineering Methods to Facilitate Derivatization	78
4.2 Design and Synthesis of Tirandamycin Analogues)	80
4.2.1 Previous Synthetic Strategies Towards a Tirandamycin Analogue	80

4.2.2 Asymmetric Allylation Optimization: A 5-Step Synthesis Towards a Tirandamycin Analogue	80
4.2.3 Synthesis and Coupling of Synthetic Anchors to a Tirandamycin Analogue	81
4.3 Substrate Engineering Efforts to Expand Substrate Scope	83
4.3.1 Enzymatic Reactions with Synthetic Tirandamycin Analogues and Computational Investigations	83
4.4 Discussion	87
4.5 References	89
Chapter 5: Microbiome Mapping of the Peruvian Boiling River	91
5.1 Background and Introduction	91
5.1.1 The Amazon Rainforest and its Macro- and Microdiversity	91
5.1.2 The Boiling River of the Peruvian Amazon	93
5.1.3 Ecological Importance of Studying the Boiling River	94
5.1.4 Amazonian Extremophiles as a Source of Natural Products and Thermostable Biocatalysts	95
5.2 Boiling River Sampling Expedition	98
5.2.1 Obtaining Research Permits from the Peruvian Government	98
5.2.2 Design and Execution of Sampling Strategy	99
5.3 Characterization of the Boiling River Microbiome	100
5.3.1 Isolation of Metagenomic DNA and Sequencing the 16S rRNA gene	100
5.3.2 Microbial Diversity Meta-Analysis	101
5.3.2.1 Taxonomic Composition	101
5.3.2.2 Clustering and Correlation Analyses	105
5.3.2.3 Samples with Unique Microbial Configurations	106
5.3.3 Biomarker Evidence for Bacteria Involved in Sulfur Cycling	108
5.4 Discussion	110
5.5 References	113
Chapter 6: Summary, Discussion and Future Directions	116
6.1 Structural Characterization and Engineering of P450 TamI	116
6.1.1 Summary and Conclusions	116

6.1.2 Future Directions	117
6.2 Mechanistic Insights into the Multiple Oxidant Species of P450 TamI	118
6.2.1 Summary and Conclusions	118
6.2.2 Future Directions	119
6.3 Chemical Synthesis of a Tirandamycin Analogue to Explore Substrate Scope	119
6.3.1 Summary and Conclusions	119
6.3.2 Future Directions	120
6.4 Characterizing the Microbial Diversity of the Boiling River	121
6.4.1 Summary and Conclusions	121
6.4.2 Future Directions	122
Appendices	123

List of Figures

Figure 1.1 Natural product drug discovery workflow	2
Figure 1.2 Biosynthesis of tirandamycin antibiotics	4
Figure 1.3 General P450 structure	8
Figure 1.4 P450-catalyzed systems	11
Figure 1.5 Engineering plant P450s	16
Figure 1.6 P450 involved in rare triterpenoid biosynthesis	18
Figure 2.1 Tirandamycin congeners reported to date	24
Figure 2.2 Structure of P450 TamI in complex with tirandamycin C (1)	26
Figure 2.3 Mutagenesis analysis with P450 TamI	28
Figure 2.4 Computational analysis of TamI iterative oxidation	30
Figure 2.5 Active site residues of TamI selected for an alanine-valine scan	34
Figure 2.6 End-point assays with TamI mutants	36
Figure 2.7 Computational analysis of TamI L101A_L295I	39
Figure 2.8 DFT calculations of TamI L101A_L295I regioselectivity	41
Figure 2.9 Computational analysis of iterative TamI L244A_L295V	44
Figure 2.10 Total turnover numbers (TTN) for TamI WT and mutants with tirandamycins	46
Figure 3.1 Molecular Dynamics (MD) simulations of TamI L295A with 1	60
Figure 3.2 End-point assays with TamI variants	62
Figure 3.3 TamI-catalyzed oxidations using H ₂ O ₂ as a cosubstrate	64
Figure 3.4 End-point assays with P450 TamI in H ₂ O vs D ₂ O buffers	66
Figure 3.5 DFT calculations on P450 TamI and 7' (<i>epi</i> -tirandamycin M)	68
Figure 3.6 Oxidizing species involved in TamI L295A catalysis	70
Figure 4.1 Bioactivity of tirandamycin increases as oxidation levels increase	78
Figure 4.2 Simplified representation of substrate-bound crystal structure of TamI	79
Figure 4.3 Design of synthetic anchors for substrate engineering efforts with TamI	82
Figure 4.4 Enzymatic reactions with synthetic tirandamycin analogues	84

Figure 4.5 MD simulations of TamI with a truncated tirandamycin	85
Figure 4.6 MD simulations with tirandamycin analogue 33	86
Figure 4.7 Molecular docking of TamI with substrate 36	86
Figure 5.1 Amazon Rainforest map	92
Figure 5.2 Peruvian Boiling River photos	93
Figure 5.3 Extremophiles and their optimal environmental temperature ranges	96
Figure 5.4 Peruvian Boiling River map	98
Figure 5.5 Krona visualization of bacterial diversity	102
Figure 5.6 Distribution of bacterial community	103
Figure 5.7 Archaeal composition and distribution	104
Figure 5.8 Clustering patterns of Boiling River microbes	107
Figure 5.9 Krona chart of microbial diversity in Plaza de Azufre (station 16BR)	109
Figure A.1 Comparing substrate-bound and substrate-free TamI	151
Figure A.2 Equilibrium dissociation constant binding curves for TamI WT and mutants with tirandamycin (1) substrate	151
Figure A.3 Qualitative spectral shift assays	152
Figure A.4 QM calculations for stereoselectivity in TamI chemistry	153
Figure A.5 MD simulations of iterative oxidation cascade with TamL	154
Figure A.6 CO difference spectroscopy for TamI mutants designed to explore the structural basis of TamI	155
Figure A.7 Analytical enzymatic reactions with TamI WT, TamI variants and tirandamycin C substrate	156
Figure A.8 Analytical enzymatic reactions with TamI mutants and tirandamycin C substrate	157
Figure A.9 Suspected H-bonding interaction between His99 and tirandamycin D ketone	158
Figure A.10 PyMOL simulations of TamI mutants	159
Figure A.11 MD simulations of TamI mutants	160
Figure A.12 Endpoint assays with TamI L295A	161
Figure A.13 Endpoint assays with TamI L101A_L295I	162
Figure A.14 Endpoint assays with TamI L295V	162
Figure A.15 The stepwise oxidative cascade of iterative TamI L244A_L295V and TamI WT	163
Figure A.16 Bicyclic ketal distance from Fe(IV)-oxo radical cation	164

Figure A.17 Assays with the flavoprotein TamL	165
Figure A.18 Stereochemical configuration of tirandamycins O and O'	165
Figure A.19 ¹ H-NMR experiments with tirandamycin O (9)	166
Figure A.20 Endpoint assays with TamI WT and variants using tirandamycin M substrate ..	167
Figure A.21 Endpoint assays with TamI WT and variants using tirandamycin N substrate ..	167
Figure A.22 Endpoint assays with TamI WT and variants using tirandamycin O and O' substrates (5:1 ratio of tirandamycin O to O')	168
Figure A.23 Equilibrium dissociation constant binding curves for TamI WT and variants with different tirandamycin substrates	169
Figure A.24 CO difference spectroscopy for TamI mutants designed to alter the selectivity of TamI	170
Figure A.25 New tirandamycin compound characterization - HRMS	171
Figure A.26 Compound characterization – tirandamycin L (6)	172
Figure A.27 Compound characterization – tirandamycin M (7, also known as TAM E)	178
Figure A.28 Compound characterization – tirandamycin N (8)	184
Figure A.29 Compound characterization – tirandamycin O (9)	189
Figure A.30 Compound characterization – tirandamycin O' (10)	194
Figure B.1 Molecular Dynamics simulations of TamI variants and tirandamycin congeners where C11/12 alkene epoxidation takes place	213
Figure B.2 Additional end-point assays with TamI mutants and 1	214
Figure B.3 Hydrogen peroxide formation in TamI-catalyzed reactions	214
Figure B.4 TamI reactions supplemented with Reactive Oxygen Species (ROS) scavengers ..	215
Figure B.5 TamI reactions with tirandamycin C (1) substrate and supplemented with Reactive Oxygen Species (ROS) scavengers in a dose-dependent manner	216
Figure B.6 TamI reactions with tirandamycin L (6) substrate and supplemented with Reactive Oxygen Species (ROS) scavengers in a dose-dependent manner	216
Figure B.7 Additional TamI-catalyzed oxidations via peroxide shunt pathway	217
Figure B.8 End-point assays with P450 TamI mutants and tirandamycin C (1) substrate initiated with iodosobenzene	218
Figure B.9 End-point assays with P450 TamI mutants and tirandamycin L (6) substrate initiated with iodosobenzene	219

Figure B.10 End-point assays with P450 TamI mutants and tirandamycin C (1) substrate initiated with NADPH versus H ₂ O ₂	220
Figure B.11 End-point assays with P450 TamI mutants and tirandamycin L (6) substrate initiated with NADPH versus H ₂ O ₂	221
Figure B.12 End-point assays with P450 TamI WT and other various tirandamycin substrates initiated with NADPH versus H ₂ O ₂	222
Figure B.13 End-point assays with P450 TamI L295A and various tirandamycin substrates initiated with NADPH versus H ₂ O ₂	223
Figure B.14 Endpoint assays with TamI L295A	224
Figure B.15 Uncatalyzed hydration of congeners 7 and 7' epoxides	225
Figure B.16 Equilibrium dissociation constant binding curves for TamI WT and mutants with tirandamycin C (1)	225
Figure B.17 Oxidized spectra for P450 TamI and variants	226
Figure B.18 CO difference spectroscopy for TamI mutants	226
Figure B.19 MD simulations of TamI L295A and tirandamycin L (6)	227
Figure C.1 Substrate engineering strategy to expand the substrate scope of P450 TamI	240
Figure C.2 Computational docking of simplified anchoring groups	240
Figure C.3 Enzymatic reactions with synthetic substrate 36	241
Figure C.4 Compound characterization – Compound 20	242
Figure C.5 Compound characterization – Compound 25	244
Figure C.6 Compound characterization – Compound 30	246
Figure C.7 Compound characterization – Compound 33	248
Figure C.8 Compound characterization – Compound 34	250
Figure C.9 Compound characterization – Compound 35	251
Figure C.10 Compound characterization – Compound 36	253
Figure D.1 Sequence quality control	266
Figure D.2 Alpha diversity metrics	267
Figure D.3 Beta diversity analyses	269
Figure D.4 PERMANOVA analyses between stations at the phyla level	270
Figure D.5 Correspondence analysis (no distance matrix)	271
Figure D.6 Total bacterial phyla composition in the Boiling River	272

Figure D.7 Composition and distribution of bacterial phyla in Boiling River samples	273
Figure D.8 Composition and distribution of less abundant bacterial phyla in Boiling River samples	275
Figure D.9 Composition of bacterial phyla in samples from station 16BR at the phyla level ·	276
Figure D.10 Composition of bacterial phyla in samples from station 16BR at the genera level ···	277
Figure D.11 Distribution of bacterial phyla in the Boiling River stations.....	278
Figure D.12 Boiling River stations	288

List of Tables

Table 1.1 Select iterative P450 enzymes	12
Table 2.1 Tirandamycin C (1) equilibrium dissociation constant (K_d) values	27
Table 2.2 Select Michaelis-Menten kinetic values	47
Table 2.3 Select MIC values	48
Table 5.1 Boiling River microclimates	100
Table A.1 X-ray data collection and refinement statistics	142
Table A.2 NMR characterization for tirandamycin L (6) in CD_2Cl_2	143
Table A.3 NMR characterization for tirandamycin M (7) in CD_2Cl_2	144
Table A.4 Comparison of 1H -NMR spectra between tirandamycin M (<i>in vitro</i>) and the previously published spectra (<i>in vivo</i> , named TAM E)	145
Table A.5 NMR characterization for tirandamycin N (8) in CD_2Cl_2	146
Table A.6 NMR characterization for tirandamycin O (9) in CD_2Cl_2	147
Table A.7 NMR characterization for tirandamycin O' (10) in CD_2Cl_2	148
Table A.8 Total turnover numbers (TTN) for enzymatic reactions involving TamI WT and variants and different tirandamycin substrates	149
Table A.9 Kinetic parameters for TamI WT, iterative L244A_L295V and selective L101A_L295I	149
Table A.10 Substrate binding analysis of TamI WT and variants with different tirandamycin substrates	150
Table A.11 Antimicrobial activity of tirandamycin congeners against select strains	150
Table B.1 Substrate binding analysis of P450 TamI with different tirandamycin substrates ..	211
Table B.2 Determination of turnover frequency (TOF), coupling efficiencies (%) and NADPH consumption rate	212
Table C.1 Asymmetric allylation optimization	239
Table D.1 Boiling River stations and GPS coordinates	261

Table D.2 Sequence length statistics	261
Table D.3 Frequency per ASV feature	262
Table D.4 Demultiplexed sequence counts summary	262
Table D.5 Bacterial abundance at the order level	263
Table D.6 Bacterial abundance at the family level	264
Table D.6 Bacterial abundance at the genera level	265

List of Schemes

Scheme 1.1 C-H functionalization strategies	6
Scheme 1.2 P450 catalytic cycle	10
Scheme 1.3 Iterative oxidative cascades by P450 enzymes	14
Scheme 2.1 Iterative late-stage oxidation reactions catalyzed by P450 TamI	25
Scheme 2.2 TamI mutants generating novel tirandamycin congeners with divergent selectivity from wild-type enzyme	40
Scheme 3.1 General cycle for cytochrome P450 catalysis.....	56
Scheme 3.2 Late-stage oxidation catalyzed by P450 TamI variants.....	57
Scheme 3.3 Proposed mechanisms for the oxidative cascade catalyzed by TamI L295A.....	58
Scheme 4.1 Total synthesis of tirandamycin	77
Scheme 4.2 Synthetic strategies towards a tirandamycin bicycle analogue	81
Scheme 4.3 Synthesis of tirandamycin analogues harboring simplified anchoring groups	83
Scheme A.1 Possible mechanisms for iterative TamI L244A_L295V to install the C10 ketone	141
Scheme B.1 DFT calculations on P450 TamI (peroxo-iron shown as species E in Scheme 3.1) and intermediates 7 and 7'	208
Scheme B.2 DFT calculations on ferredoxin [Fe ₂ S ₂] cluster model, showing charge -3/multiplicity 10 as representative structure	208
Scheme B.3 DFT calculations on P450 TamI (superoxo-iron, shown as species D in Scheme 3.1) and intermediate 7'	209
Scheme B.4 DFT calculations on P450 TamI heme model (superoxo-iron, shown as species D in Scheme 3.1) and intermediate 7'	209
Scheme B.5 DFT calculations on P450 TamI heme model (Cpd I, shown as species G in Scheme 3.1) and intermediate 7.....	210
Scheme B.6 DFT calculations on P450 TamI heme model (Cpd I, shown as species G in Scheme 3.1) and intermediate 7'.....	210
Scheme C.1 Compound 20	230

Scheme C.2 Compound 30	230
Scheme C.3 Compound 25	231
Scheme C.4 Compound 33	232
Scheme C.5 Compound 34	233
Scheme C.6 Compound 35	234
Scheme C.7 Compound 36	234
Scheme C.8 First-generation total synthesis of a tirandamycin bicyclic analogue	238
Scheme D.1 Denoising procedures using DADA2	259

List of Appendices

Appendix A: Experimental Procedures and Supplemental Information for Chapter 2	123
A.1 Materials and General Methods	123
A.2 Fermentations	124
A.3 Protein Expression and Purification	125
A.4 Protein Crystallization	126
A.5 CO-Bound Reduced Difference Spectra of TamI Enzymes	127
A.6 Site-Directed Mutagenesis	128
A.7 Enzymatic Assays	128
A.8 TamI Large-Scale Enzymatic Reactions and Isolation of Products	133
A.9 Computational Methods	135
A.10 Computational Energies and Coordinates of all Optimized Structures	137
A.11 Supplemental Schemes	141
A.12 Supplemental Tables	142
A.13 Supplemental Figures	151
A.14 References	199
Appendix B: Experimental Procedures and Supplemental Information for Chapter 3	200
B.1 Materials and General Methods	200
B.2 Isolation and Purification of Tirandamycin	201
B.3 Protein Expression and Purification	201
B.4 Enzymatic Assays	202
B.5 Computational Methods	207
B.6 Supplemental Schemes	208
B.7 Supplemental Tables	211
B.8 Supplemental Figures	213
B.9 References	227
Appendix C: Experimental Procedures and Supplemental Information for Chapter 4	229

C.1 Materials and General Methods	229
C.2 Synthesis of Tirandamycin Analogues	229
C.3 Protein Expression and Purification	235
C.4 Analytical Enzymatic Assays.....	235
C.5 Computational Methods.....	236
C.6 Supplemental Schemes	238
C.7 Supplemental Tables.....	239
C.8 Supplemental Figures.....	240
C.9 References.....	255
Appendix D: Experimental Procedures and Supplemental Information for Chapter 5	256
D.1 Boiling River Expedition	256
D.2 Metagenomics Methods	257
D.3 Supplemental Schemes.....	259
D.4 Supplemental Tables.....	260
D.5 Supplemental Figures.....	266
D.6 References	278

Abstract

Nature provides us with a wealth of natural product molecules and biological catalysts that represent a rich source of new life-saving pharmaceutical medicines and biotechnological tools for green chemistry. The synthetic challenges associated with derivatizing natural product molecules to access new and improved analogues present a new opportunity for protein and substrate engineering to develop robust biocatalytic methods. Bacterial iterative cytochrome P450 enzymes can perform late-stage C-H oxidation reactions on complex natural product scaffolds with exquisite levels of regio-, stereo- and chemoselectivity. These biocatalytic systems represent chemically useful tools for selective and multi-step oxidation cascades to facilitate access to structural diversification. Furthermore, tapping into unexplored biologically rich environments can enable the discovery of novel and structurally intriguing natural product scaffolds and the biocatalytic enzymes responsible for their biosynthesis and late-stage functionalization. Hence, discovery, characterization and engineering of new bioactive molecules and biocatalysts can advance medicine and biotechnology for improvement of human health and sustainable chemical transformations.

Most of the studies presented in this thesis focus on characterizing and engineering the multifunctional bacterial P450 TamI to elucidate its substrate binding mechanism and apparent stepwise reaction order, as well as alter the step sequence, native regio-, chemo- and stereoselectivity, and number of reactions catalyzed. This was completed successfully with the design of a toolbox of TamI biocatalysts that override the innate substrate reactivity in a catalyst-controlled fashion, catalyzing the biosynthesis of five novel tirandamycin congeners that display previously unreported bioactivities against human pathogens.

The ability of P450 TamI for catalyzing selective and mechanistically divergent oxidative pathways prompted us to elucidate TamI's catalytic cycle. This work revealed that the enzyme uses multiple catalytically active oxidant species for its iterative oxidative cascade towards the formation of two of the new tri-oxidized tirandamycin congeners. This versatility motivated us to explore TamI's scope for oxidizing unnatural bicyclic scaffolds using a substrate engineering

approach. Guided by structural and biochemical results, synthetic analogues were made and tested in enzymatic assays to learn that P450 TamI can recognize, bind and turnover substrates with tirandamycin-like bicyclic moieties that harbor a cyclic ring at the tail of the molecule, even if missing the tetramic acid functionalities, but cannot oxidize a different bicyclic scaffold. The in-depth investigation of P450 TamI-mediated catalysis provides a foundation for future protein and substrate engineering efforts to manipulate iterative P450 enzymes to generate powerful biocatalysts to produce structurally novel natural product molecules.

The last part of this work focuses on characterizing the microbial composition of a unique, extreme, and biologically rich ecosystem in the Peruvian Amazon, the Boiling River, for future genome mining of versatile tailoring enzymes, such as P450 TamI, that display thermostable properties for the development of robust biocatalytic systems. Using metagenomics and bioinformatic tools, we elucidate the bacterial and archaeal communities in this river, and identify samples with unique ecological roles that may translate into novel biosynthetic gene clusters. The continued investigation of these meso-, thermo- and hyperthermophiles expands our understanding of the biological diversity of fragile environments and builds the groundwork for future genomics and metabolic mining to discover new natural product molecules and their biosynthetic enzymes.

Chapter 1*

Introduction and Background

1.1 Natural Products in Medicine

1.1.1 Medicinal Natural Products: History and Sources

Medicinal natural products have been used for thousands of years and throughout multiple living generations. Some of the earliest records date back to the period where *Homo neanderthalensis*, also known as Neanderthals, inhabited Europe and Western Asia.^{1, 2} Closely related to humans with up to 99.7% DNA similarity, Neanderthals' dental plaque DNA provided clues that they may have been auto-medicating when ill. Genetic analyses showed co-occurrence between dental abscess and diarrhea-causing parasites, and fungal and plant species that produce secondary metabolites with pharmaceutical properties including penicillin (anti-bacterial) and aspirin (anti-inflammatory).¹ These results suggested that Neanderthals may have had traditional knowledge on the use of medical natural sources to treat various illnesses.

Natural product (NP) molecules and their derivatives occupy a vast chemical space with significant structural and chemical diversity. With novel mechanisms of action and potent drug-like properties, NP molecules have revolutionized modern medicine. As an example, scientists Y. Tu, W. C. Campbell, and S. Omura received the 2015 Nobel Prize in Physiology or Medicine for discovering plant-derived artemisinin and microbe-derived avermectins, natural product drugs that have become essential antiparasitic medicines.³ Additionally, the unique structural scaffolds of NP molecules provide continuous inspiration for the development of therapeutic synthetic small molecules.

As one of the most successful sources of potential drug leads, NPs and their derivatives represent approximately 45% of the best-selling pharmaceutical drugs.⁴⁻⁶ These molecules have been isolated from multiple natural sources such as animals, marine organisms, plants, fungi, and microorganisms (Figure 1.1).⁷ Famous examples of FDA-approved drugs isolated or derived

* Chapter 1 is included in a manuscript. **Espinoza, R.V.** and Sherman, D.H. Exploring the molecular basis for selective C-H functionalization in plant P450s. *Synth. Syst. Biotechnol.*, **2020**, 5(2), 97.

from plants include the antitumor agent taxol (originally obtained from *Taxus brevifolia*) and the antimalarial quinine (*Cinchona ledgeriana*).⁶ From animal sources, some examples include the anticoagulant heparin (from canine liver cells), the gastroenterologic drug ursodiol (from polar bear bile) and the blockbuster hypertension-treating drug captopril (inspired on a NP from *Bothrops jararaca*, also known as the Yarara or viper snake).⁸ From marine organisms, the non-narcotic ziconotide for severe pain (from venom of sea snail *Conus magnus*) and the antiviral vidarabine (from Caribbean sponge *Tethya crypta*).⁹ From fungi, the chemotherapeutic agent paclitaxel (from *Penicillium raistrickii*) and the family of cholesterol drugs statins (from *Penicillium citrinum*). Lastly, from bacteria, the antibiotics chloramphenicol (*Streptomyces venezuelae*) and tetracycline (*Streptomyces aureofaciens*).^{10, 11} NP compounds have evolved over time as drug-like molecules with complex structures and unusual functionalities, providing new opportunities for successful natural-based drug discovery as exemplified above.

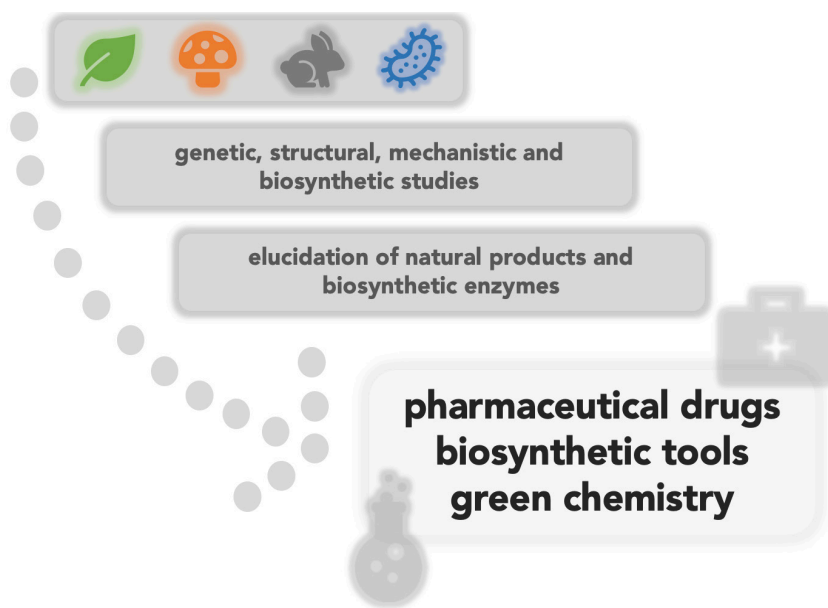


Figure 1.1. Natural product drug discovery workflow.

NP compounds have evolved over time as drug-like molecules with complex structures and unusual functionalities, providing new opportunities for successful natural-based drug discovery as exemplified above.

1.1.2 The Present and Future of Natural Products in Drug Discovery

Although NP molecules are notoriously abundant in natural environments, less than 10% of the world's biodiversity has been explored for NP bioactivity. This may be due to a combination of factors including rediscovery of known compounds and difficulties in obtaining sample collection permits that comply with international ethical standards to access unexplored biodiversity hotspots. The rise of combinatorial chemistry in the 1980s also led to a critical decline in natural product drug discovery research and an increase in synthetic libraries of small molecules.⁵ Additionally, there is evidence that only a minuscule percentage of microbes have been successfully grown in the laboratory.¹² Current microbial culture techniques have only unveiled a small fraction of the world's microbial biodiversity possible due to the challenge in

mimicking the naturally occurring microclimate where these microorganisms exist. There is an abundant number of microorganisms left to be discovered and studied for their role in biological processes and their potential for developing new medicines and biotechnological tools.

Recent advances in synthetic biology, bioinformatics, and genomics and metagenomics technologies are providing new routes towards previously inaccessible genetic and chemical data (Figure 1.1).¹³ At the genetic level, studying the entire microbial community present in one sample (metagenomics), in addition of solely focusing on capturing a small fraction of bacterial isolates from the sample (genomics), enables access to rare microbes with intriguing chemical profiles. These previously undocumented microorganisms may harbor potentially unknown biosynthetic gene clusters that are responsible for producing new molecular scaffolds with novel connectivities and functional groups. These techniques also facilitate the discovery of “cryptic” gene clusters, genes that are silenced under laboratory growth conditions and that may encode for the synthesis of an array of NPs of clinical importance.¹⁴ At the biochemical level, we are now able to manipulate and engineer specific enzymes associated with the biosynthesis of bioactive NPs to facilitate derivatization of the molecules, expand the chemical diversity of structurally related compounds, and develop potent catalytic tools for sustainable and green chemistry. Considering a frightening increase in antibiotic resistance worldwide, viral pandemics shutting down economies, concerning climate changes, and irreversible losses of fragile natural environments and their rich biodiversity, it is particularly important that we continue supporting NP drug discovery and monitoring of the producing natural sources, including microbes. Nature has evolved and adapted over millions and millions of years and exploring it from a medicinal and ecological perspective at the microbial level can inspire us to fight for its conservation while helping make our lives better.

1.2 Biocatalysis and C-H Functionalization

1.2.1 General Biosynthesis of Secondary Metabolites

Secondary (or specialized) metabolites are organic NP compounds that provide a competitive survival and evolutionary advantage to the producing organism including plants, animals, fungi, and microbes. Unlike primary metabolites (amino acids, nucleobases and sugar residues), secondary metabolites are not considered critical for the growth and reproduction of the

natural producer.¹⁵ These NPs have a broad array of ecological functions including the use of pheromones as social signaling compounds or siderophores as protective agents that act as repellants or toxins. Secondary metabolites are often produced selectively and based on what the producer is exposed to during its lifetime and environment.¹⁶

The biosynthesis of secondary metabolites is derived from primary metabolic pathways including the shikimic acid pathway and the tricarboxylic acid cycle, where intracellular intermediates are condensed into more complex scaffolds. With over 2 million secondary metabolites known to science,^{3, 15} they are commonly classified based on their structure, function,

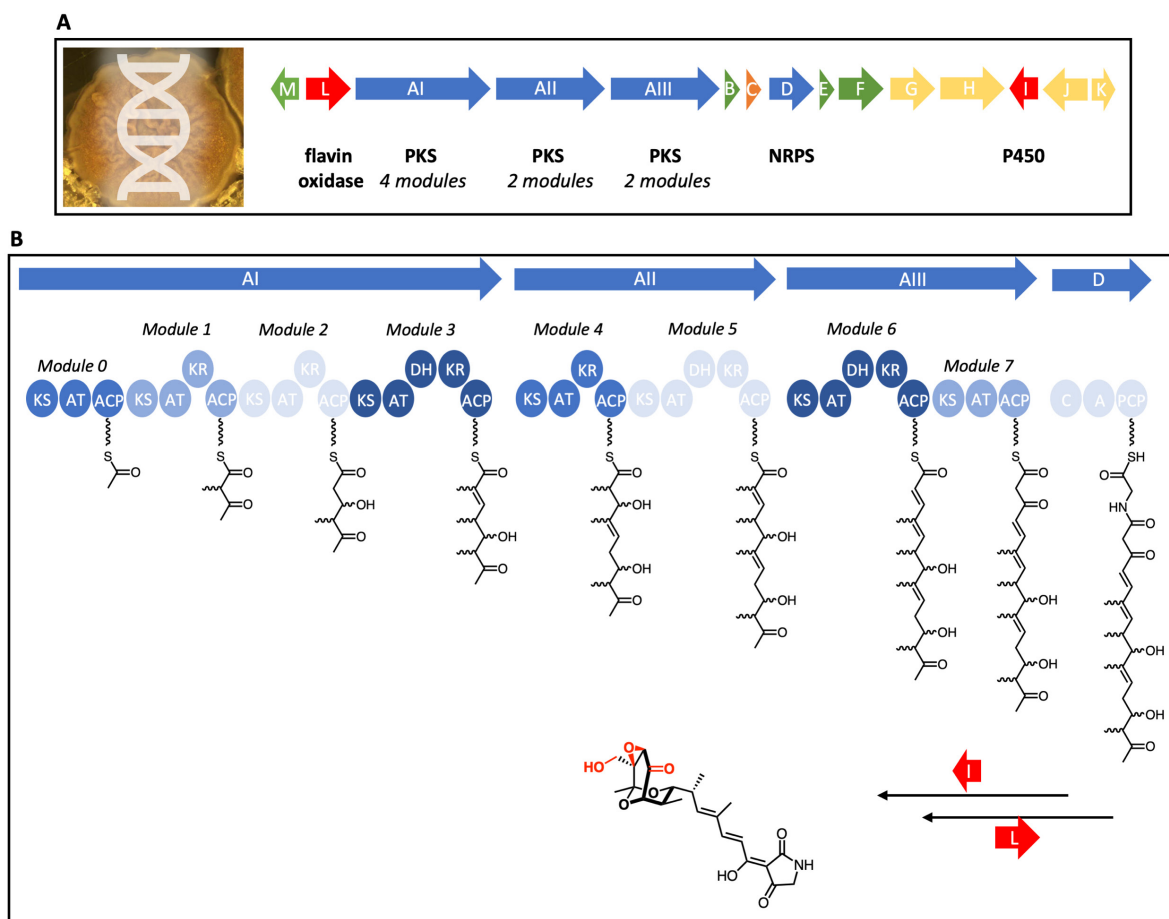


Figure 1.2. Biosynthesis of tirandamycin antibiotics. A) Genome sequencing enabled elucidation of the tirandamycin biosynthetic gene clusters. Genes are color-coded to indicate enzyme classification (blue = PKS or NRPS, yellow = regulatory and resistance genes, red = tailoring enzymes, green = unknown function, orange = Dieckmann cyclase). B) Biosynthetic assembly line responsible for producing tirandamycin incorporates methylmalonyl-CoA, malonyl-CoA and glycine building blocks. Tailoring biocatalysts TamI and TamL (red) oxidize tirandamycin at a late-stage.

and biosynthesis. Five major classes exist including nonribosomal peptides, linear and aromatic polyketides, alkaloids, terpenoids and steroids and enzyme cofactors.

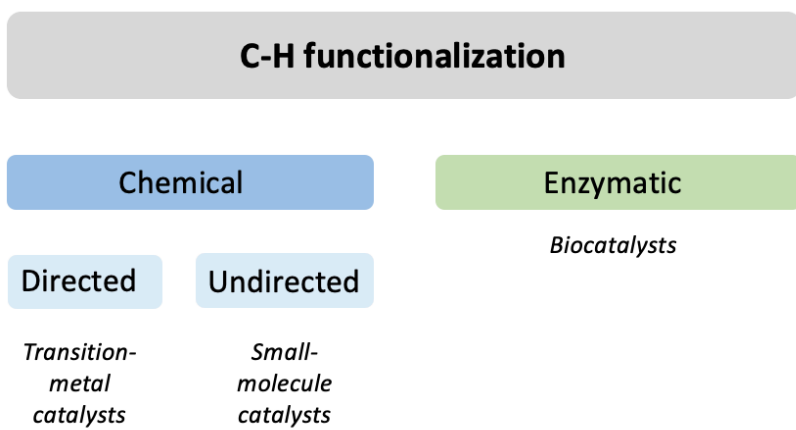
Biosynthetic gene clusters (BGC) found within the producer's genome encode for biosynthetic machineries and tailoring enzymes that build these complex secondary metabolites using the controlled assembly of biological building blocks (Figure 1.2A). For example, a hybrid system comprising large, multi-domain enzymes known as polyketide (PKS) and nonribosomal peptide synthases (NRPS) is responsible for building the backbone structure of the microbe-derived antibiotic tirandamycin (Figure 1.2B).^{17, 18} Once the scaffold is released from the assembly line, free standing and tailoring enzymes located near the core BGC functionalize the molecule and install biologically relevant hydroxyl, epoxide, and ketone groups. Insights into the structure and mechanisms of these biocatalysts open the possibility for direct (or late-stage) functional group modification, including C-H bond functionalization, for the derivatization of NPs to access new analogues with improved medicinal properties and molecular stability or novel modes of action. Manipulating these biocatalysts using protein or substrate engineering strategies also provides new avenues for biocatalytic and environmentally-friendly routes towards making new, bioactive molecules.

1.2.2 Chemical and Enzymatic C-H Functionalization

Many biologically relevant NPs, including pharmacologically important molecules, are oxidized hydrocarbons.¹⁹ Consequently, achieving selective functionalization of C-H bonds within a complex scaffold is one of the most attractive strategies in organic and medicinal chemistry to streamline the diversification and derivatization of molecules.¹⁹ Hartwig and Larsen describe C-H functionalization as the “replacement of an unactivated C-H bond with a functional group”¹⁹ (Scheme 1.1). Using transition metal catalysts,²⁰ chemists can achieve *directed* C-H functionalization, where a pre-existing functional group in the substrate is required to direct regio- and stereoselectivity. Diversely, *undirected* C-H functionalization is accomplished using small-molecule catalysts^{21, 22} and does not require coordination or chelation prior functionalization, although tuning regioselectivity can be a challenge. The field of chemical C-H bond activation and functionalization has successfully evolved over the last decade,¹⁹ however, the downsides of current methods are that they often require harsh conditions, the incorporation of a directing group into the target molecule,²³ or the use of protection/deprotection strategies. These inconveniences

add additional time, effort and resources. Additionally, selectivity often results from inherent substrate reactivity limiting access to new points of derivatization within the molecule. Thus, late-stage and selective functionalization of a C-H bond within a complex chemical structure continues challenging synthetic and medicinal chemists.

Alternatively, nature has evolved to overcome these challenges through the use of enzymes that play critical roles in a variety of physiological functions and achieve biocatalytic/enzymatic C-H functionalization.²⁴ Enzymes, nature's catalysts, perform specific reactions on particular substrates with high degrees of chemo-, stereo-, and regioselectivity and under mild conditions.²⁵ ²⁶ To achieve exquisite selectivity and efficiency, enzymes bind and position the substrate in a precise orientation and geometry towards the oxidizing group, often overriding the inherent substrate reactivity. In the words of Breslow, "biochemical selectivity usually reflects such orientation, rather than the intrinsic reactivity of the substrate molecule".²⁷ Unlike small-molecule catalysts, biocatalytic technologies can be optimized using rational mutagenesis or directed evolution of the enzyme for a specific application, including expanding the substrate scope,



increasing substrate turnover or altering the native selectivity. Thus, enzymatic C-H functionalization enables direct access to structural core diversification by accessing unreactive C-H bonds, generating new sets of analogues with novel points of derivatization and maximizing product yield.

Scheme 1.1. C-H functionalization strategies.

Some challenges related to enzymatic catalysis include enzyme instability, the need for expensive cofactors, and the high costs associated with enzyme engineering. However, despite of these difficulties, industries are increasingly adopting biocatalytic approaches towards the synthesis of biologically relevant compounds. A stellar example includes the use of an engineered ketoreductase for the large-scale asymmetric reduction of a ketone towards production of montelukast sodium, also known as Singulair, a leukotriene receptor antagonist used for treating asthma and seasonal rhinitis.²⁸ With a directed evolution campaign, the ketoreductase (CDX-026)

improved catalytic activity from 0.1 g to 100 of product per liter using a lyophilized cell lysate, 70% organic solvent and 45 °C conditions, at a > 200 kg scale. This enzymatic process provided an eco-friendly substitute to the original synthetic route that used a moisture-sensitive chemical catalyst with poor atom economy, high levels of corrosivity, and tedious workup conditions. Another excellent example involves using a tuned transaminase/PLP enzyme for the generation of sitagliptin phosphate, a pharmaceutical used for treatment of type II diabetes, via enamine formation and asymmetric hydrogenation. Initially, the desired product was made using a chemocatalytic route with 97% e.e. where high pressures were needed for the Rh catalyst to function properly and remaining trace amounts of Rh had to be removed to minimize side effects. Using a combination of substrate walking, modeling and directed evolution, the engineered biocatalyst achieved turnover of 200 g of substrate per liter, with a 99.9% e.e, while reducing manufacturing costs and total waste.²⁹

Biocatalysts represent a renewable, biodegradable and non-toxic alternative approach to chemical C-H functionalization methods for the regio- and stereoselective functionalization of C-H bonds, where atom economy is optimized, and waste minimized. Consequently, biocatalytic C-H functionalization is a valuable tool for green and sustainable chemistry and holds great promise for chemical synthesis.³⁰⁻³³

1.3 Cytochrome P450 Enzymes as Biocatalysts

1.3.1 History and Structure of P450s

Highly prevalent in nature, cytochrome P450 enzymes are powerful biocatalysts for selective C-H functionalization.¹⁹ These enzymes contain a cysteine-ligated heme-iron center and efficiently catalyze the activation of inert C-H bonds within complex molecules for a plethora of transformations with exquisite regio-, chemo- and stereoselectivity.³⁴ Few synthetic strategies approach the levels of efficiency achieved by P450s, inspiring chemists to engineer these enzymes into potent biocatalysts with maximized synthetic utility for the production of valuable molecules.²⁷

Identified in the 1950-60s,³⁵ P450s were first described as an unknown CO-binding pigment within liver microsomes of rabbits that is responsible for the oxidation of xenobiotic

molecules and that absorbs at 450 nm. Subsequent research establishes that cytochrome P450 enzymes are found in all forms of life including humans, plants, fungi, viruses, insects, bacteria, yeasts, nematodes, and mollusks, with over 350,000 CYP genes reported to date.³⁶ P450s are divided into families based on sequence identity at the family and subfamily levels.³⁷ This P450 nomenclature system facilitates communication across various disciplines involving the study of P450 enzymes and shows evolutionary relationships among P450 genes.

Cytochrome P450s catalyze essential chemical transformations that play important roles in xenobiotic degradation, hormone metabolism, drug interactions and secondary metabolism responsible to produce natural product molecules. For example, the catalytic activity of P450s is key in plant metabolism where they have been reported to participate in herbicide detoxification

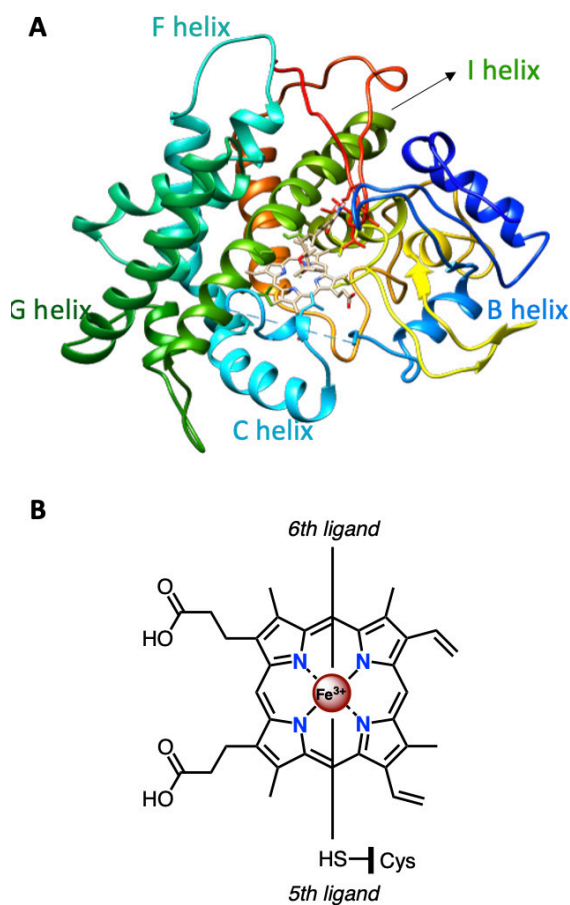


Figure 1.3. General P450 structure. A) The solved crystal structure of P450 TamI with important helix motifs labeled. B) Heme ligand in P450s. The 5th ligand refers to the highly conserved cysteine residue, and the 6th ligand indicates a non-covalently bound ligand like water.

via dealkylation or hydroxylation reactions, enabling herbicide tolerance to develop in crops.³⁸ Their role in human metabolism is also noteworthy. For instance, human CYP3A4 displays high level of substrate promiscuity and metabolizes many xenobiotic pharmaceuticals, making this P450 a key target candidate for further pharmacodynamics studies.³⁹ The repertoire of chemical reactions that P450s catalyze for their key role in biological processes is vast and includes C-H hydroxylation, desaturation and nitration, aromatic hydroxylation, N-hydroxylation, alkene epoxidation, dehydrogenation, dealkylation, rearrangement, cyclopropanation, C-C and C-N bond formation, and C-S and C-O bond formation. Additionally, some P450-catalyzed transformations, such as stereoselective hydroxylation, can be PCP or ACP-dependent.⁴⁰

Starting with the first solved structure of a cytochrome P450 enzyme in 1985 (bacterial

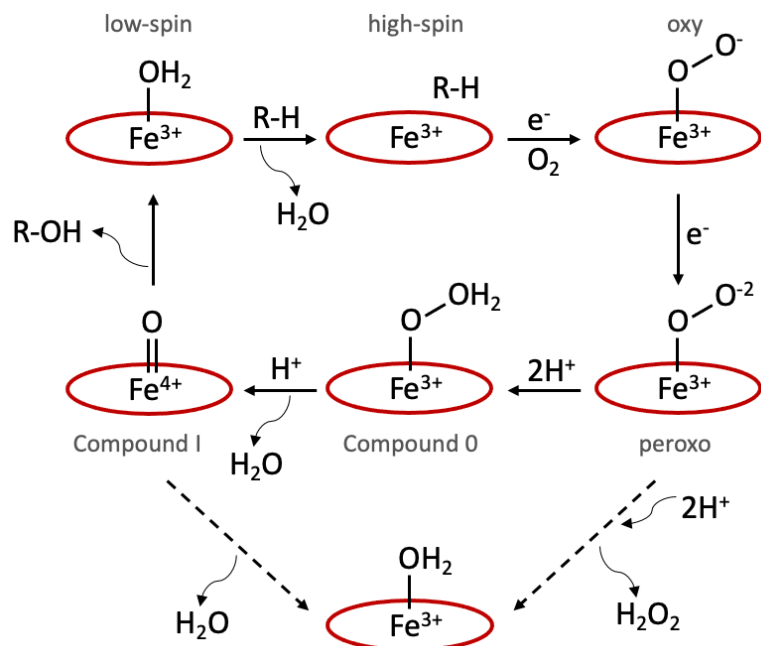
P450cam, CYP101A1),³⁵ investigation of the crystal structures of P450s shows that the structural overall fold of these enzymes is maintained. Even when sharing a low sequence identity, P450s display a triangular prism shape with the active site buried in the center. The secondary structure elements of the enzyme comprise approximately 12 or 13 α -helices and 4 or 5 β -sheets.³⁵ Helices F and G and its connecting F/G loop are known to be critical for closure of the P450 active site upon substrate binding while the I helix, the longest helix, typically contains the conserved acid-alcohol pair that is important for oxygen activation. The B and C helices and B/C loop are highly divergent and can influence substrate binding and selectivity.

In 1964, Omura and Sato identified the heme ligand in P450s as heme β containing an iron protoporphyrin IX ring moiety.³⁵ The four equatorial and inward facing pyrrole nitrogen atoms of the planar protoporphyrin ring stabilize the centrally bound iron atom (Figure 1.3B). Two additional ligands located at axial positions, including the highly conserved thiol ligand, are also available for interaction and contribute to the heme iron stabilization. Known as the fifth ligand, the negatively charged thiolate sulfur atom of the cysteine residue links to the positively charged iron atom and is essential for the “push-pull” mechanism in P450s. The donating thiolate plays a role in regulating the rate constant and energetic barriers required for oxygen activation and hydrogen abstraction and influences the various redox states that the iron may access during the P450 catalytic cycle.⁴¹ The sixth coordinate site is often occupied by a non-covalently bound and replaceable ligand such as water, molecular oxygen, a P450 substrate or inhibitor.

1.3.2 P450 Mechanism

The P450 catalytic cycle involves a myriad of transient iron intermediate species that alternate mostly between the ferric (Fe^{3+}) and the ferrous states (Fe^{2+}), and spend a short time in the hypervalent Fe^{4+} oxidation state.⁴¹ First, the substrate binds to the active site in the heme domain, displacing the water molecule that acts as a sixth ligand (Scheme 1.2) This binding step stimulates a conformational change in the active site and induces transition of the iron from a low to a high spin ferric species, where the outer orbital of the iron contains one or more unpaired electrons. The shift to a high spin state can be observed spectrophotometrically upon substrate binding and measured to calculate the % spin shift of the enzyme. This ferric iron form facilitates the first electron transfer from the redox protein partner(s) to the heme group in the presence of NAD(P)H leading to reduction of the iron to the ferrous state. Once in this form, molecular oxygen

binds covalently to the iron at the sixth axial position activating the oxygen and generating the ferric-superoxy species. Uncoupling can occur at this step from dissociation of the iron-oxygen bond releasing anion superoxide radicals (auto-oxidation shunt pathway). The second electron is transferred reducing the dioxygen to the negatively charged, nucleophilic and short-lived peroxide iron form. Next, a proton relay system (comprising a



Scheme 1.2. P450 catalytic cycle. Uncoupling pathways are shown in dashed arrows.

conserved acid-alcohol pair, amino acid side chains and water molecules in the active site) facilitates the delivery of two protons in a sequential fashion. Upon the first proton transfer, the ferric hydroperoxyl species (known as Compound 0 or Cpd 0) is formed. If protonation of Cpd 0 occurs, hydrogen peroxide gets released as an uncoupled product (peroxide shunt pathway). The second proton delivery facilitates cleavage of the O-O bond, releasing a water molecule and generating the highly reactive, ferryl oxo Compound I species (Cpd I). This iron form is considered the main oxidizing species in the P450 catalytic cycle responsible for C-H hydroxylation via the oxygen bound mechanism, where the active oxidant abstracts a hydrogen from the substrate followed by radical recombination and generation of the oxygenated product.⁴¹ Double reduction and double protonation of Cpd I can take place forming water as a side product (oxidase shunt pathway). Finally, water occupies the sixth coordination site again, turning the enzyme to its original resting state.

1.3.3 P450 Classification and Redox Partner Systems

Cytochrome P450 enzymes employ redox partner enzymes to sequentially transport electrons from NAD(P)H to the P450 heme center for oxygen activation and catalytic activity.³¹ Different protein domain systems have been described in the literature for P450-catalyzed reactions including three-, two- and one-component architectures (Figure 1.4). Most of the known

bacterial P450s, including P450cam, utilize a three-protein system where each enzyme is expressed from a different gene but function synergistically for substrate turnover.³⁵ Some eukaryotic mitochondrial P450s have also been described where the P450 is bound to the inner mitochondrial membrane through the N-terminus. This system includes a ferredoxin reductase domain, a ferredoxin containing an iron-sulfur cluster and a P450 constituent. As an example, P450cam has an FAD-dependent ferredoxin reductase (putidaredoxin reductase) and a 2Fe-2S ferredoxin iron-sulfur cluster (putidaredoxin). The two-component system uses a cytochrome P450 reductase (also known as CPR) that binds to FAD and FMN for electron transport. The CPR component evolved from a fusion between a ferredoxin reductase and a 2Fe-2S ferredoxin, where FMN binding replaced ferredoxin binding.⁴² P450s using a two-protein architecture are most found in the endoplasmic reticulum membranes of eukaryotic species such as rabbit-derived P450 CYP2B4,⁴³ and hence, their recombinant expression is often challenging. Some examples of self-sufficient P450s have been reported where the P450 domain is joined to the electro-donating partner CPR via a peptide linker, such as in the case of bacterial P450-BM3 from *Bacillus megaterium*.²⁵ Although less abundant, P450s using a one-protein system offer distinctive advantages for industrial biocatalytic applications.

Given the challenging feat of finding and characterizing the genes encoding for the native electron transport proteins within a genome, heterologous redox partner constructs are often utilized. An example includes the common use of ferredoxin and ferredoxin reductase enzymes from *Spinacia oleracea* (spinach) in the laboratory for transporting

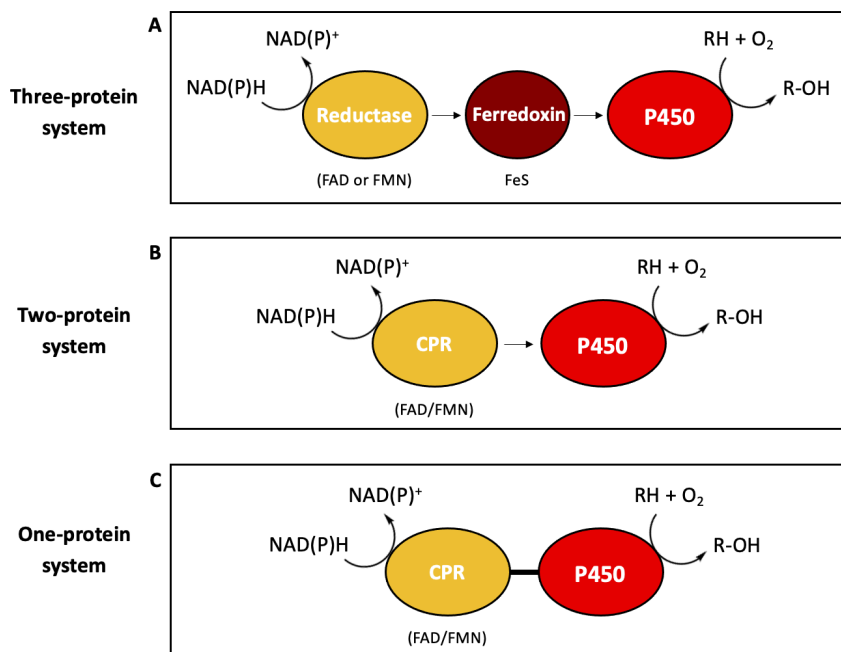


Figure 1.4. P450-catalyzed systems. A) Three-protein domain architecture includes a ferredoxin reductase, a ferredoxin containing an iron-sulfur cluster, and a P450. B) Two-protein system includes a P450 domain and a cytochrome P450 reductase (CPR). C) Self-sufficient P450 systems where the CPR constituent is covalently linked to the P450.

electrons to a free-standing P450.⁴⁴ Alternatively, the requirement for expensive separate protein partners can be bypassed by covalently genetically linking the P450 to a reductase system. The Sherman Laboratory at University of Michigan has previously generated catalytically self-sufficient biosynthetic versions of bacterial P450s by fusing the P450 constituent to the reductase domain (RhFRED) of P450RhF from *Rhodococcus sp.* NCIMB 9784.40.³⁴ This genetic engineering strategy facilitates harvesting and purification of proteins required for enzymatic reactions, lowering costs and manual labor efforts.

1.3.4 P450s in Late-Stage Natural Product Diversification

1.3.4.1 Iterative Oxidation Cascades in Microbial P450s

The hallmark reaction of cytochromes P450 is the single oxidation of a C-H bond,^{25, 45} but a few multifunctional P450s that carry out iterative oxidation reactions have been identified. Multifunctional P450s can perform multiple consecutive oxidation reactions on the same core scaffold (Table 1.1).

P450	Organism of origin	Natural product molecule	# of reactions	Reaction type
TamI	<i>Streptomyces sp. 307-9</i>	tirandamycins	3	Hydroxylation and epoxidation
MycG	<i>Micromonospora griseorubida</i>	mycinamycin	2	Hydroxylation and epoxidation
GfsF	<i>Streptomyces graminofaciens</i>	FD891	2	Epoxidation and hydroxylation
AurH	<i>Streptomyces thioluteus</i>	aureothin	2	Hydroxylation and ether formation
LovA	<i>Aspergillus terreus</i>	lovastatin	2	Hydroxylation
P450-1 (KAO)	<i>Gibberella fujikuroi</i>	gibberellins	4	Hydroxylation, ring contraction and aldehyde oxidation
Tri4	<i>Fusarium graminearum</i>	isotrichotriol	4	Hydroxylation and epoxidation

Table 1.1. Select iterative P450 enzymes. These multifunctional enzymes perform 2 or more consecutive oxidation reactions on a single substrate.

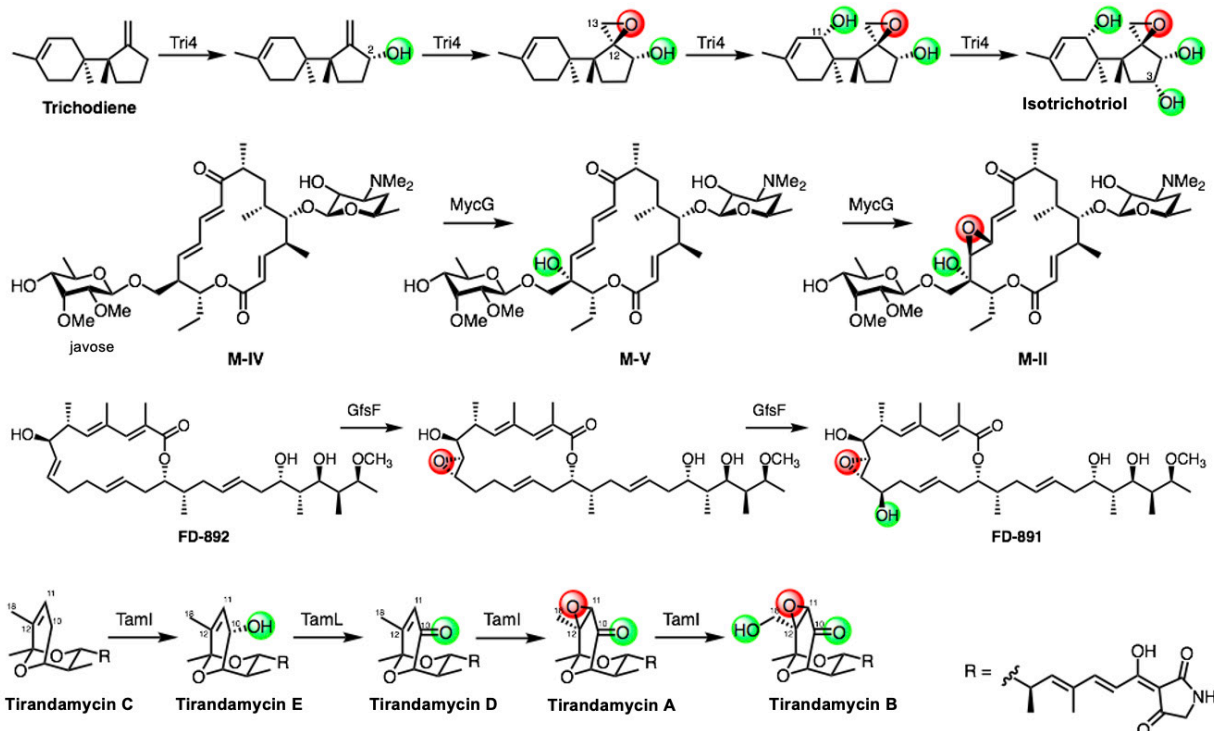
A large variety of multifunctional P450s able to catalyze three or more oxidation steps are monooxygenases of plant⁴⁶ or fungal^{47, 48} origin such as P450 Tri4 (also known as CYP58) from the *Fusarium sporotrichioides* fungal plant pathogen. P450 Tri4 catalyzes four stepwise oxidation

reactions including three methylene and allylic hydroxylation steps and one epoxidation reaction on the trichodiene natural product (Figure 1.5) to generate isotrichotriol towards formation of the mycotoxin trichothecene skeleton. Although versatile, these P450s are often challenging to heterologously express and structurally characterize.⁴⁹ Therefore, investigating their more accessible bacterial counterparts can provide insights into the oxidative mechanism and versatile capabilities of multifunctional P450s. Shining examples of bacterial multifunctional P450s include MycG70 and GfsF,⁵⁰ both of which can catalyze up to two successive C-H oxidation reactions (Scheme 1.3). P450 MycG performs an allylic hydroxylation on the mycinamicin macrolide antibiotic (in the carbon atom located beta from the O-linked javose) followed by an alkene epoxidation in neighboring carbons. Work by Yang et. al. showed that this dual cascade occurs in a strict order given that MycG could not oxidize a substrate analogue harboring an epoxy group but lacking a hydroxyl moiety.⁵¹ Contrarily, P450 GfsF first epoxidizes macrolide FD-892 followed by a hydroxylation step to yield FD-891 without substrate promiscuity detected.⁵⁰

To date, only one bacterial multifunctional P450 enzyme able to perform up to three oxidation reactions on the same scaffold has been reported (Figure 1.5).¹⁸ This exceptionally versatile P450, TamI, performs successive epoxidation and hydroxylation reactions in an iterative cascade with the flavin oxidase TamL, leading to the oxidative tailoring of the bicyclic ketal moiety of tirandamycin antibiotics in the actinomycete *Streptomyces* sp. 307-9 (Figure 1.2).¹⁸ Although collectively they are capable of oxidizing a broad range of substrates with precise selectivity, individual P450s often show high levels of substrate specificity,⁵² limited stability, and low turnover rates, particularly towards unnatural substrates, thus limiting the scope of their applications in synthetic chemistry.^{33, 53} Extensive studies with P450-BM3 and P450cam have shown that the regioselectivity profile and substrate scope of P450s may be altered through mutagenesis,^{54, 55} highlighting the ability of these powerful enzymes to address challenges faced by many small-molecule catalysts.²⁵ However, besides these model cases, a very limited number of bacterial P450s have been extensively studied and re-engineered to improve their catalytic properties and broaden their substrate scope.^{31, 56, 57}

Inspired by prior efforts to engineer bacterial P450s to partially overcome their limitations,^{52, 58} the Sherman and Montgomery laboratories in University of Michigan have previously exploited the active site selectivity, substrate anchoring mechanism, and substrate promiscuity of the bacterial P450 PikC from the pikromycin biosynthetic pathway.^{56, 57} Using

protein and substrate engineering methods, they generated a potent C-H bond activation catalyst that displays remarkable regio- and stereoselectivity and exceptionally broad substrate scope.⁵⁶ In spite of the development of these P450-based C-H functionalization biocatalysts, they mostly remain limited to catalyzing mono-oxidation reactions. Therefore, investigating multifunctional monooxygenases capable of performing multiple oxidative steps is necessary to further probe and expand the utility of P450s as biocatalysts.⁵⁹ While the versatile capabilities of multifunctional P450s hold great promise for employment as C-H bond activation catalysts, they remain underexplored in terms of function and structure.⁴¹ Thus, exploring the unique catalytic promiscuity of bacterial multifunctional P450 enzymes and expanding their substrate scope via protein and substrate engineering efforts can broaden the synthetic utility of these enzymes and further probing the utility of P450s as biocatalysts. The fine-tuned P450s may then be utilized as late-stage iterative biocatalysts to perform selective C-H functionalization reactions and produce complex molecules via chemoenzymatic methods,⁶⁰ thereby facilitating access to new structural motifs³⁴ that may be extremely challenging to acquire by conventional synthetic methods.



Scheme 1.3. Iterative oxidative cascades by P450 enzymes. A) Fungal P450 Tri4 generates isotrichotriol in four consecutive oxidation reactions. B-C) Dual-function bacterial P450 enzymes MycG and GfsF catalyze epoxidation and hydroxylation of macrolide antibiotics. D) Bacterial P450 TamI catalyzes two hydroxylation and one epoxidation of tirandamycin natural products generating bioactive tirandamycin B.

1.3.4.2 Selective C-H Functionalization in Plant P450s

Many plant P450s are of industrial relevance in the production of pharmaceuticals, fragrances, pesticides, and vitamins, yet few have been employed commercially.⁶¹ Key examples include the P450 CYP71AV1 used for large-scale production of the antimalarial drug artemisinin, and CYP75 enzymes exploited for their differential hydroxylation abilities to manipulate color patterns in top-selling flowers.⁶² These cases illustrate a minuscule portion of the synthetic potential of P450s.

Unveiling the molecular basis for selective C-H oxidation in plant-derived P450s can facilitate their engineering towards increased commercial applications. However, despite their widespread occurrence,^{61, 63} little is known about the basis for their late-stage selectivity given the limited number of crystal structures available.⁶⁴ Contrary to bacterial P450s, which are generally soluble and expressed in high levels,^{62, 63} most plant P450s are found in low yields in native tissues complicating their isolation. Moreover, they are typically membrane-bound through the N-terminus in the endoplasmic reticulum (Figure 1.5), and insoluble when produced in bacterial expression systems such as *Escherichia coli*.⁶¹ Other challenges include the need for co-expression of native reductase partners, proper incorporation of the heme co-factor, differences in codon preference and genetic instabilities when large or multiple plasmids are employed.⁶⁵ Eukaryotic-based heterologous hosts exist, including *Saccharomyces cerevisiae* and *Pichia pastoris*, yet similar challenges remain.

In a recent *ACS Catalysis* article,⁶⁶ authors Chun Li and co-workers highlight an interesting avenue towards the rational engineering of plant P450s that are difficult to express *in vitro*. This work deepens our understanding of the selectivity and iterative mechanism involved in the biosynthesis of medicinally relevant molecules. Elucidating the role P450s play in functionalizing plant natural products is fundamental for guiding engineering efforts towards improving the performance of these enzymes. This may include redirecting the selectivity of the biocatalyst, designing self-sufficient P450-redox partner chimeric fusions, allowing the use of light-driven cofactor regeneration processes, or enabling new-to-nature chemical reactions.⁶⁷

The work of Chun Li et al.⁶⁶ was focused on decoding and fine-tuning the molecular factors controlling selectivity and iterative oxidation in the CYP72A63 from *Medicago truncatula* towards the synthesis of high-value, bioactive licorice triterpenoids. When efforts towards expressing

CYP72A63 in *E. coli* failed, Chun Li et al. engineered a yeast strain to produce 11-oxo- β -amyrin *in vivo* by introducing the genes responsible for its biosynthesis along with CYP72A63, which was previously determined to synthesize glycyrrhetic acid from 11-oxo- β -amyrin. Although efficient, the P450 does not catalyze formation of glycyrrhetic acid selectively as other intermediates and by products, including glycyrrhetol, glycyrrhetaldehyde and 29-OH-11-oxo- β -amyrin, are also generated. Based on this catalytic promiscuity, Chun Li et al. postulated that key residues in the P450 active site may govern selectivity.

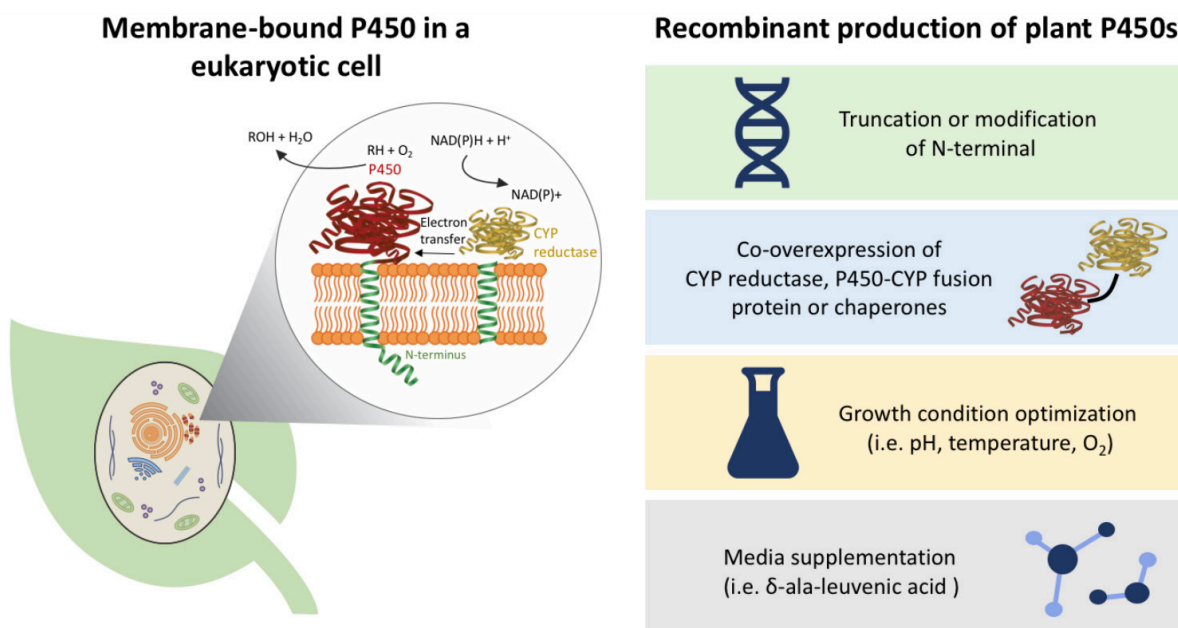


Figure 1.5. Engineering plant P450s. A) Location of a membrane-bound P450 (catalytic domain in red and membrane anchor in green) and CYP reductase within an eukaryotic cell. B) Methods to enable production of recombinant plant P450 enzymes.

Although recombinantly producing plant P450s is challenging, there is an increasing number of strategies to do so (Figure 1.5). Truncating the membrane-anchoring segment can increase the solubility of a P450 as exemplified by Nagano et al. in their work to crystallize the plant CYP90B1.⁶⁴ If this deletion approach adversely impacts the ability of the P450 to couple with CYP reductase for electron transfer, modifying the transmembrane helix is an alternative option.⁵ In some cases, the need of P450s for NADPH, which is provided by the reductase, can stress an organism leading to metabolic imbalances. Although often low yielding, this can be

resolved by co-overexpressing the reductase or creating the corresponding P450 fusion protein. Co-expressing chaperones to facilitate protein folding can also promote protein expression. Additional approaches include media supplementation and optimization of cultivation parameters such as pH, temperature and dissolved oxygen concentrations.⁵ Generally, the *S. cerevisiae* expression system leads to higher yields of P450s in comparison to *E. coli* and *P. pastoris*, although the use of novel photosynthetic hosts such as cyanobacteria is becoming more prevalent.⁶⁸ It is important to consider that prior to rigorously characterizing P450s *in vitro*, transient expression systems can provide insightful, rapid information into the biochemical function of these enzymes.⁶⁹

In the absence of crystallographic data, indirect methods such as molecular modeling serve to investigate substrate docking in a P450 active site.^{67, 69, 70} To enable structure-based engineering of CYP72A63, Chun Li et al.⁶⁶ created homology models of the enzyme for docking with 11-oxo- β -amyrin and glycyrrhetol. This led to the identification of nine active site residues, of which two had been previously reported,⁷⁰ hypothesized to impact the differential oxidation patterns observed. Using site-directed and site-saturation mutagenesis, the investigators gained valuable information on the structure-function relationship of the P450 and engineered variants that fine-tune catalytic activity, regio- and chemoselectivity. In comparison to wild type CYP72A63, the T338S mutant increases the distance between the C-29 hydrogen and the heme iron, resulting in complete regioselectivity for the C-30 position and enhanced iterative abilities, producing glycyrrhetic acid at a 7-fold higher level (Figure 1.6). Reversing regioselectivity to favor C-29 oxidation, the L398I mutant facilitates a 180° rotation of the substrate leading to exclusive formation of 29-OH-11-oxo- β -amyrin. With increased hydrophobicity, the CYP72A63 mutant L509I changes substrate accessibility into the active site, altering the chemical output and selectively producing glycyrrhetol in a 90.6% yield. Given that efforts to design a variant that selectively generates glycyrrhetaldehyde were not fruitful, the researchers screened heterologous CYP reductases and determined that co-overexpressing GuCPR2 from *Glycyrrhiza uralensis* with CYP72A63 selectively produces the aldehyde in an 82.3% yield.⁶⁶ By exploring the versatile catalytic potential of CYP72A63, the authors accentuate the value of plant P450s as essential tools for late-stage oxidation of triterpenoid molecules, providing an excellent starting point for additional engineering of related enzymes.

Increasing technologies for the *in silico*, *in vitro* and *in planta* characterization of plant-derived P450s^{61, 65, 68, 69} will continue to unveil the molecular factors governing their selectivity in C-H functionalization. The article of Chun Li et al.⁶⁶ demonstrates the potential of plant P450s for protein engineering and expanding the synthetic abilities of these biocatalysts. Future work will undoubtedly reveal their increasingly important role for introducing late-stage structural diversity in complex chemical scaffolds.

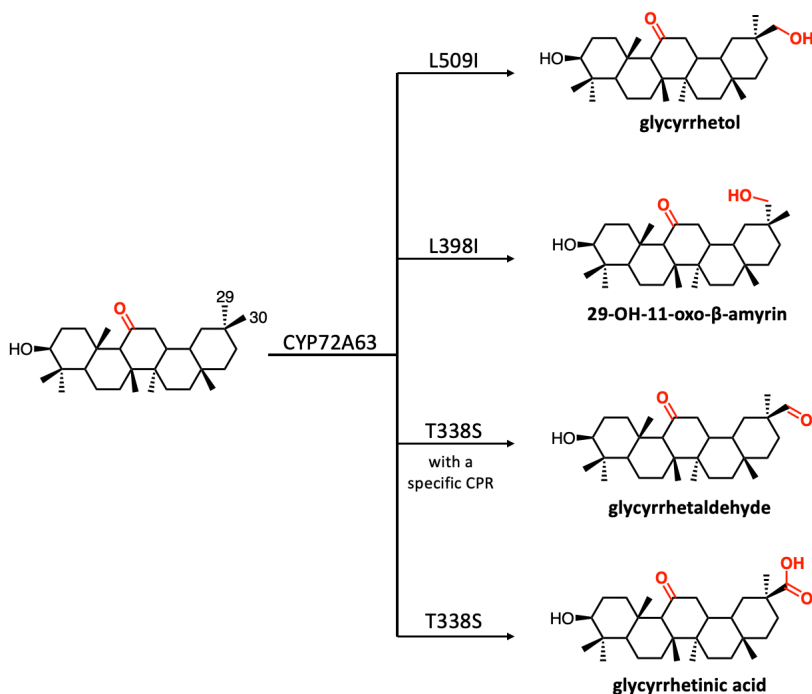


Figure 1.6. P450 involved in rare triterpenoid biosynthesis. Plant P450 CYP72A63 was engineered towards the selective synthesis of licorice triterpenoid natural products.

1.4 Thesis Outline

My doctoral thesis has focused on the investigation of key tailoring enzymes involved in the biosynthesis of pharmaceutically relevant NPs, and their producing microbial organisms. Most of my efforts were dedicated to studying the bacterial cytochrome P450 enzyme TamI involved in tirandamycin biosynthesis. Using protein and substrate engineering strategies, I manipulated the enzyme to alter its native oxidative reactivity and selectivity, expand its substrate scope, access new bioactive tirandamycin analogues, and develop sustainable biocatalytic technologies for challenging oxidation reactions. Structural and biochemical data led us to hypothesize that P450 TamI could be manipulated to evolve unique selectivities, enhance iterative capabilities and generate novel compounds that could not be accessed using current synthetic methods. I also pursued a follow-up study to identify the active oxidant species in P450 TamI responsible for its multi-step catalytic nature involving mechanistically distinct oxidative pathways. This work involved in depth kinetic and computational characterization, leading to the discovery that iterative

P450 TamI uses multiple oxidizing species to produce a unique tri-oxidized decorative pattern in tirandamycin. With detailed mechanistic and structural elucidation, my research work has built a framework for future engineering efforts of P450 enzymes as powerful biocatalysts for late-stage oxidations.

Lastly, I was grateful to receive a competitive research grant from National Geographic Society to build upon and continue a long-standing collaboration with Universidad Nacional Agraria La Molina in Lima Perú, a partnership that was initiated by my mentor Dr. David H. Sherman over a decade ago. I collaborated with a Peruvian graduate student in the Laboratory of Mycology and Biotechnology, Felipe Huanachín Carahuano, to perform the first large-scale microbial biodiversity assessment of a unique and fragile extreme natural environment within the Peruvian Amazon Rainforest. We also collaborated with geologist and director of the Boiling River Project,⁷¹ Andrés Ruzo, to measure physico-chemical parameters, inform our studies with environmentally relevant geological hypotheses and safely and respectfully access the area of study for sample collection. Additionally, the Life Sciences Institute at University of Michigan, Dr. Sherman and I worked closely with the Peruvian Government to obtain all collection and study permits complying with international ethical standards of research (further described in Chapter 5). Collaboratively, we described the microbial structure and composition of the Peruvian Boiling River, and how the unique microclimates found along the course of the river (including elevated temperatures and sulfur-rich geochemical features) influence the local microdiversity. With a geogenome mining approach, this work led us to identify key environmental samples that may harbor microbial species new to science with previously unknown gene clusters that may produce structurally intriguing secondary metabolites. This project builds a foundation for future microbial biodiversity monitoring of Amazonian ecosystems for targeted NP drug discovery research and conservation efforts.

1.5 References

1. Weyrich, L. S.; Duchene, S.; Soubrier, J.; Arriola, L.; Llamas, B.; Breen, J.; Morris, A. G.; Alt, K. W.; Caramelli, D.; Dresely, V.; Farrell, M.; Farrer, A. G.; Francken, M.; Gully, N.; Haak, W.; Hardy, K.; Harvati, K.; Held, P.; Holmes, E. C.; Kaidonis, J.; Lalueza-Fox, C.; de la Rasilla, M.; Rosas, A.; Semal, P.; Soltysiak, A.; Townsend, G.; Usai, D.; Wahl, J.; Huson, D. H.; Dobney, K.; Cooper, A. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **2017**, *544*, 357-361.
2. Krings, M.; Stone, A.; Schmitz, R. W.; Krainitzki, H.; Stoneking, M.; Paabo, S. Neandertal DNA sequences and the origin of modern humans. *Cell* **1997**, *90*, 19-30.

3. Shen, B. A new golden age of natural products drug discovery. *Cell* **2015**, 163, 1297-1300.
4. Patridge, E., Gareiss, P., Kinch, M.S. and Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov* **2016**, 21(2), 204-207.
5. Dias, D. A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012**, 2, 303-336.
6. Schmidt, B. M.; Ribnicky, D. M.; Lipsky, P. E.; Raskin, I. Revisiting the ancient concept of botanical therapeutics. *Nat Chem Biol* **2007**, 3, 360-366.
7. Renfrey, S.; Downton, C.; Featherstone, J. The painful reality. *Nat Rev Drug Discov* **2003**, 2, 175-6.
8. Newman, D. J. a. C., G.M. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* **2016**, 79(3), 629-661.
9. McChesney, J. D. Natural products in drug discovery--organizing for success. *P R Health Sci J* **2002**, 21, 91-95.
10. Abdel-Razek, A. S., El-Naggar, M.E., Allam, A., Morsy, O.M. and Othman, S.I. Microbial natural products in drug discovery. *Processes* **2020**, 8(4), 470.
11. Bansal, H.; Singla, R. K.; Behzad, S.; Chopra, H.; Grewal, A. S.; Shen, B. Unleashing the potential of microbial natural products in drug discovery: focusing on *Streptomyces* as antimicrobials goldmine. *Curr Top Med Chem* **2021**.
12. Stewart, E. J. Growing unculturable bacteria. *J Bacteriol* **2012**, 194, 4151-4160.
13. Schloss, P. D.; Handelsman, J. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol* **2005**, 6, 229.
14. Rutledge, P. J.; Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* **2015**, 13, 509-523.
15. Thirumurugan, D., Cholarajan, A., Raja, S.S. and Vijayakumar, R. An introductory chapter: secondary metabolites. *Second metab—sources Appl* **2018**, 1-21.
16. Maplestone, R. A.; Stone, M. J.; Williams, D. H. The evolutionary role of secondary metabolites—a review. *Gene* **1992**, 115, 151-157.
17. Carlson, J. C.; Fortman, J. L.; Anzai, Y.; Li, S.; Burr, D. A.; Sherman, D. H. Identification of the tirandamycin biosynthetic gene cluster from *Streptomyces* sp. 307-9. *ChemBiochem* **2010**, 11, 564-572.
18. Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H. Tirandamycin biosynthesis is mediated by co-dependent oxidative enzymes. *Nat Chem* **2011**, 3, 628-33.
19. Hartwig, J. F.; Larsen, M. A. Undirected, homogeneous C-H bond functionalization: challenges and opportunities. *ACS Cent Sci* **2016**, 2, 281-292.
20. Lyons, T. W.; Sanford, M. S. Palladium-catalyzed ligand-directed C-H functionalization reactions. *Chem Rev* **2010**, 110, 1147-1169.
21. Gomez, L.; Canta, M.; Font, D.; Prat, I.; Ribas, X.; Costas, M. Regioselective oxidation of nonactivated alkyl C-H groups using highly structured non-heme iron catalysts. *J Org Chem* **2013**, 78, 1421-1433.
22. Gormisky, P. E.; White, M. C. Catalyst-controlled aliphatic C-H oxidations with a predictive model for site-selectivity. *J Am Chem Soc* **2013**, 135, 14052-14055.
23. Ochiai, H.; Uetake, Y.; Niwa, T.; Hosoya, T. Rhodium-catalyzed decarbonylative borylation of aromatic thioesters for facile diversification of aromatic carboxylic acids. *Angew Chem Int Ed Engl* **2017**, 56, 2482-2486.
24. Tao, J.; Xu, J. H. Biocatalysis in development of green pharmaceutical processes. *Curr Opin Chem Biol* **2009**, 13, 43-50.
25. Whitehouse, C. J.; Bell, S. G.; Wong, L. L. P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev* **2012**, 41, 1218-1260.
26. Groves, J. T. Models and mechanisms of cytochrome P450 action. *Cytochrome P450* **2005**, 1-43.
27. R, B. Biomimetic control of chemical selectivity. *Acc Chem Res* **1980**, 13 (6), 170-177.
28. Liang, J., Lalonde, J., Borup, B., Mitchell, V., Mundorff, E., Trinh, N., Kochrekar, D.A., Nair Cherat, R. and Pai, G.G. Development of a biocatalytic process as an alternative to the (–)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Org Process Res Dev* **2010**, 14(1), 193-198.
29. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, 329, 305-309.
30. Loughlin, W. A. Biotransformations in organic synthesis. *Bioresourc Technol* **2000**, 74(1), 49-62.
31. O'Reilly, E., Köhler, V., Flitsch, S.L. and Turner, N.J. Cytochromes P450 as useful biocatalysts: addressing the limitations. *ChemComm* **2011**, 47(9), 2490-2501.

32. Kumar, S. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert Opin Drug Metab Toxicol* **2010**, 6, 115-131.
33. Urlacher, V. B.; Lutz-Wahl, S.; Schmid, R. D. Microbial P450 enzymes in biotechnology. *Appl Microbiol Biotechnol* **2004**, 64, 317-325.
34. O'Reilly, E., Mark Corbett, Shahed Hussain, Paul P. Kelly, Dominique Richardson, Sabine L. Flitsch, and Nicholas J. Turner. Substrate promiscuity of cytochrome P450 RhF. *Catal Sci Technol* **2013**, 3(6), 1490-1492.
35. Cook, D. J.; Finnigan, J. D.; Cook, K.; Black, G. W.; Charnock, S. J. Cytochromes P450: history, classes, catalytic mechanism, and industrial application. *Adv Protein Chem Struct Biol* **2016**, 105, 105-126.
36. Urlacher, V. B.; Girhard, M. Cytochrome P450 monooxygenases in biotechnology and synthetic biology. *Trends Biotechnol* **2019**, 37, 882-897.
37. Nelson, D. R. Cytochrome P450 nomenclature, 2004. *Methods Mol Biol* **2006**, 320, 1-10.
38. Werck-Reichhart, D.; Hehn, A.; Didierjean, L. Cytochromes P450 for engineering herbicide tolerance. *Trends Plant Sci* **2000**, 5, 116-123.
39. Yuan, R., Madani, S., Wei, X.X., Reynolds, K. and Huang, S.M. Evaluation of cytochrome P450 probe substrates commonly used by the pharmaceutical industry to study *in vitro* drug interactions. *Drug Metab Dispos* **2002**, 30(12), 1311-1319.
40. Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. The catalytic pathway of cytochrome P450cam at atomic resolution. *Science* **2000**, 287, 1615-1622.
41. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and chemistry of cytochrome P450. *Chem Rev* **2005**, 105, 2253-2277.
42. Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R. Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta* **2007**, 1770, 330-344.
43. Lehnerer, M.; Schulze, J.; Achterhold, K.; Lewis, D. F.; Hlavica, P. Identification of key residues in rabbit liver microsomal cytochrome P450 2B4: importance in interactions with NADPH-cytochrome P450 reductase. *J Biochem* **2000**, 127, 163-169.
44. Bohme, H. On the role of ferredoxin and ferredoxin-NADP⁺ reductase in cyclic electron transport of spinach chloroplasts. *Eur J Biochem* **1977**, 72, 283-289.
45. Meunier, B. a. B., J. Active iron-oxo and iron-peroxo species in cytochromes P450 and peroxidases; oxo-hydroxo tautomerism with water-soluble metalloporphyrins. *Metal-Oxo and Metal-Peroxo Species in Catalytic Oxidations* **2000**, 1-35.
46. Nelson, D.; Werck-Reichhart, D. A P450-centric view of plant evolution. *Plant J* **2011**, 66, 194-211.
47. Cresnar, B.; Petric, S. Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta* **2011**, 1814, 29-35.
48. Durairaj, P.; Hur, J. S.; Yun, H. Versatile biocatalysis of fungal cytochrome P450 monooxygenases. *Microb Cell Fact* **2016**, 15, 125.
49. Kim, Y. H.; Kwon, T.; Yang, H. J.; Kim, W.; Youn, H.; Lee, J. Y.; Youn, B. Gene engineering, purification, crystallization and preliminary X-ray diffraction of cytochrome P450 p-coumarate-3-hydroxylase (C3H), the Arabidopsis membrane protein. *Protein Expr Purif* **2011**, 79, 149-155.
50. Kudo, F.; Motegi, A.; Mizoue, K.; Eguchi, T. Cloning and characterization of the biosynthetic gene cluster of 16-membered macrolide antibiotic FD-891: involvement of a dual functional cytochrome P450 monooxygenase catalyzing epoxidation and hydroxylation. *Chembiochem* **2010**, 11, 1574-1582.
51. Yang, S.; DeMars, M. D., 2nd; Grandner, J. M.; Olson, N. M.; Anzai, Y.; Sherman, D. H.; Houk, K. N. Computational-based mechanistic study and engineering of cytochrome P450 MycG for selective oxidation of 16-membered macrolide antibiotics. *J Am Chem Soc* **2020**, 142, 17981-17988.
52. Farinas, E. T., Schwaneberg, U., Glieder, A. and Arnold, F.H. Directed evolution of a cytochrome P450 monooxygenase for alkane oxidation. *Adv Synth Catal* **2001**, 343(6-7), 601-606.
53. Fasan, R. Tuning P450 enzymes as oxidation catalysts. *ACS Cat* **2012**, 2(4), 647-666.
54. Graham-Lorence, S., Truan, G., Peterson, J.A., Falck, J.R., Wei, S., Helvig, C. and Capdevila, J.H. An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S, 15R)-arachidonic acid epoxygenase. *J Biol Chem* **1997**, 272(2), 1127-1135.
55. Meinhold, P., Peters, M.W., Hartwick, A., Hernandez, A.R. and Arnold, F.H. Engineering cytochrome P450 BM3 for terminal alkane hydroxylation. *Adv Synth Catal* **2006**, 348(6), 763-772.
56. Narayan, A. R.; Jimenez-Oses, G.; Liu, P.; Negretti, S.; Zhao, W.; Gilbert, M. M.; Ramabhadran, R. O.; Yang, Y. F.; Furan, L. R.; Li, Z.; Podust, L. M.; Montgomery, J.; Houk, K. N.; Sherman, D. H. Enzymatic

- hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. *Nat Chem* **2015**, *7*, 653-660.
57. Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J Biol Chem* **2006**, *281*, 26289-26297.
58. Li, Q. S.; Schwaneberg, U.; Fischer, M.; Schmitt, J.; Pleiss, J.; Lutz-Wahl, S.; Schmid, R. D. Rational evolution of a medium chain-specific cytochrome P-450 BM-3 variant. *Biochim Biophys Acta* **2001**, *1545*, 114-21.
59. Prier, C. K.; Arnold, F. H. Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. *J Am Chem Soc* **2015**, *137*, 13992-14006.
60. Lewis, J. C., Bastian, S., Bennett, C.S., Fu, Y., Mitsuda, Y., Chen, M.M., Greenberg, W.A., Wong, C.H. and Arnold, F.H. Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450BM3 demethylases. *PNAS* **2009**, *106*(39), 16550-16555.
61. Shang, Y.; Huang, S. Engineering plant cytochrome P450s for enhanced synthesis of natural products: Past achievements and future perspectives. *Plant Commun* **2020**, *1*, 100012.
62. Renault, H.; Bassard, J. E.; Hamberger, B.; Werck-Reichhart, D. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Curr Opin Plant Biol* **2014**, *19*, 27-34.
63. Guengerich, F. P. Mechanisms of cytochrome P450-catalyzed oxidations. *ACS Catal* **2018**, *8*, 10964-10976.
64. Fujiyama, K., Hino, T., Kanadani, M., Watanabe, B., Lee, H.J., Mizutani, M. and Nagano, S. Structural insights into a key step of brassinosteroid biosynthesis and its inhibition. *Nat Plants* **2019**, *5*(6), 589-594.
65. Hausjell, J.; Halbwirth, H.; Spadiut, O. Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Biosci Rep* **2018**, *38*.
66. Sun, W., Xue, H., Liu, H., Lv, B., Yu, Y., Wang, Y., Huang, M. and Li, C. Controlling chemo- and regioselectivity of a plant P450 in yeast cell toward rare licorice triterpenoid biosynthesis. *ACS Catal* **2020**, *10*(7), 4253-4260.
67. Chen, K. a. A., F.H. Engineering new catalytic activities in enzymes. *Nat Catal* **2020**, *3*(3), 203-213.
68. Mellor, S. B.; Nielsen, A. Z.; Burow, M.; Motawia, M. S.; Jakubauskas, D.; Moller, B. L.; Jensen, P. E. Fusion of ferredoxin and cytochrome P450 enables direct light-driven biosynthesis. *ACS Chem Biol* **2016**, *11*, 1862-1869.
69. Dang, T. T. T., Franke, J., Carqueijeiro, I.S.T., Langley, C., Courdavault, V. and O'Connor, S.E. Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat Chem Biol* **2018**, *14*(8), 760-763.
70. Fanani, M. Z.; Fukushima, E. O.; Sawai, S.; Tang, J.; Ishimori, M.; Sudo, H.; Ohyama, K.; Seki, H.; Saito, K.; Muranaka, T. Molecular basis of C-30 product regioselectivity of legume oxidases involved in high-value triterpenoid biosynthesis. *Front Plant Sci* **2019**, *10*, 1520.
71. Ruzo, A. The Boiling River of the Amazon: adventure and discovery in the Amazon. *TED Books, Simon & Schuster Inc. New York* **2016**.

Chapter 2*

Investigating and Engineering P450 TamI as an Iterative Biocatalyst for Late-Stage C-H Oxidation of Complex Natural Products

2.1 Introduction to the Tirandamycin Family of Tetramic Acid Antibiotics

2.1.1 Discovery of Tirandamycin Natural Products

Tirandamycin comprises a small family of structurally intriguing dienoyltetramic acid-containing natural products¹ and is structurally similar to the antibiotic streptolydigin. Originally discovered in 1971,² a total of 15 tirandamycin congeners have been reported from various sediment and terrestrial derived-bacterial species including the marine isolate *Streptomyces* sp. 307-9 (Figure 2.1). Most tirandamycin molecules differ in the extent of oxidative decorations in the bicyclic ketal moiety with a few congeners harboring an O-linked sugar motif. We previously characterized new intermediates tirandamycin C (**1**), E (**2**) and D (**3**), in search of lead molecules against the vancomycin resistant *Enterococcus faecalis* (VRE) and determined that the heavily oxygenated bicyclic ketal moiety of tirandamycin as seen in tirandamycin A (**4**) and B (**5**) is key to bioactivity.³

2.1.2 Biological Applications and Mode of Action of Tirandamycin

Biological properties reported for tirandamycin metabolites include potent *in vitro* activity against lymphatic filariasis-causing *Brugia malayi* adult parasites,⁴ anti-VRE properties,⁵ inhibition of the futasoline pathway (an operative pathway in *Helicobacter pylori*)⁵ and specific anticancer effects in a *Drosophila* tumor model by inhibition of asparaginyl-tRNA synthetase.⁶ The mode of action of tirandamycin mimics that of streptolydigin. Similar to the latter, previous

* Chapter 2 is included in two manuscripts. 1) Newmister, S.A., Srivastava, K.R., **Espinoza, R.V.**, Caddell Haatveit, K., Khatri, Y., Martini, R.M., Garcia-Borràs, M., Podust, L.M., Houk, K.N. and Sherman, D.H. Molecular basis of iterative C–H oxidation by TamI, a multifunctional P450 monooxygenase from the tirandamycin biosynthetic pathway. *ACS Cat.*, **2020**, *10*(22), 13445-13454.; 2) **Espinoza, R. V.**, Haatveit, K.C., Grossman, S.W., Tan, J.Y., McGlade, C.A., Khatri, Y., Newmister, S.A., Schmidt, J.J., Garcia-Borràs, M., Montgomery, J., Houk, K.N. and Sherman, D. H. Engineering P450 TamI as an Iterative Biocatalyst for Selective Late-Stage C–H Functionalization and Epoxidation of Tirandamycin Antibiotics. *ACS Cat.*, **2021**, *11*, 8304-8316.

work demonstrated that tirandamycin inhibits *de novo* bacterial transcription by disrupting function of the RNA polymerase enzyme.⁷ This RNA polymerase inhibitory nature is highly specific to bacterial systems as tirandamycin does not interrupt the function of mammalian RNA polymerase. It was also shown that in addition to interrupting chain initiation, tirandamycin inhibits RNA chain elongation without interrupting the formation of stable template DNA-RNA polymerase complexes. Further studies with synthetic analogues indicated that the tetramic acid moiety of tirandamycin is a critical structural requirement for the proper recognition and interaction with RNA polymerase.

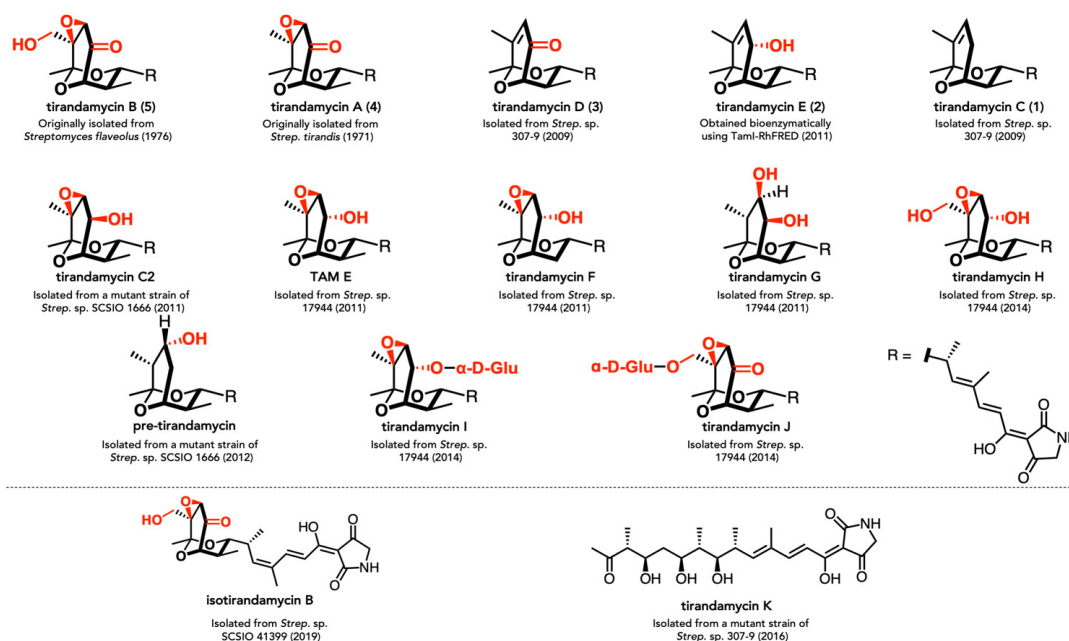
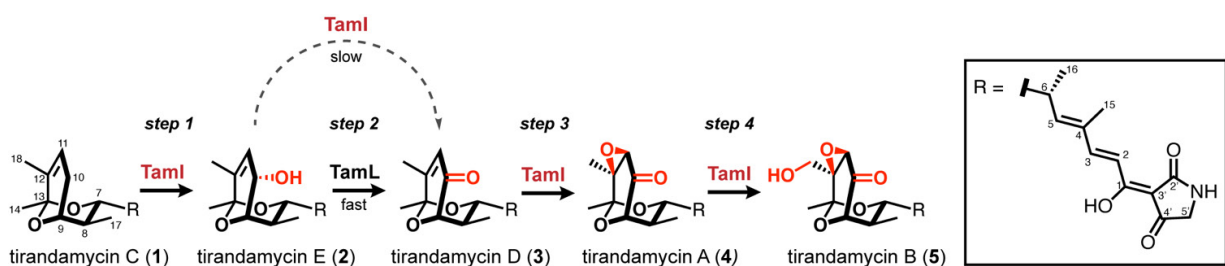


Figure 2.1. Tirandamycin congeners reported to date.

2.1.3 Elucidation of the Tirandamycin Biosynthetic Pathway

Tirandamycin natural products are formed through a hybrid polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) system with a colinear domain organization.¹ The 2,4-pyrrolidinedione ring system of tirandamycin is derived from condensation of an amino acid to a polyketide-derived acyl chain to generate the dienoyltetramic acid motif.³ Gene knock-out experiments revealed the presence of two tailoring enzymes that are responsible for the sequential late-stage installation of oxygen functionalities during tirandamycin biosynthesis: the multifunctional cytochrome P450 TamI and the flavin oxidase TamL (Scheme 2.1). First, TamI

abstracts the C10-(*S*) hydrogen of **1** to install an allylic hydroxyl (*step 1*) generating **2**, which undergoes oxidation by TamL (*step 2*) to yield **3**. Next, TamI catalyzes the formation of a C11/C12 (*R/S*) epoxide (*step 3*) producing **4**, followed by methyl hydroxylation at C18 (*step 4*) generating the terminal product **5**. TamI has been shown to oxidize **1** → **2** → **3** → **4** at low levels without the need for TamL, suggesting the possibility of a more complex iterative mechanism in some instances. Based on gene deletion studies and isolation of biosynthetic intermediates, the unusual iterative cascade catalyzed by TamI occurs in a strict sequence, with each oxidation being a prerequisite for the next.^{3, 8}



Scheme 2.1. Iterative late-stage oxidation reactions catalyzed by P450 TamI. In *step 1*, TamI catalyzes an allylic hydroxylation at C10 of **1**. In *step 2*, the flavoprotein TamL oxidizes the C10 hydroxyl of **2** to a ketone. In *step 3*, TamI performs an alkene epoxidation at C11/12 of **3**. In *step 4*, TamI installs a primary hydroxyl at C18 of **4** yielding the terminal product **5**.

2.2 Molecular and Structural Basis of Iterative C-H Oxidation by TamI

2.2.1 Background

Although earlier enzymatic and kinetic experiments informed the efficiency and order of oxidative steps catalyzed by P450 TamI,³ the structural and molecular basis for the iterative nature of the enzyme remained unknown. This gap in knowledge motivated us to elucidate the substrate-binding mechanism of TamI and the inherent reactivity of tirandamycin substrates using structural biology and computational methods. The first hypothesis we decided to examine was whether TamI uses a tetramic acid based-binding mechanism to anchor substrate **1** in the correct geometry for the consecutive oxidation reactions to take place. A discrete substrate anchoring mechanism has been previously observed in the TamI homologue PikC (47% sequence identity), where the dimethylamino sugar desosamine allows for productive substrate recognition enabling

hydroxylation at a distal carbon site of the substrate. This anchoring mechanism inspired the synthesis of suitable dimethylamino sugar desosamine “anchor” analogues that enable the oxidation of non-native cyclic scaffolds with P450 PikC.⁹ Understanding the basis for substrate binding in TamI could provide avenues for substrate engineering efforts to achieve oxidative diversification of non-natural bicyclic-containing scaffolds and expand the synthetic utility of the enzyme.

2.2.2 Structure of P450 TamI in Complex with Tirandamycin C and Identification of Key Residues

Using molecular replacement and the substrate-free structure as a model, Sean Newmister in the Sherman group at the University of Michigan successfully solved the structure of TamI bound to **1** at 2.7 Å resolution (Figure 2.2).⁸ Analysis of the structure showed that the C10 atom of **1** is at a 4.1 Å distance from the heme iron of TamI, a reasonable distance for the first oxidation reaction to occur (*step 1*). Sean observed various interactions between the tetramic acid moiety of **1** and distinct loop regions of the protein including direct polar contact with residues Pro43 (β 1-1/2 loop), Ser397, Thr398 and Leu399 (β 3-3/2 loop) and hydrophobic interactions with Val44 (β 1-1/2 loop) and Phe92 (BC loop) (Figure 2.3A). Moreover, a hydrophobic pocket adjacent to the heme center where the bicyclic ketal moiety and conjugated polyene chain sit was identified. Specifically, the polyene was found to interact with the side chains of Leu399 and Ile400 (β 3-3/2 loop), Val185 (FG region) and Phe92 (BC loop) constricting the binding site in this region. The bicyclic ketal moiety was surrounded by the hydrophobic side chains of Leu101, Leu244 and Leu295 as well as other residues including His102, Gly248, Thr252 and Thr299 (Figure 2.3B). We hypothesized that the various polar contacts with the tetramic acid moiety of **1** may be consistent with a tetramic acid-based anchoring mechanism that enables iterative oxidation at distinct carbon sites of the bicyclic ketal moiety.

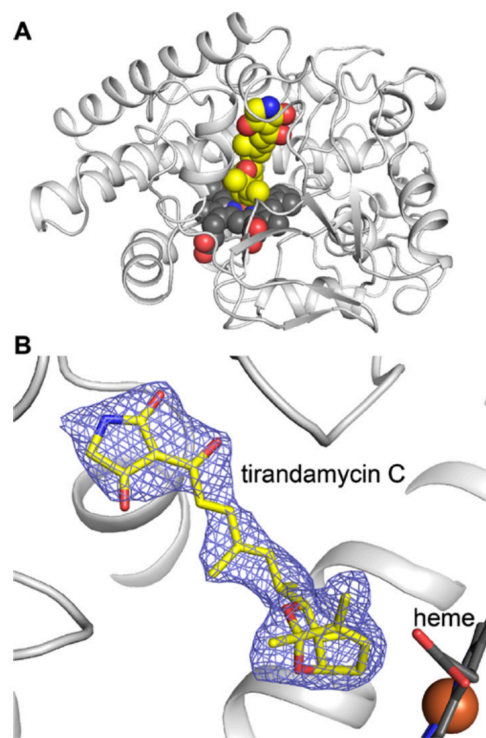


Figure 2.2. Structure of P450 TamI in complex with tirandamycin C (1**).** (A) Cartoon representation of bound structure where **1** is shown in yellow and heme is shown in black. (B) Omit map (Fo-Fc) for **1**. The bicyclic ketal moiety of **1** is close in proximity to the heme iron.

2.2.3 Mutagenesis and Biochemical Characterization

2.2.3.1 Probing a Tetramic Acid-Based Anchoring Mechanism

Guided by the enzyme structure, Kinshuk Srivastava in the Sherman group at the University of Michigan performed site-directed mutagenesis of active site residues in P450 TamI to probe their role in binding, catalysis, and selectivity. First, key polar active site residues interacting with the tetramic acid moiety of **1** were targeted to probe their role in substrate binding and interrogate if these residues could act as a substrate anchor analogous to the desosamine sugar moiety in narbomycin and YC-17 substrates in the PikC system.^{9, 10} Single (S397A, T398A, and V185A) and double (P43A_V44A) TamI variants were prepared and enzymatic assays were performed using optimized conditions in a three-component system comprising TamI, spinach ferredoxin, and ferredoxin reductase. Contrary to our hypothesis, the mutations introduced did not abolish enzymatic activity or alter the product profile compared to TamI WT. Most TamI variants showed similar conversion to WT when tested with substrate **1** (Figure 2.3) despite of showing minor perturbations in substrate binding affinities (Table 2.1). These experiments revealed that the polar interactions between the different loop regions of the enzyme and the tetramic acid moiety of **1** have no direct impact on the allylic C10 hydroxylation of **1** to **2**, discarding a tetramic acid-based substrate anchoring mode.

Mutant	K_d	R^2
WT	0.17 ± 0.02	0.98
F92A	n.s.*	--
F92A/L101V/H102S	n.s.	--
L101V/H102S	n.s.	--
S397A	1.57 ± 0.11	0.98
T398A	4.50 ± 0.08	0.99
S397A/T398A	4.67 ± 0.23	0.99
L399A/I400A	n.s.	--
S397A/T398A/L399A/I400A	n.s.	--
I400A	n.s.	--
V185A	2.81 ± 0.15	0.99
P43A/V44A	39.3 ± 3.6	0.98

Table 2.1. Tirandamycin C (1**) equilibrium dissociation constant (K_d) values.** “n.s.” denotes no spectral shift observed.

2.2.3.2 Probing a Hydrophobic Pocket in the Active Site of TamI

We decided to explore alternative interactions between TamI’s active site and substrate **1** and analyze the hydrophobic and aromatic residues lining the conjugated polyene moiety of **1** including amino acids Leu399, Ile400, and Phe92. Double, triple, and quadruple mutants were prepared and tested in end-point assays with **1**. The variants L399A_I400A and L101V_H102S showed decreased conversion to product compared to WT (Figure 2.3). Interestingly, the alanine mutant of Phe92 led to a drastic loss of catalytic activity of TamI, forming **2** as a trace product

only (Figure 2.3B). These mutants did not show any spin shift with substrate **1** indicating substrate binding has been disturbed (Table 2.1).

We performed Molecular Dynamics (MD) simulations of TamI WT with **1** and the variants F92A, L399A and I400A to analyze the effect of these single point mutants on substrate orientation within the enzyme's active site. We found that substrate **1** occupies multiple different positions in the active sites of the single point mutants compared to WT where it was the lowest root-mean-square deviation value. These results suggest that the residues mutated are essential for dictating the correct substrate geometry for TamI catalysis (Figures 2.4). Moreover, we also noted that the bicyclic ketal moiety remains close to the heme center during the entire length of the MD simulations, while the polar contacts observed with the tetramic acid tail in the crystal structure rapidly relax within the first few nanoseconds of the simulation (See Appendix A for details). These results demonstrate that the hydrophobic interactions around the polyene region of the substrate are most essential for productive substrate binding and C10 hydroxylation catalysis in vP450 TamI.

Moreover, amino acids surrounding the bicyclic ketal of **1** were targeted for mutagenesis to probe their role in substrate binding and selectivity. Specifically, residues Leu101, His102, Leu295, and Thr299 were mutated and tested in end-point assays with **1**. Double variant L101V_H102S led to decreased product conversion (Figure 2.3B) and no change in product profile suggesting that increasing the active site volume does not impact site selectivity for *step 1* but it does alter substrate binding. We hypothesize that residues Leu101 and His102 may be critical for productive substrate binding and positioning for the C10 hydroxylation step to occur. Interactions between these residues and the C14 methyl group of the substrate were observed in the crystal structure where the C δ atom of Leu101 is at ~ 3.4 Å and N ϵ of His102 is at ~ 4 Å in relation to the oxygen atoms of the bicyclic ketal moiety in **1**. The double mutant L295A_T299A showed WT activity indicating these residues do not impact TamI catalysis. Overall, these mutagenesis studies showed that the conjugated polyene chain in tirandamycin and active site residues Leu399, Ile400 and Phe92 are essential structural and molecular features for productive substrate binding and orientation in TamI oxidative catalysis.

2.2.4 Computational Investigations of Iterative Oxidations

2.2.4.1 Density Functional Theory on Intrinsic Reactivity

In collaboration with the Houk Lab at UCLA, we performed density functional theory (DFT) and MD simulations to elucidate the innate reactivity of tirandamycin substrates for the competing stepwise oxidation reactions catalyzed by P450 TamI and understand how the enzyme controls the selectivity observed. The reactive Fe(IV)-oxo radical cation and a truncated model of substrates **1**, **3**, and **4** containing the bicyclic ketal moiety were used for DFT calculations to compare the energies of the transition state barriers for *step 1* (C10 hydroxylation), *step 3* (C11/12 epoxidation) and *step 4* (C18 hydroxylation) (Figure 2.4). The C-H abstraction barrier for C10 hydroxylation is lowest in energy for substrate **1** with a value of 14.5 kcal/mol (Figures 2.4A) matching the experimentally observed regioselectivity. No difference was observed for abstracting the *R* or *S* hydrogen at the C10 site suggesting that the TamI active site is critical for driving the selectivity between these diastereomeric transition states. The transition state barrier for driving the selectivity between these diastereomeric transition states. The transition state barrier for the (*R/S*) epoxidation at C11/12 double bond of **3** is lowest in energy by 1.9 kcal/mol, matching the chemo- and stereoselectivity observed (Figures 2.4A), while the energetic barrier for the C18

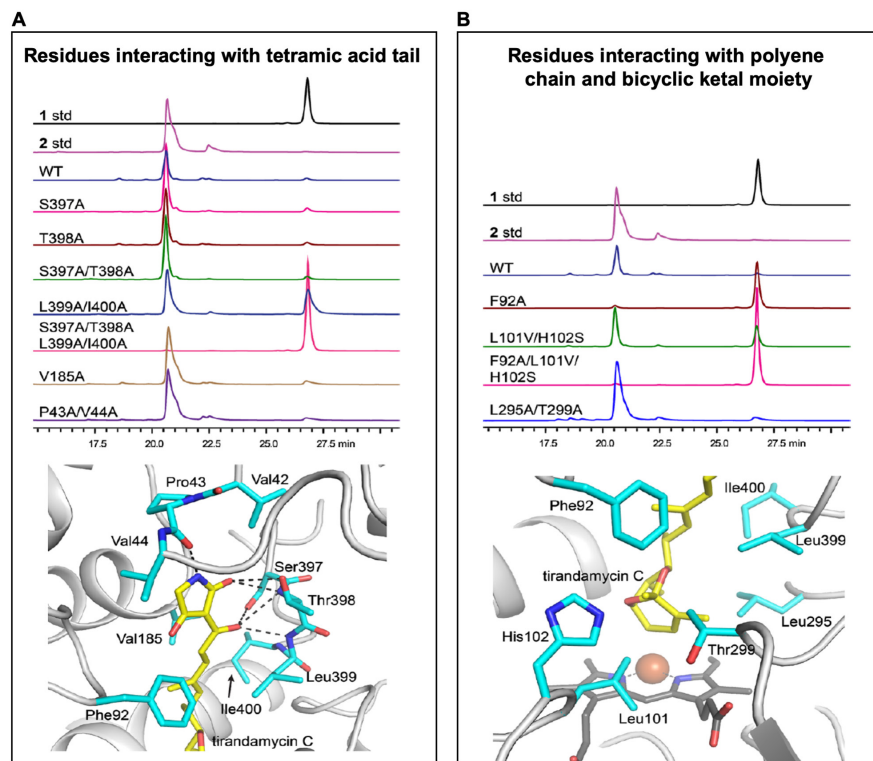


Figure 2.3. Mutagenesis analysis with P450 TamI (1**).** (A) Key interactions with the tetramic acid moiety of **1** are displayed in a stick representation. HPLC traces for enzymatic assays with TamI mutants probing the substrate binding mechanism were extracted at 340nm, signature UV of tirandamycin. (B) Key interactions with the conjugated polyene and bicyclic ketal moiety of **1** are shown in a stick representation. HPLC traces for endpoint assays with TamI mutants probing a hydrophobic pocket mediates binding were extracted at 340nm, signature UV of tirandamycin.

hydroxylation on **4** is 17.5 kcal/mol. The trend observed in the calculated transition state energy barriers with substrates **1**, **3** and **4** is analogous to the trend in catalytic efficiency reported for the TamI-RhFRED chimeric fusion enzyme, where k_{cat} lowers from 83.8 min⁻¹ to 40.5 min⁻¹ to 0.11 min⁻¹ for *step 1*, *step 3* and *step 4*, respectively.

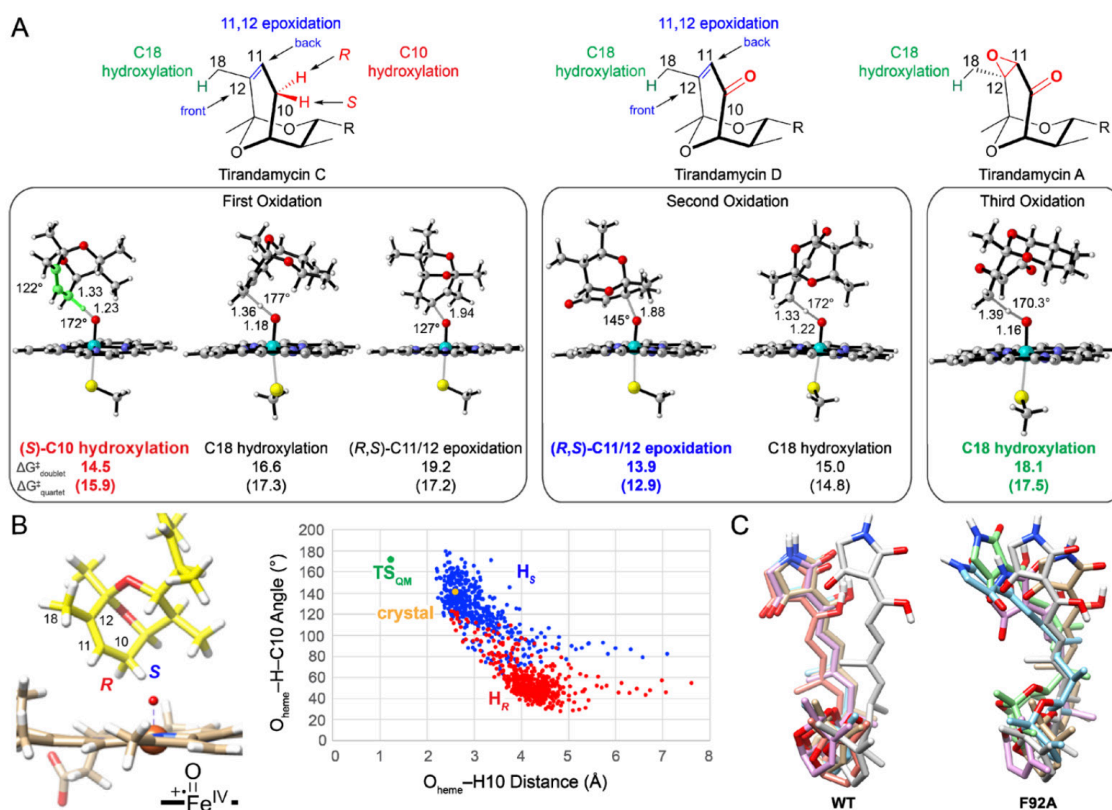


Figure 2.4. Computational analysis of TamI iterative oxidation (1). (A) Inherent reactivity for each competing oxidation step in tirandamycin biosynthesis. (B) MD simulations show P450 TamI interacting with **1**. The pro-*S* hydrogen remains closer to the QM ideal transition state (green dot) geometry in the crystal structure (orange dot) and throughout the entire simulation. (C) Overlays of top five occupied clusters for substrate **1** with TamI WT and mutant F92A.

2.2.4.2 Molecular Dynamics with Tirandamycin Congeners

Next, we assessed the influence of TamI in the regio- and stereoselectivity observed for the C10 hydroxylation reaction. MD simulations were performed using the substrate-bound crystal structure of TamI and **1**. The C10 and C11 atoms remained the closest to the iron-oxo species throughout the 500 ns simulation, consistent with C10 hydroxylation. The $\text{O}_{\text{heme}}-\text{C10}$ hydrogen distance and $\text{O}_{\text{heme}}-\text{C10}$ hydrogen-C10 angle geometries for both the *S* and *R* hydrogens from the

MD simulations were compared to the ideal QM calculated transition state to assess how TamI influences the substrate geometry. The *S* hydrogen was found closer to the transition state geometry for the entire simulation, indicating that TamI controls the orientation of substrate **1** in the active site and drives the stereoselectivity for C-H oxidation at C10 (Figure 2.4B). Similar MD simulation experiments were performed with substrates **3** and **4**, revealing that the orientation of tirandamycin changes as the oxidation cascade progresses, with each substrate occupying the correct geometry for the observed reactivity in TamI's active site.

2.2.5 Discussion

Following a structural and computational approach, we elucidated 1) the structure of P450 TamI WT bound to substrate **1**; 2) the molecular and structural basis for substrate binding and C10 hydroxylation catalysis and 3) the innate reactivity of **1** for the sequential oxidation pattern catalyzed by the enzyme. Close examination of the bound crystal structure and experimental mutagenesis analysis revealed a hydrophobic pocket in the active site and key residues critical for productive substrate binding and catalysis, discarding a tetramic acid-based anchoring mechanism. Specifically, active site amino acids Phe92, Leu399, and Ile400 seemed to mediate essential steric and hydrophobic interactions with the conjugated polyene chain of **1**. MD simulations showed that substrate **1** adopts multiple binding orientations losing the correct geometry for catalysis when these three key residues are mutated to alanine and the hydrophobic π - π interactions are lost. Moreover, QM calculations were performed to reveal the innate reactivity of substrate **1** for the competing oxidation reactions catalyzed by P450 TamI. A clear preference for secondary allylic C10 hydroxylation of **1** over C11/12 epoxidation and primary C18 hydroxylation is observed. However, no energetic preference for the *S* or *R* hydrogen abstraction at C10 is detected, suggesting the enzyme drives the exquisite stereoselectivity observed *in vitro* for the formation of the (*S*) hydroxy. This observation was further validated with MD simulations that showed how TamI orients **1** in a specific geometry to maintain the pro-*S* hydrogen closer to the ferryl oxygen for abstraction. Additional QM calculations with substrate **3** and **4** demonstrated that the substrate binding orientations stay consistent with idealized geometries for the experimentally observed order of oxidation steps. Overall, this work provides important insights into the molecular and structural principles guiding substrate binding and regio- and stereoselectivity for the multistep C-H oxidation and epoxidation cascade of TamI. Precisely, we revealed key hydrophobic residues and substrate structure features critical for binding and catalysis in P450 TamI. Our investigation

establishes a foundation for future work focused on engineering TamI for altered selectivity and chemical outcome, and expanding the enzyme's substrate scope to facilitate late-stage C-H functionalization and alkene epoxidation of complex chemical scaffolds.

2.3 Engineering P450 TamI for Selective C-H Functionalization and Epoxidation of Tirandamycin Antibiotics

2.3.1 Introduction and Background

Ubiquitous in all domains of life, cytochrome P450 enzymes represent a versatile platform for the controlled functionalization of specific, nonactivated C-H bonds within complex organic molecules.^{11, 12} These heme-containing biocatalysts introduce late-stage functionality with precise regio-, chemo- and stereoselectivity, expanding the synthetic abilities of chemical oxidation catalysts and reagents for structural diversification of biologically important molecules.^{13, 14} In rare cases for select secondary metabolite pathways, the iterative control of C-H oxidation and epoxidation is observed, typically in plant or fungal systems.¹⁵⁻¹⁸ Understanding the function of iterative P450s is a first step toward gaining the ability to tune these multi-step biocatalytic oxidations through protein engineering, a challenging goal for development of efficient oxidative tailoring strategies.^{19, 20}

Eukaryotic iterative P450s are often bound to the endoplasmic reticulum membrane, and frequently challenging to express as recombinant proteins *in vitro*, which has limited our mechanistic understanding of their reactivity and selectivity.¹⁵ Investigating their soluble bacterial counterparts can address this gap in knowledge and lead to the development of biocatalysts with high synthetic utility. Currently, the number of characterized iterative bacterial P450s is scarce, and most are restricted to catalyzing two oxidation events. Examples include P450s MycG^{21, 22} and Gfsf,²³ in the biosynthesis of the macrolide antibiotics mycinamicin and FD-891, respectively, and the P450 AurH,²⁴ involved in the biosynthesis of the nitroaryl-substituted polyketide aureothin. The cytochrome P450 TamI, from the tirandamycin biosynthetic system, represents a rare example of a bacterial P450 naturally catalyzing three highly selective, successive oxidation reactions at distinct carbon atoms of the substrate (Scheme 2.1).^{8, 25}

Based on our previous work, we envisioned that structure-based engineering of the trifunctional P450 TamI could provide important new insights into the controlling elements of its

oxidative cascade, allowing us to manipulate the enzyme to develop new tools for catalyst-controlled iterative C-H oxidation and epoxidation with altered selectivity and step sequence.

Herein, we report the design and characterization of a toolbox of P450 TamI biocatalysts, generated by point mutations, that control the sequence and site-selectivity of epoxidation and C-H oxidation of tirandamycin molecules, achieving divergent reactivity compared to wild-type (WT) enzyme. Using these tuned catalysts, a series of novel tirandamycin derivatives were generated from a single molecular scaffold and evaluated for activity against a panel of human bacterial pathogens. Utilizing active-site engineering, a TamI variant with remarkably enhanced catalytic abilities was developed, overriding the need for the oxidative partner TamL, and enabling a direct four-step biocatalytic cascade for the construction of **5** from **1**. Furthermore, quantum mechanics (QM) and molecular dynamics (MD) simulations were conducted to examine the basis for TamI mutant selectivity and provided new insights into reaction mechanisms of this catalyst-controlled system.

2.3.2 Design and Engineering of a Toolbox of P450 TamI Biocatalysts

2.3.2.1 Identifying Key Active Site Residues for Mutagenesis

We recently reported the structure of TamI in complex with **1** and found that active-site hydrophobic residues have a significant influence in substrate binding and step sequence reactivity, favoring C10 C-H hydroxylation (Scheme 2.1, *step 1*) as the first catalytic reaction.⁸ Computational experiments suggested that TamI's unique active site geometry is key to discriminating between the diastereomeric transition states leading to **2**. However, preliminary mutagenesis efforts failed to reveal how TamI modulates substrate orientation to favor the (*S*) enantiomer, or how the enzyme controls the strict sequence of oxidation. Thus, the specific molecular factors controlling this selectivity and order of steps remained elusive despite a high-resolution x-ray structure of the substrate-TamI complex.

We envisioned identifying key active site residues that dictate the selectivity and step sequence displayed by TamI. Thus, amino acids within 5Å of the tirandamycin bicycle were selected for an alanine/valine scan by single point mutation (Figure 2.5), generating a total of 13 TamI mutants: L101A, L101V, L244A, L244V, L295A, L295V, I247A, I247V, G248A, T251A, T299A, T299V and H102V (Figure 2.6). Mutagenesis efforts to obtain G248V, T251V and H102A did not yield the desired substitutions. Enzymatic conversions of **1** were conducted *in vitro* using

a previously optimized three-component system comprising spinach ferredoxin, ferredoxin reductase and P450 TamI in the presence of a NADPH cofactor regeneration system. Initial analytical evaluation of the 13 mutants revealed nearly WT activity for all but five specific positions, revealing potential key residues in both the binding and positioning of the tirandamycin substrate. Of the five, the TamI H102V variant resulted in severe loss of catalytic activity highlighting its likely role in substrate binding as previously described,⁸ through a proposed interaction between the N ϵ of His102 and the oxygen atoms of the ketal group (Figure 2.6A, lane 11).

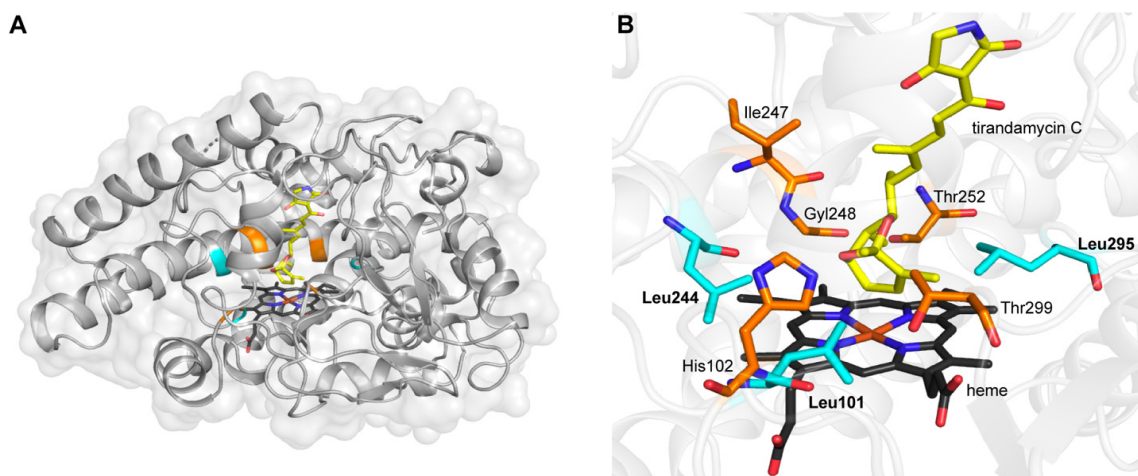


Figure 2.5. Active site residues of TamI selected for an alanine-valine scan. (A) Cartoon representation of TamI in complex with **1** (shown in yellow). Active-site residues of TamI located within 5Å of the tirandamycin bicycle are highlight. The three amino acids found to be critical in controlling TamI's selectivity and iterative oxidation are highlighted in cyan; the remaining amino acids mutated are shown in orange. (B) Stick representation.

2.3.2.2 TamI Mutants with Divergent Selectivity from Wild-Type

Analysis of products formed from the TamI L244A mutant revealed trace amounts of a new double oxidation congener (**11**) (Figure 2.6A, lane 8). The selectivity of L101A was poor, producing a mixture of an unknown single oxidation intermediate (**6**), and two new double oxidation congeners with variant polarities (**7** and **8**) (Figure 2.6A, lane 7). The alanine substitution may hamper the ability of the substrate to adopt one distinct orientation within the enzyme binding pocket, leading to multiple oxidation products. On the other hand, this catalytic flexibility makes L101A a promising starting point toward engineering selective TamI variants for the production

of each of these new congeners. TamI L295V led to the preferential formation of **7**, the least polar double oxidation congener indicated above (Figure 2.6A, lane 10), while TamI L295A selectively produced two unique triple oxidation products that eluted as a single peak (**9** and **10**) (Figure 2.6A, lane 9). This divergent reactivity suggests that subtle steric factors at the Leu295 position significantly alter the substrate binding orientation with respect to the heme center.

TamI mutants displaying non-selective product profiles (TamI L101A and TamI L244A) were further engineered to improve selectivity and maximize production of the new tirandamycin congeners. The double variant L101A_L295V favored production of **8** that single mutants L101A and L295V generated in small amounts. Although increased formation of the desired product was observed, this double variant expressed poorly in *E. coli*. We addressed this complication by introducing L295I instead, which was one of the highest producing single mutant forms of TamI. Thus, TamI L101A_L295I selectively produced **8** as the major product (Figure 2.6A, lane 12). Following a similar strategy, the L244A_L295I double mutant was generated in an effort to improve formation of **11**. This double mutant catalyzed increased conversion of **1** to the desired product (Figure 2.6A, lane 14) compared to single variant L244A, albeit in poor yield. Moreover, adjusting reaction parameters (enzyme concentration and time) with mutants L101A_L295I and L295A favored accumulation of the single oxidation product **6** (Figure 2.6A, lane 13), suggesting it may be an intermediate to the more highly oxidized derivatives. Triple mutant L101A_L295I_L244A (See Appendix A for LC-MS traces) was engineered to explore the effect of simultaneously introducing mutations at these three key leucine sites and it displayed a nonselective product profile. Additional mutations were introduced to probe the effect of different substituents at the Leu101, Leu295 and Leu244 positions with the goal of further improving the selectivity of TamI variants. Most substitutions at Leu101 led to non-selective catalysis while nearly all Leu295 and Leu244 mutants resulted in wild-type product profiles and were not further investigated. The best performing TamI mutants with divergent selectivity from WT (L295A, L295V, L101A_L295I and L244A_L295I) were chosen for further experimental work. This included biochemical characterization and large-scale enzymatic reactions to generate sufficient amounts of the new tirandamycins for structure elucidation and mechanistic studies.

2.3.2.3 TamI Mutants with Enhanced Iterative Capabilities

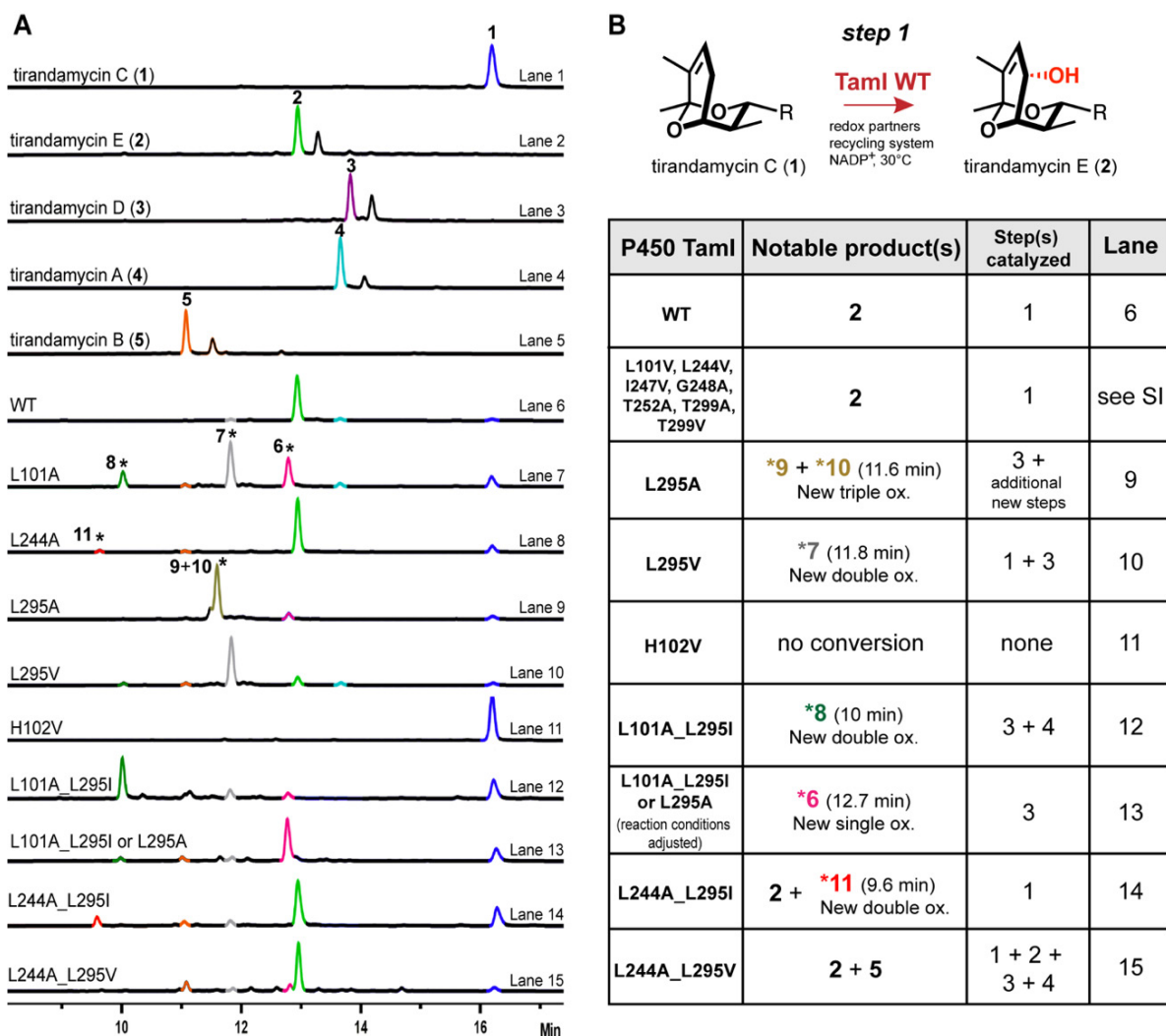


Figure 2.6. Endpoint assays with TamI mutants. (A) LC-MS traces (extracted at 354nm, signature UV of tirandamycin) of select reactions. Authentic standards include **1-5** (lanes 1-5). Compounds **2-5** show a sister peak due to isomerization in methanol. Peaks of new tirandamycin congeners are marked with an asterisk (lanes 7-9). All reactions were performed following standard conditions (see Methods) except for lane 13 reaction that included 1 μ M P450 L101A_L295I catalyst loading instead of 2 μ M. The new single oxidation congener (12.7 min) was also generated as the major product with P450 L295A as the catalyst when adjusting reaction parameters. (B) Summary of *in vitro* analytical enzymatic reactions with TamI variants and **1**. The steps catalyzed by each mutant are shown in sequential order of reactivity where *step 1* is C10 hydroxylation, *step 2* is C10 ketone formation, *step 3* is C11/12 epoxidation and *step 4* is C18 hydroxylation as shown in Scheme 2.1.

Our earlier work demonstrated that the flavin oxidase TamL serves as a co-dependent participant in conversion of **2** to **3** (Scheme 2.1, *step 2*).²⁵ We envisioned engineering a TamI biocatalyst capable of more complete oxidation, forming **5** from **1** in a 4-step cascade approach in

the absence of TamL. LC-MS analysis revealed that mutants L101A, L244A and L295V formed **5** in trace amounts (Figure 2.6A, lanes 7, 8 and 10). Notably, P450 L244A_L295V converted **1** to **5** in larger amounts than the corresponding single mutant forms (Figure 2.6A, lane 15), and represents a TamI variant able to catalyze up to an eight-electron oxidation cascade. The sterically less hindered replacements may relax the conformational constraints of the substrate in the active site, favoring productive binding and orientation of the more highly oxidized tirandamycin intermediates, enabling a more continuous oxidation in the absence of the flavoprotein TamL. Moreover, mutants that were previously generated also revealed altered patterns of iterative oxidation activity. Specifically, mutant L101Y generated **3** as the major product albeit with significant reduced activity while mutant I247A catalyzed increased formation of the more extensively oxidized intermediate **4** compared to TamI WT (See Appendix A for LC-MS traces). However, these variants expressed poorly in *E. coli* and were not further investigated. Based on these results, TamI L244A_L295V (also referred to as the iterative mutant) was selected for additional experimental studies.

2.3.2.4 Catalytically Self-Sufficient TamI Mutant Biocatalysts

To facilitate preparative scale enzymatic reactions, we envisioned designing self-sufficient TamI biosynthetic systems as we have previously reported. L295V-RhFRED and L101A_L295I-RhFRED were successfully generated, while attempts to obtain L295A-RhFRED were unfruitful. The L101A_L295I-RhFRED system did not yield soluble protein. Incubating **1** with L295V-RhFRED led to the formation of **7**, however at greatly diminished proportions compared to the three-component control reaction. Hence, we decided to proceed with optimized ferredoxin/ferredoxin reductase-containing reactions for the generation of the new tirandamycin.

2.3.3 Elucidation of New Tirandamycin Congeners and Mechanisms

2.3.3.1 Optimization of Large-Scale Fermentation Conditions

Optimization of fermentation conditions of a *Streptomyces sp.* 307-9 Δ tamI P450 mutant strain enabled reliable production of **1** in a 10-20mg/L yield, a 10-fold improvement from initial fermentation efforts. Following a media-optimization approach, MD2 media was found optimal for 1L-scale growths using a 2.8L baffled flask and following a 3% inoculation from a single colony and a 7-day growth protocol. Ribosome engineering was attempted to further improve production of **1** but efforts were unsuccessful. The tirandamycin material was employed for

subsequent preparative-scale enzymatic reactions to obtain sufficient starting material for structural characterization using TamI and its variants. NMR analysis, MD simulations, QM calculations and end-point assays revealed that the engineered TamI mutants have altered the selectivity and order of steps programmed into the WT enzyme to create new compounds. These studies provide additional knowledge on the TamI mechanism of iterative oxidation and represent new tools for selectivity in C-H functionalization and alkene epoxidation.

2.3.3.2 TamI Mutants L101A_L295I and L295A Interrupt the Stepwise Oxidative Cascade Generating Tirandamycin L

Diverging from native reactivity where C10 allylic hydroxylation (*step 1*) is the first oxidation event to occur, TamI L101A_L295I and L295A catalyzed the C11/12 epoxidation (*step 3*) of **1** generating a new single oxidation congener tirandamycin L (**6**) (Scheme 2.2; for LC-MS traces of the reactions described, refer to Appendix A). The stereochemistry of **6** was determined based on the observed NOE correlations of H-11 to H-7. The characterization of **6** indicates that the apparent strict stepwise oxidative cascade of TamI has been interrupted. Previous density functional theory (DFT) calculations comparing the transition state barriers for C10 hydroxylation, C11/12 (*R/S*) epoxidation and C18 hydroxylation starting from **1** revealed that the olefin epoxidation is highest in energy by 2.6 kcal/mol.⁸ This suggests that subtle variations in the catalytic environment of TamI L101A_L295I and L295A are critical for catalyzing the least favored reaction, by potentially reorienting the substrate to lower the activation barrier. MD simulations of TamI L101A_L295I with **1** were analyzed and compared to those of TamI WT. These experiments revealed that L101A_L295I demonstrated two major binding poses: the first similar to the WT with the C10-(*S*) hydrogen oriented in the correct geometry for C-H abstraction reactivity and the second with the correct geometry for the olefin epoxidation to occur (Figure 2.7A). These two poses are highlighted in a comparison of the MD snapshots to the ideal transition state geometry from DFT. One analyzes *step 1*, which compares the O_{heme}-C10 H distance and O_{heme}-H10-C10 angle and demonstrates a newly introduced competition with the C10-(*R*) that was not observed in the WT, as more of the MD snapshots show the correct geometry for C10-(*S*) reactivity in the WT. The other analyzes *step 3*, which compares the O_{heme}-C11 distance and O_{heme}-C11-C12-C18 dihedral angle and shows the change in facial selectivity between the binding poses. Thus, some of the snapshots are close to the ideal transition state dihedral angle and the rest are the opposite, allowing the possibility of the epoxidation to occur with the second

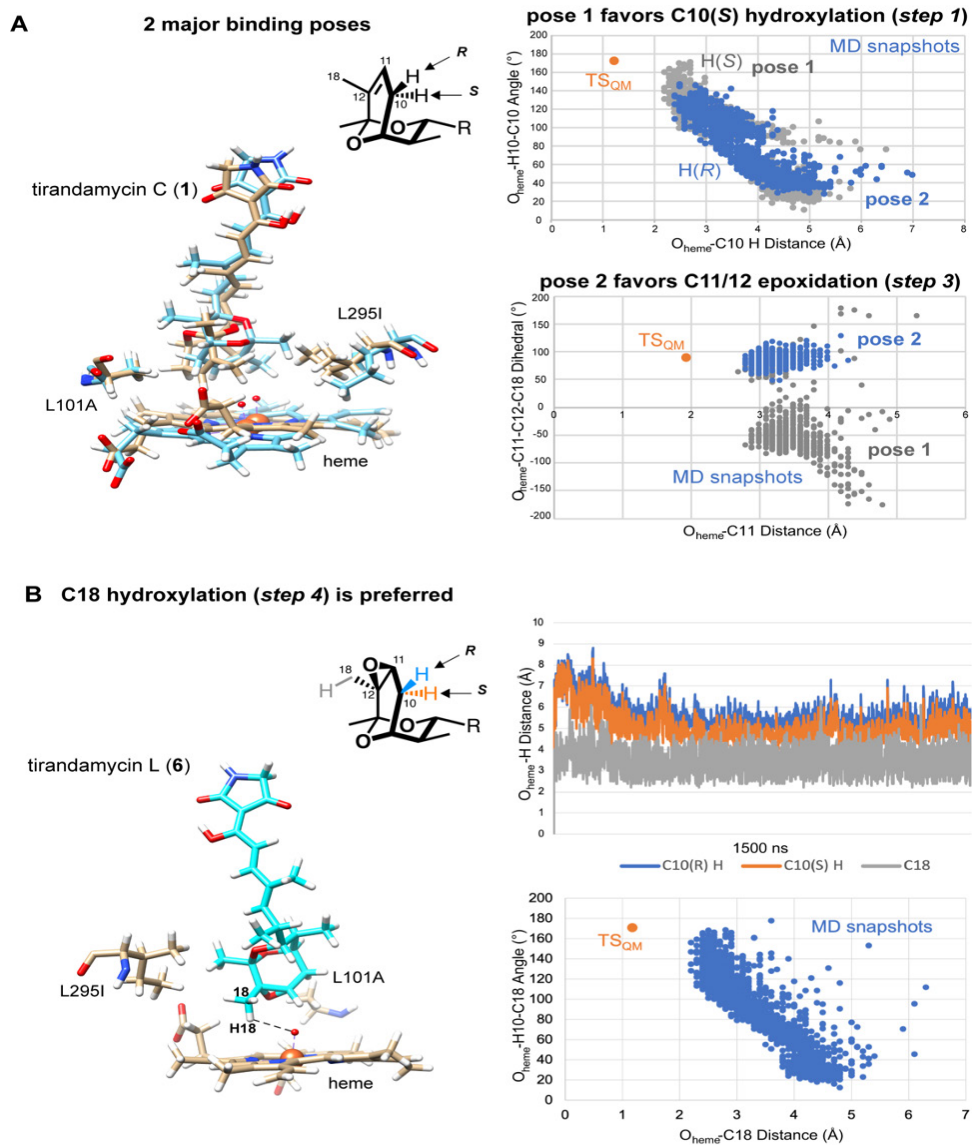
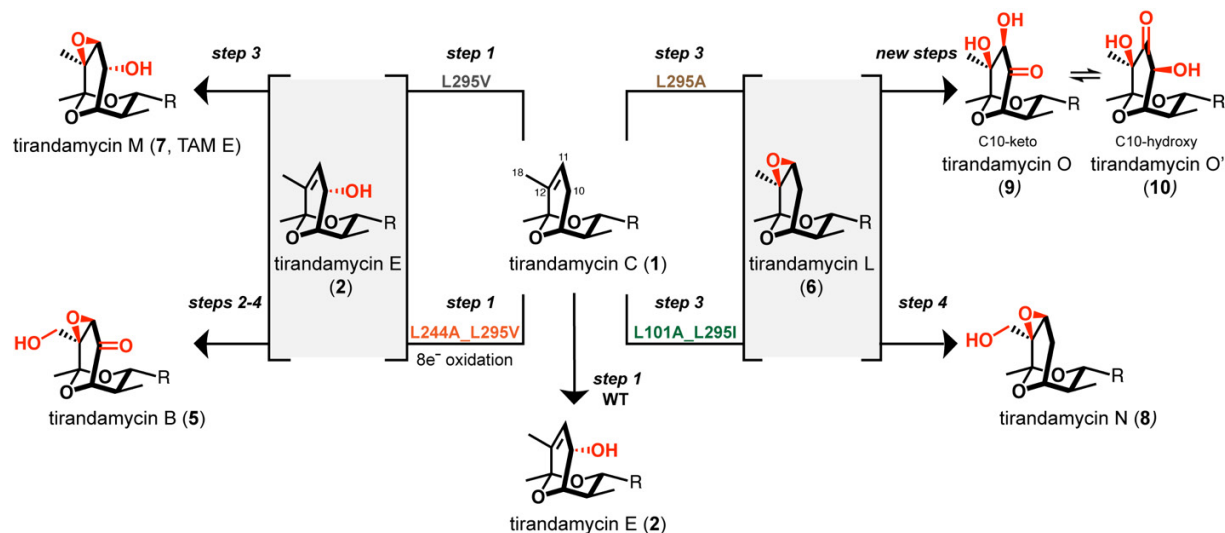


Figure 2.7. Computational analysis of TamI L101A_L295I. (A) MD simulations of TamI variant with **1** demonstrate two binding modes for the substrate that allow for competition for reactivity for C11/12 epoxidation (*step 3*) over C10 hydroxylation (*step 1*), as opposed to that of the WT. The second binding pose (blue) promotes a geometry that is suitable for epoxidation, supporting the experimentally observed product. (B) MD simulations with **6** shows preference for the C18 position over C10.

binding pose. In addition, MD simulations of TamI L295A with **1** showed that proximity of the C11 atom to the heme center is maintained for the majority of the simulation, consistent with C11/12 epoxidation. These observations indicate that the hydrophobic interactions around the

bicyclic ketal are most critical for redirecting oxidative pathways in TamI from an allylic C-H oxidation to an epoxidation event.



Scheme 2.2. TamI mutants generating novel tirandamycin congeners with divergent selectivity from wild-type enzyme. TamI L244A_L295V performs a 4-step oxidation cascade. New tirandamycin biosynthesized include tirandamycin L (6), M (7), N (8), O (9) and O' (10). For complete NMR characterization data including NOE correlations, refer to SI. For PyMOL and MD simulations of TamI mutants, see Appendix A.

2.3.3.3 Dual-Function TamI L295V Produces Tirandamycin M

Analogous to WT activity, TamI L295V first catalyzes *step 1* forming **2** from **1**. However, the next tailoring step for this variant is *step 3* on substrate **2** to generate the double oxidation congener tirandamycin M (7, TAM E) (Scheme 2.2). Although this molecule has been previously isolated from *Streptomyces* sp. 17944,⁴ our work demonstrates its production *in vitro* using a TamI P450 mutant as a biocatalyst. MD was similarly performed with TamI L295V and **2**, which showed that a lower energy conformation was achieved after initial minimization. It also revealed that the C11 remained closest relative to other reactive positions and the oxo-iron species throughout the 1500 ns simulation, consistent with C11/12 epoxidation. Moreover, HPLC analysis of the culture broth of a *Streptomyces* sp. 307-9 Δ tamL flavoprotein mutant strain³ revealed the presence of **2** and **7**, in approximately a 2:1 ratio, after only four days of growth. We reasoned that in the absence of the flavoprotein, the TamI WT is capable of catalyzing the epoxidation of **2** \rightarrow **7** *in vivo*. This

hypothesis is supported by the observation that when testing **2** with purified TamI WT, **7** is generated albeit as a trace product.

2.3.3.4 TamI L101A_L295I Bypasses Oxidation at C10 Forming Tirandamycin N

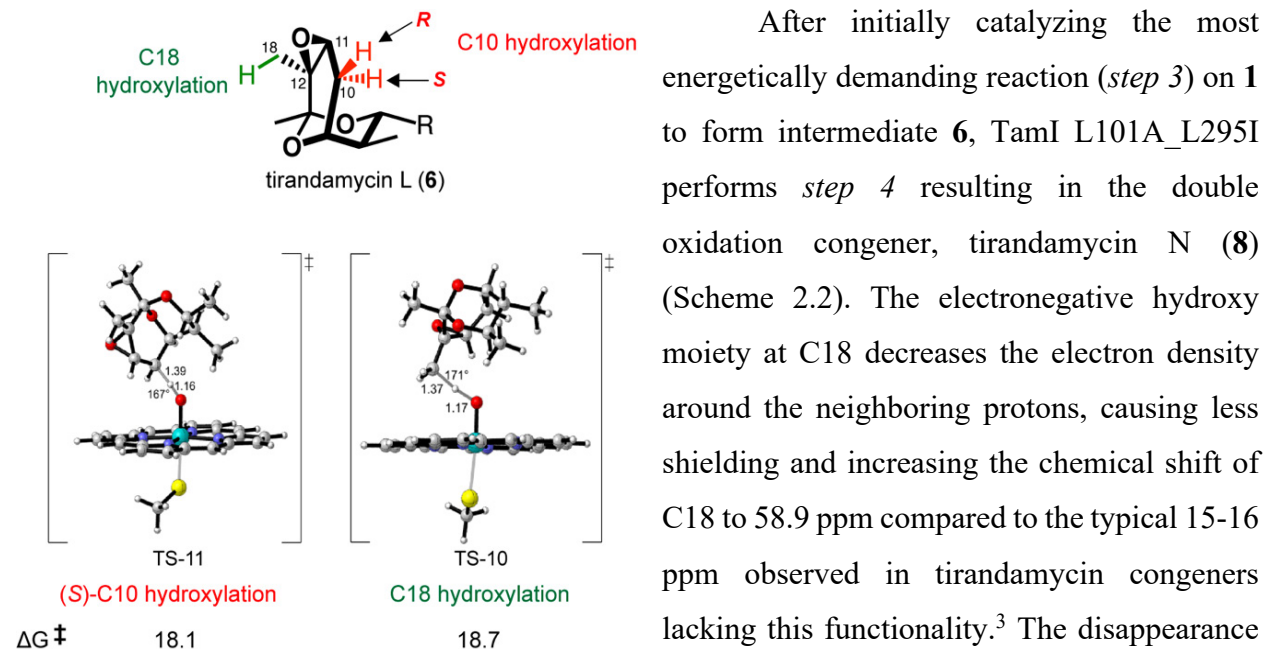


Figure 2.8. DFT calculations of TamI L101A_L295I regioselectivity. Intrinsic energy for the competing C-H oxidation reactions at C10 and C18 with **6** as substrate.

After initially catalyzing the most energetically demanding reaction (*step 3*) on **1** to form intermediate **6**, TamI L101A_L295I performs *step 4* resulting in the double oxidation congener, tirandamycin N (**8**) (Scheme 2.2). The electronegative hydroxy moiety at C18 decreases the electron density around the neighboring protons, causing less shielding and increasing the chemical shift of C18 to 58.9 ppm compared to the typical 15-16 ppm observed in tirandamycin congeners lacking this functionality.³ The disappearance of the singlet corresponding to protons of the C18 methyl group, and the presence of new signals relating to a methylene group corroborates this assignment. DFT calculations were performed to determine the transition state barrier for competing hydroxylation reactions at C18 and C10 starting from **6**. The C-H abstraction barrier for the C10(S) hydroxylation and C18 hydroxylation had essentially no energy difference at 0.6 kcal/mol with the former being lower in energy (Figure 2.8). This contradicts the experimentally observed regioselectivity with TamI L101A_L295I, where **8** is exclusively formed from **6**. MD simulations performed with the variant and **6** showed that the C18 is closest to the reactive heme iron-oxo throughout the entire 1500 ns simulation, consistent with *step 4* (Figure 2.7B). The $O_{\text{heme}}\text{-C18}$ hydrogen distance and $O_{\text{heme}}\text{-C18}$ hydrogen-C18 angle geometries from the MD simulations were compared to the ideal QM calculated transition state. This indicated that the active site geometry of TamI L101A_L295I controls the orientation to prefer reactivity of C18 hydroxylation and therefore, is crucial in discerning the selectivity between these regioisomeric transition states.

2.3.3.5 Multifunctional TamI L295A Catalyzes an Unexpected Oxidative Cascade Generating Tirandamycin O and O'

Similar to TamI L101A_L295I, TamI L295A first catalyzed *step 3* on substrate **1** producing **6**. However, divergent from the double mutant selectivity, TamI L295A catalyzes a unique series of oxidation steps leading to the formation of triple oxidation products tirandamycin O (**9**) and tirandamycin O' (**10**) (Scheme 2.2) that eluted as a single peak during HPLC purification. In the analytical scale, a small shoulder on the major product peak is observed when incubating TamI L295A with **1** and **6**, separately, suggesting the formation of both congeners *in vitro*.

The tri-functionalized congener **9** displays an unusual oxidation pattern on the bicyclic core including a C10 keto group, a C11(*R*) hydroxyl and a C12(*S*) hydroxyl, with corresponding chemical shifts at 204 ppm, 72.1 ppm and 78.3 ppm. The relative stereochemistry of the C11(*R*) and C12(*S*) hydroxy moieties was determined by irradiation of the H-11 and H-7 protons resulted in a NOE to H-7 and H-11, respectively, while H-18 showed an NOE with H-14 and H-7, and vice versa. The ¹H-NMR spectrum suggests the presence of an isomer in a 5:1 mixture with **9** (C10-keto) as the predominant form. Moreover, we learned that upon storing this mixture in DMSO, the ratio of molecules changed to 3:1 with **10** (C10-hydroxy) as the major form. In **10**, the keto functionality is present at the C11 site that now displays a 205 ppm shift while the C12(*S*) hydroxyl is conserved. The C10 position harbors a hydroxy moiety in the (*S*) configuration, assigned based on the H-10 proton showing an NOE with H-17. Treatment of compound **9** (5:1) with D₂O did not show deuterium incorporation, even after extended periods of incubation (See Appendix A for NMR spectra), suggesting that protic-mediated tautomerization is not occurring. Furthermore, to elucidate the molecular basis for the exquisite stereoselectivity of the hydroxyl groups, the $\Delta\Delta G$ of each of the four theoretically possible stereoisomers was calculated. The experimentally observed isomer **10** was higher in energy than its C10 epimer while compound **9** was lowest in energy (See Appendix A for energy values). Currently, efforts are underway to decipher the mechanistic pathway towards the formation of tri-oxidized **9** and **10**, which will be reported in due course. We reason that the diol moiety in **9** does not result from epoxide ring opening in aqueous solution, but rather from one of three other potential mechanisms including: 1) H-abstraction from iron oxo (Cpd I) generating a free radical that leads to epoxide ring opening and eventual formation of **9** and **10**; 2) Incorporation of hydroperoxo (Cpd 0) to hydroxylate the C11 position of the

epoxide yielding the α -hydroxy ketone; 3) Nucleophilic attack to open epoxide at C11 catalyzed by a peroxo-iron species that leads to the formation of the ketone at C11.

2.3.3.6 *Iterative TamI L244A_L295V Enables a Continuous δ^- Oxidation*

The iterative TamI L244A_L295V catalyzed sequential C-H oxidation and epoxidation pathways leading to compounds 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 in a strict order as observed in the WT cascade that includes TamL (Scheme 2.1, *step 2*).³ However, in the absence of TamL, TamI L244A_L295V is responsible for transforming the C10(S) hydroxyl of **2** into a ketone, possibly through initial formation of a geminal-diol product. To further understand this catalytic activity, we sought to explore the mechanism in which the enzyme installs the C10 keto group using DFT methods. Previous work^{27, 28} provided a framework for our analysis that includes three possible mechanisms (Figure 2.9A). In Route 1, TamI iteratively hydroxylates at C10 forming a geminal-diol that exists predominantly in the keto form. In Route 2, the enzyme catalyzes a C10-H abstraction followed by an O-H abstraction and radical coupling to yield the ketone. Route 3 resembles Route 2 except it suggests an inverted order of proton abstraction. Analysis of the calculated free energy profiles suggests the preference of Routes 1 and 2 over Route 3 as the rate-limiting transition state is 2.8 kcal/mol higher in Route 3. In distinguishing the plausibility between Routes 1 and 2, there are virtually no energetic differences between the gem-diol pathway (Route 1) and the double hydrogen abstraction (Route 2), due to computation constraints that the second transition states are essentially barrierless. Thus, it is possible that either or both mechanisms are plausible with this P450 variant. Using previously described methods,⁸ MD simulations of TamI L244A_L295V were performed with **2** as the substrate to assess the influence of the mutant on the selectivity within the oxidative cascade and compared with that of the TamI WT. The TamI L244A_L295V displays a higher degree of flexibility of substrate **2** with the active site compared to the WT (Figure 2.9B). This suggests that these mutations enable movement that is required for the C10 oxidation to occur, as our DFT mechanisms show that a shift in the binding pose is required to promote the second mechanistic steps.

2.3.4 *Biochemical and Kinetic Characterization of TamI Mutants*

2.3.4.1 *Substrate Scope, Total Turnover Capacity and Substrate Binding Properties of TamI Mutants*

Effectively applying the biocatalytic potential of the engineered TamI P450s for late-stage C-H oxidation and epoxidation requires a comprehensive investigation of their substrate specificity, total turnover capacity and substrate binding affinity. We surveyed the substrate scope of TamI variants using various tirandamycin molecules and calculated the corresponding total turnover numbers (TTN) (Figure 2.10). Additionally, the effect of TamI mutations on the equilibrium dissociation constants (K_d) of tirandamycin substrates was assessed via substrate-induced heme spin shift experiments.

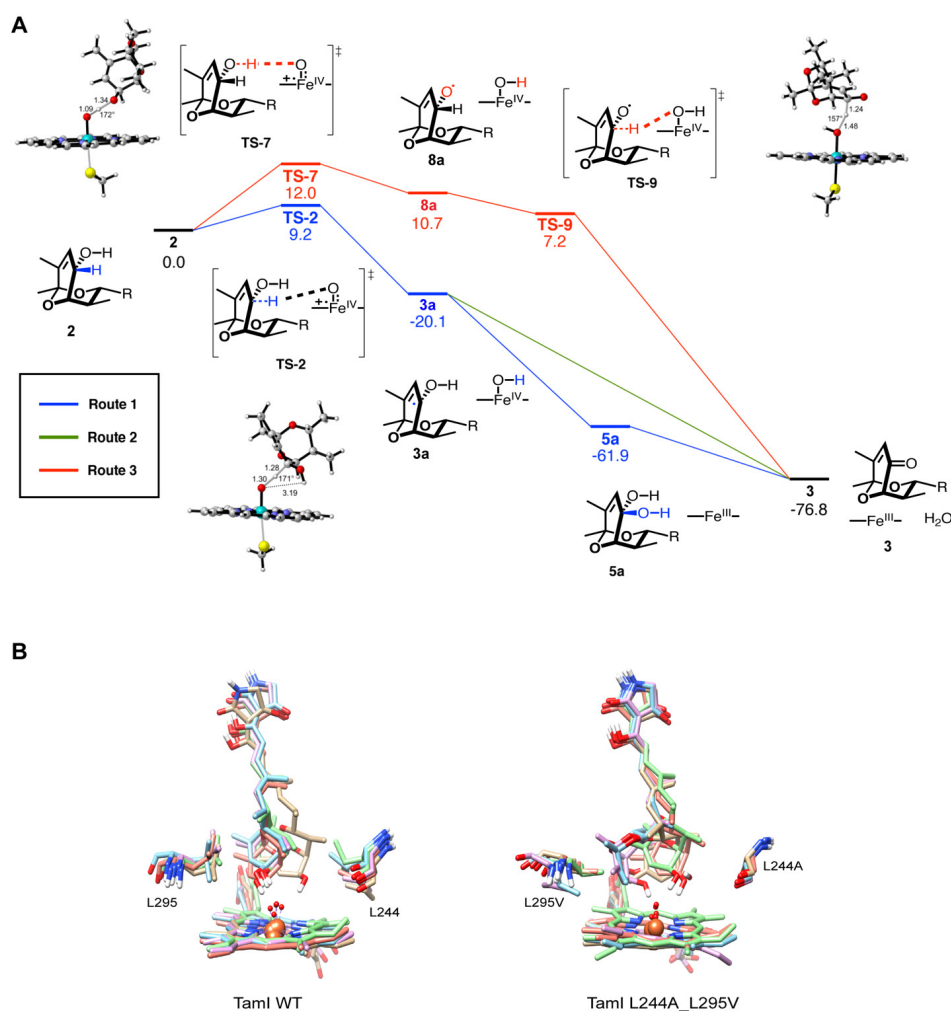


Figure 2.9. Computational analysis of iterative TamI L244A_L295V. (A) Free energy profile of potential mechanisms for the C10 ketone formation. Initial abstraction of the OH-10 proton requires a higher energetic barrier compared to H-10 abstraction, discarding Route 3 as a viable mechanism pathway. There are virtually no energetic differences that discriminate Route 1 from Route 2, suggesting that either mechanism is plausible.. (B) Overlays of top five occupied clusters for 2 with TamI WT and TamI L244A_L295V.

Substrate 1: Despite catalyzing distinct oxidation pathways, TamI WT, L295V and L101A_L295I have comparable total turnovers (TTN = 368-369) and show tight binding affinity to **1** with the double mutant displaying the strongest affinity ($K_d = 0.09 \mu\text{M}$). The TamI H102V mutant that fails to generate any product when tested with **1** shows a weak binding affinity to the substrate ($K_d = 10.27 \mu\text{M}$).

Substrate 2: In addition of TamI L295V displaying epoxidase activity forming **7** from **2**, TamI WT, L295A, L101A_L295I and L244A_L295I were also found capable of catalyzing *step 3* on **2**. The double mutant L101A_L295I showed the highest TTN value of 324 and a relatively tight affinity to the substrate ($K_d = 4.75 \mu\text{M}$), while no spin shift was observed with the other P450s including the parent enzyme.

Substrate 3: Similar to WT, all variants tested with **3** catalyzed formation of the more oxidized **4** and **5**, with comparable turnover (TTN = 313-340) and binding affinities ($K_d = 6.89$ - $8.32 \mu\text{M}$). TamI L244A_L295V and L244A_L295I led to the highest ratios of **5** to **4**, compared to TamI WT that produced the least amount of the terminal product **5**.

Substrate 4: All P450s showed selectivity for *step 4*, converting **4** to **5** with iterative TamI P450 L244A_L295V displaying improved turnover capacity and binding affinity (TTN = 167; $K_d = 7.09 \mu\text{M}$) compared to the parent enzyme (TTN = 143; $K_d =$ no shift observed).

Substrate 6: All P450s tested, except TamI L295A, catalyzed *step 4* on substrate **6** forming **8**. In contrast, TamI L295A generated **9** and **10** as the main products and displayed the highest TTN value of 252. No spin shift is observed when testing **6** with the P450s, except with the double mutant L101A_L295I that shows a weak binding affinity of $11.98 \mu\text{M}$.

Substrate 7: No oxidative activity was observed when testing the substrate with TamI biocatalysts. Surprisingly, the control reaction containing TamL flavoprotein yielded complete conversion of **7** to **4**. Isothermal Titration Calorimetry (ITC) was employed to measure the binding affinities of **7** and native substrate **2** with the TamL enzyme. The resulting affinities were comparable with K_d values of 1.09 and $2.26 \mu\text{M}$ for **7** and **2**, respectively.

Substrates 8, 9 and 10: Finally, reactions with **8**, **9** and **10** failed to yield products suggesting that the engineered TamI enzymes have a limited scope for substrates with higher degrees of oxidation.

2.3.4.2 Michaelis-Menten Kinetic Parameters of TamI Variants

Michaelis-Menten kinetic parameters were determined for the various oxidation routes catalyzed by TamI WT, selective TamI L101A_L295I and iterative TamI L244A_L295V (Table 2.2). These enzymes were selected for testing based on their abilities to catalyze divergent iterative oxidation cascades and their enhanced binding and turnover properties.

Consistent with previously described results,³ TamI WT is most efficient at catalyzing *step 1* converting **1** to **2** with a k_{cat}/K_M value of $1.87 \mu\text{M}^{-1}\text{min}^{-1}$ (Table 2.2, lane 1). Interestingly, TamI

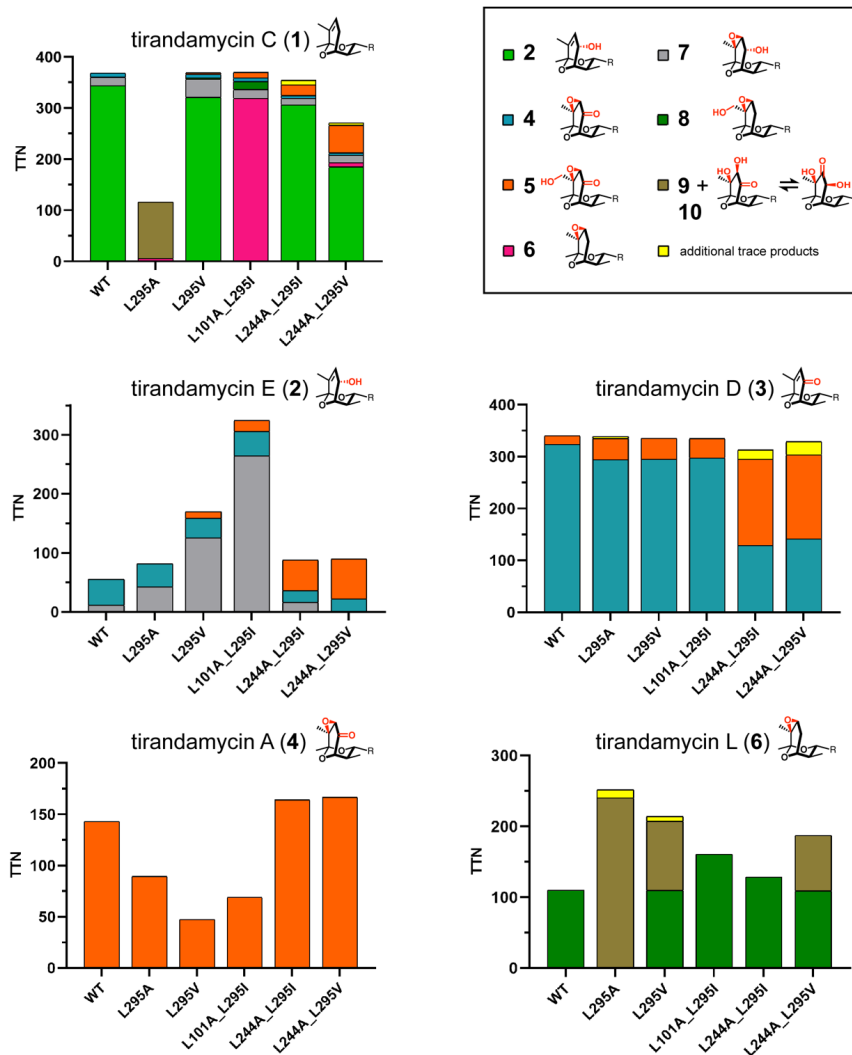


Figure 2.10. Total turnover numbers (TTN) for TamI WT and mutants with tirandamycins. Reaction substrates are displayed above each bar graph. TTN = mol substrate consumed/mol P450. Ratios of product formation were estimated based on HPLC data. Experiments were performed in triplicate. For definition of R see Scheme 2.1.

L101A_L295I catalyzes *step 3* on **1** forming **6** and the iterative oxidation of **1** to **7** (*steps 1* and *3*) with virtually the same efficiency ($k_{cat}/K_M = 1.71 \mu\text{M}^{-1}\text{min}^{-1}$) (Table 2.2, lane 2). As expected,

TamI WT is notoriously inefficient at catalyzing the continuous oxidation of **2** to **4** (*steps 2 and 3*) ($k_{cat}/K_M = 0.0003 \mu\text{M}^{-1}\text{min}^{-1}$) (Table 1, lane 3). Compared to the WT enzyme, iterative TamI L244A_L295V shows a 36-fold increase of catalytic efficiency ($k_{cat}/K_M = 0.011 \mu\text{M}^{-1}\text{min}^{-1}$) when converting **2** to **3**, **4** and **5** (*steps 2, 3 and 4*) (Table 2.2, lane 4). This substantial improvement in efficiency for oxidizing **2** highlights the significance of this variant in overcoming the bottleneck reaction of the cascade (*step 2*) without catalytic assistance from the TamL flavoprotein. The calculated catalytic efficiencies for *step 3* on substrate **3** forming **4** with TamI WT and TamI L101A_L295I are 0.19 and $0.47 \mu\text{M}^{-1}\text{min}^{-1}$ (Table 2.2, lanes 5 and 6), while the k_{cat}/K_M value for *step 3* and *step 4* on **3** generating **4** and **5** with iterative TamI L244A_L295V is $0.06 \mu\text{M}^{-1}\text{min}^{-1}$.

Moreover, the iterative TamI L244A_L295V is 23-times more efficient than the WT enzyme at catalyzing *step 4* on substrate **4** to generate the terminal product **5** with k_{cat}/K_M values of 0.102 and $0.004 \mu\text{M}^{-1}\text{min}^{-1}$, respectively (Table 2.2, lanes 8 and 7). TamI L101A_L295I also outperforms TamI WT with a k_{cat}/K_M of $0.0375 \mu\text{M}^{-1}\text{min}^{-1}$, suggesting an 8-fold increase in efficiency to perform the primary C-H oxidation event. Finally, TamI L101A_L295I is a more efficient catalyst than TamI WT for hydroxylation of **6** to **8** with k_{cat}/K_M values of 0.0044 and $0.0001 \mu\text{M}^{-1}\text{min}^{-1}$ (Table 2.2, lane 9), respectively. Iterative TamI L244A_L295V shows an even higher k_{cat}/K_M value ($0.0196 \mu\text{M}^{-1}\text{min}^{-1}$) for catalyzing multiple oxidative reactions forming a mixture of **8**, **9** and **10** from substrate **6** (Table 2.2, lane 10).

Substrate	P450 TamI	K_M (μM)	K_{cat} (min^{-1})	K_{cat} / K_M ($\mu\text{M}^{-1}\text{min}^{-1}$)	Products observed and tracked	Step(s) catalyzed	Lane
1	WT	16.0 ± 2.6	29.8 ± 1.9	1.87	2	1	1
	L101A_L295I	3.8 ± 0.6	6.6 ± 0.3	1.71	6 + 7	3, 1	2
2	WT	108.5 ± 25.9	0.04 ± 0.01	0.0003	4 + traces of 7	2, 3	3
	L244A_L295V	115.6 ± 22.4	1.27 ± 0.10	0.0110	3 + 4 + 5 + traces of 7	2, 3, 4	4
3	WT	36.5 ± 7.1	6.8 ± 0.5	0.19	4	3	5
	L101A_L295I	120.1 ± 33.5	56.4 ± 9.6	0.4693	4	3	6
4	WT	8.2 ± 1.0	0.036 ± 0.001	0.0044	5	4	7
	L244A_L295V	7.0 ± 0.9	0.716 ± 0.023	0.1022	5	4	8
6	WT	34.9 ± 5.5	0.005 ± 0.001	0.0001	8	4	9
	L244A_L295V	55.6 ± 9.9	1.092 ± 0.066	0.0196	8 + 9 + 10	4, additional new steps	10

Table 2.2. Select Michaelis-Menten kinetic values.

2.3.5 Antimicrobial Activities of New Tirandamycin

The naturally occurring and newly synthesized tirandamycin congeners were tested against a panel of human bacterial pathogens and their minimum inhibitory concentrations (MICs) were determined (Table 2.3). The activities of tirandamycins **1**, **3**, **5** and the new congener tirandamycin N (**8**) were comparable to commercial antibiotic erythromycin showing strong activity against Gram positive *B. anthracis* 34f2. Tirandamycins **1**, **3**, **5** also demonstrated enhanced activity against *S. pneumonia* ATCC 49619. Tirandamycin A (**4**) showed the strongest activity against VRE, consistent with previously described results.²⁷ A significant decrease in activity against VRE was observed in the new tirandamycin molecules. Poor activity was recorded for all tirandamycins against Gram negative pathogens, except for **1** that showed relatively weak activity against *E. coli* TolC.

	Ciprofloxacin	Erythromycin	Streptolydigin	5	4	1	3	2	9	8	7	6
Bacterial strain	MIC (μ M)											
<i>B. anthracis</i> 34f2	0.781	3.13	>100	6.25	12.5	3.13	6.25	25	>100	6.25	50	25
<i>B. cereus</i>	1.56	3.13	>100	3.13	6.25	3.13	25	>100	>100	50	>100	>100
Vancomycin-resistant <i>Enterococcus</i>	>100	3.13	>100	12.5	6.25	25	>100	>100	>100	>100	>100	>100
<i>S. pneumoniae</i> ATCC 49619	1.56	0.195	>100	3.13	25	12.5	12.5	>100	>100	25	>100	>100

Table 2.3. Select MIC values. Tirandamycin **9** contained a minor presence of congener **10** in the sample in a 5:1 ratio.

2.3.6 Discussion

TamI is an iterative bacterial P450 capable of catalyzing three consecutive and highly selective oxidation reactions in a strict stepwise manner.³ The work described herein demonstrates that this enzyme can be tuned through protein engineering to alter the selectivity, step sequence and number of reactions catalyzed, favoring either epoxidation and/or C-H activation pathways on multiple sites within a tirandamycin substrate. This tunability enabled the enzymatic synthesis of new tirandamycin derivatives with variant oxidation patterns that result from catalyst-controlled selectivity overriding innate substrate reactivity. For example, new compound **8**, which exhibits notable bioactivity against human bacterial pathogens *B. anthracis* 34f2 and *S. pneumoniae* ATCC 49619 compared to the other newly characterized congeners, is readily obtained by epoxidation and C18 methyl oxidation, whereas these oxidative decorations would normally be challenging to

access in a direct fashion without competing oxidation of the existing methylene units in the substrate structure.²⁸⁻³⁰ Our work also illustrates that TamI can be manipulated to effectively catalyze a four-step oxidative cascade without the assistance of the oxidative partner TamL, opening new avenues for the development of multifunctional oxidation catalysts with enhanced iterative properties. These results demonstrate the power of iterative bacterial P450s to develop new tools for selectivity and cascade reactions, expanding enzyme-controlled iterative C-H functionalization and epoxidation of complex scaffolds and streamlining the synthesis of high-value target compounds.

Using structure-guided mutagenesis, end-point assays and NMR analysis, active-site residues Leu101, Leu244 and Leu295 were found to be critical in dictating the selectivity features and direct oxidation cascade of TamI (Scheme 2.2). Subtle steric alterations at the Leu295 position tuned the ability of TamI to favor either epoxidation or C-H activation pathways with exquisite control of the timing and site-selectivity of the cascade reactions from a single molecular scaffold. Specifically, starting with **1** as the substrate, TamI L295V catalyzed *step 1* followed by *step 3* yielding **7**, while TamI L295A catalyzed *step 3* generating intermediate **6** followed by successive oxidations to form tri-oxidized **9** and **10**. The TamI L101A_L295I catalyst also favored the initial epoxidation of **1** to **6** followed by *step 4* producing **8** with exquisite regioselectivity.

The ability of TamI mutants for crafting and further processing the epoxide-containing intermediate **6** yielding the more highly functionalized **8**, **9** and **10** is noteworthy. These results provide evidence that 1) *step 1* is not a prerequisite for *step 3* to occur and 2) *step 3*, as the first reaction catalyzed, does not block further oxidative reactions with TamI, catalytic limitations that are often observed with other multifunctional P450s. For example, in the mycinamicin P450 MycG,²⁰ an analogue substrate harboring an epoxide moiety but lacking a hydroxyl cannot undergo oxidation by the catalyst due to a strict hierarchy in order of catalytic steps, where hydroxylation is imperative for subsequent epoxidation. Moreover, in the iterative P450 GfsF, epoxidation is a prerequisite for hydroxylation of the 16-membered macrolactone, with no flexibility detected during the oxidative steps.²² In both P450 MycG and P450 GfsF dual-function cascades, the step sequence seems to be fully controlled by inherent substrate reactivity and conformation. The capacity of TamI for catalyzing multi-step oxidative cascades at four distinct sites on substrate **1** (C10-C11-C12-C18) is also worth highlighting. Contrary to TamI, many iterative P450s oxidize a single site of the substrate. Examples include bacterial P450s RosC28

and XiaM29 that convert a methyl group to a carboxylic acid through the corresponding alcohol and aldehyde intermediates, although more examples of P450s with similar catalytic activities are found in plants³¹ and animals.³² While same-site iterative oxidation cascades install functional groups that are critical for the bioactivity of secondary metabolites, the ability of TamI to oxidize multiple carbon atoms within a substrate enhances direct access to core structure diversification, generating new sets of analogues with different points of derivatization.

We performed computational analyses to investigate if the altered biocatalyst selectivity and chemical output observed with TamI mutants was a result of innate substrate reactivity or driven by a catalyst-controlled system. DFT calculations with **1** had previously shown that *step 3* is highest in energy compared to *step 1* and *step 4*,⁸ making it the least favored reaction to occur in the absence of TamI L101A_L295I and L295A. MD simulations of the enzyme variants with **1** showed that the P450s reorient the substrate maintaining close proximity of the C11 atom to the heme center, thus facilitating *step 3* catalysis over *step 1*. Moreover, DFT calculations with **6** revealed that *step 1* is preferred over *step 4* (Figure 2.8). Once again, this contradicts the experimentally observed regioselectivity and indicates that TamI L101A_L295I overrides the innate reactivity of the substrate, achieving activation of the least electronically favored C-H bond generating **8**, a novel congener with notable bioactivity. MD simulations of TamI L101A_L295I with **6** showed that the C18 remains closer to the oxo-iron compared to C10 throughout the 1500 ns simulation, thus supporting that the enzyme environment is key in controlling regioselectivity (Figure 2.7B). These results illustrate that the hydrophobic interactions around the bicyclic ketal of tirandamycin, especially at the Leu295 and Leu101 sites, are most critical for redirecting oxidative pathways in TamI from an allylic C-H oxidation to an epoxidation event when starting with **1**, and for controlling the regioselectivity favoring primary over secondary hydroxylation when **6** is the substrate. It is also important to highlight that the precise control of TamI L101A_L295I for catalyzing a divergent oxidation pathway (compared to WT) does not sacrifice catalytic efficiency as the variant catalyzes *step 3* on **1** with virtually the same K_{cat}/K_M as TamI WT catalyzes *step 1* on **1** (Table 2.2).

Furthermore, TamI L244A_L295V was engineered to catalyze an unprecedented four-step oxidative cascade in the absence of oxidative partner TamL. The Michaelis-Menten model kinetics displayed enhanced rates of continuous oxidation of **2** (up to 36-fold) and of C18 methyl oxidation of **4** (up to 23-fold) for the TamI L244A_L295V compared to WT. These kinetic values

demonstrate that by facilitating *step 2*, the TamI variant overcomes the bottleneck reaction of the pathway becoming a more efficient catalyst. Previous work with the bacterial P450cam demonstrated that an L244A mutation in the enzyme increased the active site malleability, enabling the oxidation of larger substrates than those accepted by the WT enzyme.³³ The authors hypothesized that the smaller alanine residue expanded the substrate-binding cavity providing a higher degree of freedom of motion for substrate positioning and catalytic function.³⁵ Analogous to the L244A P450cam system, the L244A substitution in TamI L244A_L295I generates a larger space and a higher degree of flexibility to better accommodate **2** enabling a more continuous oxidation cascade than the WT (Figure 2.9B). In efforts to elucidate the basis of improved iterative oxidation in TamI L244A_L295I, we propose two mechanisms for oxidation of the C10 allylic hydroxyl of **2** to a ketone either through gem-diol formation or double C-H abstraction (Figure 2.9A). Although a similar pattern of reactivity has been reported for the iterative bacterial P450 DoxA from the daunorubicin pathway,^{27, 36} the free energy profile for each potential mechanism has not been previously calculated to support the mechanistic hypotheses of allylic hydroxyl oxidation in an iterative bacterial P450.

Investigation of TamI has provided new insights into the underlying biocatalytic mechanisms for the diverse selectivity and ordered reaction process of multifunctional P450s on complex scaffolds. Minor alterations of Leu244, Leu295 and Leu101 in the active site of TamI override substrate-controlled reactivity to enable catalyst-controlled iterative oxidation cascades, involving chemically and mechanistically distinct processes. TamI variants are able to differentiate the reactivity of the three competing sites in the unfunctionalized core structure of **1** that contains an internal 1,2-disubstituted alkene flanked by allylic methyl and allylic methylene groups. Specifically, TamI is able to selectively oxidize the C18 methyl site over the C10 methylene (observed in the production of **8**), reverse the site selectivity between an alkene and allylic C-H bond (e.g. production of **2** and **6**), and oxidize non-allylically-activated C-H bonds (as seen in the formation of **5**, and **8-10**, where alkene epoxidation precedes C-H oxidation). This approach enabled the enzymatic synthesis of new tirandamycins that display potent bioactivity against Gram positive human pathogens and that could be further explored for their antiparasitic activity to combat lymphatic filariasis.⁴ Thus, TamI represents a compelling example for bio-inspiration given that small molecule catalysts with the potential for catalyzing both epoxidation and C-H activation are typically difficult to utilize in controlled iterative processes and present substantial

challenges for catalyst-controlled site-selectivity reversals.³⁴⁻³⁶ Additionally, our work provides a potential pathway for further protein engineering efforts with P450 homologs catalyzing multi-step oxidative cascades on a common substrate. For example, as an equivalent to TamI Leu101, dual-function P450s MycG11 and GfsF13 have active site non-polar residues Leu84 and Ala100, respectively, that have not yet been explored for their potential role in modulating selectivity and reaction step sequence. Similarly, equivalent to TamI Leu295, P450 GfsF contains an Ala residue at the 297 active site position, which could be modified to assess its function in controlling the native oxidation cascade and for modifying selectivity. Finally, in addition to investigating the mechanism towards formation of tri-oxidized **9** and **10**, future work will focus on substrate engineering efforts to expand the substrate scope of TamI as a versatile and highly selective P450 biocatalyst for iterative late-stage C-H oxidation and epoxidation of complex bicyclic and polycyclic scaffolds.

2.4 References

1. Carlson, J. C.; Fortman, J. L.; Anzai, Y.; Li, S.; Burr, D. A.; Sherman, D. H. Identification of the tirandamycin biosynthetic gene cluster from *Streptomyces* sp. 307-9. *ChemBiochem* **2010**, *11*, 564-572.
2. Meyer, C. E. Tirandamycin, a new antibiotic isolation and characterization. *J Antibiot (Tokyo)* **1971**, *24*, 558-560.
3. Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H. Tirandamycin biosynthesis is mediated by co-dependent oxidative enzymes. *Nat Chem* **2011**, *3*, 628-633.
4. Yu, Z.; Vodanovic-Jankovic, S.; Ledebor, N.; Huang, S.X.; Rajska, S.R.; Kron, M. and Shen, B. Tirandamycins from *Streptomyces* sp. 17944 inhibiting the parasite *Brugia malayi* asparagine tRNA synthetase. *Org Lett* **2011**, *13*(8), 2034-2037.
5. Ogasawara, Y.; Kondo, K.; Ikeda, A.; Harada, R.; Dairi, T. Identification of tirandamycins as specific inhibitors of the futasoline pathway. *J Antibiot (Tokyo)* **2017**, *70*, 798-800.
6. Yeom, E.; Kwon, D. W.; Lee, J.; Kim, S. H.; Lee, J. H.; Min, K. J.; Lee, K. S.; Yu, K. Asparaginyl-tRNA synthetase, a novel component of hippo signaling, binds to salvador and enhances yorkie-mediated tumorigenesis. *Front Cell Dev Biol* **2020**, *8*, 32.
7. Reusser, F. Tirandamycin, an inhibitor of bacterial ribonucleic acid polymerase. *Antimicrob Agents Chemother* **1976**, *10*, 618-622.
8. Newmister, S. A.; Srivastava, K. R.; Espinoza, R. V.; Caddell Haatveit, K.; Khatri, Y.; Martini, R. M.; Garcia-Borràs, M.; Podust, L. M.; Houk, K. N.; Sherman, D. H. Molecular basis of iterative C–H oxidation by TamI, a multifunctional P450 monooxygenase from the tirandamycin biosynthetic pathway. *ACS Catal* **2020**, *10*, 13445-13454.
9. Narayan, A. R.; Jimenez-Oses, G.; Liu, P.; Negretti, S.; Zhao, W.; Gilbert, M. M.; Ramabhadran, R. O.; Yang, Y. F.; Furan, L. R.; Li, Z.; Podust, L. M.; Montgomery, J.; Houk, K. N.; Sherman, D. H. Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. *Nat Chem* **2015**, *7*, 653-660.
10. Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J Biol Chem* **2006**, *281*, 26289-26297.
11. Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem Rev* **2004**, *104*, 3947-3980.

12. Ortiz de Montellano, P. R. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem Rev* **2010**, *110*, 932-948.
13. Fessner, N. D. P450 Monooxygenases enable rapid late-stage diversification of natural products via C-H bond activation. *ChemCatChem* **2019**, *11*, 2226-2242.
14. Lowell, A. N.; DeMars, M. D., 2nd; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A.; Schultz, P. J.; Sherman, D. H. Chemoenzymatic total synthesis and structural diversification of tylactone-based macrolide antibiotics through late-stage polyketide assembly, tailoring, and C-H functionalization. *J Am Chem Soc* **2017**, *139*, 7913-7920.
15. Cochrane, R. V.; Vederas, J. C. Highly selective but multifunctional oxygenases in secondary metabolism. *Acc Chem Res* **2014**, *47*, 3148-3161.
16. Mizutani, M.; Sato, F. Unusual P450 reactions in plant secondary metabolism. *Arch Biochem Biophys* **2011**, *507*, 194-203.
17. Espinoza, R. V.; Sherman, D. H. Exploring the molecular basis for selective C-H functionalization in plant P450s. *Synth Syst Biotechnol* **2020**, *5*, 97-98.
18. Hang, L.; Liu, N.; Tang, Y. Coordinated and iterative enzyme catalysis in fungal polyketide biosynthesis. *ACS Catal* **2016**, *6*, 5935-5945.
19. Podust, L. M.; Sherman, D. H. Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep* **2012**, *29*, 1251-1266.
20. Roiban, G. D.; Agudo, R.; Reetz, M. T. Cytochrome P450 catalyzed oxidative hydroxylation of achiral organic compounds with simultaneous creation of two chirality centers in a single C-H activation step. *Angew Chem Int Ed Engl* **2014**, *53*, 8659-8663.
21. Yang, S.; DeMars, M. D., 2nd; Grandner, J. M.; Olson, N. M.; Anzai, Y.; Sherman, D. H.; Houk, K. N. Computational-based mechanistic study and engineering of cytochrome P450 MycG for selective oxidation of 16-membered macrolide antibiotics. *J Am Chem Soc* **2020**, *142*, 17981-17988.
22. Anzai, Y.; Li, S.; Chaulagain, M. R.; Kinoshita, K.; Kato, F.; Montgomery, J.; Sherman, D. H. Functional analysis of MycCI and MycG, cytochrome P450 enzymes involved in biosynthesis of mycinamicin macrolide antibiotics. *Chem Biol* **2008**, *15*, 950-959.
23. Miyanaga, A.; Takayanagi, R.; Furuya, T.; Kawamata, A.; Itagaki, T.; Iwabuchi, Y.; Kanoh, N.; Kudo, F.; Eguchi, T. Substrate recognition by a dual-function P450 monooxygenase GfsF involved in FD-891 biosynthesis. *Chembiochem* **2017**, *18*, 2179-2187.
24. Richter, M. E.; Traitcheva, N.; Knupfer, U.; Hertweck, C. Sequential asymmetric polyketide heterocyclization catalyzed by a single cytochrome P450 monooxygenase (AurH). *Angew Chem Int Ed Engl* **2008**, *47*, 8872-8875.
25. Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H. Tirandamycin biosynthesis is mediated by co-dependent oxidative enzymes. *Nat Chem* **2011**, *3*, 628-633.
26. Grandner, J. M.; Cacho, R. A.; Tang, Y.; Houk, K. N. Mechanism of the P450-catalyzed oxidative cyclization in the biosynthesis of griseofulvin. *ACS Catal* **2016**, *6*, 4506-4511.
27. Bellucci, G.; Chiappe, C.; Pucci, L.; Gervasi, P. G. The mechanism of oxidation of allylic alcohols to alpha,beta-unsaturated ketones by cytochrome P450. *Chem Res Toxicol* **1996**, *9*, 871-874.
28. Carlson, J. C.; Li, S.; Burr, D. A.; Sherman, D. H. Isolation and characterization of tirandamycins from a marine-derived *Streptomyces* sp. *J Nat Prod* **2009**, *72*, 2076-2079.
29. Yoshimura, H.; Takahashi, K.; Ishihara, J.; Hatakeyama, S. Unified synthesis of tirandamycins and streptolydigin. *Chem Commun (Camb)* **2015**, *51*, 17004-17007.
30. Takahashi, K.; Harada, R.; Hoshino, Y.; Kusakabe, T.; Hatakeyama, S. and Kato, K. Formal synthesis of tirandamycin B. *Tetrahedron* **2017**, *73*(25), 3548-3553.
31. Yadav, J. S., Dhara, S. and Mohapatra, D.K. Stereoselective total synthesis of 10-epi-tirandamycin E. *Tetrahedron* **2017**, *73*(10), 1358-1366.
32. Ro, D. K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J. and Chang, M.C. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *440*(7086), 940-943.
33. Pikuleva, I. A.; Bjorkhem, I.; Waterman, M. R. Expression, purification, and enzymatic properties of recombinant human cytochrome P450c27 (CYP27). *Arch Biochem Biophys* **1997**, *343*, 123-130.
34. Verras, A.; Alian, A.; de Montellano, P. R. Cytochrome P450 active site plasticity: attenuation of imidazole binding in cytochrome P450(cam) by an L244A mutation. *Protein Eng Des Sel* **2006**, *19*, 491-496.
35. De Voss, J. J.; Ortiz de Montellano, P. R. Substrate docking algorithms and the prediction of substrate specificity. *Methods Enzymol* **1996**, *272*, 336-347.

36. Dickens, M. L.; Priestley, N. D.; Strohl, W. R. *In vivo* and *in vitro* bioconversion of epsilon-rhodomyconone glycoside to doxorubicin: functions of DauP, DauK, and DoxA. *J Bacteriol* **1997**, 179, 2641-2650.

Chapter 3*

Multiple Activated Oxygen Species in P450 TamI Iterative Catalysis

3.1 Background

3.1.1 Involvement of Various Oxidizing Species in P450 Catalysis

The large superfamily of heme-containing cytochrome P450 enzymes catalyze a myriad of important chemical reactions including, but not limited to, aromatic and aliphatic hydroxylations, epoxidations, ring formation and C-C bond cleavage. Not only are the P450-catalyzed reactions critical in the degradation of xenobiotic compounds, but they are essential for the biosynthesis of structurally complex bioactive molecules. These versatile monooxygenases can perform precise chemical reactions with high levels of selectivity while overcoming innate substrate reactivity.

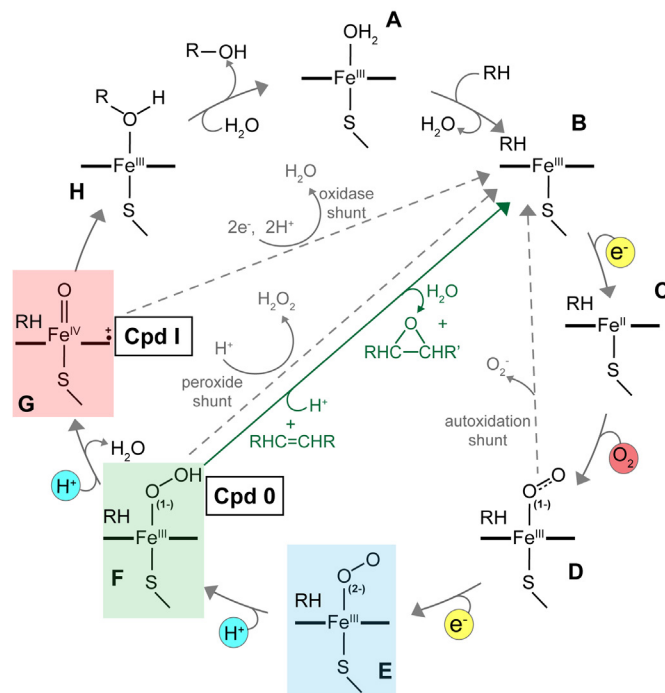
When a small molecule substrate binds to the P450 active site, it displaces the water molecule acting as a sixth ligand of the heme iron and shifts the enzyme from a low-spin system to a high-spin ferric state (Scheme 3.1, iron species A to B). This shift increases the redox potential (100-230 mV) and facilitates the first electron transfer from the redox partners (species C). Molecular oxygen then binds to the ferrous enzyme generating the oxy-ferrous intermediate (species D) that can accept a second electron transfer to form the key ferric-peroxo anion intermediate (FeO^{2-} , species E). Subsequent proton transfer (typically supplied by the bulk solvent) to the distal oxygen atom is facilitated by a highly conserved active site acid-alcohol pair and bound water molecules leading to the ferric hydroperoxide Compound 0 (Cpd 0, FeO_2H). A second protonation event facilitates the O-O cleavage step generating the high-valent iron-oxo species Compound I (Cpd I), resulting in release of a water molecule. Found in the oxygen-binding groove of the I-helix,¹ the acid-alcohol pair is thought to be essential for the two sequential protonation events necessary to form Cpd I from the ferric-peroxo species, Cpd 0. However, a few exceptions exist where an alanine or asparagine occur in place of the alcohol residue such as in the case of

* Chapter 3 is included in a manuscript. **Espinoza, R.V.**, Maskeri, M.A, Turlik, A., Nangia, A., Khatri, Y., Montgomery, J., Houk, K.N. and Sherman, D.H. Epoxidation and late-stage C-H functionalization by P450 TamI is mediated by variant heme-iron oxidizing species. *ChemRxiv*, **2021**. Under review in *ACS Cat*.

P450eryF² and P450cin³ where the substrate hydroxyl is thought to act as the alcohol residue and asparagine is believed to retain a water molecule, respectively. Moreover, throughout the P450 catalytic cycle, the peroxo states can undergo uncoupling and release peroxides byproducts while regenerating the ferric form of the enzyme.

The remarkable myriad of chemical reactions catalyzed by P450 enzymes poses the question of whether these biocatalysts use different oxidizing species for each different oxidation or whether the same catalytically active species is utilized. Thus, identification of the oxidant species in P450

catalysis remains a topic of high interest. Nowadays there are a variety of approaches, including kinetics, substrate atom labeling, spectroscopy and computational calculations that provide evidence for the contribution of the different oxygen species to the mechanistic pathway under examination. For example, the ferric-peroxo anion intermediate (species E) can perform a nucleophilic attack on the substrate leading to C-C bond cleavage as observed in the P450 CYP17A1 involved in human steroid hormone biosynthesis.⁴ Alternatively, the ferric peroxide species can be protonated to form Cpd 0 (species F), which enables double bond oxygen insertion for epoxidation reactions such as in the case of P450 PimD in the pimaricin biosynthesis where the substrate's topology does not favor catalysis via Cpd I.⁵ Alternatively, double protonation leads to the high-spin intermediate Cpd I (species G), the catalytic species considered to be the primary oxidant for most of P450-catalyzed oxidations, which include C-H hydroxylation by P450 CYP19A1.⁶⁻⁸ Fewer experimental and computational efforts have been dedicated to interrogate the active oxidizing species in P450-catalyzed iterative oxidation cascades involving mechanistically distinct pathways and multiple carbon sites within a substrate. In essence, it could be hypothesized that iterative P450 enzymes may have the ability to shift from one catalytically active species to

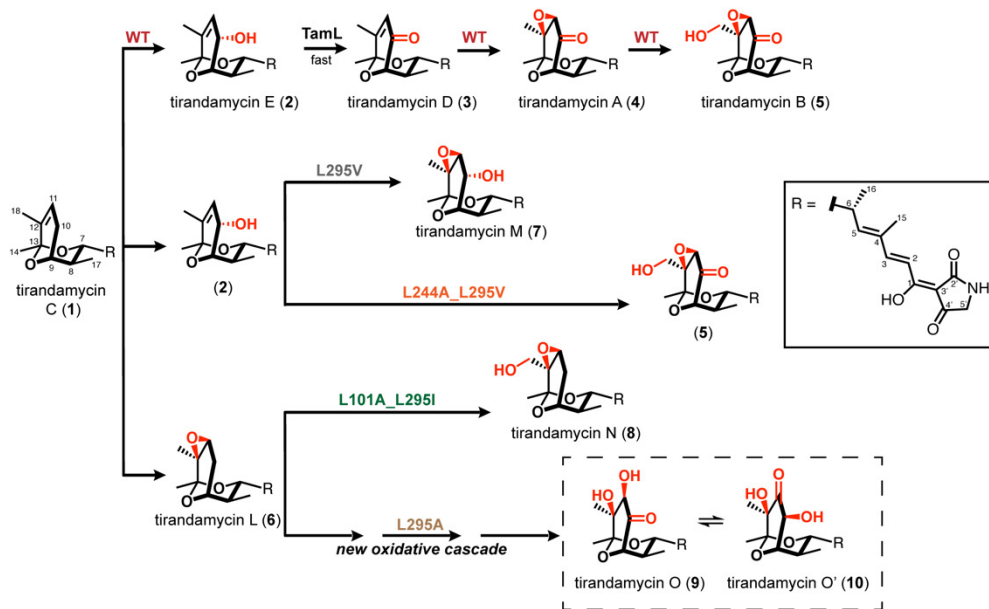


Scheme 3.1. General cycle for cytochrome P450 catalysis. The three unproductive uncoupling pathways are shown in dashed arrows.

another to modulate their iterative oxidation sequence with exquisite selectivity on a common substrate. Elucidating the oxidant species involved in P450-based multistep biocatalytic systems may inspire the engineering of homologous enzymes to expand the range of oxidation reaction types they can catalyze in a cascade fashion and driven by catalyst-controlled reactivity.

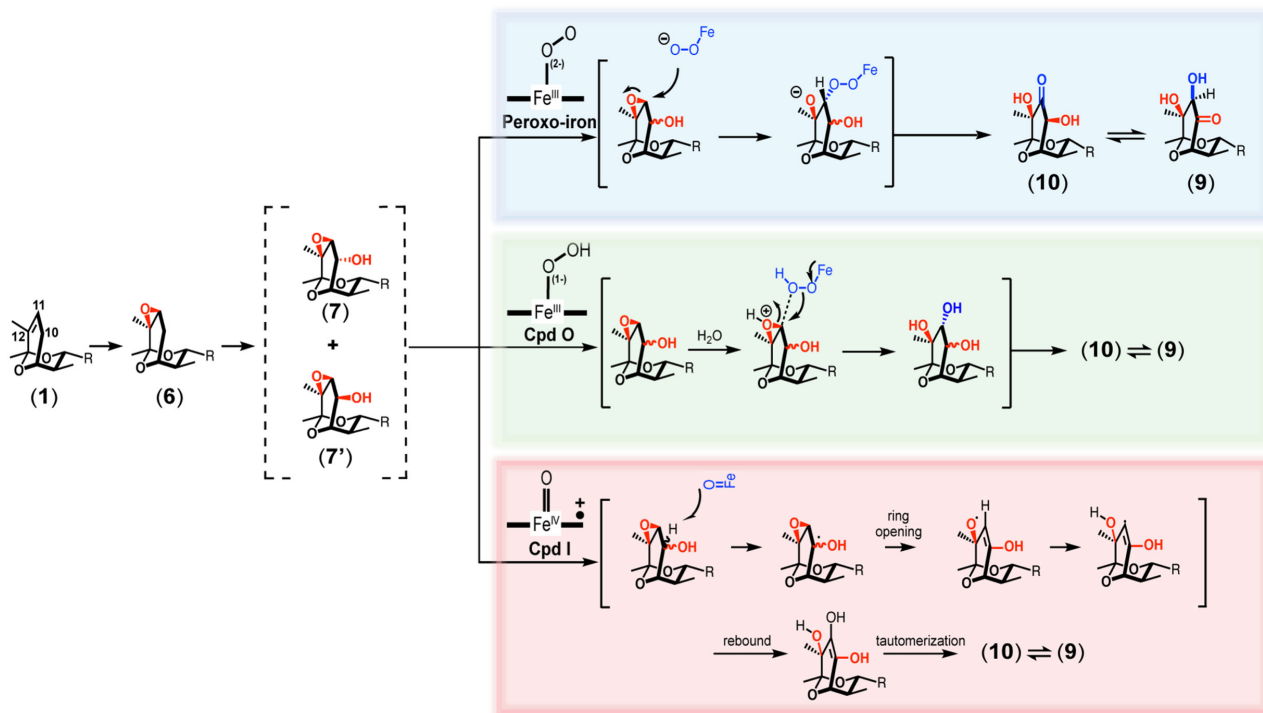
3.1.2 Proposed Mechanisms in P450 TamI Oxidative Cascade Supporting Distinct Functional Oxidants

The *Streptomyces*-derived P450 TamI is an iterative enzyme responsible for the late-stage oxidation of tirandamycin C (**1**) to B (**5**), the terminal product of the tirandamycin biosynthetic pathway and the most potent congener in bioactivity against an array of human pathogens including vancomycin resistant-*Enterococci faecalis* (Scheme 3.2). As a result of intrinsic substrate reactivity, P450 TamI performs methylene C10 hydroxylation, C11/12 alkene epoxidation and C18 methyl hydroxylation, in a strict sequence.⁹ We recently described the engineering of P450 TamI biocatalysts that alter the iterative process, site-selectivity and number of reactions performed in a catalyst-controlled fashion, leading to the enzymatic synthesis of five novel tirandamycin congeners including tri-oxidized tirandamycin O (**9**) and O' (**10**) (Scheme 3.2).¹⁰ The role that variant reactive oxygen species play in mediating the iterative sequence and flexibility of TamI and its mutants for catalyzing mechanistically divergent oxidative reactions has not been previously established. It is this gap in knowledge that motivated the current study, with the expectation that deeper insights may facilitate rational design for bioinspired catalysis.



Scheme 3.2. Late-stage oxidation catalyzed by P450 TamI variants. Naturally-occurring tirandamycin include **1**, **2**, **3**, **4** and **5**. New metabolites generated from TamI mutants include **6**, **7**, **8**, **9**, and **10**.

Based on the stepwise oxidative sequence exclusively catalyzed by P450 TamI L295A to convert **1** \rightarrow **6** \rightarrow \rightarrow **9** and **10** (Scheme 3.2) and the unusual oxidation pattern of the tri-oxidized terminal products, we hypothesized that this TamI variant may have the ability to shift from one catalytically active oxidant species to another. First, TamI L295A catalyzes a C11/12 double bond epoxidation on **1** rather than a C10-(*S*) hydroxylation as seen with the WT, indicating a possible Cpd 0-mediated mechanism.^{5, 11} The inherent reactivity of **6** was explored for primary C18 and secondary C10 hydroxylation reactions.¹⁰ Hydroxylation at both positions exhibited similar energetic barriers, suggesting formation of either C-H functionalization product is favorable when using **6** as substrate, with a slight preference for C10 oxidation (0.6 kcal/mol difference). We rationalized that hydroxylating the C10 position on **6** makes the C11 site more electrophilic and prone to opening of the epoxide. As previously established, C10 hydroxylation via Cpd I-mediated catalysis is anticipated.¹² Thus, we hypothesized that TamI L295A catalyzes a C10 hydroxylation on **6** to form **7** and/or **7'** (the C10 epimer of **7**, Scheme 3.3). However, these proposed transient intermediates are not detected in *in vitro* enzymatic assays when incubating **1** or **6** with the enzyme.¹⁰ This could imply that once **6** is oxidized to **7** or **7'**, the hydroxy-containing intermediate is not released from the enzyme active site pocket but is instead instantly oxidized to obtain the



Scheme 3.3. Proposed mechanisms for the oxidative cascade catalyzed by TamI L295A. Three mechanisms are described using the nucleophilic peroxo-iron (blue), Cpd 0 (green), and Cpd I (red) as oxidant species.

final products. Given that congener **7** is available as an authentic standard (obtained from preparative-scale enzymatic reactions),¹⁰ enzymatic reactions with **7** and TamI L295A were performed. Incubating **7** with TamI L295A at increased catalyst loading conditions led to low-level formation of **9** and **10** (Figure S14, lane 14).¹⁰ These results provide compelling evidence that **7** is a transient intermediate towards **9** and **10** under these experimental conditions. The nominal production of the terminal products and the lack of spin shift observed upon titration of **7** to the enzyme suggest that **7** might not be the preferred transient intermediate or that the additional oxidative decorations in **7** impact the substrate binding affinity making it a poorer fit in TamI L295A active site. Computational analysis showed that **7'** exhibits an internal hydrogen bond from the hydroxy to the epoxide, stabilizing its configuration and activating the epoxide for opening (Figure S15). This might suggest that **7'** could be a preferred transient intermediate in the L295A-catalyzed pathway, however this congener is not available as an authentic standard. We suspected that **4** might also be an intermediate in this oxidative cascade, however no product was observed when incubating **4** with TamI L295A, even at increased catalyst loading. Hence, we postulate that TamI L295A hydroxylates **6** at C10 forming intermediates **7** and/or **7'** prior the oxidative cascade leading to **9** and **10**.

After C10 hydroxylation, TamI L295A catalyzes an epoxide-opening cascade not previously observed with P450 TamI chemistry. This may indicate the potential involvement of different oxidants en route to **9** and **10**. Hence, we proposed three potential mechanisms for the formation of **9** and **10** by TamI L295A including: 1) Attack by the nucleophilic ferric-peroxo to open the epoxide at the C11 site yielding the ketone (Scheme 3.3, blue box); 2) Incorporation of the hydroperoxide Cpd 0 leading to C11 hydroxylation and formation of the α -hydroxy ketone (green box); 3) H10-abstraction by the electrophilic higher valent metal-oxo species Cpd I generating a free radical that leads to epoxide ring opening (red box). Thus, we sought to explore the identity of the active oxidant species in P450 TamI L295A responsible for catalyzing the iterative oxidation cascade from **1** \rightarrow **6** \rightarrow **7** or **7'** \rightarrow **9** and **10**, and three proposed mechanisms towards the exclusive formation of **9** and **10** from **6**. In this effort, we 1) disrupted the proton relay network required for dioxygen activation with site-directed mutagenesis, 2) assessed the utility of exogenous oxygen surrogates and reactive oxygen species (ROS) scavengers to support P450 TamI catalysis and 3) calculated kinetic solvent isotope effects (KSIEs) on the steady-state turnover of P450 TamI in its hydroxylation, epoxidation and unusual oxidative cascade

functionalities. Additionally, quantum mechanics (QM) calculations using density functional theory (DFT) calculations were conducted to provide insights into the contribution of the plausible reaction mechanisms and different oxidant species in the P450 catalytic cycle for generating the structurally unique metabolites **9** and **10**.

3.2 Mechanism of Epoxidation via Cpd **0** to Form Tirandamycin L (**6**)

3.2.1 Structure-Derived Mechanism

In P450 TamI, the WT enzyme catalyzes the C10 allylic hydroxylation of **1** to generate **2**. The substrate-bound crystal structure of the enzyme revealed that the C10-H (*S*) σ -bond points towards the heme iron within a 4 Å distance, prone for hydroxylation to occur via hydrogen abstraction and oxygen rebound mechanism.⁹ This C-H functionalization reaction occurs likely through involvement of the electrophilic species Cpd I, as previous experimental and theoretical work has undoubtedly concluded for the mechanism of P450-catalyzed hydroxylation reactions.¹² However, a subtle variation in the active site of mutant TamI L295A changed reactivity from C10 hydroxylation to C11/12 epoxidation on **1** to form **6** suggesting the enzyme is controlling this divergence in chemical output. Previous DFT calculations with **1** revealed that the transition state barrier for C11/12 (*R/S*) epoxidation is highest in energy (19.2 kcal/mol) compared to hydroxylation at C10 or C18 sites (14.5 and 16.6 kcal/mol, respectively).⁹ Thus, we

hypothesized that TamI L295A may reorient the substrate to lower the activation barrier and/or favor a different catalytically active oxidant species to facilitate the least favored epoxidation reaction over the C-H functionalization product.

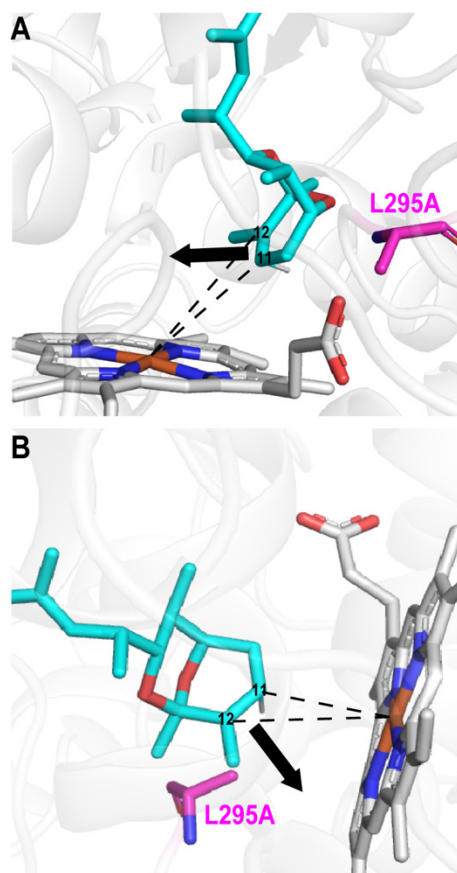


Figure 3.1. Molecular Dynamics (MD) simulations of TamI L295A with **1.** Two different views emphasizing the orientation of the C11/12 alkene are depicted. Black arrows show the orientation of the C11/12 π -orbitals.

Close examination of the 5 top occupied clusters of the MD simulations with TamI L295A and **1** revealed that the π -orbitals of the conjugated C11/12 system point away from the iron at a 115.9° angle to the heme (Figure 3.1).¹² This differs from the angles calculated in simulations with TamI WT and **3** and TamI L101A_L295I and **1** where diastereoselective C11/12 alkene epoxidation is also observed experimentally. In the latter cases, the C11/12 alkene of **3** and **1** points towards the iron at a 90-93° angle to the heme facilitating double bond epoxidation, in agreement with *in vitro* results (Figure S1). Contrarily, in TamI L295A, the geometry of **1** does not favor epoxidation by a synchronous trajectory of oxygen insertion or a charge-transfer complex via a radical or cationic pathway. Instead, the proximity (< 4 Å) of the C11-H σ -bond to the heme iron seems favorable for Cpd I-mediated hydroxylation. However, as noted with P450 PimD and the pimaricin substrate, the energy required to abstract a hydrogen from a vinyl carbon is prohibitive. This led us to hypothesize that the hydroperoxo Cpd 0 species may be the active oxidant species in TamI L295A for the first reaction step of its iterative oxidative cascade (Scheme 3.2), where a concerted insertion of the distal oxygen into the C11/12 double bond of **1** occurs to generate **6**. The lengthened O-O bond in the hydrogenated peroxide species (Cpd 0) would bring the distal oxygen even closer to the C11/12 π -orbitals of **1** enabling a productive topology of the substrate for C11/12 epoxidation to take place over C10 hydroxylation.

3.2.2 Catalytic Evidence and Mutagenesis Experiments

In P450 TamI, the conserved acid-alcohol pair is composed of Glu251 and Thr252, yet another Thr253 resides at the adjacent position that could contribute to the I-helix proton relay network or supplant Thr252 if absent.¹³ Mutation of the conserved alcohol residue in P450s is thought to interrupt the second proton transfer, accumulating the Cpd 0 species, and diminishing substrate oxidation rates while increasing uncoupling of the enzyme and hydrogen peroxide formation.^{9, 11, 13} For example, P450cam typically hydroxylates its native substrate camphor with high levels of efficiency. However, its variant T252A showed little to no hydroxylated product and instead, was found capable of performing olefin epoxidation.¹¹ This result was presented as experimental evidence for the reactivity of a hydroperoxo-ferric Cpd 0 species in the P450cam system.

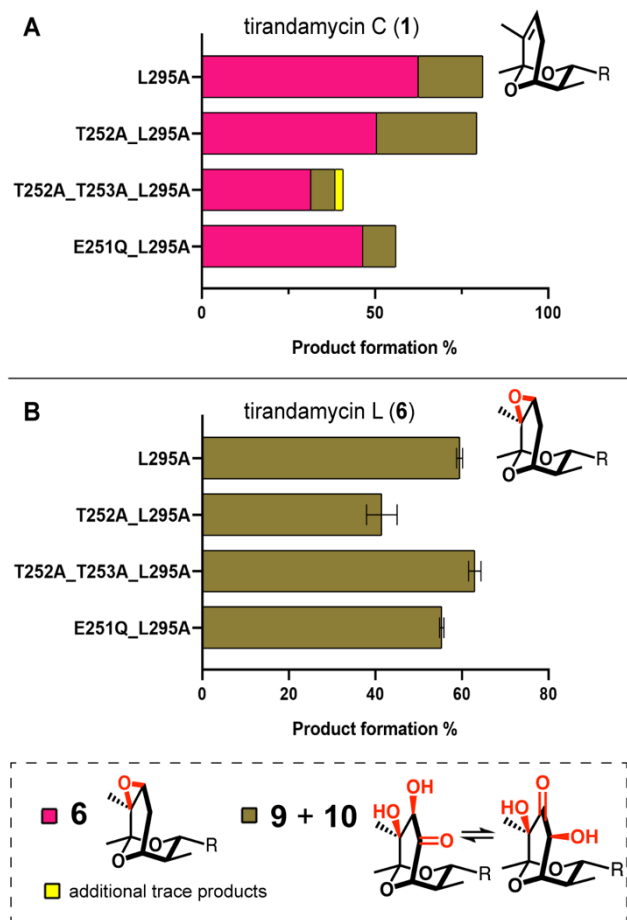


Figure 3.2. End-point assays with TamI variants. (A) Reactions with substrate **1**; (B) Reactions with substrate **6**. Substrates are shown next to each bar graph. Ratios of product formation were estimated based on HPLC traces. For definition of R, see Scheme 3.2. For additional reactions, see SI.

Moreover, the double and triple variants displayed tighter substrate binding affinity values compared to TamI L295A (See Appendix B), showing that mutating the threonine residues does not impact substrate binding in an unproductive manner.

These mutagenesis results show that altering the conserved alcohol pair in TamI to an alanine that cannot longer accept a hydrogen bond from Cpd 0 to drive O-O cleavage and form Cpd I does not inhibit catalytic activity. In fact, similar product formation to L295A is observed for C11/12 epoxidation of **1** to **6**, suggesting that Cpd 0 may be a viable oxidant for L295A-catalyzed epoxidation. This is further supported by the observation that increased hydrogen

We created threonine mutants of P450 TamI L295A to probe their role in oxygen activation and catalysis (Figure 3.2). When incubated with **1**, variant T252A_L295A showed a similar product formation to TamI L295A generating a mix of **6**, **9** and **10**, while triple mutant T252A_T253A_L295A showed decreased yields. Both threonine mutants displayed reduced hydrogen peroxide formation compared to TamI L295A suggesting less free radicals occur in these systems (See Appendix B). This was further supported by the observation that the TamI L295A-catalyzed oxidation of **1** to **6**, **9** and **10** increments in the presence of ROS scavengers (ascorbate, catalase, and dismutase) in a dose-dependent fashion, indicating uncoupled formation of free radicals in TamI L295A (See Appendix B). Nonetheless, this increase in product formation could also derive from the fact that these chemical scavengers are able to protect the enzyme, substrate and/or products

peroxide formation was detected for L295A, which could derive from accumulated Cpd 0 undergoing uncoupling through the peroxide shunt pathway (Scheme 3.1).

3.2.3 Use of Oxygen Surrogates in TamI-Catalyzed Reactions

Artificial oxygen surrogates that may distinguish among P450 mechanisms were employed. Hydrogen peroxide can provide the oxygen atom and electrons necessary to overcome the need for molecular oxygen and NADPH-derived reducing equivalents in P450s.¹⁴ Therefore, direct addition of hydrogen peroxide to ferric P450 is expected to form ferric hydroperoxide Cpd 0. However, it is important to note that Cpd 0 can subsequently and irreversibly convert to Cpd I or undergo reversible proton transfer to form peroxo-iron (species E). Therefore, its role in TamI catalysis can only be postulated. Alternative organic peroxides such as iodosobenzene would not be able to generate Cpd 0 and insert the distal oxygen atom in the substrate. Thus, we hypothesized that if Cpd 0 is the preferred active oxidant then the H₂O₂-initiated reactions would form noticeable amounts of epoxide-containing **6** when incubating TamI L295A with **1**.

As anticipated, significant production of **6** was detected for reactions initiated with H₂O₂ and TamI L295A (Figure 3.3). Triple variant T252A_T253A_L295A also generates product **6** under these reaction conditions, albeit in diminished yields. Reactions with freshly prepared iodosobenzene led to nominal trace product formation of **6** further supporting our hypothesis. To interrogate whether the ability to activate H₂O₂ as a cosubstrate for epoxidation transformations is exclusive to biocatalyst TamI L295A or not, additional reactions with variant TamI L101A_L295I and **1** were tested. Control reactions containing redox partners yielded formation of **6** and **8** from **1** as previously observed (Scheme 3.2),¹⁰ while no oxidative activity was observed with H₂O₂-initiated reactions. This indicated that unlike TamI L295A, variant L101A_L295I is not capable of using Cpd 0 for epoxidation catalysis. Together with computational calculations and mutagenesis results, our catalytic data suggests that TamI L295A has the flexibility to use Cpd 0 as a viable oxidant for efficient alkene epoxidation of **1** generating intermediate **6**.

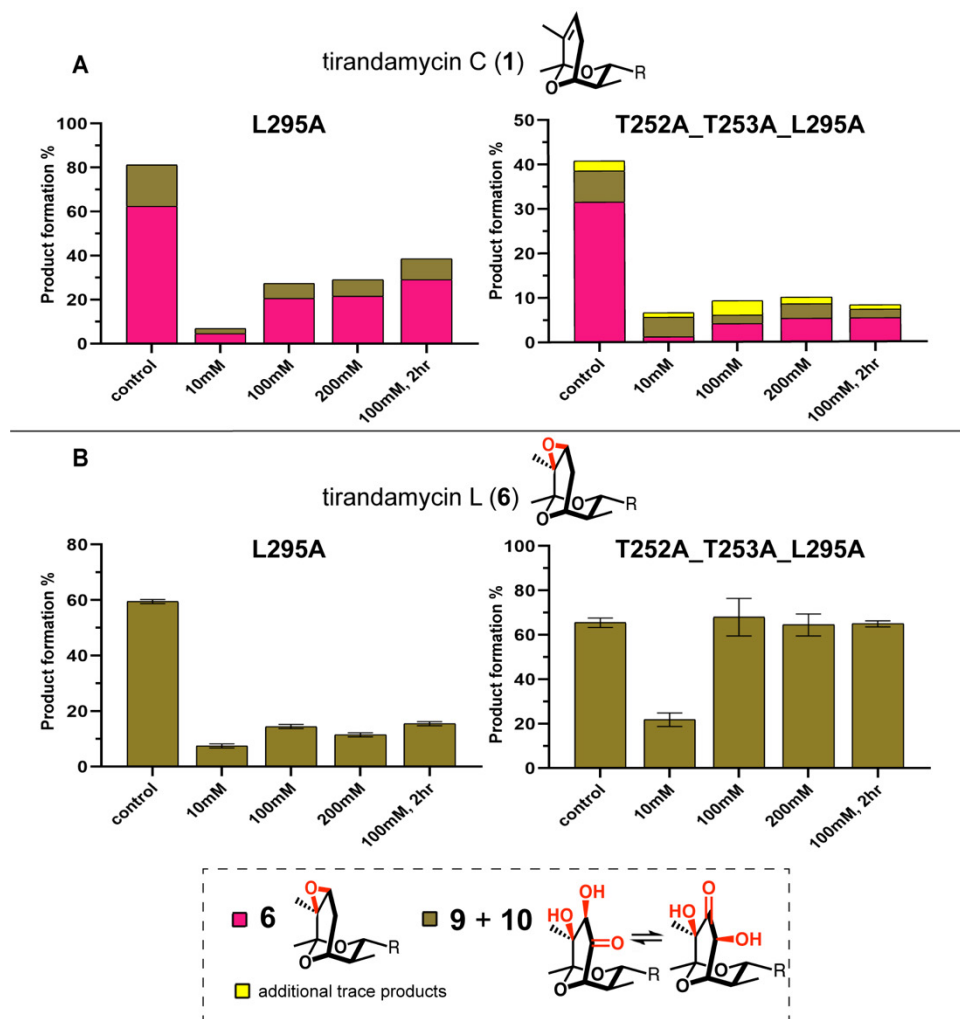


Figure 3.3. TamI-catalyzed oxidations using H₂O₂ as a cosubstrate. Reactions were initiated with increasing concentrations of H₂O₂ and run for 30 min unless otherwise stated. Substrate is shown above bar graphs. For reactions initiated with iodosobenzene, see SI.

3.3 Iterative Oxidation Cascade to Produce Tirandamycin O and O'

3.3.1 Evidence for a Reactive, Nucleophilic Peroxo-Iron Oxidant

We hypothesize that biocatalyst TamI L295A uses the ferric hydroperoxo Cpd 0 to catalyze epoxidation of **1** → **6**, followed by Cpd I species catalyzing **6** → **7** or **7'**, and then shifts to a different oxidant species to open the epoxide of **7** or **7'** to form **9** and **10**. Inspired by previous literature,⁴ we hypothesized the ferric-peroxo anion (species E) could perform a nucleophilic attack at C11 on intermediate **7** or **7'**, opening up the epoxide, retaining stereoselectivity at C12 and forming a C11 ketone either through radical or proton abstraction mechanisms (Scheme 3.3, blue

box), generating **9** and **10**. To interrogate the plausibility of a peroxo anion-catalyzed oxidative cascade, we performed catalytic assays using **6** as the substrate.

Previous work with P450cam, P450cin, CYP19A1, P450-BM3, and CYP17A1 suggest that mutating the conserved Glu251 acid residue disrupts the first proton transfer in the P450 catalytic cycle, favoring formation of the nucleophilic peroxo-iron species (Scheme 3.1, species E) and drastically reducing substrate oxidation activity. Variant E251Q_L295A converts **6** to tri-oxidized **9** and **10** in comparable yields to L295A (Figure 3.2). The glutamic acid mutant displays a lower rate of hydrogen peroxide formation compared to L295A with **6**, while no spin shift is observed with either enzyme. Adding ROS scavengers to the reaction of E251Q_L295A and **6** led to a modest decrease in conversion compared to L295A. This could suggest that when using **6** and the glutamic acid variant as a biocatalyst, products **9** and **10** may result from shunt pathways. However, the response is not dose-dependent. Mutating the conserved acid residue to a neutral glutamine does not abolish catalytic activity. Instead, product formation is comparable to TamI L295A. This data indicates that the glutamic acid residue is not critical for the oxidative cascade catalyzed by TamI L295A, suggesting that the peroxo-iron could compensate as a viable oxidant for the formation of **9** and **10** when Cpd I may be compromised.

Moreover, kinetic solvent isotope effects (KSIE) experiments were performed to address the identity of the iron-oxo species and variant mechanisms involved. Formation of Cpd I in the P450 catalytic cycle is dependent on the sequential transfer of two protons, a process that typically slows down when hydrogen has been replaced by deuterium. Hence, comparing the steady-state kinetic parameters of TamI-catalyzed reactions in H₂O versus D₂O-based buffers can be mechanistically informative.⁶ Using saturating substrate concentrations, product formation and NADPH oxidation rates of P450 TamI were measured at pH/pD 7.4 (see Appendix B for details). Deuterated samples were obtained by exhaustive exchange of proteins in D₂O. The production of **2** from **1** occurred at a rate of 9.70 min⁻¹ for TamI WT in H₂O, while the rate lowered to 3.17 min⁻¹ in D₂O (Figure 3.4A). These values correspond to a significant KSIE of $K_h/K_d = 3.06$ for C10 hydroxylation, similar to what has been previously reported for other P450 systems catalyzing hydroxylation chemistry via Cpd I.⁶ When testing **6** with TamI L295A and triple variant T252A_T253A_L295A, the rate of product formation increased in the D₂O-based buffer, yielding a repeatable, small inverse KSIE of 0.89 and 0.79, for the multi-step cascade responsible for **9** and **10** (Figure 3.4A). Although the potential influence of deuterium solvent on enzyme function are

complex, these results implicate the role of peroxo-iron as a potential oxidant species catalyzing this cascade.¹⁶ Alternative explanations include the possibility that protonation and deprotonation can occur throughout the amino acid chains of the P450 or that the substrate's hydroxy group in the tetramic acid moiety exchanges with deuterium influencing substrate binding.

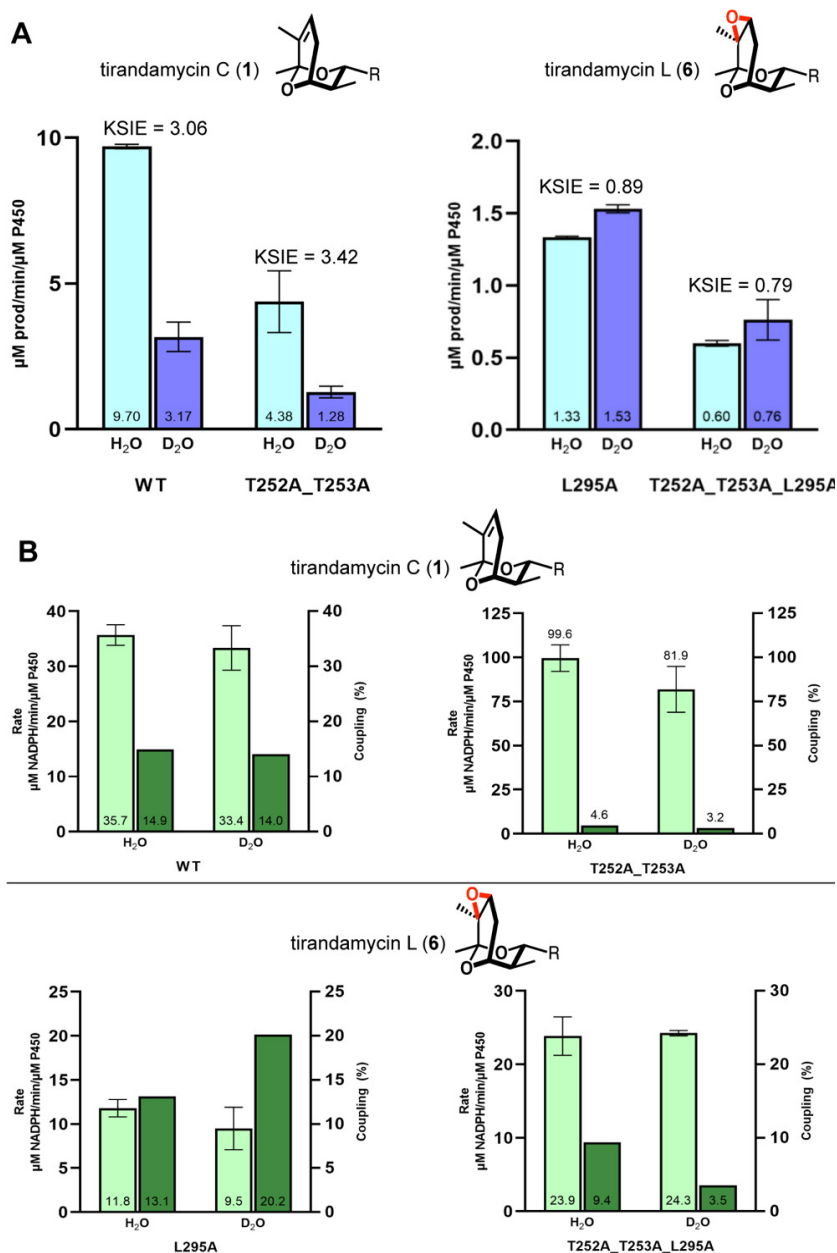


Figure 3.4. End-point assays with P450 TamI in H₂O vs D₂O buffers. Substrates are shown above bar graphs. (A) Steady-state Kinetic Solvent Isotope Effects (KSIE) observed for P450 TamI with **1** and **6**. $KSIE = K_h/K_d$. The number inside the bar displays the value of catalytic activity. (B) Comparison of NADPH oxidation (light green) and coupling efficiency (dark green) by P450 TamI. The number inside the bar represents the corresponding calculated value. For exact values, refer to Table S2.

TamI L295A shows an approximate 50% increase in coupling efficiency in D₂O versus in H₂O with slightly lowered NADPH consumption values when using **6** (Figure 3.4B). Meanwhile, triple mutant T252A_T253A_L295A displayed a two-fold increase in NADPH consumption compared to L295A, with lowered coupling values. Together with KSIE data, a slight decrease in NADPH oxidation rates for the iterative oxidation of **6** → **9** and **10** with TamI L295A do not fully support a Cpd I-driven mechanism. Instead, these results suggest that inhibiting protonation of the peroxy-iron intermediate can promote productive oxidation rather than uncoupling to peroxide byproducts.

We sought to further investigate the potential involvement of a peroxy-iron species using computational methods. We conducted DFT calculations on a previously-validated truncated model of the P450 TamI heme and tirandamycin spiroketal.¹⁰ Oxidation of congeners **7** and **7'** at the UB3LYP-D3(BJ)/def2-TZVPP/SMD(diethylether)//UB3LYP-D3(BJ)/def2-SVP/SMD(diethylether) level of theory and via peroxy-iron (species E) was explored.^{4, 16-20} Computations proceeding from **7'** were found to be slightly lower in free energy (~1 kcal/mol) than those from **7** (see Appendix B). Transition state barriers for the addition of the peroxy-iron terminal oxygen into **7** and **7'** epoxides were located for the doublet (²TS2) and quartet (⁴TS2) spin states, and favor the doublet with a small preference for **7'** (13.0 kcal/mol) (Figure 3.5A). This barrier exhibits a low energy approximately equal to that located for Cpd I-facilitated pathways (Figure 3.5B). However, computational elaboration of this pathway proved challenging; while the endoperoxide intermediate **Int2** resulting from epoxide addition was successfully located for both **7** and **7'** doublet states, we were unable to converge downstream product complexes or transition structures (see Appendix B).

Moreover, we reasoned that our experimental results do not exclude the possibility of the superoxy-iron complex (species D) acting as oxidant for a radical epoxide addition.^{22, 23} Such a species is expected to persist in the TamI active site prior to an electron transfer event involving ferredoxin, and is only approximately 1.7 kcal/mol higher in energy than the peroxy-iron complex. Open-shell singlet, triplet, and quintet states were computed for the initial heme complex, of which the triplet pathway was lowest in energy. However, addition of the triplet superoxy-iron species into the **7'** epoxide exhibits a transition state barrier of 24.3 kcal/mol relative to starting material (Figure S18). As this barrier energy is significantly higher than those computed for peroxy-iron pathway, we conclude that superoxy-iron is unlikely to be a viable oxidant for this pathway.

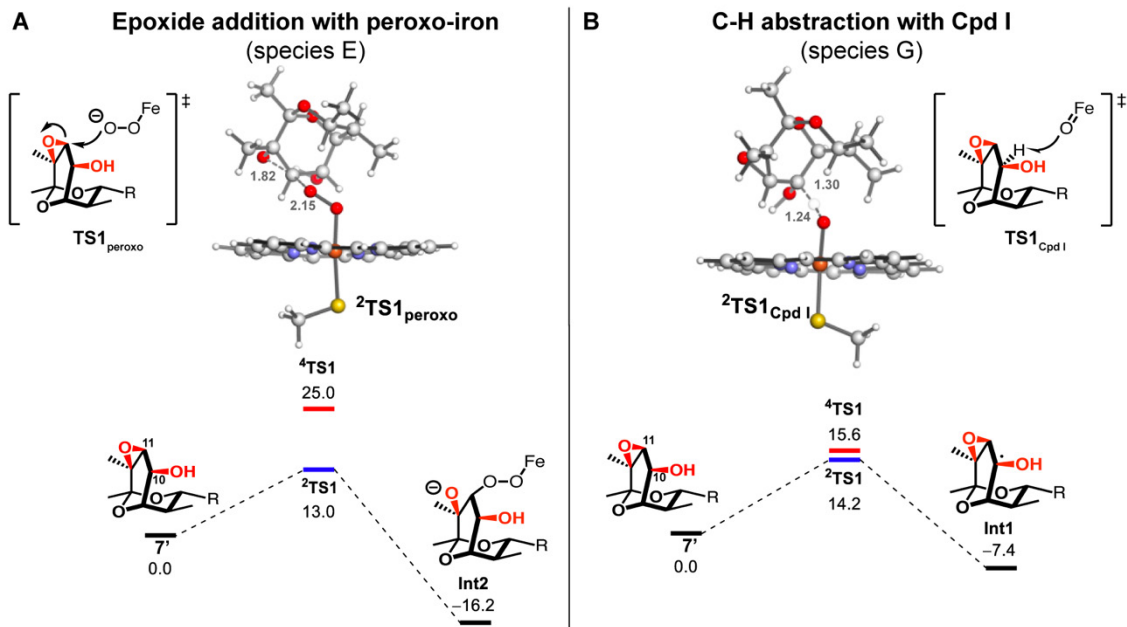


Figure 3.5. DFT calculations on P450 TamI and 7' (epi-tirandamycin M). Gibbs free energy values are shown in kcal/mol. The charge of the peroxy complex is -2. For additional calculations, see Appendix B.

3.3.2 Assessing Cpd 0 Reactivity

Another pathway that we investigated was C11 C-H hydroxylation of 7' by Cpd 0 (species E) to form 9 and 10. A reasonable transition state could not be found for the initial C-H abstraction reaction. Given the favorable energetics for the Cpd I-facilitated hydrogen abstraction process and the activity of TamI in mutants exhibiting a disrupted proton transfer chain, we considered a pathway featuring C11-H atom abstraction by a ferric-peroxy or ferric-superoxo species in both high-spin and low-spin configurations. A pathway featuring C11-H atom abstraction from 7' to drive incorporation of the distal oxygen of Cpd 0 and hydroxylation at C11 was explored. Ferric peroxy and ferric superoxo species in both high-spin and low-spin configurations were considered. In this pathway, reaction of the substrate with either iron species involves a hydrogen atom abstraction by the distal oxygen, generating a carbon-centered radical. This is followed by concerted O-O bond homolysis and OH transfer to the radical. Subsequent deprotonation of the hydroxylated intermediate with epoxide ring opening ultimately leads to the formation of 9 and 10. The barriers for C-H abstraction are calculated to be 32.8 and 31.8 kcal/mol for the triplet and

quintet ferric-superoxide species, respectively (see Appendix B). Given the favorable energetics calculated for the nucleophilic peroxo-iron (species E) or Cpd I (species G), these computations show that hydrogen atom abstraction by either superoxo species is highly unlikely. An attempt was also made to identify either a high-spin or low-spin peroxo species that participates in the initial hydrogen atom abstraction, but no converged transition-state geometries were located.

End-point assays with substrate **6** and TamI variants L295A and T252A_T253A_L295A show comparable formation of **9** and **10** (Figure 3.2B). The threonine mutant shows relatively similar hydrogen peroxide formation rates compared to TamI L295A. No spin shift is observed with any of the TamI L295A variants bound to **6**. When using H₂O₂ as a co-substrate, reactions with T252A_T253A_L295A and **6** show comparable product formation to controls, while reactions with L295A and **6** display noticeable but decreased product formation. When initiating reactions with iodosobenzene and **6**, no product is detected for any TamI variant. Adding ROS scavengers to the L295A-catalyzed reactions with **6** led to comparable product formation to controls for T252A_L295A and L295A. A slight decrease was observed for T252A_T253A_L295A indicating that the products in this reaction might result from shunt pathways, although a dose-dependent pattern was not found. Although this catalytic data does not conclusively discard Cpd 0 as a viable oxidant for catalyzing the production of **9** and **10**, the significantly high energy requirements calculated for Cpd 0-mediated catalysis do.

3.3.3 Evidence for a Cpd I-Mediated Mechanism

Although the end-point assays described above shed light to the contribution of various oxidizing species in L295A catalysis, they do not eliminate the possibility of a Cpd I-mediated mechanism. Thus, oxidative pathways via Cpd I species were explored computationally. Previous QM/MM investigations by Shaik et al. determine the Cpd I and Cpd 0 have similar energies, which enables relative comparison of the energetics of the oxidant series.²⁴ First, the pathway for C-H abstraction of **7** with Cpd I was calculated. With **7** as the substrate, the TS barrier for C-H abstraction is calculated to be 15.3 kcal/mol for the doublet state, and 16.1 kcal/mol for the quartet (Figure S20). The resulting product is 9.3 kcal/mol lower in energy than the starting material. The energy of the subsequent ring opened product is -13.1 kcal/mol in ΔG , and the following H-atom abstraction from the resulting O radical to form Int3 has $\Delta G = -2.5$ kcal/mol. OH rebound from the Fe-OH species results in Int4 ($\Delta G = -72.0$ kcal/mol). Tautomerization affords **10** and **9**, at -75.9 and -74.2 kcal/mol, respectively. The analogous C-H abstraction was also calculated for **7'**

(Figure 3.5B). This substrate is 1.9 kcal/mol lower in energy than **7**. The TS barrier for H-atom abstraction from **7'** is 14.2 kcal/mol for the doublet ($^2\text{TS1}$), and 15.6 kcal/mol for the quartet ($^4\text{TS1}$). These reasonable and low barriers and subsequent mechanism exothermicity suggest that pathways utilizing Cpd I as the oxidant are viable. While these data do not exclude the possibility of peroxo-iron as an oxidant, the low energy barrier and experimental support for the proposed Cpd I pathway suggest that Cpd I is the preferred oxidant.

3.4 Discussion

We have performed experimental and computational experiments to provide insights into the catalytically active species used by P450 TamI L295A for epoxidation and epoxide-opening oxidative cascade of tirandamycin congeners. Specifically, we aimed at shedding light on the question of a single or multiple various oxidizing species being used throughout the iterative cascade of **1** \rightarrow **6** \rightarrow **7** or **7'** \rightarrow **9** and **10** (Scheme 3.2). The experiments applied can suggest but not conclusively determine the active oxidant.

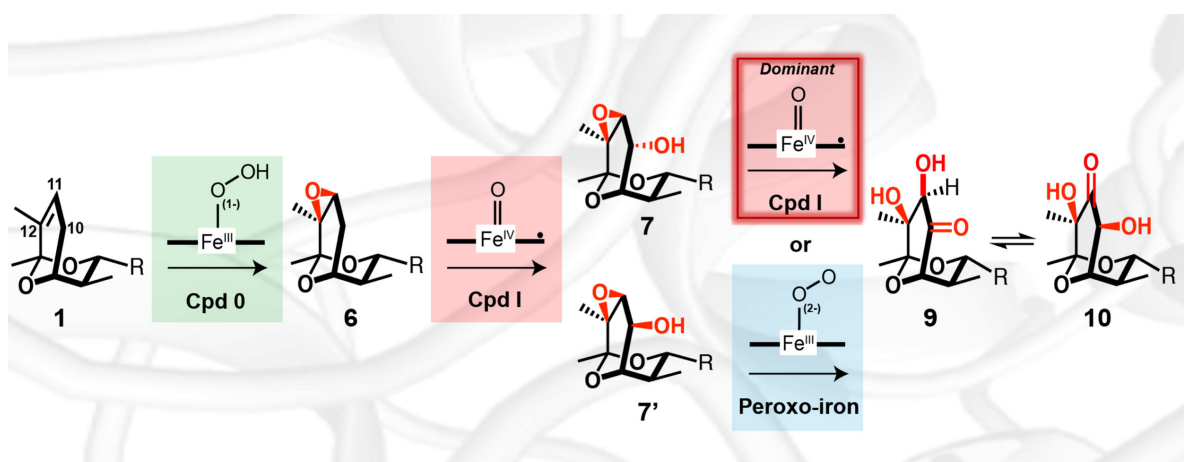


Figure 3.6. Oxidizing species involved in TamI L295A catalysis. Cpd I is the preferred oxidant for the conversion of intermediate **7** or **7'** to **9** and **10**, although kinetic solvent isotope effect experiments indicated that the nucleophilic peroxo-iron can be a viable oxidant for this epoxide-opening cascade.

On the experimental side, mutations were introduced in the highly conserved acid-alcohol pair of TamI (Glu251 and Thr252) to probe their role in oxygen activation and electron transfer (Figure 3.2). Earlier studies of bacterial P450s harboring a variation in the conserved alcohol residue displayed drastically diminished levels of hydroxylation activity and divergence to

epoxidation reactivity, indicating the involvement of a Cpd 0 (species F) oxidant species.¹¹ In L295A threonine variants, similar oxidation rates to L295A were obtained with substrate **1**, without change in the chemo- or regioselectivity patterns (Figure 3.2). These catalytic results could indicate that TamI L295A has the flexibility to exploit the use of other heme iron species besides Cpd I, including Cpd 0 for epoxidation of **1** → **6**. A H₂O₂-complexed species (the hydrogenated form of Cpd 0) was discarded based on the results obtained when adding ROS scavengers to the reactions of TamI L295A with **1**. The presence of these scavengers was expected to capture most of the naturally-occurring H₂O₂ present in the system.²⁴ An increase of product formation was observed indicating that the H₂O₂-shunt pathway to form **6** from **1** is not favorable. Mechanistic probes such as H₂O₂ and iodosobenzene were used to further interrogate the involvement of different catalytically active species (Figure 3.3). TamI L295A was able to activate H₂O₂ for epoxidation of **1** to **6**. Meanwhile, other TamI variants (including L295V, L101A_L295I and L244A_L295V) catalyzing epoxidation or hydroxylation of **1** under normal conditions were not able to use H₂O₂ as a cosubstrate for oxidative reactivity. The results implied that the unique protein environment of TamI L295A enables it to use Cpd 0 for the C11/12 diastereoselective epoxidation reaction of **1**,²⁵ inference that was supported by computational analysis. Parallel to P450 PimD's epoxidation reactivity via a Cpd 0 iron species,⁵ MD simulations of TamI L295A and **1** showed the C11/12 alkene to be at an angle that disfavored catalysis by Cpd I but that could be reached by the lengthened O-O bond in Cpd 0.

Moreover, KSIE data demonstrated that the peroxo-iron species could be a viable oxidant for the conversion of **6** → **7** or **7'** → **9** and **10** when water in the reaction buffer is replaced by deuterium, slowing down the formation of Cpd I (Figure 3.4). This was further validated by endpoint assays with the E251Q_L295A mutant that is thought to interrupt the two proton transfers required for Cpd I thus accumulating peroxo-iron (Scheme 3.1). DFT calculations for a peroxo-iron mechanism demonstrated low energy barriers for the initial TS and intermediate. Although downstream intermediates and products were not identified despite multiple efforts, computational experiments do not discard peroxo-iron as an active oxidant. A Cpd I-mediated pathway towards the generation of **9** and **10** was also explored computationally. The low barriers and mechanism exothermicity obtained for a radical-induced ring opening of the epoxide in **7** and **7'** and subsequent oxygen rebound suggest Cpd I is the preferred oxidant for TamI L295A's unique multistep oxidative pathway.

Experimental and computational methods have been recently combined to explore the active oxygen species in P450 biocatalysts catalyzing mechanistically distinct pathways on different substrates. As an example, P450-BM3 mutants performing sulfoxidation and hydroxylation on 1-thiochromanone and 1-tetralone substrates, respectively, were investigated.²⁵ Cpd I was found as the common oxidant species for both reaction processes. However, fewer efforts have been devoted to clarifying the mechanism of iterative oxidation at multiple carbon sites on a common substrate, in respect to the catalytically active oxygen species.

Elucidation of the catalytically active oxidant species in P450 TamI L295A facilitates future engineering of bio- or chemical oxidation catalysts with iterative oxidation abilities. Favoring formation of different iron oxygen species throughout a cascade reaction could enable mechanistically divergent oxidative pathways to be catalyzed on a common substrate. This could facilitate direct oxidative derivatization of multiple carbon atoms within complex molecules. For instance, it would be intriguing to explore a similar mutagenesis profile in the iterative bacterial P450 GfsF that harbors an active site Ala297 residue equivalent to the Leu295 site in TamI.²⁶ The GfsG enzyme catalyzes a stepwise dual-cascade epoxidation followed by hydroxylation to form the macrolide antibiotic FD-891. The macrolide hydroxy is located β to the epoxide, parallel to the C10 hydroxy in **7** and **7'** that is found β to the C11/12 epoxide. It is not unreasonable to imagine that engineering the Ala297 residue may favor formation of one catalytically active oxygen species over the other, facilitating opening of the epoxide to generate an α -ketone or a *cis*-diol-containing product, as observed in TamI L295A. A similar configuration for the positioning of the hydroxy group relative to the epoxide within a substrate is found in the mycinamicin natural product. In this system, the iterative P450 MycG performs a strict hydroxylation followed by epoxidation yielding a similar oxidation pattern to TamI L295A.²⁷ Although the hydroxy is located in a tertiary carbon with no hydrogen atom available (unlike the case of GfsF or TamI systems) abstraction of the epoxy-hydrogen or nucleophilic opening of the epoxide could still occur to generate products similar to **9** and **10** if a different active oxidant species may be accumulated.

The ability of TamI L295A to utilize H₂O₂ in lieu of molecular oxygen and NADPH for its iterative oxidation cascade is also noteworthy. This unexpected result represents a starting point for further design and engineering to improve the efficiency of peroxide-mediated TamI catalysis.¹⁴ For example, adding small molecules to the reaction or introducing additional

variations in the active site to stabilize H-bonding networks necessary for reactivity could facilitate an efficient peroxygenase activity for the enzyme,²⁵ maximizing its chemical utility.

Finally, elucidation of this pathway (**1** → **6** → **7** or **7'** → **9** and **10**) represents a unique oxidative outcome for P450 TamI and is also worth highlighting (Figure 9). In the native system, substrate **1** is hydroxylated at C10 to form **2** prior to epoxidation at the C11/12 alkene in a very strict sequence. TamI L295A and TamI L101A_L295I represented the first two variants capable to divert from the innate substrate reactivity to catalyze the most energetically demanding reaction first, a C11/12 epoxidation to produce **6** from **1**. With epoxide-containing **6**, double variant L101A_L295I performed C18 methyl hydroxylation to form **8**. In the present study, we demonstrate that TamI L295A is the only biocatalyst able to invert the native order of oxidative steps and install a C10 hydroxy on **6** to form **7** *in vitro*. This marks a distinct difference from iterative P450 MycG that is not able catalyze hydroxylation of a substrate that is initially epoxidized.²⁷ The ability of TamI mutants to invert reactivity in a catalyst-controlled fashion facilitates core structure diversification to produce new analogues with unique patterns of oxidation.

Our work enables us to decipher the catalytic machinery that controls the mechanism of epoxidation and epoxide-opening iterative cascade to produce tri-oxidized **9** and **10** from **1** by P450 TamI L295A (Figure 3.6). This work provides the first mechanistic picture for the catalytically active oxygen species involved in TamI. The experimental results flanked by MD simulations and DFT calculations show that TamI L295A may not have one common active oxidant in its iterative cascade. Indeed, our data suggests that TamI L295A can use the ferric hydroperoxide Cpd 0 to catalyze the epoxidation of **1** → **6**, then shift to using Cpd I for hydroxylation of **6** → **7** or **7'** followed by Cpd I-mediated epoxide opening via a radical pathway to form **9** and **10**. Interestingly, our computational and kinetic data suggests that the nucleophilic peroxy-iron species acts as a viable oxidant for processing **6** → **7** or **7'** → **9** and **10** when formation of Cpd I is compromised. We expect that these complementary approaches toward mechanistic investigations of iterative cytochrome P450 C-H functionalization and epoxidation biocatalysts will inspire engineering efforts to generate novel, bioactive metabolites.

3.5 References

1. Poulos, T. L. Heme enzyme structure and function. *Chem Rev* **2014**, 114, 3919-3962.

2. Cupp-Vickery, J. R.; Han, O.; Hutchinson, C. R.; Poulos, T. L. Substrate-assisted catalysis in cytochrome P450eryF. *Nat Struct Biol* **1996**, 3, 632-637.
3. Stok, J. E.; Yamada, S.; Farlow, A. J.; Slessor, K. E.; De Voss, J. J. Cytochrome P450(cin) (CYP176A1) D241N: investigating the role of the conserved acid in the active site of cytochrome P450s. *Biochim Biophys Acta* **2013**, 1834, 688-696.
4. Gregory, M. C.; Denisov, I. G.; Grinkova, Y. V.; Khatri, Y.; Sligar, S. G. Kinetic solvent isotope effect in human P450 CYP17A1-mediated androgen formation: evidence for a reactive peroxoanion intermediate. *J Am Chem Soc* **2013**, 135, 16245-16247.
5. Kells, P. M.; Ouellet, H.; Santos-Aberturas, J.; Aparicio, J. F.; Podust, L. M. Structure of cytochrome P450 PimD suggests epoxidation of the polyene macrolide pimaricin occurs via a hydroperoxoferric intermediate. *Chem Biol* **2010**, 17, 841-851.
6. Khatri, Y.; Luthra, A.; Duggal, R.; Sligar, S. G. Kinetic solvent isotope effect in steady-state turnover by CYP19A1 suggests involvement of compound I for both hydroxylation and aromatization steps. *FEBS Lett* **2014**, 588, 3117-3122.
7. Ortiz de Montellano, P. R. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem Rev* **2010**, 110, 932-948.
8. Ogliaro, F.; de Visser, S. P.; Cohen, S.; Sharma, P. K.; Shaik, S. Searching for the second oxidant in the catalytic cycle of cytochrome P450: a theoretical investigation of the iron(III)-hydroperoxo species and its epoxidation pathways. *J Am Chem Soc* **2002**, 124, 2806-2817.
9. Newmister, S. A.; Srivastava, K. R.; Espinoza, R. V.; Haatveit, K. C.; Khatri, Y.; Martini, R. M.; Garcia-Borras, M.; Podust, L. M.; Houk, K. N.; Sherman, D. H. Molecular basis of iterative C-H oxidation by TamI, a multifunctional P450 monooxygenase from the tirandamycin biosynthetic pathway. *ACS Catal* **2020**, 10, 13445-13454.
10. Espinoza R. V.; Haatveit, K. C.; Grossman S. W.; Tan, J. Y.; McGlade, C. A.; Khatri, Y.; Newmister, S. A.; Schmidt, J. J.; Garcia-Borràs, M.; Montgomery, J.; Houk, K. N.; Sherman, D. H. Engineering P450 TamI as an iterative biocatalyst for selective late-stage C-H functionalization and epoxidation of tirandamycin antibiotics. *ACS Catal* **2021**, 11, 13, 8304-8316.
11. Jin, S.; Makris, T. M.; Bryson, T. A.; Sligar, S. G.; Dawson, J. H. Epoxidation of olefins by hydroperoxo-ferric cytochrome P450. *J Am Chem Soc* **2003**, 125, 3406-3407.
12. Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem Rev* **2004**, 104, 3947-3980.
13. Coleman, T.; Stok, J. E.; Podgorski, M. N.; Bruning, J. B.; De Voss, J. J.; Bell, S. G. Structural insights into the role of the acid-alcohol pair of residues required for dioxygen activation in cytochrome P450 enzymes. *J Biol Inorg Chem* **2020**, 25, 583-596.
14. Joo, H.; Lin, Z.; Arnold, F. H. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **1999**, 399, 670-673.
15. Yoshimoto, F. K.; Gonzalez, E.; Auchus, R. J.; Guengerich, F. P. Mechanism of 17 α ,20-lyase and new hydroxylation reactions of human cytochrome P450 17A1. 18O labeling and oxygen surrogate evidence for a role of a perferryl oxygen. *J Biol Chem* **2016**, 291, 26587.
16. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys* **1993**, 98 (7), 5648-5652.
17. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **2011**, 32, 1456-1465.
18. Lee, C. Y., W.; Parr, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron Density. *Phys Rev B: Condens Matte* **1998**, 37 (2), 785-789.
19. Stephens, P. J. D., F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J Phys Chem* **1994**, 98 (45), 11623-11627.
20. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys Chem Chem Phys* **2005**, 7, 3297-3305.
21. Chung, L. W. L., X.; Sugimoto, H.; Shiro, Y.; Morokuma, K. ONIOM study on a missing piece in our understanding of heme chemistry: bacterial tryptophan 2,3-dioxygenase with dual oxidants. *J Am Chem Soc* **2010**, 132 (34), 11993-12005.
22. Chung, L. W.; Li, X.; Hirao, H.; Morokuma, K. Comparative reactivity of ferric-superoxo and ferryl-oxo species in heme and non-heme complexes. *J Am Chem Soc* **2011**, 133, 20076-20079.
23. Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. QM/MM study of mechanisms for compound I formation in the catalytic cycle of cytochrome P450cam. *J Am Chem Soc* **2006**, 128, 13204-13215.

24. Wang, J. B.; Huang, Q.; Peng, W.; Wu, P.; Yu, D.; Chen, B.; Wang, B.; Reetz, M. T. P450-BM3-catalyzed sulfoxidation versus hydroxylation: a common or two different catalytically active species? *J Am Chem Soc* **2020**, *142*, 2068-2073.
25. Zhang, X.; Jiang, Y.; Chen, Q.; Dong, S.; Feng, Y.; Cong, Z.; Shaik, S.; Wang, B. H-bonding networks dictate the molecular mechanism of H₂O₂ activation by P450. *ACS Catal* **2021**, *11*, 8774-8785.
26. Miyanaga, A.; Takayanagi, R.; Furuya, T.; Kawamata, A.; Itagaki, T.; Iwabuchi, Y.; Kanoh, N.; Kudo, F.; Eguchi, T. Substrate recognition by a dual-function P450 monooxygenase GfsF involved in FD-891 biosynthesis. *Chembiochem* **2017**, *18*, 2179-2187.
27. Yang, S.; DeMars, M. D., 2nd; Grandner, J. M.; Olson, N. M.; Anzai, Y.; Sherman, D. H.; Houk, K. N. Computational-based mechanistic study and engineering of cytochrome P450 MycG for selective oxidation of 16-membered macrolide antibiotics. *J Am Chem Soc* **2020**, *142*, 17981-17988.

Chapter 4*

Substrate Engineering of P450 TamI to Expand Substrate Scope

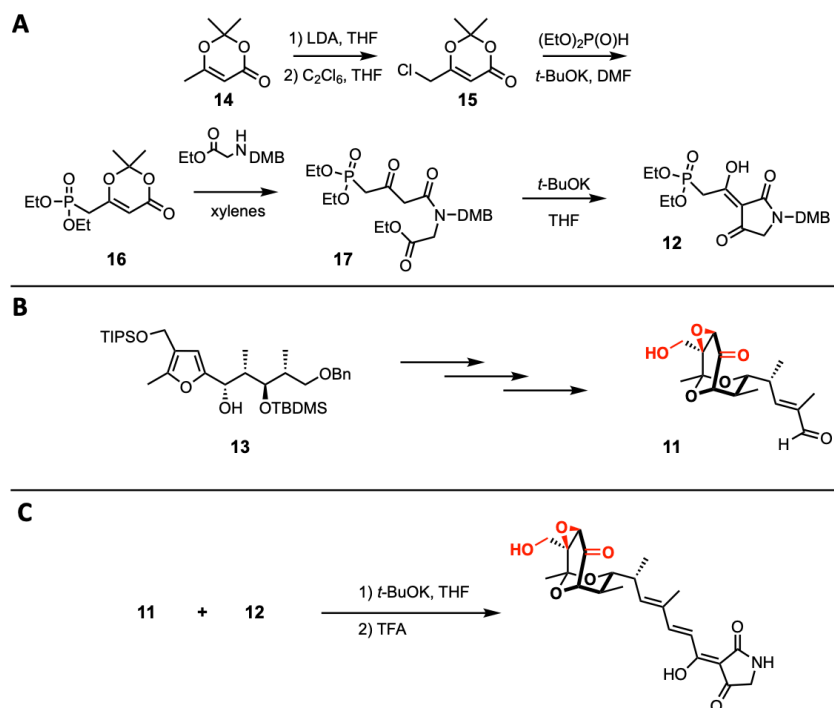
4.1 Background and Introduction

4.1.1 Previous Syntheses of Tirandamycin Natural Products

Over the last four decades, remarkable synthetic routes have been developed for the racemic and enantioselective construction of the biologically active tirandamycin natural product molecule.¹⁻³ Most of the reported syntheses focus on generating tirandamycin A (**4**), one of the most bioactive congeners (Figure 4.1), and require extensive protecting and deprotecting strategies. Additional efforts have been devoted to synthesizing the terminal product tirandamycin B (**5**),² less oxidized intermediates tirandamycin C (**1**),⁴ E (**2**)³ and D (**3**),¹ and the 2,6-dioxabicyclononane skeleton unit alone.⁵ The first reported total synthesis of tirandamycin dates to 1985,⁶ while tetramic acid analogues streptolydigin and tirandalydigin have only been totally synthesized in recently.⁵ Most of the synthetic strategies towards tirandamycin describe a similar retrosynthetic disconnection. These strategies typically share a common enal intermediate (**11**) that undergoes a Horner-Wadsworth-Emmons olefination with a protected phosphonate tetramic acid (**12**) to generate the triene system and the dienoyl tetramic acid moiety (Scheme 4.1A). Reagent **12** can be prepared from 2,2,6-trimethyl-4*H*-1,3-dioxin-4-one (**14**) in four reaction steps including allylic chlorination, dimethyl phosphite substitution of the chloro intermediate, generation of the β keto amide (**17**) and Dieckmann condensation. The commonly used protecting group for the phosphonate tetramic acid is 2,4-dimethoxybenzyl, DMB, which is essential for the tetramic acid condensation to occur successfully. Prior using DMB, the unprotected aldehyde intermediate was found to be unstable and prone to degradation. Trifluoroacetic acid (TFA) is used to remove the DMB group and yield the desired tirandamycin congener.

* Chapter 4 is included in a manuscript in preparation.

Construction of the bicyclic moiety is often achieved through oxidation of a derivatized tetrahydrofuran (THF) ring⁶ (Scheme 4.1B), addition of methyl lithium (MeLi) and ketalization of a lactone-containing intermediate.⁴ Coupling of the bicyclic moiety and the phosphonate tetramic acid then occurs upon exposure of the substrates to *t*-BuOK followed by TFA-induced deprotection of the DMB group to yield the desired final product (Scheme 4.1C).



Scheme 4.1. Total synthesis of tirandamycin. (A) Synthesis of protected phosphonate tetramic acid moiety (B) Synthesis of tirandamycin bicycle (C) Coupling of bicycle and tetramic acid via Horner-Wadsworth Emmons (HWE) olefination.

4.1.2 Biological Relevance of Derivatizing Tirandamycin

The core structure of tirandamycin is composed of a dienyl tetramic acid moiety, one triene system, and a 2,6-dioxabicyclononane skeleton. The myriad of tirandamycin congeners that have been reported in the literature (obtained either through fermentation or bioenzymatic reactions) differ in the extent of oxidation decorating the bicyclic framework that contributes to differences in their biological activities.⁷ We previously reported that a higher degree of oxidation in tirandamycin correlates to an exponential increase in bioactivity (Figure 4.1).⁸ For example, **4** and **5** show the lowest MIC values against VRE-*E. faecalis* compared to **1**, **2** and **3**, suggesting that in addition to the C10 ketone, the C11/12 epoxide and C18 primary hydroxy group are key to improving biological activity. Another study reports that the more extensively oxidized congener

5 kills the *B. malayi* adult parasite in a highly selective manner, indicating that tirandamycin congeners may serve as a new lead scaffold to treat parasitic diseases. Recent work focused on assessing the bioactivity potential of new tirandamycin congeners that display unique patterns of oxidation not found in nature⁹ (i.e. congener **8** harboring an epoxide and C18 hydroxy but no C10 hydroxy; congener **6** containing a C11/12 epoxide but no hydroxy functionality; congener **9** harboring a *cis* diol at C11 and C12 and a C10 ketone). Among the library of new congeners, tirandamycin **8** displayed the most potent bioactivity against anthrax-causing *B. anthracis* 34f2 and pneumonia-causing *S. pneumonia* ATCC 49619.

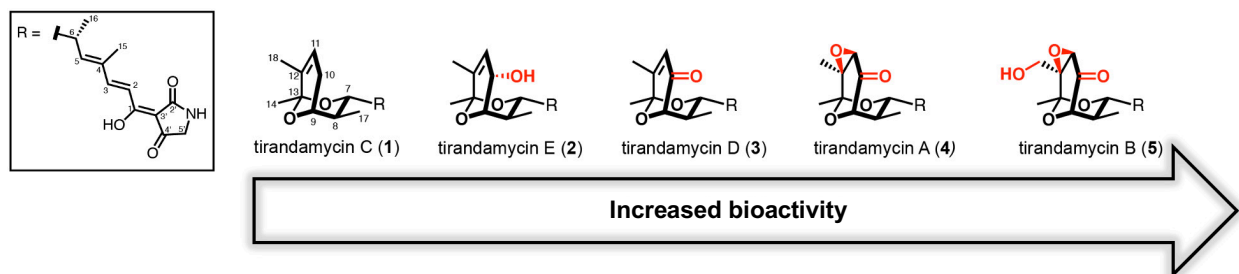


Figure 4.1. Bioactivity of tirandamycin increases as oxidation levels increase.

Synthetic efforts towards accessing tirandamycin antibiotics are often optimized to construct a specific congener, and rely on pre-defining the oxidation state of the molecule early on in the synthesis. The need for extensive protecting and deprotecting strategies and pre-determining oxidation patterns hinders a direct derivatization method of tirandamycin molecules. Consequently, achieving late-stage oxidation of tirandamycin molecules would enable efficient derivatization of these bioactive molecules for further biochemical, biological and enzymatic assays.

4.1.3 Substrate Engineering Methods to Facilitate Derivatization

Although P450 biocatalysts hold great promise as synthetic tools for the generation of complex molecules, their relatively limited substrate scope hinders their application as useful biocatalysts.¹⁰ Protein engineering has been used to tune P450s to accept unnatural substrates,^{11, 12} although extensive mutagenesis is usually required for each new substrate.¹¹ Thereupon, substrate engineering has emerged as an effective alternative strategy to broaden substrate scope.¹³ This innovative approach involves the installation of a readily removable functional group onto an unnatural substrate to control its active site binding orientation in a useful manner.^{14, 15} This

chemical modification can lead to productive substrate binding and selective oxidation of unnatural substrates using WT enzymes, therefore eliminating the need to extensively engineer the catalyst. Structural data revealing the enzyme-substrate interactions that contribute to substrate binding are critical for employing this strategy. The synthetic value of substrate engineering has been demonstrated by successfully applying this technique in the development of PikC as a powerful oxidation catalyst towards a variety of unnatural substrates harboring the “anchoring group” necessary for substrate recognition by the enzyme.¹⁶

Initial analysis of the substrate-bound crystal structure of P450 TamI¹⁷ suggested that substrate binding in the enzyme active site is facilitated by interactions with the tetramic acid moiety. However, enzymatic assays showed that these tetramic acid-based polar interactions are not essential for catalysis. Instead, experimental results indicated that the hydrophobic and π - π interactions between Phe92, Leu399 and Ile400 residues and the conjugated polyene chain of tirandamycin were critical for substrate recognition and oxidation (Figure 4.2). Exploiting the observed polyene chain-based substrate anchoring mechanism could lead to recognition and oxidation of unnatural substrates by P450 TamI, thus broadening the substrate scope of this enzyme.

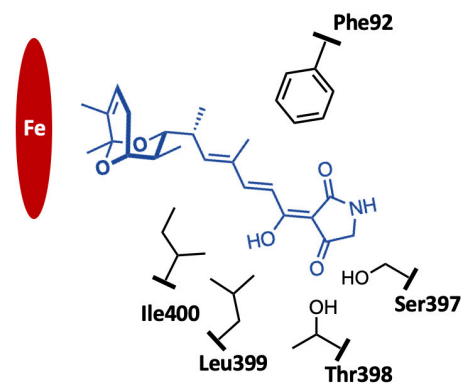


Figure 4.2. Simplified representation of substrate-bound crystal structure of TamI. Tirandamycin C (1) is shown in blue. Heme-iron is shown in red. Hydrophobic interactions between substrate and residues Phe92, Leu399 and Ile400 are essential for binding and catalysis.

To shed light into the structure-activity relationship of tirandamycin synthetic analogues and to enable a late-stage diversification strategy to access novel derivatives and expand the substrate scope of iterative P450 TamI, we report herein the short and efficient synthesis of a tirandamycin bicyclic analogue. In place of the tetramic acid tail, the truncated analogue harbors a functional hydroxy handle at the C6 position for further coupling of synthetic “anchors”. The unnatural anchoring groups are designed to mimic the conjugated polyene chain and the 5-membered ring tail of tirandamycin. Most of them are synthesized in two short steps instead of the five steps required to generate the protected phosphonate tetramic acid, facilitating access to new tirandamycin analogues. This substrate engineering strategy enables productive binding of the unnatural molecules into the active pocket of TamI for enzyme-induced iterative oxidation

chemistry. Furthermore, this work provides an initial platform to broaden the substrate scope of the biocatalyst to include oxidation of unnatural bicyclic and polycyclic moieties that would otherwise be difficult to functionalize selectively at a late-stage without extensive functional group manipulation and protecting methods. Assessing the degree of substrate flexibility of TamI towards a panel of unnatural substrates can shed light on the potential of this P450 to be developed into a useful synthetic tool to access new chemical space.

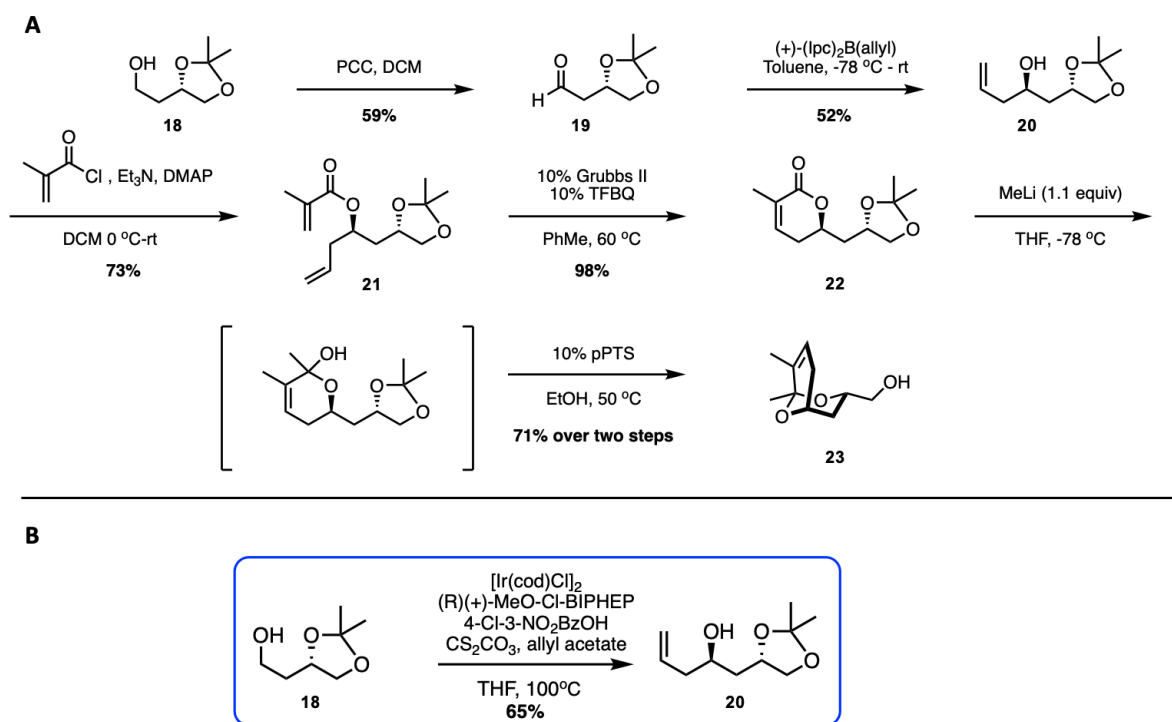
4.2 Design and Synthesis of Tirandamycin Analogues

4.2.1 Previous Synthetic Strategies Towards a Tirandamycin Analogue

Inspired by published literature on the synthesis of **1**, Jessica Stachowski in the Montgomery group at the University of Michigan developed a first- and second-generation total synthesis of a tirandamycin bicyclic analogue (**23**).¹⁸ Key elements of the first-generation approach (12 steps from a commercially available ester, see Appendix C for details) include allylboration of an aldehyde-containing intermediate to yield enantiomerically pure homoallylic alcohol, and ring closing metathesis with tetrafluoro-1,4- benzoquinone as an additive to minimize undesired isomerization often caused by ruthenium hydride. Subsequent optimizations provided a second-generation synthesis that is shorter (6 steps, Scheme 4.2A) and bypasses the need for complex asymmetric strategies. This was accomplished by excluding incorporation of the C17 methyl, which is distant from the points of oxidation in the bicycle and not expected to interrupt enzyme catalysis. The commercially available (4*S*)-(+)-4-(2-Hydroxyethyl)-2,2-dimethyl-1,3-dioxolane, **18**, was used as starting material as it already contains the desired C7 stereochemistry. Using PCC oxidation, the dioxolane alcohol is oxidized to the aldehyde **19** in lowered yields due to the volatility of the aldehyde. Allylboration followed by acylation provides the ester intermediate **21** that then undergoes ring closing metathesis to yield **22**. Treatment of **22** with MeLi followed by pPTS-induced acetonide deprotection and cyclization generates the desired truncated tirandamycin bicyclic analogue **23** in a 71% yield.

4.2.2 Asymmetric Allylation Optimization: A 5-Step Synthesis Towards a Tirandamycin Analogue

We envisioned bypassing the formation of the volatile aldehyde intermediate **19** to increase the total yield of the synthesis and lower the number of reaction steps. This was achieved by optimizing an iridium-catalyzed asymmetric allylation reaction, inspired by Krische and co-workers,¹⁹ to form the homoallylic alcohol **20** in one step from **18**. After testing various commercially available chiral ligands, the (*R*)(+)-Cl-MeO-BIPHEP ligand showed to be the most effective one leading to the formation of the desired enantiomeric product in a 65% isolated yield and shortening the synthesis from six to five steps total (Scheme 4.2B).



Scheme 4.2. Synthetic strategies towards a tirandamycin bicyclic analogue. (A) Second-generation total synthesis includes 6 reaction steps. (B) Asymmetric allylation optimization .

4.2.3 Synthesis and Coupling of Synthetic Anchors to a Tirandamycin Analogue

With a short and scalable synthesis of tirandamycin bicyclic analogue **23** optimized, we then envisioned developing a substrate engineering platform to expand the substrate scope of P450 TamI. First, we aimed at designing and synthesizing simplified anchoring groups that harbor the structural features needed for productive substrate binding in the enzyme. Inspired by previous structural and enzymatic studies,¹³ we anticipated that the conjugated polyene chain is a required structural component for substrate recognition while the tetramic acid moiety may not be as

essential. However, many questions were left unanswered. 1) What is the extent of the conjugation required (i.e. diene vs *triene*) for catalysis? 2) Can P450 TamI accept a substrate without a 5-membered ring at the tail of the molecule? 3) Can the enzyme oxidize a substrate with a hydrocarbon 5-membered ring (and no oxygen or nitrogen atoms)? Addressing these questions would guide substrate engineering efforts towards broadening the biocatalyst's substrate scope. Therefore, we decided to synthesize a series of simplified synthetic anchors containing one, two or three double bond systems, a hydrocarbon 5-membered ring or no ring at all, or a 6-membered conjugated ring (Figure 4.3). Each anchor contains a carboxylic acid to facilitate coupling to the hydroxyl-containing tirandamycin bicyclic analogue.

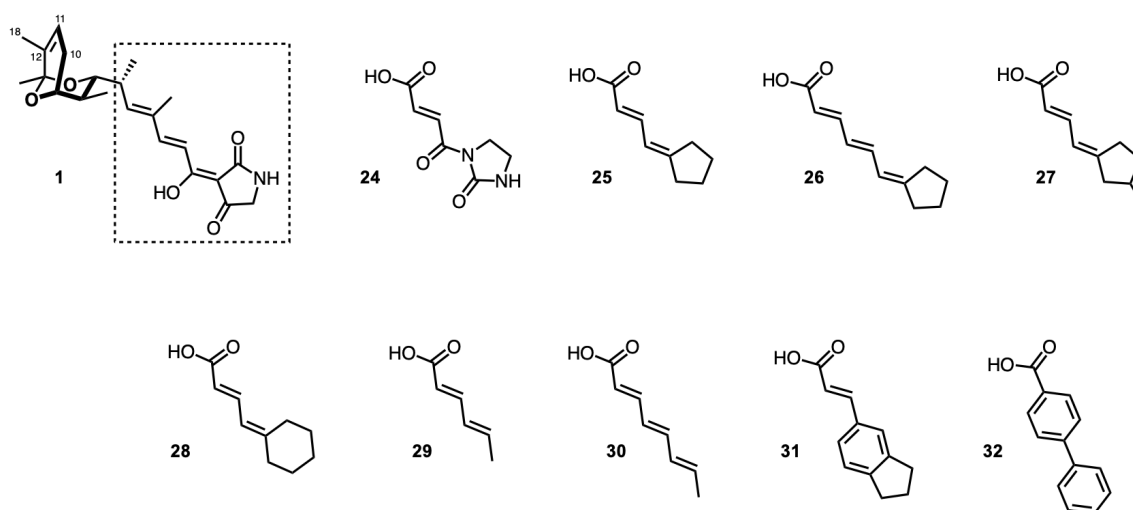
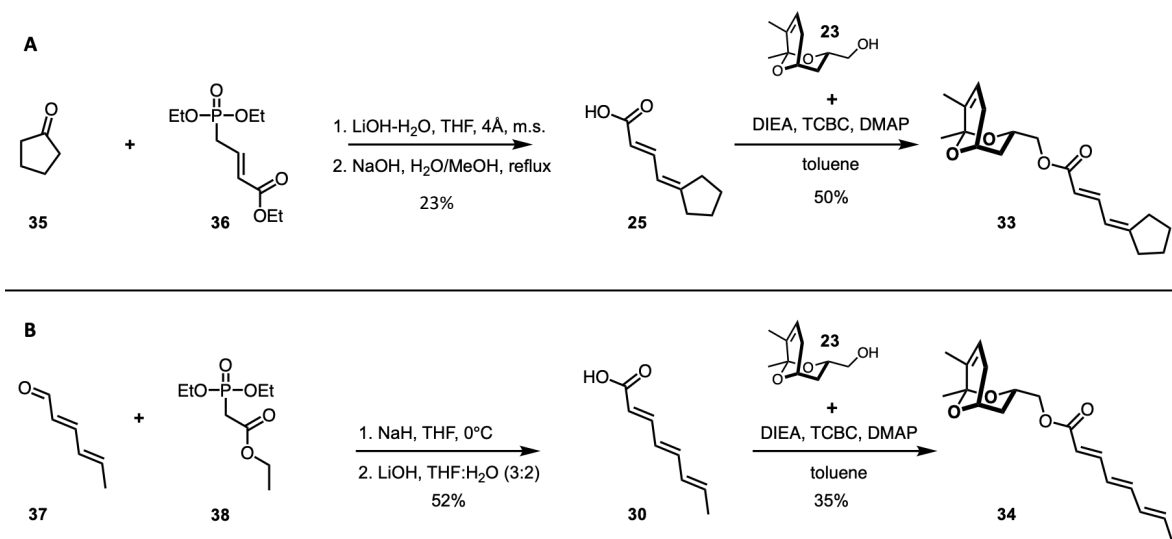


Figure 4.3. Design of synthetic anchors for substrate engineering efforts with TamI.

Docking experiments were performed using Autodock 4.0 and the P450 TamI crystal structure to prioritize anchors that would potentially lead to the tightest binding affinity in the enzyme (see Appendix C for details). Results suggested that **25** and **30** could serve as ideal anchors to replace the polyene linker of tirandamycin. The calculated binding energies for these two anchors reasonably compared with the binding energy observed in the tirandamycin system. Inspired by previously published literature, we designed a synthetic route to obtain the anchors of interest from inexpensive commercially available reagents, and in the minimum number of steps. The synthesis of **25** was previously reported,²⁰ however four steps were required to obtain the desired product. Inspired by similar work²¹ where authors used cyclohexanone as the starting material generating a 6-membered version of **25**, we were able to shorten the synthesis from four to two steps (Scheme 4.3). The desired carboxylic acid was obtained in >99% ee. The synthesis of

30 was followed as previously published²² starting with freshly distilled sorbaldehyde generating a 56:1 E/Z ratio (as determined by H-NMR) after re-crystallization and sublimation. Yamaguchi esterification conditions were then utilized to couple the carboxylic acid-containing synthetic anchors to the tirandamycin bicyclic analogue harboring a hydroxyl handle (Scheme 4.3).



Scheme 4.3. Synthesis of tirandamycin analogues harboring simplified anchoring groups.

4.3 Substrate Engineering Efforts to Expand Substrate Scope

4.3.1 Enzymatic Reactions with Synthetic Tirandamycin Analogues and Computational Investigations

The two synthetic tirandamycin analogues were tested for oxidation catalysis with P450 TamI. We reasoned that given the non-polar nature of the synthetic analogues, TamI mutants that introduce hydrophobic residues in the enzyme where the tetramic acid tail of tirandamycin usually sits may better accommodate the synthetic substrates for oxidation. Thus, TamI mutant S397A_T398A was prioritized for testing. No product formation was detected when testing the triene-containing substrate **34** (Figure 4.4A) with TamI S397A_T398A or WT. It is reasonable to imagine that the lack of a tetramic acid moiety perturbs substrate binding in the active site of the enzyme inhibiting catalysis. To interrogate this assumption, we performed MD simulations with TamI WT and a truncated tirandamycin C analogue lacking a tetramic acid ring (Figure 4.5). We found that the root-mean-square deviation of the analogue is much higher compared to

tirandamycin C substrate (**1**) that maintains a more stable geometry and orientation within the enzyme. The truncated tirandamycin occupies multiple different orientations in the enzyme's active site, suggesting that the tetramic acid ring is critical to generate the correct geometry for oxidation by P450 TamI. This computational observation agrees with the lack of oxidative reactivity detected experimentally with **34**. Previous enzymatic assays showed that the polar interactions between the tetramic acid heteroatoms and the active site are not essential for binding and catalysis.¹⁷ However, these results did not exclude the possibility that a 5-membered ring moiety is required to stabilize the orientation of the substrate to maintain the proper geometry for catalysis to take place.

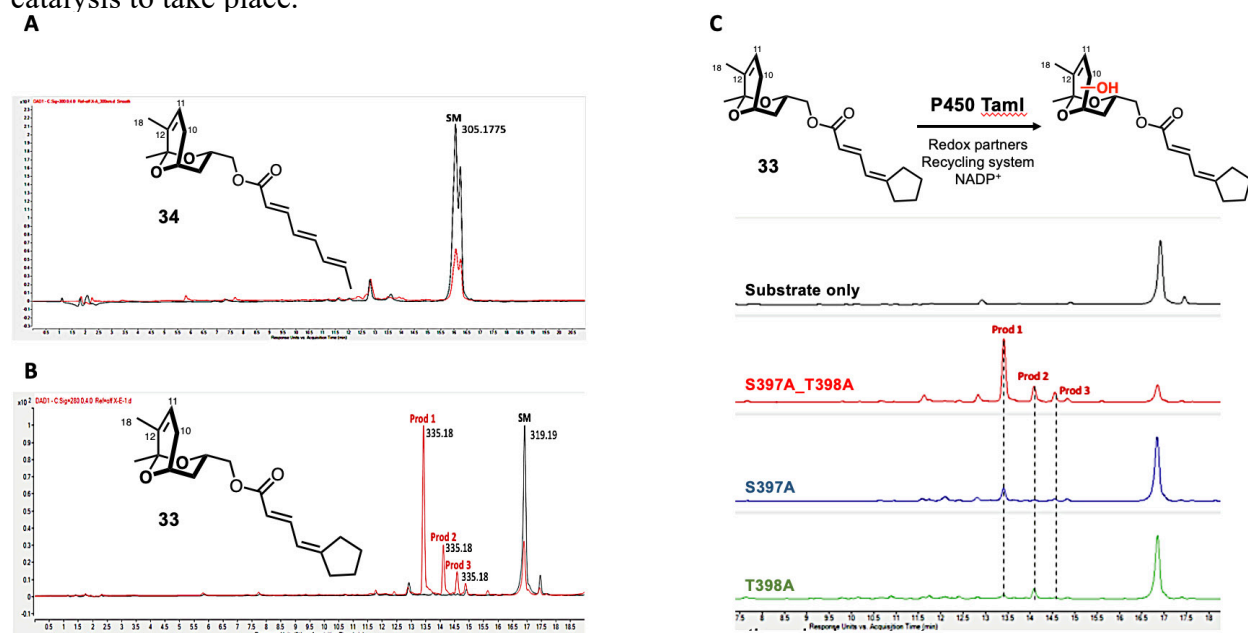


Figure 4.4. Enzymatic reactions with synthetic tirandamycin analogues. (A) Analytical reaction with substrate **34** and TamI L397A_T398A. LC-MS spectra extracted at 280nm. The substrate standard is shown in black ($M+H = 305.18$) and the crude reaction is shown in red. No product was detected. (B) Analytical reaction with **33** and TamI L397A_T398A. LC-MS spectra extracted at 280nm. The substrate standard is shown in black ($M+H = 319.19$) and the crude reaction is shown in red. Three single oxidation products were detected ($M+H = 335.18$). (C) Reactions with **33** and additional TamI mutants S397A and T398A. LC-MS spectra extracted at 280nm.

Interestingly, substrate **33** led to approximately 70% conversion to oxidative products with double mutant S397A_T398A (Figure 4.4B). No oxidative reactivity was observed with any other TamI enzyme including WT. Three single oxidation products were detected, and 60% of the total conversion corresponded to the product with retention time of 13.5 min. Unexpectedly, no double or triple oxidation products were formed unlike what we typically observe in the TamI WT system

where consecutive oxidations occur. At this point, we don't fully understand what leads to a single oxidation event rather than a multi-step oxidative cascade. We reasoned that it could be either the non-polar characteristics of the anchoring moiety in **33** (that lack the ability to form polar interactions with the enzyme's active site) or the mutations introduced that limit the ability of the enzyme to catalyze consecutive oxidation reactions. To probe if both serine and threonine mutations were required for binding and catalysis, single mutants S397A and T398A were engineered, harvested, and tested with **33**. Minimal product formation was observed with either single mutant (Figure 4.4C). Thus, we concluded that both mutations at the Ser397 and Thr398 sites are required for productive binding and catalysis with **33**. MD simulations with P450 TamI and substrate **33** were performed to further investigate the reactivity observed. Unlike **34**, substrate **33** maintains a stable geometry within the active site for the entire length of the simulation (Figure 4.6A). This indicates that the 5-membered hydrocarbon ring at the tail of **33** is key in maintaining a proper substrate orientation for catalysis. The simulations also revealed that residues Phe92 and Ile400 (critical for substrate binding) did not undergo significant changes in orientation and remained close to the substrate during the entire simulation. The distances of key carbon atoms of the **33** bicyclic ketal moiety (C10, C11, C12 and C18) to the heme-iron were measured to predict

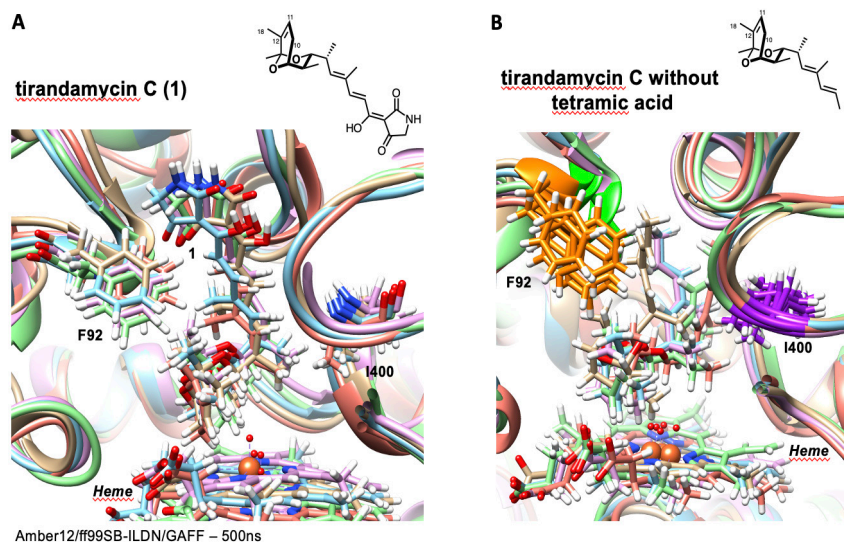


Figure 4.5. MD simulations of TamI with a truncated tirandamycin. (A) Simulations of P450 TamI WT with **1** show the substrate maintains a stable geometry throughout the 500 ns. (B) Simulations of P450 TamI WT with a truncated tirandamycin lacking the tetramic acid moiety show the substrate occupies multiple different binding poses, indicating perturbed substrate binding. Residues Phe92 and Ile400, critical for binding and catalysis, are shown in sticks.

the regioselectivity of the oxidation event (Figure 4.7B). For the first 450 ns of the simulation, C10-H (*S*) (light blue) is closest to the heme-iron (within 3Å) and thus, more likely to be abstracted. However, at around 500 ns and for the rest of the simulation, C18-H (dark blue) is closest to the heme-iron. These computational results suggest that P450 TamI may oxidize substrate **33** in an

unselective fashion, an assumption that is supported by enzymatic data. It is possible that one of the minor oxidation products detected by LC-MS refer to the predicted C10 hydroxylated product, while the major mono-oxidation product contains a C18 hydroxyl.

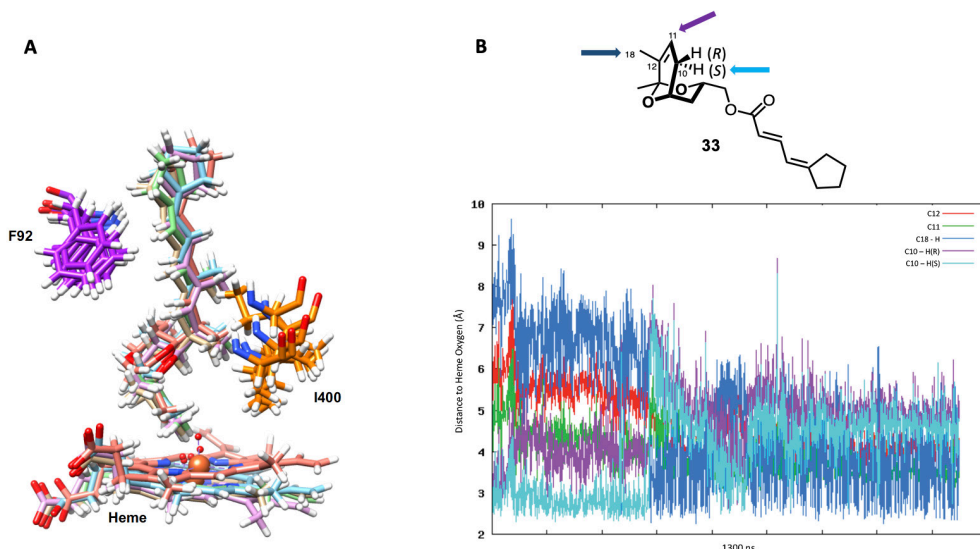


Figure 4.6. MD simulations with tirandamycin analogue 33. (A) Simulations of P450 TamI WT with **33** show the substrate maintains a stable geometry and orientation. (B) Distances of key carbon atoms (C10, C11, C12 and C18) of **33** to the heme-iron.

To further probe the chemical utility of anchor **25**, substrate **36** was synthesized from commercially available cyrene (Figure 4.7). Unlike with **33**, no oxidative reactivity was detected when testing substrate **36** with TamI S397A_T398A or WT (see Appendix C for details). This result suggests that although anchor **25** enables enzymatic oxidation of tirandamycin bicycle analogue **23**, it does not facilitate recognition of the less structurally related bicycle **36**. Examination of the five top occupied clusters of docking simulations with TamI and **36** shows the substrate in an unfavorable orientation and geometry for catalysis via the oxo-iron species (Figure 4.7). The bicyclic moiety of **36** points away from the heme iron and residue Phe92 (critical for recognition and binding via hydrophobic interactions with anchor **25**) is not close in

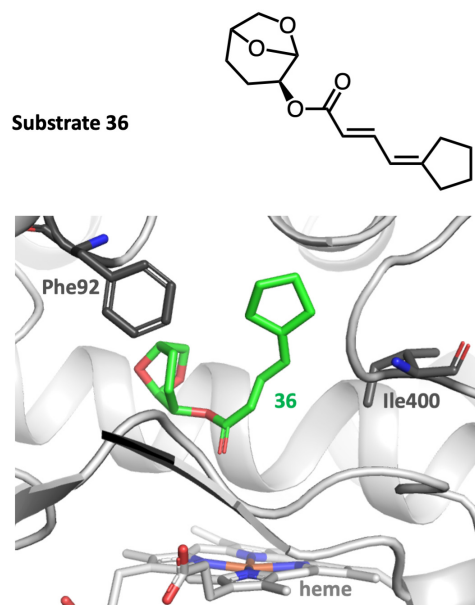


Figure 4.7. Molecular docking of TamI with substrate 36. (A) Structure of substrate 36. For details on synthesis, see Appendix C. (B) Docking simulations with P450 TamI and **36**.

proximity to the conjugated polyene moiety in **36**. These simulations suggest perturbed substrate binding for substrate **36**, which supports the lack of oxidative activity detected with P450 TamI experimentally.

4.4 Discussion

The work described herein demonstrates the ability of biocatalyst P450 TamI for accommodating and oxidizing unnatural substrates that harbor key structural features for binding in the enzyme's active site. Using a substrate engineering approach, we designed and synthesized various tirandamycin analogues that interrogate important questions towards maximizing the chemical utility of P450 TamI including 1) What are the substrate's structural requirements for binding and oxidation with P450 TamI? and 2) What is the substrate scope of the enzyme and can it be expanded by exploiting the native substrate's anchoring mechanism?

To begin our studies, the total synthesis of a tirandamycin bicycle analogue was optimized using an asymmetric allylation methodology. This optimization shortened the total number of reaction steps required to produce the bicycle analogue, facilitating its access for further testing. Guided by previous structural and enzymatic studies¹⁷ and new docking experiments, we designed and synthesized simplified anchoring groups **25** and **30** that contain a conjugated polyene system that could interact via hydrophobic and π - π interactions with active site residues Phe92, Leu399 and Ile400 and facilitate recognition and binding of unnatural substrates in P450 TamI (Scheme 4.3). Anchor **30** contains a triene without any cyclic ring tailing the molecule and thus it allows us to interrogate if a cyclic moiety is required for proper substrate positioning in the enzyme. Anchor **25** contains a diene with a 5-membered hydrocarbon ring, helping us interrogate whether a cyclic moiety is needed at all for recognition by P450 TamI and whether the oxygen and nitrogen atoms in the tirandamycin tetramic acid moiety are necessary for substrate binding and oxidation or not. Both synthetic anchors contain a carboxylic acid group that facilitate coupling to the alcohol-containing tirandamycin bicyclic analogue **23** via a Yamaguchi reaction. Double variant TamI S397A_T398A was prioritized for testing with the tirandamycin analogues given that the mutant introduces a more hydrophobic environment in the active site (near where the tail of the tirandamycin substrate typically sits) and thus may better accommodate the highly non-polar synthetic analogues.

End-point assays revealed that substrate **34** (containing linear anchor **30**) is not oxidized by P450 TamI or variants. MD simulations with a truncated tirandamycin C (**1**) analogue that also lacks the tetramic acid showed that in the absence of a ring moiety the substrate adopts multiple bindings poses within the enzyme's active site, perturbing substrate binding and oxidation. Meanwhile, analytical enzymatic reactions with **33** led to significant three single oxidation products with TamI S397A_T398A. MD simulations showed that unlike **34**, this ring-containing analogue does maintain a proper geometry and stable conformation within the enzyme's active site for the entire length of the simulation, enabling productive TamI-mediated oxidative catalysis. MD simulations with **33** predicted that hydroxylation may be occurring at the C10 and C18 sites of the substrate, with a strong preference for the primary C18 methyl group.

To probe if P450 TamI could accept a bicyclic moiety other than the tirandamycin ketal one that comprises a bicyclo[3.3.1]nonane system, substrate **36** (bicyclo[3.2.1]octane) was synthesized. No oxidative reactivity was observed when testing this substrate with TamI enzymes, indicating that P450 TamI is limited to catalyzing oxidation of a tirandamycin-like bicyclic moiety. It is reasonable to hypothesize that a proposed H-bonding interaction between the oxygen atoms of the native bicycle group and the N ϵ of active site residue His102,¹⁷ interaction that appears to be missing in **36**, is key to enable substrate recognition and oxidation by P450 TamI.

These results provide important insights into the substrate structural requirements for oxidation by P450 TamI: 1) A conjugated polyene system is required for recognition and oxidation by the enzyme, and it appears that a diene is sufficient for turnover (instead of the triene moiety found in the native tirandamycin substrates); 2) A cyclic ring moiety is also required for enzymatic oxidation with P450 TamI, and it seems as if a hydrocarbon ring lacking the oxygen and nitrogen atoms found in the native substrate may suffice; 3) A synthetic analogue harboring a bicyclo[3.2.1]octane system is not oxidized by the enzyme, suggesting that either a nonane system is needed and/or that the octane one did not allow for a key potential interaction between the bicyclic heteroatoms and the enzyme's active site required for productive binding.

Although thought-provoking, our data also prompts additional questions that future studies will focus on answering including 1) Why are only single oxidation products (and not double or triple oxidation ones) detected with substrate **33** and double variant S397A_T398A?; 2) Can the enzyme accept a tirandamycin analogue with a 4- or 6-membered ring and how does this impact selectivity?; 3) Can P450 TamI oxidize a bicyclo[3.3.1]nonane-containing substrate lacking

heteroatoms and comprising only C-H bonds?; 4) Can P450 TamI oxidize any other bicyclic or polycyclic moieties?

This work demonstrates that the tetramic acid moiety is not essential for P450 TamI to bind and oxidize substrates. Particularly, a conjugated polyene containing a 5-membered hydrocarbon ring was sufficient for oxidation to occur. Sherman and co-workers¹⁷ had previously identified that the conjugated polyene moiety of tirandamycin is needed for substrate binding and this work expands this observation to include the importance of a ring moiety in the substrate for enzyme recognition and oxidation. Although our results also indicate that P450 TamI may be limited to catalyzing oxidation of bicyclic moieties that mimic the tirandamycin native bicyclic group, more synthetic analogues should be tested prior reaching a definite conclusion. Overall, our investigations provide an initial platform for further substrate engineering efforts with P450 TamI biocatalysts to facilitate iterative oxidation catalysis of unnatural synthetic analogues harboring bicyclic groups that are difficult to selectively oxidize using current synthetic C-H functionalization methods.

4.5 References

1. Yoshimura, H.; Takahashi, K.; Ishihara, J.; Hatakeyama, S. Unified synthesis of tirandamycins and streptolydigin. *Chem Commun (Camb)* **2015**, 51, 17004-17007.
2. Takahashi, K., Harada, R., Hoshino, Y., Kusakabe, T., Hatakeyama, S. and Kato, K. Formal synthesis of tirandamycin B. *Tetrahedron* **2017**, 73(25), 3548-3553.
3. Yadav, J. S., Dhara, S. and Mohapatra, D.K. Stereoselective total synthesis of 10-epi-tirandamycin E. *Tetrahedron* **2017**, 73(10), 1358-1366.
4. Chen, M.; Roush, W. R. Total synthesis of (-)-tirandamycin C. *Org Lett* **2012**, 14, 426-428.
5. Iwata, Y.; Maekawara, N.; Tanino, K.; Miyashita, M. Tetramic acid antibiotics: stereoselective synthesis of streptolic acid and tirandalydigin. *Angew Chem Int Ed Engl* **2005**, 44, 1532-1536.
6. Schlessinger, R. H., Beberitz, G.R., Lin, P. and Poss, A.J. Total synthesis of (-)-tirandamycin A. *JACS* **1985**, 107(6), 1777-1778.
7. Carlson, J. C.; Li, S.; Burr, D. A.; Sherman, D. H. Isolation and characterization of tirandamycins from a marine-derived *Streptomyces* sp. *J Nat Prod* **2009**, 72, 2076-2079.
8. Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H. Tirandamycin biosynthesis is mediated by co-dependent oxidative enzymes. *Nat Chem* **2011**, 3, 628-633.
9. V. Espinoza, R., Haatveit, K.C., Grossman, S.W., Tan, J.Y., McGlade, C.A., Khatri, Y., Newmister, S.A., Schmidt, J.J., Garcia-Borràs, M., Montgomery, J., Houk, K.N and D.H., Sherman. Engineering P450 TamI as an iterative biocatalyst for selective late-stage C-H functionalization and epoxidation of tirandamycin antibiotics. *ACS Catal* **2021**, 11, 8304-8316.
10. O'Reilly, E., Mark Corbett, Shahed Hussain, Paul P. Kelly, Dominique Richardson, Sabine L. Flitsch, and Nicholas J. Turner. Substrate promiscuity of cytochrome P450 RhF. *Cat Sci Technol* **2013**, 3(6), 1490-1492.
11. Whitehouse, C. J.; Bell, S. G.; Wong, L. L. P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev* **2012**, 41, 1218-1260.
12. Reetz, M. T.; Kahakeaw, D.; Lohmer, R. Addressing the numbers problem in directed evolution. *Chembiochem* **2008**, 9, 1797-1804.

13. Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J Biol Chem* **2006**, 281, 26289-26297.
14. de Raadt, A.; Griengl, H.; Weber, H. The concept of docking and protecting groups in biohydroxylation. *Chemistry* **2001**, 7, 27-31.
15. Lairson, L. L., Watts, A.G., Wakarchuk, W.W. and Withers, S.G. Using substrate engineering to harness enzymatic promiscuity and expand biological catalysis. *Nat Chem Biol* **2006**, 2(12), 724-728.
16. Narayan, A. R.; Jimenez-Oses, G.; Liu, P.; Negretti, S.; Zhao, W.; Gilbert, M. M.; Ramabhadran, R. O.; Yang, Y. F.; Furan, L. R.; Li, Z.; Podust, L. M.; Montgomery, J.; Houk, K. N.; Sherman, D. H. Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. *Nat Chem* **2015**, 7, 653-660.
17. Newmister, S. A.; Srivastava, K. R.; Espinoza, R. V.; Haatveit, K. C.; Khatri, Y.; Martini, R. M.; Garcia-Borras, M.; Podust, L. M.; Houk, K. N.; Sherman, D. H. Molecular basis of iterative C-H Oxidation by TamI, a multifunctional P450 monooxygenase from the tirandamycin biosynthetic pathway. *ACS Catal* **2020**, 10, 13445-13454.
18. Stachowski, J. Investigation of cytochrome P450 enzymes as biocatalysts for multifunctional C-H oxidation; and a case study of a graduate/undergraduate laboratory exchange program. *Doctoral Dissertation, University of Michigan* **2018**.
19. Dechert-Schmitt, A. M.; Schmitt, D. C.; Krische, M. J. Protecting-group-free diastereoselective C-C coupling of 1,3-glycols and allyl acetate through site-selective primary alcohol dehydrogenation. *Angew Chem Int Ed Engl* **2013**, 52, 3195-3198.
20. Stephen A. DiBiase, B. A. L., Anthony Haag, Raymond A. Wolak, and George W. Gokel. Direct synthesis of .alpha.,.beta.-unsaturated nitriles from acetonitrile and carbonyl compounds: survey, crown effects, and experimental conditions. *J Org Chem* **1979**, 44, 25, 4640-4649
21. Takacs, J. M., Boito, S.C. and Myoung, Y.C. Recent applications of catalytic metal-mediated carbocyclizations in asymmetric synthesis. *Curr Org Chem* **1998**, 2, 233-254.
22. Sun, H.; Kong, R.; Zhu, D.; Lu, M.; Ji, Q.; Liew, C. W.; Lescar, J.; Zhong, G.; Liang, Z. X. Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis. *Chem Commun (Camb)* **2009**, 7399-7401.

Chapter 5*

Microbiome Mapping of the Peruvian Boiling River

5.1 Background and Introduction

5.1.1 The Amazon Rainforest and its Macro- and Microdiversity

Highly biodiverse areas, including the tropics, are considered a rich source of new, structurally intriguing microbial-derived molecules that could serve as important leads to develop novel clinically relevant drugs including life-saving antibiotics.¹ New molecular scaffolds often reveal the activity of powerful biocatalysts that perform complex chemistry with high levels of selectivity and turnover efficiency. For example, the remarkable iterative P450 TamI biocatalyst (that performs C-H hydroxylation and epoxidation with exquisite stereo- and regioselectivity as described in Chapters 2-4) was isolated from the marine sediment-derived *Strep.* sp. 307-9 bacterium collected from the US Virgin Islands.² Isolation and taxonomical identification of the producing bacterium, and structural elucidation of the tirandamycin antibiotic natural products enabled the discovery of P450 TamI and its role in oxidative tailoring of the tirandamycin biosynthetic pathway. Thus, exploring biodiverse-rich areas with untapped microbial ecosystems can facilitate discovery of rare bacterial species that produce new molecular scaffolds occupying a highly diverse chemical space and harbor biocatalysts useful to organic chemists for the synthesis of bioactive molecules.

The Amazon Rainforest (also referred to the forests of the Amazon Basin, Figure 5.1) is considered the largest and most biodiverse tropical forest in the planet.³ It is mainly composed of lowland and terra-firme areas (non-flooded), although other intriguing Amazonian ecological niches exist including white sand soils (known as *campinas*) and seasonally flooded forests such as *várzeas* and *igapós*.⁴ The biodiversity in this region can vary drastically among each Amazonian ecosystem, with distinct spatiotemporal patterns identified for species richness in vertebrates and trees.⁵

* Chapter 5 is included in a manuscript in preparation.

Although Amazonia is considered a global biodiversity hotspot, the biodiversity patterns elucidated in the area have been primarily focused on macroscopic organisms living above-ground including flora and fauna.⁵ Conversely, the diversity of microorganisms in the region, including bacteria and fungi, has been severely understudied and can be referred to as the “hidden biodiversity” of the Amazon Rainforest.⁴ Despite of their overwhelming abundance, most Amazonian microorganisms, their geographical distribution, the molecules they produce, and the biocatalysts they harbor remain elusive. This gap in knowledge biases our understanding of Amazonian biodiversity, hinders scientific progress in the area and can

negatively impact the implementation and execution of conservation efforts to ensure sustainable development and preservation of the Amazonian jungle.

Elucidation of the Amazonian microdiversity has been hampered due to complications in obtaining research permits, safely accessing the area for sample collection, and identifying taxonomy due to poorly characterized reference sequence databases available for the region. Thus, the majority of Amazonian microbes have yet to be identified and mapped to improve our understanding of their chemical and biological relevance in Amazonian ecosystems.⁴ Culture-independent surveys of microbial communities based on high-throughput metagenomic DNA sequencing of targeted regions of 16S rRNA genes have expanded our understanding of microbial community composition, diversity, and dynamics, and enable characterization of the microscopic biodiversity found in poorly sampled sites in the Amazon Rainforest.⁶



Figure 5.1. Amazon Rainforest map. The Amazonia region (6.7 million km²) is delineated in a dark, green line. The 9 countries that share the Amazon jungle are labeled in yellow. The course of the Amazon River is shown with a white line. The Peruvian Boiling River is shown with a red mark. Amazon deforestation hotspots (as of August 2019) are shown in dark red circles. The fire intensity is approximated by the size of the circle. The map was generated using Google Earth Pro.

5.1.2 The Boiling River of the Peruvian Amazon

One of the most unique ecosystems found in all of Amazonia is located in the heart of the



Figure 5.2. Peruvian Boiling River photos. A) The hottest sections of the Boiling River reach over 200°F in temperature. B) Areas rich in flora diversity flank the river. C) Large algae mats can be found in some sections of the Boiling River during specific times of the year. Photos were taken by Stephanie King.

Central Peruvian Amazon, the Shanay-timpishka Geothermal System, popularly known as the “Boiling River of the Amazon” (Figure 5.2).⁷ Culturally, the Boiling River has been a site of great significance to the local Amazonians for many generations, and many still regard it as a sacred place of healing and great spiritual power. It is located in Peru’s Huánuco Department, Puerto Inca Province, and Honoria District; roughly 50 km southwest of the city of Pucallpa (the principal city of Peru’s Central Amazon, population circa 330, 000). A detailed discussion of the Boiling River, its ecosystem, and history can be found in Ruzo (2016)⁷ and *Ruzo et al., in progress*.⁸

Summarized, the Boiling River is located on the northern end of a geologic structure known as the Agua Caliente Dome, found near the center of the Ucayali Sedimentary Basin, and part of the tectonic deformation that created the Sira Mountains.⁸ The river flows into the Pachitea River, which in turn flows into the Ucayali River, and ultimately the Amazon River. It is considered one of the world’s largest thermal rivers, in spite of

being located over 700 km away from the nearest active volcanic center. The river flows hot for roughly 6.24 km, and reaches a maximum (non-flood) width and depth of circa 30 m and 4.5 m, respectively. The hottest temperature measured was a geothermal spring at 99.1 °C, and the river itself can reach over 90 °C in sections, particularly during the dry season when rain-input is at a minimum. It should be noted that though the river's water temperature may vary with rain-inputs, the geothermal springs remain stunningly consistent in their temperatures as geochemical parameters, independent of season.

The Boiling River starts off as a small, ambient-temperature stream, which flows for roughly 3 km before encountering multiple series of fault-fed hot springs, which inject geothermal waters into the stream, raising its temperature and volume until it becomes the imposing “Boiling River. The total river system, ambient and thermal, runs roughly 9 km. At specific points the temperature lowers enough for locals to shower and play in the river. It is likely that each habitat hosts microbial communities with distinct structures.⁹ Geologically, the section of fault-fed hot spring injection is strongly associated with the Agua Caliente Formation, a Cretaceous-age unit, which in this area is defined by very clean sandstones, originally deposited in a shallow-marine environment, and it's known as an ideal aquifer.

The geochemistry of the Boiling River and its springs is exceedingly clean, with most of its waters exhibiting circa-neutral pH, low electrical conductivities, and low total dissolved solids, to the degree that some elemental fingerprinting work has been difficult to complete due to a lack of natural geochemical tracers in the waters and minerals. Although extensive geoscientific and anthropological investigations are ongoing, the river's microbiome remains unexplored.

5.1.3 Ecological Importance of Studying the Boiling River

Deforestation associated with legal and illegal development pose a critical risk to the health and well-being of the Amazon and Boiling River ecosystems. Deforestation dramatically reduces the microbial diversity of a particular environment, affecting the dynamics of the local ecosystem in a long-lasting way.¹⁰ Local deforestation in the Boiling River area already appears to be impacting this once-pristine ecosystem, as thinned or cleared jungles result in increased run-off sedimentation. Given the increasing impacts of human activity on the Boiling River, the importance of this study becomes even more pressing as we face the risk of potentially losing microbial life in the Boiling River Area before we get an opportunity to document and understand it, and its natural product biosynthetic potential. Unfortunately, this devastating truth is not isolated

to our area of investigation as deforestation rates in the entire Amazon Rainforest continue skyrocketing. More than a quarter of the Amazon biome is predicted to vanish by 2030 if no action is taken.¹¹ As we continue to lose Amazonian biodiversity, it is tragic to note that although the Amazon is home to one in ten known species on the planet and one of the most biodiverse areas on earth, the bacterial diversity of this fascinating ecosystem is one of the least understood.¹⁰ Satellite-image based deforestation losses for the region by the Boiling River Project (conservation focused non-profit) estimated that roughly half a soccer of Peruvian Amazon is lost every minute. This estimate was determined prior to the COVID-19 pandemic, and deforestation losses are expected to have increased significantly.

Besides aiding in the study of Boiling River, documenting the microscopic life thriving in this area before further deforestation occurs will further general understanding of Amazonian microbial biodiversity and the unique extremophiles that have evolved to survive in the most challenging conditions at the center of the Peruvian Amazon. Understanding the bio-assets possible from the Boiling River Area's microbes, will also be key in informing the government on how to best sustainably manage this unprotected, exploitable area. We must explore, document, and investigate the Boiling River Area and the uncommon forms of life thriving within and along its rivers before the area is too highly impacted by human activity and before the fires of deforestation reach sensitive parts of this remarkable microscopic universe.

5.1.4 Amazonian Extremophiles as a Source of Natural Products and Thermostable Biocatalysts

Despite its inhospitable conditions, many forms of life, including blue-green algae, flourish in and around the Boiling River. Local microorganisms have evolved their genomes to include unique survival advantages so they can adapt and reproduce in this highly selective environment. It is also reasonable to imagine that previously undocumented species have emerged over time as a result of the external stress.^{12, 13} The heat-loving microorganisms in the Boiling River must have particular genotypes that allow them to live in temperature ranges inhospitable for many organisms, or, as is regularly seen in the Boiling River, kill terrestrial species that fall into it. Using cutting-edge technologies¹⁴ to unveil the remarkable microbiome of the Boiling River Area will provide unparalleled insights into the creation and development of each river and advance scientific knowledge about unique aquatic ecosystems that harbor extremophiles.

Unique microbes capable of thriving under extreme conditions have been found at the base of the tree of life.¹⁵ In fact, hyperthermophiles (organisms living at temperatures higher than 140°F) are considered to be the ancestor of bacteria, the earliest known life form.^{16, 17} These findings may support the Thomas Gold’s “deep, hot biosphere” theory.^{18, 19} He hypothesized that life on earth started with microbial life growing many kilometers below the surface, in the pores of grains in rocks, where temperatures are highly elevated. Despite the increasing interest in studying extremophiles to gain new insights into the origin of life, further investigation of the subsurface microbiome is needed.¹⁸ We know that the Boiling River “boils” due to fault-fed hot springs whose are likely surging up from at least a mile below the surface.⁷ Can the Boiling River Area hyperthermophiles serve as models of primordial organisms? Can they provide key new evidence to support the theory of a hot origin of life? Can the study of these organisms give us a glimpse at life in the deep-subsurface of the Amazon? Phylogenetic analysis of the local microbiome will reveal their evolutionary lineage and provide valuable insights into how life on the Amazon began.

Moreover, extremophiles have arguably been humanity’s “best-friend” in the modern era. They have taught and helped us “hack,” copy, and manipulate genetic codes, thus allowing us to make new medicines, materials, and even figure out our ancestry.²⁰ Simply studying one

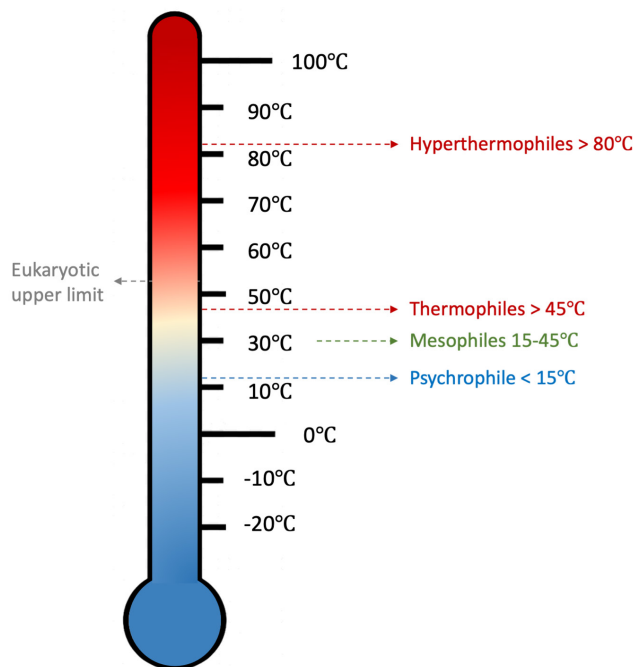


Figure 5.3. Extremophiles and their optimal environmental temperature ranges.

extremophile from the geothermal pools of Yellowstone National Park (*Thermus aquaticus*) allowed us to open the modern genetic age, and even opened up multi-billion-dollar industries.²¹ If this was possible at Yellowstone, what can happen when we study a geothermal “hot bed of life,”²² in the heart of one of the worlds mega-biodiverse region? Studying these organisms may yield valuable biotechnology tools and research probes with applications in commonly used laboratory techniques such as the Polymerase Chain Reaction (PCR) method,²³ or in the billion-dollar industry of bioremediation and

oil-spill cleanup.²⁴ These unique microbes can also lead to the discovery of intriguing secondary metabolites with biological properties against an array of life-threatening conditions including cancer and infections.^{25, 26} Additionally, analyzing the molecular machineries of these bacteria can result in the identification of powerful enzymes that perform intrinsic chemistry unmatched by current synthetic methods, such as cytochrome P450 enzymes, improving access to biologically relevant molecules.^{27, 28} Our specific objectives involve leveraging cutting-edge metagenomics, genomics and bioinformatics technologies to perform a meta-analysis of the microbial samples collected and generate the first local reference bacterial database of the Boiling River Area. Advancing scientific knowledge of the unexplored extremophiles that exist in the Boiling River Area will benefit the global community in four ways including: 1) Expanding the Earth's microbial biodiversity by discovering and characterizing previously undocumented²⁹ and biotechnologically relevant extremophiles and extremoenzymes before they go extinct; 2) Elucidating the mechanisms upon which natural selection enables local bacteria to survive under conditions traditionally considered detrimental; 3) Providing insightful details about the evolution of Amazonian life as well as the development of natural extreme rivers within the jungle using phylogenetic and taxonomic analysis; 4) Expanding the world's chemical and biocatalytic diversity by investigating extremophiles for their potential to generate unusual metabolites, and their corresponding biochemical assembly processes.

In the present study, we use metagenomics techniques to estimate bacterial and archaeal diversity in environmental sediment samples from 19 survey collection points (referred to as "stations") across the Boiling River of the Peruvian Amazon using ribosomal 16S rRNA gene sequences and Amplicon Sequence Variants (ASVs) as proxies. Our work provides the first large-scale biodiversity assessment across the Peruvian Boiling River including mapping of the soil microbial communities and elucidation of the geochemical and geophysical factors impacting the ASV patterns observed for bacterial richness and composition. This work serves as a framework for future genome mining of intriguing sediment samples to identify potential thermostable biocatalysts that perform difficult chemical reactions, including late-stage C-H functionalization, and new molecular scaffolds that could serve as leads to develop new antibiotic pharmaceuticals.

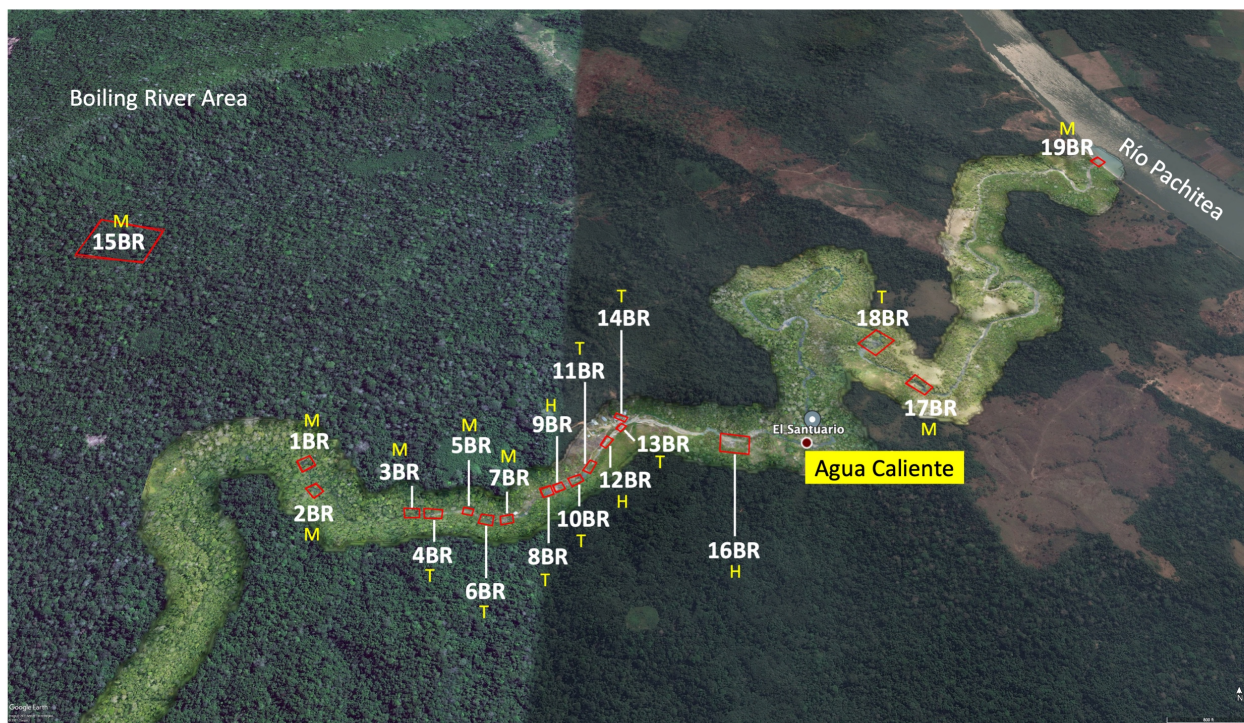


Figure 5.4. Peruvian Boiling River map. The perimeters for each sampling station are depicted in red. A total of 19 stations were sampled. For a close-up view and precise location of each station, refer to Appendix D. For specific values, refer to Table 5.1. Samples were collected from inside the water, riverbank or from a few feet from riverbank. The station's microclimate temperature range is shown in yellow as M (mesophilic, 15- 45 °C), T (thermophilic, > 45°C) and H (hyperthermophilic, > 80 °C).

5.2 Boiling River Sampling Expedition

5.2.1 Obtaining Research Permits from the Peruvian Government

The sample collection process and research work described herein are carried out in collaboration with the Universidad Nacional Agraria La Molina (UNALM), located in Lima, Perú and with the Boiling River Project led by geologist Andrés Ruzo from Southern Methodist University, located in Dallas, Texas. Specifically, the Sherman Laboratory at the University of Michigan and the Laboratory of Mycology and Biotechnology at UNALM have had an ongoing, synergistic research collaboration for many years built upon the development of a collaboration agreement (also known as Memorandum of Understanding, MOU). This international partnership has resulted in two scientific publications reporting high-quality, whole genome sequences from distinct fungal strains isolated from Amazon Rainforest soil.³⁰

To follow best practices, ensure fairness and safety, and comply with the ethical standard of research including the Nagoya Protocol on Access and Benefit Sharing,³¹ we obtained research and collection permits from the Ministry of Production (PRODUCE) in Lima, Peru (N° 540-2019-PRODUCE/DGPCCHI and N° 00151-2020-PRODUCE/DGPCCHI). Lasting for multiple months, this process included 1) Renewing the MOU between University of Michigan and UNALM; 2) Writing a research proposal in Spanish including a detailed description of our objectives and methodologies and following PRODUCE's formatting requirements; 3) Obtaining high-quality geographical maps of the Boiling River Area describing the sampling sites (Stations) with GPS coordinates (developed by a contracted local geospatial engineer) ; 4) Creating a model field tab following PRODUCE's requirements for field notes collection; 5) Attending regular meetings with decision-maker officials within PRODUCE to provide further details on our research proposal and methodology as needed; and 6) Providing signed documentation that shows the Life Sciences Institute at University of Michigan supports our research work and collection expedition in Peru. As a result of this work, the Sherman Laboratory became the first academic research laboratory in the United States to obtain a research and collection permit to study free-living microbial species (not living in symbiosis with higher forms of life such as plants or animals) found in Peruvian territory. The process we initiated, followed, and completed now serves as a framework to facilitate future permit applications for scientific research and collection of microbial samples in Peru.

5.2.2 Design and Execution of Sampling Strategy

The sampling design covered the lower, middle, and upper courses of the Boiling River. A total of 19 sampling sites (stations) were prioritized based on accessibility and unusual geographical features including noticeable smells of sulfur-like gases. The 19 stations varied in temperature including warm to hyperthermophilic environments (22.0 – 93.7 °C), pH including acidic to alkalophilic microclimates (6.44 – 8.48) and electrical conductivity or EC (123.5 – 3051.0 uS/cm). Sediment samples were collected from different points within each station including under water, riverbank and from a few feet from the riverbank. Samples were qualified as wet or dry based on qualitative field observations. Some wet samples contained a small amount of inseparable vegetation litter or cyanobacterial filaments. After collection, samples were stored in secured, previously labeled, heat-resistant plastic tubes. GPS coordinates and general field observations, such as water and odors, were documented. High-quality photographs and videos were taken at

each sampling site to showcase current sampling conditions and document the macroscopic view of our fieldwork. When applicable, sediment was collected as surface samples, discrete-depth samples, and depth-integrated samples. The sampling device was made of stainless steel to avoid contamination of organic compounds present in the sample. Samples were immediately stored at 40°F after their arrival at the Sherman Laboratory at University of Michigan (USDA Permit N° P526P-18-04189, regulated by 7 CFR 330). Strict safety guidelines were followed to ensure the security of each team member during sample collection.

Stations' associated microclimates					
Station	Station name	Water temp. (°C)	Extremophiles temp. ranges	pH	Electrical conductivity (uS/cm)
1BR	Poza Las Cortinas	23.4	M	6.94 @ 26.1°C	144.5 @ 28.9°C
2BR	Escaleras del Kakataibo	27.1	M	8.13 @ 25.6°C	310.9 @ 25.6°C
3BR	Tambos Yacumama	22.0	M	7.25 @ 24.3°C	156.6 @ 25.2°C
4BR	Yacumama & Paredes Fumonas	76.8	T	7.99 @ 24.2°C	297.7 @ 24.2°C
5BR	Plaza de las Fallas & KM 6	38.8	M	6.44 @ 25.8°C	226.0 @ 25.8°C
6BR	Poza Escalera de Raíces	61.3	T	7.00 @ 26.1°C	261.0 @ 30.2°C
7BR	Poza Dos Cataratas	41.3	M	7.60 @ 32.5°C	123.5 @ 33.8°C
8BR	Poza del Sheripiari	66.3	T	6.54 @ 25.8°C	293.0 @ 25.8°C
9BR	Cauce del Sheripiari	89.7	H	7.10 @ 26.0°C	356.0 @ 30.5°C
10BR	Catarata del Sumiruna	64.5	T	n.d.	n.d.
11BR	Cañon del Sumiruna	79.3	T	7.81 @ 24.2°C	236.3 @ 25.4°C
12BR	Las Aguas Sagradas	93.7	H	7.90 @ 24.0°C	508.0 @ 24.7°C
13BR	Rápidos de Mayantuyacu	79.3	T	7.96 @ 24.2°C	481.3 @ 24.5°C
14BR	Quebradita Mayantuyacu	49.9	T	6.97 @ 39.4°C	377.0 @ 39.9°C
15BR	Poza Salada	24.5	M	7.52 @ 24.3°C	2127.0 @ 24.8°C
16BR	Plaza de Azufre	82.2	H	7.58 @ 24.2°C	3051.0 @ 25.3°C
17BR	Catarata Impicciatori	43.1	M	n.d.	n.d.
18BR	Escaleras de Morán	47.6	T	8.48 @ 24.3°C	852.0 @ 25.0°C
19BR	Boca del Shanay-Timpishka	42.5	M	7.75 @ 24.3°C	842.0 @ 25.3°C

Table 5.1. Boiling River microclimates. The physicochemical values shown in this table represent the microclimate conditions of each sampling station, with the closest associated Boiling River water parameters included. Samples were collected from inside the water, riverbank or from a few feet from riverbank. The station's microclimate temperature range is shown as M (mesophilic, 15- 45 °C), T (thermophilic, > 45°C) and H (hyperthermophilic, > 80 °C). The annotation n.d. refers to no data available.

5.3 Characterization of the Boiling River Microbiome

5.3.1 Isolation of Metagenomic DNA and Sequencing the 16S rRNA gene

A total of 114 sediment samples (six samples per station) were processed for total DNA extraction using the DNeasy PowerSoil Kit and following the manufacturer's instructions with minor modifications (see Appendix D for details). We targeted, amplified, and sequenced the V3-

V4 regions (approximately 460 bases) of the ribosomal 16S rRNA gene using GENEWIZ metagenomics services, which followed standard protocols using the Illumina MiSeq 2x250 platform. A bioinformatic pipeline was developed *in-house* to process the metagenomic 16S rRNA gene sequences and generate ASV data using DADA2, QIIME 2.0 and RStudio (see Appendix D for details).

5.3.2 Microbial Diversity Meta-Analysis

We obtained a total of 13,583,684 reads that resulted in 60,296 unique ASVs using the DADA2 pipeline and strict parameters (99% similarity cutoff using SILVA SSU 138.1),^{32, 33} with 59,880 ASVs for bacteria, 183 ASVs for archaea, and 47 ASVs for eukaryota at the domain level (see Appendix D for details including rarefaction curves and sequence quality). A total of 186 ASVs were unassigned. After filtering, denoising, and merging, the resulting ASVs were 443 bp long in average, with 98% of sequences consisting of 467 bp total. In the bacterial domain, a total of 10 established and 1 candidate phyla taxa were detected in 1% or more of the 114 samples and 56 phyla groups (including 31 candidate phyla) were found in < 1%. Approximately 0.74% of bacterial ASVs could not be further classified, indicating that a considerable proportion of novel microorganisms were detected that are different from those included in bacterial reference databases.

5.3.2.1 Taxonomic Composition

The taxonomic composition of the bacterial domain (1% cutoff) depicts that the phylum with the highest number of ASVs is *Proteobacteria* (37.4 % of the bacterial taxa identified in the 114 samples, most of which belong to *Alphaproteobacteria* and *Gammaproteobacteria*, Figure 5.5). This agrees with previous studies describing the microbial composition of other temperate and hot thermal springs around the world where *Proteobacteria* dominates.³⁴⁻³⁶ In general, this taxon is well distributed among Boiling River samples with a slightly higher abundance in stations 11BR (2.59 %), 19BR (2.54 %), 17BR (2.34 %), 3BR (2.26 %) and 1BR (2.25 %), compared to stations 8BR (1.42 %) and 16BR (1.17 %) (Figure 5.6). The next most numerous phyla are *Actinobacteriota* (15.5 %, with *Actinobacteria* and *Thermoleophilia* as the most abundant classes), *Bacteroidota* (11.8 %, most of which belong to *Bacteroidia*), *Firmicutes* (10.1 %, with *Clostridia* as the predominant class), *Acidobacteriota* (8.6 %, with dominant levels of *Vicinamibacteria* and *Acidobacteriae*) and *Chloroflexi* (7.2 %, with *Chloroflexia* as the most prevalent class group). The

least frequent phyla include *Cyanobacteria* (3.7 %, *Cyanobacteriia*), *Myxococcota* (2.3 %, *Polyangia*), *Gemmatimonadota* (2.0 %, *Gemmatimonadetes*), *Desulfobacterota* (1.3 %, *Desulfuromonadia*) and candidate *Methylomirabilota* (1.0 %, *Methylomirabilia*). Most of these phyla groups are unevenly distributed across the river, and candidate *Methylomirabilota*, considered essential in evolutionary history and biogeochemical processes as its members are denitrifying methanotrophs, is virtually absent from many samples but significantly present in P105 (4BR).³⁷ At the family level (1 % cutoff), the most abundant taxon is *Chitinophagaceae* (6.07 %, *Bacteroidota*). Members of this group prevail in station 12BR (0.66 %) and are virtually absent from station 16BR (0.03 %). Family *Microscillaceae* (3.55 %, *Cytophagales*) dominates station 19BR (0.85 %) with nominal representation in stations 10BR and 11BR (0.03 % in both locations). At the genus level (0.5 % cutoff) *Pseudomonas* (*Proteobacteria*) and *Brevundimonas*



Figure 5.5. Krona visualization of bacterial diversity. The 16S rRNA amplicon-based diversity of bacteria in Boiling River samples. Each circle represents the phylum and class taxa from inside to outside. The relative abundance of bacteria representing each group is depicted by the percent identity. For microbial composition at the order, class, and genus level, refer to Appendix D.

(*Proteobacteria*) are the most numerous groups, with numerous ASVs found in station 11BR for both genera.

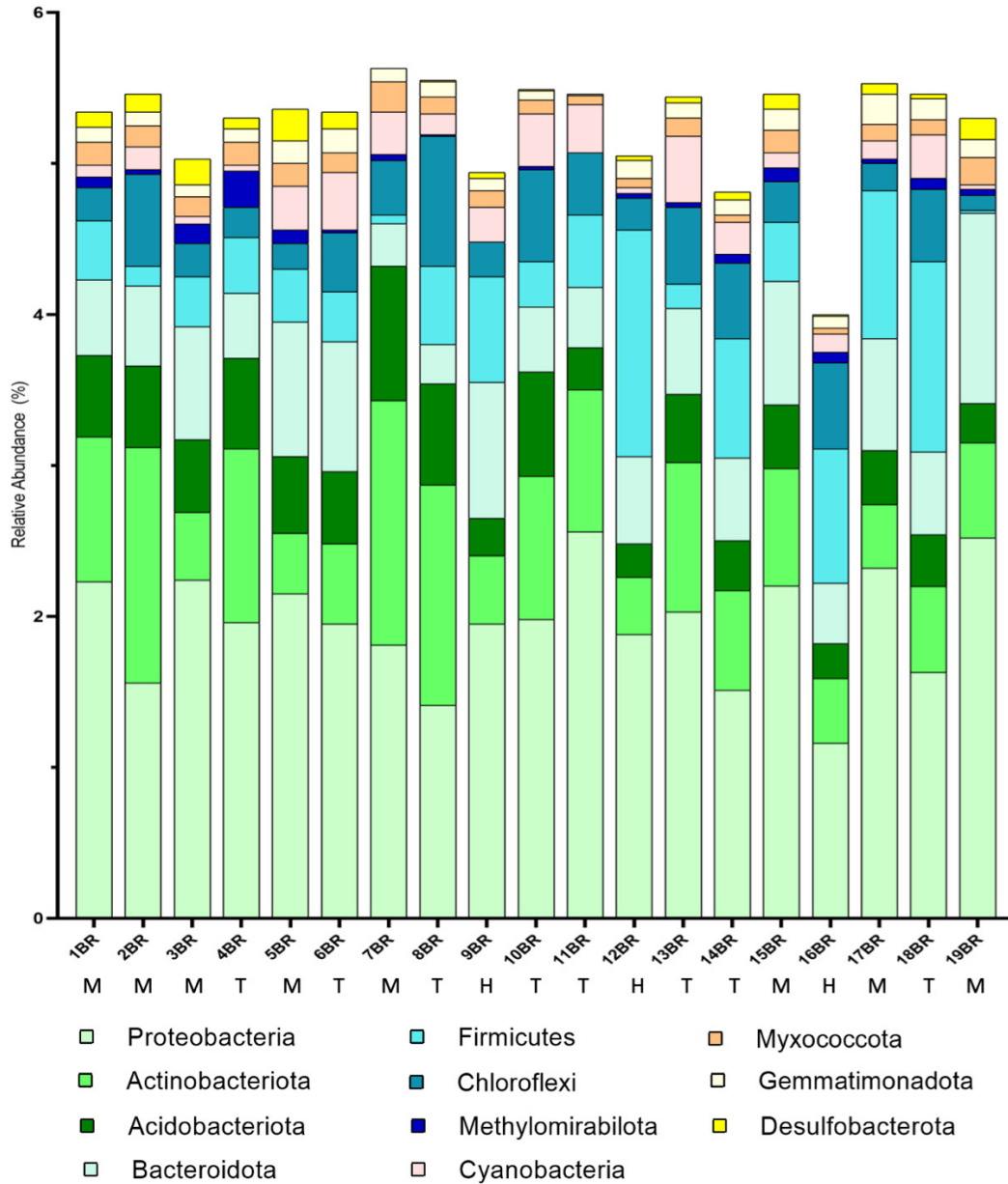


Figure 5.6. Distribution of bacterial community. Relative abundance of total bacterial phyla (1% cutoff) across sampling stations in the Boiling River. Each bar contains data for the six samples collected and processed per station. The station’s microclimate temperature range is shown as M (mesophilic, 15- 45 °C), T (thermophilic, > 45°C) and H (hyperthermophilic, > 80 °C). To see relative abundance at the phylum level per sample, refer to Appendix D. For a map of the river, refer to Figure 5.3

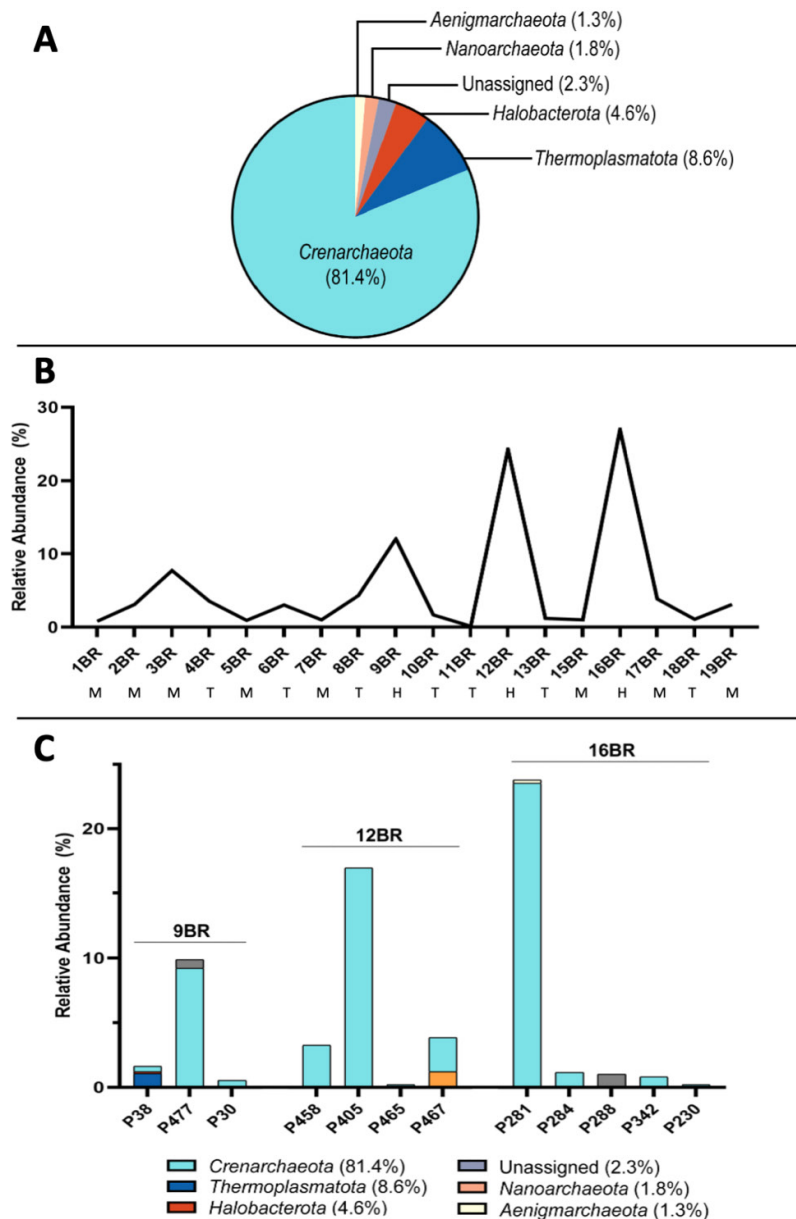


Figure 5.7. Archaeal composition and distribution. A) Total relative abundance of archaeal phyla in the Boiling River (1% cutoff). B) Archaeal relative abundance across stations. Not every sample analyzed contained archaeal ASVs including all six samples from station 14BR. The station's microclimate temperature range is shown as M (mesophilic, 15- 45 °C), T (thermophilic, > 45°C) and H (hyperthermophilic, > 80 °C). C) Stations with the most numerous archaeal ASVs.

Most of the eukaryotic ASVs are unassigned. Contribution of archaeal ASVs to the whole microbial communities is lower than 0.2%. For archaea, the group with the highest number of ASVs is *Crenarchaeota* (81.4 % of the archaeal taxa in the 114 samples, most of which belonged to *Nitrososphaeria*, Figure 5.7A) followed by *Thermoplasmata* (8.6 %). Most of the archaeal ASVs for all taxa are relatively limited to specific sampling sites, occurring primarily in stations 3BR, 9BR, 12BR and 16BR (Figure 5.7B). Interestingly, these sampling sites also show a slightly lower abundance in bacterial phyla, suggesting a small inverse correlation between the domains. A closer examination of the archaeal distribution revealed that the elevated average abundance observed for specific stations derives primarily from one sample harboring most of the station's archaeal ASVs (Figure 5.7C) at each site. This observation indicates that the archaeal

abundance in the Boiling River varies widely within and across stations, even within a few meters distance.

Contrarily, bacterial ASVs were widespread across all the stations. We observed a small non-linear reduction in the number of relative abundances of microbial phyla as some of the stations' microclimates (9BR, 12BR and 16B) increased to near boiling temperatures (79 °C or above, Figure 5.5). This indicates that the thermal characteristics of the Boiling River microclimates have an influence in shaping the local microbial community composition. The lowest abundance of total bacterial phyla (near 4 %) is found in station 16BR, where sulfide odor, white sulfur precipitates and the largest EC values have been detected.

5.3.2.2 Clustering and Correlation Analysis

Actinobacteriota, the second most numerous bacterial taxon (15.5 %), is correlated with the distribution of phyla *Acidobacteriota* (8.6 %), *Myxococcota* (2.3 %) and *Cyanobacteria* (3.7%). For all four bacterial groups, stations 7BR, 10BR and 13BR show some of the highest levels of ASV abundance on average (see Appendix D) while stations 12BR and 16BR display some of the lowest.

Bacteroidota, the third largest bacterial taxon (11.8 %), displays affiliation with *Myxococcota* (2.3%) and *Desulfobacterota* (1.3 %), with stations 3BR, 5BR, 9BR, 15BR and 19BR exhibiting the highest abundance on average. Unexpectedly, *Desulfobacterota*, which comprises the taxa previously classified as *Thermodesulfobacteria*³⁸ and includes thermophilic sulfate-reducing bacteria, is barely present in station 16BR. Meanwhile, *Firmicutes* (10.1 %) is primarily present in stations 12BR, 16BR, 17BR and 18BR and does not show any specific correlation to other phyla. *Chloroflexi* (7.2 %) is predominant in stations 2BR, 8BR, 10BR, 13BR, 14BR, 16BR and 18BR and does not display any specific pattern of clustering with other groups. The noticeable abundance of *Firmicutes* and *Chloroflexi* in station 16BR, compared to other taxa groups, is noteworthy. Previous studies report the ability of *Firmicutes* bacteria to produce endospores resistant to extreme environments³⁹ and *Chloroflexi* has shown to include representative members that are aerobic thermophiles and anaerobic halorespirers,⁴⁰ both of which could explain their increased presence in station 16BR. Lastly, *Gemmatimonadota* (2.0 %) is most abundant in station 17BR and shows a discrete affiliation with *Bacteroidota* and

Cyanobacteria in stations 5BR and 6BR where the high ASV abundance levels are maintained for all three groups.

5.3.2.3 Samples with Unique Microbial Configurations

For unconstrained ordination of microbial communities, a measure of β diversity known as Non-Metric Multidimensional Scaling (NMDS) using the Bray-Curtis dissimilarity method was performed. This indirect gradient analysis indicates the amount of variance in phyla microbial diversity found in each station, which can unravel relational patterns among microbial composition profiles from phyla-station data. (Figure 5.8A). The NMDS results were quantitatively evaluated with permutational multivariate analysis of variance (PERMANOVA), which found no significant difference in the stations' microbial composition at the phyla level. This outcome indicates that location within the river is not the major determinant of the stations' bacterial community structure.

However, NMDS analysis of the phyla abundance detected in each sample rather than in each station was found to be statistically significant (Figure 5.8B).³⁸ A closer examination of these results revealed samples with a unique microbial configuration. While most samples cluster together within the same two-dimensional space (indicating they are similar in phyla composition), samples P284 and P281 (station 16BR), P477 (9BR), and P214 (14BR) present as outliers. For P477, the taxa that positions the sample as an outlier with a significant distinct microbial composition include *Candidatus Fervidibacteria* (0.05 % of the absolute abundance of bacterial taxa identified for all 114 samples), and an unassigned *Archaea* group (2.3 % of the archaeal taxa). For P281 and P284, the globally distributed *Candidatus Acetothermia* (0.28 %)⁴¹ and the thermophilic *Thermotogota* (0.20 %).⁴² For P281 and P214, *Caldisericota* (0.06 %) and the candidatus phylum *TA06* (0.06 %) that was first discovered in a hypersaline microbial mat.⁴³ For P284, the group *Aquificota* (0.52 %) that is known for its strict hyperthermophilic growth requirements,⁴⁴ for P214 only, *Coprothermobacterota* (0.01 %) and for P281, archaeal group *Aenigmarchaeota* (1.3 %). Refer to Appendix 5 for a detailed view of the phyla configuration of each sample described above.

To validate our observations, a related technique known as Correspondence Analysis (CA) was used (Appendix D). As detected with NMDS, P284 is farthest away from most samples with phyla *Aquificota*, *Thermotogota* and *Candidatus Acetothermia* having a significant impact on its microbial community structure. Although not as dissimilar as P284, sample P477 continues

appearing distant to most samples. A hierarchical cluster dendrogram was constructed with UPMG and repeated 999 times to further observe the patterns of clustering and relatedness of bacterial phyla in each sample with statistical significance. In addition of showing P284 and P477 as outliers, most samples clustered in six major groups (Figure 5.8C) that did not display any distinct station similarity among them except for most samples from stations 7BR, 12BR and 19BR.

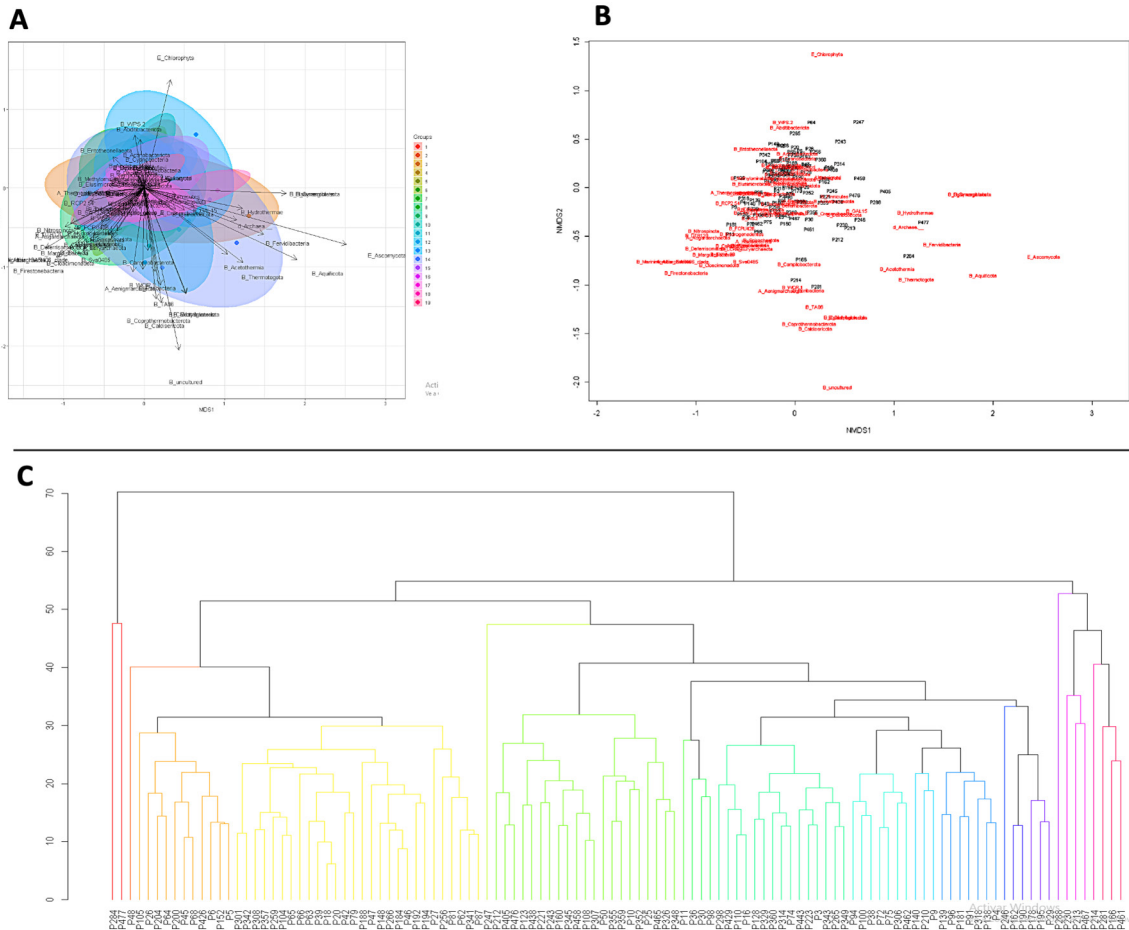


Figure 5.8. Clustering patterns of Boiling River microbes. A) Non-Metric Multidimensional Scaling (NMDS) analysis for all 19 stations at the phyla level. Each station is shown in a different color. Each arrow represents a phylum. The longer the arrow, the more representative the phylum is in the samples included in its trajectory. No statistically significant difference in the stations' microbial composition was detected. B) Statistically significant NMDS analysis for all 114 samples at the phyla level. Most samples cluster together suggesting a similar microbial configuration, while some are shown as outliers indicating a dissimilar and unique phyla composition. C) A dendrogram showing taxonomic relationships among Boiling River samples. Each color indicates a cluster formed. P284 and P477 appear distant to most samples. Samples that show association include P62, P63, P65, P66 (7BR, yellow cluster); P405, P476, P438, P458, P465 (12BR, green) and P429, P329, P360, P343, P349 (19BR, turquoise).

5.3.3 Biomarker Evidence for Bacteria Involved in Sulfur Cycling

To elucidate the biological fingerprint of station 16BR and identify groups potentially involved in sulfur cycling, the microbial configuration of the six samples collected in this site were further examined (Figure 5.9 and Appendix D). The uncultured bacteria GAL15 was the most numerous genus found in this station (8.57 %, after 0.5 % cutoff) and prevailed in samples P288, P281 and P342, with no ASVs found in P284 and P160. Little is known about the representative members of this phylum, but candidate GAL15 has been primarily found in thermal hot springs with circumneutral pH and low levels of sulfide.⁴⁵ Additionally, this phylum is often more abundant in oxic subsurface sediments.⁴⁶ Consistent with previous studies⁴⁷ where the presence of GAL15 increases with soil depth, this candidate group is most abundant in P288, which was collected from a deep hole in the soil of station 16BR (approximately 12-16 inches below ground level).

P284 was dominated by *Sulfurihydrogenibium* (*Hydrogenothermaceae*) with a 34.6% abundance. This genus was virtually absent from all other Boiling River samples. First detected by culture-free DNA analysis,⁴⁸ members of the genus *Sulfurihydrogenibium* are neutrophilic and thermophilic, commonly found in circumneutral pH and play an essential role in iron mineralization and sulfur cycling. They have been reported to use various electron donors (thiosulfate, sulfite, elemental sulfur, Fe (II), arsenite, selenite and H₂) and acceptors (O₂, nitrate, arsenate, selenate and Fe (III)).⁴⁸ This sulfur-oxidizing genus has been found in sulfate-rich areas in hot springs from Iceland, Japan, and Yellowstone National Park.⁴⁹ Other sulfur-oxidizing and thermophilic genera were also prevalent and exclusively found in P284 including *Thiofaba* (*Halothiobacillaceae*, 6.7 %)⁵⁰ and *Hydrogenobacter* (*Aquificaceae*, 7.3 %).⁵¹ The co-dominance of genera *Sulfurihydrogenibium*, *Thiofaba* and *Hydrogenobacter* in P284 reveals the importance of sulfur metabolism in this section of the Boiling River and indicates the existence of a sulfur cycle in station 16BR between the three dominant genera.

Genus *Sulfurospirillum* (*Sulfurospirillaceae*), known for preferentially growing in areas rich in arsenate or selenate (or other organohalides such as tetrachloroethene, PCE)⁵² dominated the microbial composition of sample P230 (collected in vicinity to P284, further implying these minerals may be present in the area. Further examination revealed that *Thiofaba*, *Hydrogenobacter* and *Sulfurospirillum* are present in various samples from thermophilic stations 12BR and 14BR

(P212-214, P458, P405, P476, P438) raising the possibility that sulfur cycling may occur in these areas despite the absence of sulfur gases detected empirically.

Lastly, sulfur-reducing genera including *Desulfobacca* (*Desulfobaccaceae*), Candidate *PB19* (*PB19*), *Syntrophobacter* (*Syntrophobacteraceae*), *Desulfofundulus* (*Desulfotomaculales*), *Desulfotomaculum* (*Desulfotomaculales*), *Desulfitibacter* (*Desulfitibacteraceae*), and *Desulfovirgula* (*Thermoanaerobacteraceae*),⁵³ were found in less than 0.5 % of samples from station 16BR and most of them were exclusive to P281.



Figure 5.9. Krona chart of microbial diversity in Plaza de Azufre (station 16BR). The 16S rRNA amplicon-based diversity of bacteria in Boiling River samples collected from station 16BR. Each circle represents phylum to genus from inside to outside. The relative abundance of bacteria representing each group is depicted by the percent identity.

5.4 Discussion

In this work, the bacterial and archaeal communities of 19 sections (denominated sampling stations) in the Boiling River of the Peruvian Amazon Rainforest were investigated for the first time through amplicon sequencing and molecular examination. Each station harbors a unique microclimate (°C, pH, and electrical conductivity) with wide water temperature ranges including mesophilic to hyperthermophilic conditions. Samples collected and analyzed included dry and wet sediment samples, and in some cases wet samples contained inseparable amounts of vegetation litter or cyanobacterial filaments, and they were collected from inside the river, riverbank or within a few meters from the riverbank. This study provides a large-scale biodiversity elucidation of the Boiling River and its associated microclimates to map the local bacterial communities in and around the thermal system, elucidate any physicochemical factors impacting the microbial composition and identify samples harboring potentially new microbial species that could translate into novel and intriguing biosynthetic gene clusters.

Using ribosomal 16S rRNA gene sequences amplified from environmental metagenomic DNA, Amplicon Sequence Variants (ASVs) were generated as proxies of microbial diversity. This ASV approach can distinguish sequences with single-nucleotide resolution providing a more precise measure of sequence variation that does not rely on clustering between sequencing reads (de-novo OTUs or operational taxonomic units) or against a reference database (closed-reference OTUs). Different statistical and bioinformatic analyses were conducted to characterize the microbial composition, distribution, and diversity across stations at higher and lower taxonomic levels.

Proteobacteria was the dominant phylum in the Boiling River with a relative abundance of 37.4 % of the bacteria. Other phyla groups detected in 1% or more of the samples include *Actinobacteriota* (15.5 %), *Bacteroidota* (11.8 %), *Firmicutes* (10.1 %), *Acidobacteriota* (8.6 %), *Chloroflexi* (7.2 %), *Cyanobacteria* (3.7 %), *Myxococcota* (2.3 %), *Gemmatimonadota* (2.0 %) *Desulfobacterota* (1.3 %) and *Methylomirabilota* (1.0 %). From a continental perspective, a similar taxonomical dominance has been reported for other hot springs around the world (South Africa, Malaysia, Pakistan, India, Tibetan)⁵⁴ where *Proteobacteria* was found to be the most numerous taxa in conditions typically averaging 45 °C. In water systems with thermo- and hyperthermophilic conditions (45 - 100 °C) such as the Mirror Pool in Yellowstone National Park, and Flúðir and Hurðarbak in Iceland, *Chloroflexi* was the most abundant taxon followed closely by

Cyanobacteria.⁵⁴ Taxa *Bacteroidetes* was also significantly numerous in the Icelandic thermal systems. Hot springs located in the Tibetan Plateau that share similar extreme conditions show a prevalence of *Cyanobacteria*, *Aquificota*, *Thermotogae* (also known as *Thermotogota*), and *Thermodesulfobacteria* in their bacterial compositions. In the Boiling River, the presence of *Chloroflexi* and *Cyanobacteria* vary across the river without any specific environmental pattern, while *Aquificota* (0.52 % absolute abundance) is primarily found in samples P284 (station 16BR) and P477 (9BR), and *Thermotogota* (0.20 %) is mostly present in P214 (14BR). The hypothesis that Boiling River water emerges from underground could explain the relatively low amounts of *Cyanobacteria* found in the area compared to the other hot spring communities, despite of the river having a large solar exposure throughout its course.

Nitrospirae is also predominant in most other⁵⁵ thermal springs, but it is found in low levels (0.83 % absolute abundance) in the Boiling River. *Actinobacteriota* is often detected in metagenomic surveys of hot springs with varying pH and temperatures, although usually in a lower abundance of 0.5 – 4% compared to its numerous 15.5 % in the Boiling River. These differences could be attributed to the specific physicochemical parameters and geographical location of this spring. Lastly, disproportionally abundant unassigned-phyla sequences were detected in samples P213-214 (14BR), P281 (16BR), P467 (12BR) and P461 (4BR) indicating potential of novel phylotypes. It is also reasonable to imagine that occasional earthquakes and a rapidly changing ecological environment could contribute towards the evolution of potentially new phyla or movement and exposure of unidentified subsurface microbes.

We did not find any clear spatial patterns in the microbial community composition of the Boiling River and its associated microclimates. Instead, we found a general trend of lowered microbial abundance with increasing microclimate temperature, and this is accordance with other hot spring metagenomic amplicon analyses.⁵⁶⁻⁵⁸ Many of the abundance patterns likely result from increased responses to the environmental stressors found in thermo- and hyperthermophilic microclimates.

Albeit found in nominal amounts, archaeal taxa were highly variable among the different locations. *Crenarchaeota*, the most numerous phyla group (81.4 %), dominated samples P281 (16BR), P405 (12BR) and P477 (9BR). This is in accordance with Yellowstone and Icelandic springs that also show *Crenarchaeota* as the dominant archaeal phylum in metagenomic analyses. However, previous investigations show that environments with high temperatures are often

associated with a high abundance of *Archaea* in the microbial composition.⁵⁹ Thus, it is unexpected that the *Archaea* population of the Boiling River is significantly low or even absent in some sampling stations. This may be due to the circumneutral to alkaline pH conditions in the area, as many *Archaea* groups are reported to prefer low pH environments, or biases in PCR amplification. Furthermore, similar to other thermal systems worldwide, a negative correlation between *Archaea* and bacteria ratio was detected.⁶⁰

Different statistical and bioinformatic analyses were conducted to characterize the microbial diversity across station and samples at the phyla taxa level. Geographical station location, other than station 16BR, did not seem to determine phyla community structure as shown with β diversity analyses including the Non-Metric Multidimensional Scaling (NMDS). However, an elevated abundance of a single phyla in one sample will influence clustering of samples based on phyla level abundances. No specific pattern was found for the uniqueness of phyla groups across sampling sites, except for the presence of bacteria involved in sulfur cycling and hydrocarbon degradation. Hydrocarbon-degrading genera including *Achromobacter* (*Alcaligenaceae*), *Acinetobacter* (*Moraxellaceae*), *Mycobacterium* (*Mycobacteriaceae*), *Pseudomonas* (*Pseudomonadaceae*), *Stenotrophomonas* (*Xanthomonadaceae*), *Aeromonas* (*Aeromonadaceae*), *Novosphingobium* (*Sphingomonadaceae*), *Sphingomonas* (*Sphingomonadaceae*), *Brevundimonas* (*Caulobacteraceae*) were found primarily in stations 11BR, 12BR and 15BR. Moreover, bacterial genera participating in sulfur cycling were common in 16BR, the one section of the river where sulfur gases were detected through a noticeable odor. Sulfur-oxidizing genera *Sulfurihydrogenibium*, *Thiofaba*, *Hydrogenobacter* and *Sulfurospirillum* were notably present in the station (Figure 5.9), while sulfur-reducing groups were detected in low abundances and appeared endemic to sample P281. Further geochemical studies to assess the presence of petroleum and sulfur in these sites could inform on the use of these genera as proxies for environmental monitoring and for the discovery of potential bioremediation biocatalytic tools.

This work describes the first elucidation of Boiling River bacteria and archaea found throughout this thermal system, encompassing mesophilic to hyperthermophilic microclimate environments. The local microscopic diversity is similar to that found in other hot spring systems around the world for most phyla groups, with some differences that could be explained with more in depth geochemical and geophysical investigations. Our study highlights that the “hidden” microdiversity (subterranean, subwater and superficial) of fragile environmental ecosystems need

to be taken into consideration to characterize a broader proportion of the total diversity patterns. This work serves as a framework and starting point for future genome mining of intriguing sediment samples to assemble and recover population-specific genomes from metagenomes for posterior bioinformatic analyses to identify thermostable biocatalysts that perform difficult chemical reactions, including late-stage C-H functionalization, and new molecular scaffolds that could serve as leads to develop new pharmaceuticals.

5.5 References

1. Ghirga, F., Quaglio, D., Mori, M., Cammarone, S., Iazzetti, A., Goggiamani, A., Ingallina, C., Botta, B. and Calcaterra, A. A unique high-diversity natural product collection as a reservoir of new therapeutic leads. *Org Chem Front* **2021**, 8(5), 996-1025.
2. Carlson, J. C.; Li, S.; Burr, D. A.; Sherman, D. H. Isolation and characterization of tirandamycins from a marine-derived *Streptomyces* sp. *J Nat Prod* **2009**, 72, 2076-2079.
3. Hansen, M. C.; Potapov, P. V.; Moore, R.; Hancher, M.; Turubanova, S. A.; Tyukavina, A.; Thau, D.; Stehman, S. V.; Goetz, S. J.; Loveland, T. R.; Kommareddy, A.; Egorov, A.; Chini, L.; Justice, C. O.; Townshend, J. R. High-resolution global maps of 21st-century forest cover change. *Science* **2013**, 342, 850-853.
4. Ritter, C. D., Zizka, A., Barnes, C., Nilsson, R.H., Roger, F. and Antonelli, A. Locality or habitat? Exploring predictors of biodiversity in Amazonia. *Ecography* **2019**, 42(2), 321-333.
5. Ritter, C. D.; Faurby, S.; Bennett, D. J.; Naka, L. N.; Ter Steege, H.; Zizka, A.; Haenel, Q.; Nilsson, R. H.; Antonelli, A. The pitfalls of biodiversity proxies: differences in richness patterns of birds, trees and understudied diversity across Amazonia. *Sci Rep* **2019**, 9, 19205.
6. Rodrigues, J. L.; Pellizari, V. H.; Mueller, R.; Baek, K.; Jesus Eda, C.; Paula, F. S.; Mirza, B.; Hamaoui, G. S., Jr.; Tsai, S. M.; Feigl, B.; Tiedje, J. M.; Bohannan, B. J.; Nusslein, K. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc Natl Acad Sci U S A* **2013**, 110, 988-993.
7. Ruzo, A. The Boiling river of the Amazon: adventure and discovery in the Amazon. *TED Books, Simon & Schuster Inc. New York* **2016**.
8. Ruzo, A. The Boiling river of the Amazon – characterizing the Shanay-timpishka geothermal system. *In preparation*.
9. Friedrich, M. W. Microbial communities, structure, and function. *Encyclopedia of Geobiology* (book) **2011**.
10. Kroeger, M. E., Delmont, T.O., Eren, A.M., Meyer, K.M., Guo, J., Khan, K., Rodrigues, J.L., Bohannan, B.J., Tringe, S.G., Borges, C.D. and Tiedje, J.M. New biological insights into how deforestation in Amazonia affects soil microbial communities using metagenomics and metagenome-assembled genomes. *Front Microbiol* **2018**, 9, 1635.
11. The World Wild Fund website, Amazon Deforestation site.
12. Lozupone, C. A.; Knight, R. Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **2007**, 104, 11436-11440.
13. Lynch, R. C., King, A.J., Fariás, M.E., Sowell, P., Vitry, C. and Schmidt, S.K. The potential for microbial life in the highest-elevation (> 6000 masl) mineral soils of the Atacama region. *J Geophys Res Biogeosci* **2012**, 117(G2).
14. Gilbert, J. A.; Dupont, C. L. Microbial metagenomics: beyond the genome. *Ann Rev Mar Sci* **2011**, 3, 347-371.
15. Di Giulio, M. The universal ancestor and the ancestor of bacteria were hyperthermophiles. *J Mol Evol* **2003**, 57(6), 721-730.
16. Rampelotto, P. H. Extremophiles and extreme environments. *Life (Basel)* **2013**, 3, 482-485.
17. Islas, S., Velasco, A.M., Becerra, A., Delaye, L. and Lazcano, A. Extremophiles and the origin of life. *Physiology and biochemistry of extremophiles* (book) **2007**, 1-10.

18. Colman, D. R.; Poudel, S.; Stamps, B. W.; Boyd, E. S.; Spear, J. R. The deep, hot biosphere: twenty-five years of retrospection. *Proc Natl Acad Sci U S A* **2017**, 114, 6895-6903.
19. Gold, T. The deep, hot biosphere. *Proc Natl Acad Sci U S A* **1992**, 89, 6045-9.
20. Coker, J. A. Extremophiles and biotechnology: current uses and prospects. *F1000Res* **2016**, 5.
21. Brock, T. D. The value of basic research: discovery of *Thermus aquaticus* and other extreme thermophiles. *Genetics* **1997**, 146, 1207-1210.
22. Laursen, L. Why the tropic are an evolutionary hotbed. *Nature News* **2013**.
23. Champdore, M. D., Staiano, M., Rossi, M. and D'Auria, S. Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest. *J R Soc Interface* **2007**, 4(13), 183-191.
24. Peeples, T. L. Bioremediation using extremophiles. *Microbial Biodegradation and Bioremediation* (book) **2014**, 251-268.
25. Wilson, Z. E.; Brimble, M. A. Molecules derived from the extremes of life: a decade later. *Nat Prod Rep* **2021**, 38, 24-82.
26. Zhang, X.; Li, S. J.; Li, J. J.; Liang, Z. Z.; Zhao, C. Q. Novel natural products from extremophilic fungi. *Mar Drugs* **2018**, 16.
27. Antranikian, G. a. E., K. Extremophiles, a unique resource of biocatalysts for industrial biotechnology. *Physiology and Biochemistry of Extremophiles* (book) **2007**, 359-406.
28. Niehaus, F., Bertoldo, C., Kähler, M. and Antranikian, G. Extremophiles as a source of novel enzymes for industrial application. *Appl Microbiol Biotechnol* **1999**, 51(6), 711-729.
29. Borneman, J.; Triplett, E. W. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* **1997**, 63, 2647-2653.
30. Paul, S.; Zhang, A.; Ludena, Y.; Villena, G. K.; Yu, F.; Sherman, D. H.; Gutierrez-Correa, M. Insights from the genome of a high alkaline cellulase producing *Aspergillus fumigatus* strain obtained from Peruvian Amazon rainforest. *J Biotechnol* **2017**, 251, 53-58.
31. Smith, D.; da Silva, M.; Jackson, J.; Lyal, C. Explanation of the Nagoya Protocol on access and benefit sharing and its implication for microbiology. *Microbiology (Reading)* **2017**, 163, 289-296.
32. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F. O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **2013**, 41, D590-D596.
33. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F.O. The SILVA and “all-species living tree project (LTP)” taxonomic frameworks. *Nucleic Acids Res* **2014**, 42(D1), D643-D648.
34. Guo, L., Wang, G., Sheng, Y., Sun, X., Shi, Z., Xu, Q. and Mu, W. Temperature governs the distribution of hot spring microbial community in three hydrothermal fields, Eastern Tibetan Plateau Geothermal Belt, Western China. *Sci Total Environ* **2020**, 720, 137574.
35. Tekere, M., Lötter, A., Olivier, J., Jonker, N. and Venter, S. Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. *Afr J Biotechnol* **2011**, 10(78), 18005-18012.
36. Chan, C. S.; Chan, K. G.; Ee, R.; Hong, K. W.; Urbietta, M. S.; Donati, E. R.; Shamsir, M. S.; Goh, K. M. Effects of physiochemical factors on prokaryotic biodiversity in Malaysian circumneutral hot springs. *Front Microbiol* **2017**, 8, 1252.
37. He, Z.; Cai, C.; Wang, J.; Xu, X.; Zheng, P.; Jetten, M. S.; Hu, B. A novel denitrifying methanotroph of the NC10 phylum and its microcolony. *Sci Rep* **2016**, 6, 32241.
38. Bray, J. R. a. C., J.T. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **1957**, 27(4), 326-349.
39. Filippidou, S.; Wunderlin, T.; Junier, T.; Jeanneret, N.; Dorador, C.; Molina, V.; Johnson, D. R.; Junier, P. A combination of extreme environmental conditions favor the prevalence of endospore-forming *Firmicutes*. *Front Microbiol* **2016**, 7, 1707.
40. Hemp, J.; Ward, L. M.; Pace, L. A.; Fischer, W. W. Draft genome sequence of *Ardenticatena maritima* 110S, a thermophilic nitrate- and iron-reducing member of the *Chloroflexi* Class *Ardenticatena*. *Genome Announc* **2015**, 3.
41. Hao, L.; McIlroy, S. J.; Kirkegaard, R. H.; Karst, S. M.; Fernando, W. E. Y.; Aslan, H.; Meyer, R. L.; Albertsen, M.; Nielsen, P. H.; Ducholm, M. S. Novel prosthecate bacteria from the candidate phylum *Acetothermia*. *ISME J* **2018**, 12, 2225-2237.
42. Connors, S. B.; Mongodin, E. F.; Johnson, M. R.; Montero, C. I.; Nelson, K. E.; Kelly, R. M. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microbiol Rev* **2006**, 30, 872-905.

43. Ley, R. E., Harris, J.K., Wilcox, J., Spear, J.R., Miller, S.R., Bebout, B.M., Maresca, J.A., Bryant, D.A., Sogin, M.L. and Pace, N.R. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **2006**, 72(5), 3685-3695.
44. Griffiths, E. a. G., R.S. Molecular signatures in protein sequences that are characteristics of the phylum *Aquificae*. *Int J Syst*, **2006**, 56(1), 99-107.
45. Kochetkova, T. V.; Toshchakov, S. V.; Zayulina, K. S.; Elcheninov, A. G.; Zavarzina, D. G.; Lavrushin, V. Y.; Bonch-Osmolovskaya, E. A.; Kublanov, I. V. Hot in cold: microbial life in the hottest springs in permafrost. *Microorganisms* **2020**, 8.
46. Lin, X., Kennedy, D., Fredrickson, J., Bjornstad, B. and Konopka, A. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environ Microbiol* **2012**, 14(2), 414-425.
47. Brewer, T. E.; Aronson, E. L.; Arogyaswamy, K.; Billings, S. A.; Botthoff, J. K.; Campbell, A. N.; Dove, N. C.; Fairbanks, D.; Gallery, R. E.; Hart, S. C.; Kaye, J.; King, G.; Logan, G.; Lohse, K. A.; Maltz, M. R.; Mayorga, E.; O'Neill, C.; Owens, S. M.; Packman, A.; Pett-Ridge, J.; Plante, A. F.; Richter, D. D.; Silver, W. L.; Yang, W. H.; Fierer, N. Ecological and genomic attributes of novel bacterial taxa that thrive in subsurface soil horizons. *mBio* **2019**, 10.
48. Nakagawa, S.; Shtaih, Z.; Banta, A.; Beveridge, T. J.; Sako, Y.; Reysenbach, A. L. *Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subterraneum* and *Sulfurihydrogenibium azureense*. *Int J Syst Evol Microbiol* **2005**, 55, 2263-2268.
49. Takacs-Vesbach, C., Inskeep, W.P., Jay, Z.J., Herrgard, M.J., Rusch, D.B., Tringe, S.G., Kozubal, M.A., Hamamura, N., Macur, R.E., Fouke, B.W. and Reysenbach, A.L. Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three Aquificales lineages. *Front Microbiol* **2013**, 4, 84.
50. Mori, K.; Suzuki, K. *Thiofaba tepidiphila* gen. nov., sp. nov., a novel obligately chemolithoautotrophic, sulfur-oxidizing bacterium of the *Gammaproteobacteria* isolated from a hot spring. *Int J Syst Evol Microbiol* **2008**, 58, 1885-18891.
51. Arai, H.; Kanbe, H.; Ishii, M.; Igarashi, Y. Complete genome sequence of the thermophilic, obligately chemolithoautotrophic hydrogen-oxidizing bacterium *Hydrogenobacter thermophilus* TK-6. *J Bacteriol* **2010**, 192, 2651-2652.
52. Kruse, S.; Goris, T.; Westermann, M.; Adrian, L.; Diekert, G. Hydrogen production by *Sulfurospirillum* species enables syntrophic interactions of *Epsilonproteobacteria*. *Nat Commun* **2018**, 9, 4872.
53. Plugge, C. M.; Zhang, W.; Scholten, J. C.; Stams, A. J. Metabolic flexibility of sulfate-reducing bacteria. *Front Microbiol* **2011**, 2, 81.
54. Podar, P. T.; Yang, Z.; Bjornsdottir, S. H.; Podar, M. Comparative analysis of microbial diversity across temperature gradients in hot springs from Yellowstone and Iceland. *Front Microbiol* **2020**, 11, 1625.
55. Singh, P., Jain, K., Desai, C., Tiwari, O. and Madamwar, D. Microbial community dynamics of extremophiles/extreme environment. *Microbial diversity in the genomic era* (book) **2019**, 323-332.
56. Lewin, A.; Wentzel, A.; Valla, S. Metagenomics of microbial life in extreme temperature environments. *Curr Opin Biotechnol* **2013**, 24, 516-25.
57. Valverde, A.; Tuffin, M.; Cowan, D. A. Biogeography of bacterial communities in hot springs: a focus on the actinobacteria. *Extremophiles* **2012**, 16, 669-79.
58. Inskeep, W. P.; Rusch, D. B.; Jay, Z. J.; Herrgard, M. J.; Kozubal, M. A.; Richardson, T. H.; Macur, R. E.; Hamamura, N.; Jennings, R.; Fouke, B. W.; Reysenbach, A. L.; Roberto, F.; Young, M.; Schwartz, A.; Boyd, E. S.; Badger, J. H.; Mathur, E. J.; Ortmann, A. C.; Bateson, M.; Geesey, G.; Frazier, M. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* **2010**, 5, e9773.
59. Chapelle, F. H.; O'Neill, K.; Bradley, P. M.; Methe, B. A.; Ciuffo, S. A.; Knobel, L. L.; Lovley, D. R. A hydrogen-based subsurface microbial community dominated by methanogens. *Nature* **2002**, 415, 312-315.
60. Chaudhuri, B.; Chowdhury, T.; Chattopadhyay, B. Comparative analysis of microbial diversity in two hot springs of Bakreshwar, West Bengal, India. *Genom Data* **2017**, 12, 122-129.

Chapter 6

Summary, Discussion, and Future Directions

6.1 Structural Characterization and Engineering of P450 TamI

6.1.1 Summary and Conclusions

After several years of elucidating the biosynthetic pathway towards the formation of tirandamycin natural products, including reporting the tirandamycin biosynthetic gene cluster, discovering key tirandamycin intermediates, and characterizing P450 TamI and flavin oxidase TamL, the two oxidative tailoring enzymes responsible for derivatization of these molecules, we have taken this work further. We have solved the structure of P450 TamI in complex with its least oxidized substrate tirandamycin C at a 2.7 Å resolution and identified essential substrate and enzyme structural requirements for a productive substrate binding mechanism. Using site-directed mutagenesis and end-point assays, we learned that a hydrophobic active site pocket mediates substrate binding, enabling a correct geometry for catalysis. Mutating the side chain of amino acids Phe92, Leu399 and Ile400 to alanine led to disruption of product formation and accumulation of starting material. MD simulations revealed that in the WT system, the substrate adopts one main binding pose in the enzyme's active site facilitating efficient catalysis and selective oxidation. However, in the mutant F92A system, the π - π interactions between Phe92 and the conjugated polyene linker of the tirandamycin substrate are interrupted, causing the substrate to adopt many binding poses and losing the ideal geometry for catalysis. These results contradicted an initially hypothesized tetramic acid-based anchoring mechanism, where observed hydrogen bonding interactions between the tetramic acid functionality and polar active site residues Ser397 and Thr398 were thought to mediate substrate binding. Instead, altering these amino acids did not impact substrate turnover or product profile, and only caused a minor perturbation in substrate binding affinity as shown with tirandamycin equilibrium dissociation studies. Additionally, QM calculations showed that the inherent reactivity of tirandamycin substrates matches the selectivity of products obtained *in vitro*, indicating that the substrate drives regioselectivity. However, MD simulations demonstrated that the enzyme orients tirandamycin C substrate to favor abstraction of the pro-*S* hydrogen, dictating stereoselectivity for the first allylic hydroxylation step.

This structural characterization work served as the foundation for rational engineering of P450 TamI to identify the key residues dictating hydroxylation stereoselectivity and assess how the enzyme's active site structure enables an iterative oxidation cascade. An alanine-valine scan of residues neighboring the bicyclic ketal moiety of the substrate led to the discovery of three important leucine residues (Leu101, Leu244 and Leu295) that control the selectivity, order of steps and number of reactions catalyzed. A toolbox of TamI biocatalysts were engineered to improve any novel selectivity observed and facilitate preparative scale reactions for the isolation and characterization of new tirandamycin congeners detected. In a catalyst-controlled fashion, TamI L101A_L295I overrides the innate substrate reactivity, bypassing oxidation at C10 and installing a C11/12 alkene epoxide to generate new tirandamycin L as an intermediate. This biocatalyst then performs one more reaction, a methyl hydroxylation at C18, to generate tirandamycin N, a compound with potent bioactivity against human pathogen *B. anthracis*. In both instances and as shown with DFT calculations and MD simulations, this double mutant catalyzes the more energetically demanding reactions by controlling the orientation of the substrate in the active site to favor the observed reactivity. TamI L295V performs a dual-oxidation cascade with TamI WT selectivity for the C10 hydroxylation and C11/12 epoxidation steps, bypassing C10 ketone formation and forming tirandamycin M. Double variant TamI L244A_L295V displays WT selectivity too but it also shows enhanced iterative capabilities, performing a continuous 8 e⁻ oxidative cascade generating the terminal product of the pathway, tirandamycin B. Michaelis-Menten model kinetics revealed that TamI L244A_L295V overcomes the bottleneck reaction of the oxidative cascade, drastically accelerating the rate at which the enzyme forms the C10 keto group. Lastly, TamI L295A performs an unprecedented oxidative cascade, generating tri-oxidized tirandamycin O and O' that harbor a unique and structurally intriguing oxidative decoration pattern. Further characterization of these TamI biocatalysts was performed to assess their substrate scope, substrate binding affinities, kinetics, and biological activities.

6.1.2 Future Directions

The reactions involved in the engineered tirandamycin biosynthetic systems all involve efficient biocatalysts that direct selectivity and control the order and number of steps catalyzed. Although the self-sufficient chimera fusion constructs obtained for the TamI variants did not perform as effectively as the three-component reactions, these enzymes represent an interesting opportunity to develop potent biocatalysts for stereo- and regio-controlled oxidative cascades of

complex bicyclic scaffolds. Solving the crystal structures of these mutant enzymes could provide valuable insights into how the subtle steric differences alter the environment of the active site to favor alternative substrate binding poses and geometries that override the substrate's inherent reactivity. This knowledge could inform further engineering strategies of P450 homologs or small molecule catalysts that perform multi-step oxidation pathways to manipulate selectivity and iterative capabilities. For example, bacterial iterative P450s MycG and GfsF have active site non-polar amino acids that are equivalent to Leu101 and Leu295 and have not been studied for their potential involvement in controlling selectivity and the native oxidation cascade. I propose that mutating these motifs may result in altered product profiles, providing a frame of reference for future mutagenesis campaigns of bacterial iterative P450s. Additionally, developing whole cell systems for the biosynthesis of the novel tirandamycin congeners could facilitate access to the material and enable further derivatization of these promising bioactive compounds.

6.2 Mechanistic Insights into the Multiple Oxidant Species of P450 TamI

6.2.1 Summary and Conclusions

The structurally intriguing oxidation pattern of tirandamycin O and O' motivated us to elucidate its mechanistic formation using a combination of kinetic, mutagenesis and computational strategies. Using the new tirandamycin congeners as substrates to assess reactivity, we showed that TamI L295A can selectively reverse the order of steps observed in TamI WT without altering stereoselectivity, converting tirandamycin C to L (C11/12 epoxide-containing intermediate) to M (epoxide and C10 hydroxy), to then catalyze an unprecedented epoxide-opening cascade to generate O and O'. This versatility is exclusively observed with biocatalyst TamI L295A, and MD simulations did not provide any initial information to formulate a hypothesis that may explain its catalytic activity. This prompted us to explore a question we had not interrogated before and investigate the catalytically active species used by P450 TamI WT and TamI L295A to assess if it plays a role in dictating the divergent selectivity observed. Using H₂O₂ as a mechanistic probe indicated that TamI L295A may be capable of using Cpd 0 species for the epoxidation reaction, proposition that was further validated by in depth structural analysis and mutagenesis of the highly conserved acid-alcohol pair responsible for the formation of Cpd 0 and Cpd I species. Furthermore, kinetic solvent isotope effect experiments, point mutations and QM calculations revealed that

TamI L295A uses Cpd I as the preferred oxidant for a radical-induced epoxide opening cascade, and uses the peroxy iron species as a viable oxidant for a nucleophilic epoxide opening mechanism in the absence of Cpd I. Our work provided the first mechanistic insight for the active oxidant species involved in the P450 TamI cycle and showed that TamI L295A uses multiple oxidizing species to catalyze the unique formation of tri-oxidized tirandamycin O and O'.

6.2.2 Future Directions

Unlike other TamI variants including TamI L295V, TamI L295A can use H₂O₂ in lieu of molecular oxygen and NADPH for its multi-step oxidative cascade. This result is a starting point for additional engineering campaigns to improve the efficiency of peroxide-mediated P450 TamI iterative catalysis. Further investigation of the structural basis for this reactivity can serve as inspiration for the design and engineering of P450 homologs to develop or improve the ability to use peroxide as a reaction initiator. Furthermore, we detected a similar configuration for the location of the hydroxy relative to the epoxide in the substrates for P450 homologs GfsF and MycG. I propose that investigation of the catalytically active oxidant species in these P450 homologs could enable engineering strategies to favor the formation of one species over the other to induce epoxide opening of the substrate to generate new product congeners that may be difficult to access with synthetic chemistry.

6.3 Chemical Synthesis of a Tirandamycin Analogue to Explore Substrate Scope

6.3.1 Summary and Conclusions

The versatility of P450 TamI biocatalysts for catalyzing mechanistically divergent oxidative pathways prompted us to explore substrate engineering efforts to expand their substrate scope. Inspired on previous knowledge gained with structural and biochemical studies of P450 TamI, we envisioned exploiting the substrate binding mechanism of the enzyme to facilitate oxidation of unnatural bicyclic or polycyclic scaffolds in a selective manner and without extensive protein engineering. Similar to the successful strategy followed with P450 PikC to expand substrate scope, we imagined synthesizing simplified synthetic anchors that enable substrate recognition and binding in TamI, and that can be readily coupled to unnatural cyclic substrates for enzyme-catalyzed oxidation. To this end, we optimized the total, short and scalable synthesis of a

truncated tirandamycin C analogue using asymmetric allylation methods, in a total of five reaction steps. We designed a suite of simplified synthetic anchors that mimic the substrate structural requirements for binding and catalysis, and used computational docking experiments to prioritize which anchors to synthesize first for coupling with the truncated tirandamycin analogue. This led to the synthesis of two tirandamycin analogues, one containing a triene linear moiety and one harboring a diene and a 5-membered hydrocarbon ring. Using engineered P450 TamI 397A_T398A that introduces a more non-polar environment in the active site to better accommodate the synthetic analogues, we observed no substrate turnover for the triene-containing compound and high levels of single oxidation of the ring-containing analogue, with C10 and C18 as potential sites for C-H hydroxylation as shown with MD analyses. Further computational studies indicated that a truncated tirandamycin lacking the tetramic acid moiety adopts multiple binding conformations within the enzyme's active site, perturbing substrate binding and losing the correct geometry for catalysis. The ring-containing anchor was coupled to an unnatural bicyclic moiety that is unrelated to the tirandamycin structure to assess its chemical utility as a probe for substrate engineering. No oxidative product was detected when testing this substrate suggesting that the P450 TamI's active site environment may be highly selective to tirandamycin-like bicyclic scaffolds. This work expanded our knowledge on the substrate structural features needed for enzyme recognition and turnover in the TamI biocatalytic platform. Our results demonstrated that a cyclic ring moiety is needed for catalysis in addition of a conjugated polyene system as mutagenesis studies had previously indicated.

6.3.2 Future Directions

This substrate engineering work revealed that P450 TamI has high levels of substrate selectivity with restricted flexibility to accommodate unnatural bicyclic moieties that are structurally unrelated to tirandamycin. The one synthetic substrate that was successfully oxidized by TamI 397A_T398A led single oxidation products only, with no iterative oxidative activity detected. I propose that losing the observed hydrogen bonding interactions between the tetramic acid moiety and the polar residues Ser397 and Thr398 in TamI may inhibit the enzyme's ability to catalyze an iterative oxidative cascade resulting in single oxidation events exclusively. This hypothesis could be furthered investigated by introducing these alanine variations in the engineered TamI biocatalysts that show enhanced iterative abilities, such as TamI L244A_L295V that enables an 8 e⁻ oxidation cascade. If doing so restricts the iterative characteristics of

L244A_L295V to catalyze a C10 hydroxylation only, as seen in TamI WT, instead of catalyzing four consecutive oxidative steps, this would demonstrate that although the polar interactions with the tetramic acid are not required for substrate binding, they are needed to mediate an iterative cascade. Computational analyses could then shed additional light on this phenomenon to provide new knowledge on the structural basis for efficient iterative oxidation of bacterial P450 enzymes and inspire engineering efforts to effectively expand their substrate scope.

6.4 Characterizing the Microbial Diversity of the Boiling River

6.4.1 Summary and Conclusions

The identification, genome mining and metabolite profiling of rare microbial species thriving in unexplored and highly biodiverse regions enables the discovery of new potent biocatalysts, such as P450 TamI, and their natural product substrates. Mapping previously unidentified microorganisms in fragile ecosystems can also facilitate biodiversity monitoring at a more comprehensive level and provide insightful information for the effective conservation of the areas. Building upon a long-standing collaboration with Universidad Nacional Agraria La Molina (Lima, Peru) and the Boiling River Project (Southern Methodist University, Dallas, Texas), we obtained research and collection permits to explore and investigate one of the most unique ecosystems of the Amazon Rainforest, the Boiling River of the Peruvian Amazon. Partially funded by the National Geographic Society, we performed the first large-scale biodiversity assessment of the Boiling River and its associated mesophilic to hyperthermophilic microclimates. Using 16S rRNA gene sequences and amplicon sequence variants, we found that the local microbial diversity resembles that of other hot spring found around the world including Yellowstone National Park, Iceland, South Africa and Tibetan. However, some microbial patterns were endemic to the Boiling River including elevated abundance of *Actinobacteriota* and low presence of *Cyanobacteria*, with the latter feature suggesting the river water may emerge from the subsurface. *Proteobacteria* was the most numerous phyla followed by *Actinobacteriota*, *Bacteroidota*, *Firmicutes*, *Acidobacteriota* and *Chloroflexi*. Although no specific environmental pattern was identified for pH and electrical conductivity parameters, a negative correlation between bacterial phyla abundance and microclimate temperature (in reference to the closest associated Boiling River water conditions) was identified for most stations, as previously seen with other hot springs'

metagenomic studies. Moreover, contrary to what is typically observed in thermal systems, a very low number of *Archaea* ASVs were detected in the area, with most of the sequences identified in specific stations where bacterial abundance was slightly diminished. Lastly, we identified bacteria involved in sulfur cycling and hydrocarbon degradation to inform on the use of these groups as specialized proxies for environmental surveillance and as biocatalytic remediation tools. This work represents the first characterization of the “hidden” Boiling River bacteria and archaea flourishing in a wide range of temperatures including hyperthermophilic environments ($> 80\text{ }^{\circ}\text{C}$), and demonstrates that including microorganisms in biodiversity assessments expands our understanding of diversity patterns in unique and fragile environmental ecosystems.

6.4.2 Future Directions

Elucidating the Boiling River microbiome sets the foundation for future long-read sequencing and bioinformatic efforts to generate metagenomics data sets to assemble population-specific genomes from samples with unique configurations of bacterial consortia. The recovered genomes will be mined for their biosynthetic gene clusters to look for thermostable biocatalysts that perform highly selective and complex chemical reactions at a late-stage, such as the cytochrome P450 TamI. Accessing the genetic profile of unique Boiling River samples, including those with high abundance of unclassified microbes, can lead to the characterization of previously unidentified bacteria and novel gene clusters that encode for extremozymes catalyzing intriguing and difficult reactions. Future work will include integrating our results into a Google Earth Pro map project to develop a virtual and interactive “microbial” map where the bacterial composition of each Boiling River station is described along with the microclimate physicochemical parameters and aerial images of the location. Additionally, we will create a library of bacterial isolates from Boiling River samples for fermentation growths, transcriptomics assays and metabolite profiling to identify potential new natural product scaffolds and the enzymes responsible for their biosynthesis. We anticipate this work to serve as a framework for future elucidation studies of the microscopic universe of other unique environments in the Amazon Rainforest. Our data can also serve as a fundamental framework for the development and implementation of conservation policies and laws to protect the Boiling River Area and its water system.

APPENDIX A:

Experimental Procedures and Supplemental Information for Chapter 2

A.1 Materials and General Methods

Unless otherwise stated, chemical reagents, solvents, and media components for *Escherichia coli* growth were purchased from Sigma-Aldrich and Thermo Fisher Scientific. Kanamycin sulfate, isopropyl- β -D-thiogalactopyranoside (IPTG), and dithiothreitol (DTT) were obtained from Gold Biotechnology. Nicotinamide adenine dinucleotide phosphate (NADP⁺) was purchased from Chem-Impex Int'l INC. δ -aminolevulinic acid was purchased from Oakwood Chemical. Lysozyme was purchased from dot Scientific Inc. Imidazole was purchased from AK Scientific. Amicon Ultra centrifugal filters used for protein concentration were from EMD Millipore. PD-10 columns were purchased from GE Healthcare. Glucose-6-phosphate (G6P) was from Biosynth and glucose-6-phosphate dehydrogenase (G6PDH) (yeast) was from Alfa Aesar. Deionized water was obtained from a Milli-Q system (EMD Millipore) using QGard 2/Quantum Ex Ultrapure organex cartridges. For media, buffers, and other solutions, pH was monitored using a VWR sympHony SB70P pH meter calibrated according to the manufacturer's specifications. Media solutions were autoclaved or filter-sterilized prior use.

DNA oligonucleotides for mutagenesis were purchased from Integrated DNA Technologies. PCR was performed using a Bio-Rad iCycler thermal cycler system. Invitrogen plasmid miniprep kits were purchased from Thermo Fisher Scientific and all DNA extractions/purifications were accomplished following the manufacturer's protocols. DNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer. DNA sequencing was performed at the University of Michigan DNA Sequencing Core. Optical density (OD₆₀₀) was measured using an Eppendorf BioPhotometer. All centrifugations were accomplished using a Beckman Coulter Avanti J-20 XP centrifuge, and sonication was performed by a Fisherbrand Model 705 Sonic Dismembrator.

Reversed-phase high-performance liquid chromatography (RP-HPLC) purification was performed using Waters XBridge 5 mm C18 columns and a solvent system of acetonitrile and H₂O

supplemented with 0.1% formic acid (FA) unless stated otherwise. All NMR spectra were acquired on a Bruker 800MHz spectrometer at the BioNMR Core of University of Michigan. NMR spectra were processed using MestReNova software. High-resolution ESI-MS spectra was acquired and performed on a quadrupole time-of-flight spectrometer (Agilent Q-TOF 6500 series) using a XBridge™ C18 3.5 μm 50 mm column with a MeCN and H₂O solvent system supplemented with formic acid (0.1% v/v). The same column and solvent system were utilized for analysis of enzymatic reactions using a Shimadzu analytical HPLC.

A.2 Fermentations

A.2.1 *Streptomyces sp. 307-9 growth and extraction of tirandamycin A and B*

The isolation and purification procedures were adapted from Carlson et al.¹ The strain was maintained on ISP2 agar plates and as spore stocks. All culture incubations were at 30°C and with 150rpm shaking for liquid cultures. Seed cultures of 30mL TSB media were inoculated with a loopful of vegetative cells from plate culture and grown for 7 days. Fermentation cultures were baffled flasks of Md2 media (10 g dextrose, 2 g NZ-Amine A, 1 g yeast extract, 0.77 g meat extract, 30 g NaCl per 1 L H₂O) seeded with a 3% inoculation of TSB culture and grown for 6 days. To each flask, 20g of autoclaved XAD-16 resin (Supelco, Bellefonte, PA) were added. Production cultures were grown for one additional day, after which the resin was collected for extractions. Resin was rinsed with water, filtered and batch extracted with 90% CH₂Cl₂/10% MeOH twice. The combined extract was dried by rotary evaporation to yield a dark brown oil. This extract was purified by RP-HPLC on a 45 min gradient of 30-100% MeCN and H₂O supplemented with 0.1% FA and followed by UV/Vis photo-diode array detection. The purity of the compounds was determined by LC-MS analysis and ¹H-NMR.

A.2.2 *Streptomyces sp. 307-9 ΔtamI P450 and ΔtamL flavoprotein growth and extraction of tirandamycin C, E and M*

The culturing and extraction procedures were adapted from Carlson et al.² The mutant strains were maintained on ISP2 agar plates supplemented with 50 $\mu\text{g}/\text{mL}$ apramycin and as spore stocks. Fermentation and purification was carried out as described earlier for the wild-type strain with the following exceptions: all liquid media was supplemented with 50 $\mu\text{g}/\text{mL}$ apramycin;

fermentation cultures for Δ tamL flavoprotein were grown for 4 days prior adding the autoclaved XAD-16 resin and grown for 2 additional days; fermentation cultures for Δ tamI P450 were grown for 7 days prior adding the resin and grown for 2 additional days. Approximately 10-20 mg of pure tirandamycin C were consistently isolated from 1 L of culture. Fractions containing tirandamycin E and M were analyzed via LC-MS using 354 nm UV detection, positive/negative ion MS detection and the following conditions (A = H₂O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-100% B over 15 min, 100% B for 2 min, 10% B for 2 min; flow rate 0.2 mL/min.

A.3 Protein Expression and Purification

A.3.1 TamI protein expression and purification

Expression of TamI and TamI mutants: The pET28b_TamI plasmid previously generated² was used to express N-terminally His-tagged TamI. *E. coli* BL21 (DE3) was transformed with the plasmid. Kanamycin (50 μ g/mL) was added to 1 L of Terrific Broth (TB) media, which was then inoculated with the transformed *E. coli* cells. For crystallization, TamI was subcloned into a modified pET28 vector containing a tobacco etch virus (TEV) protease cleavable N-terminal His8 tag using the “Quikchange” cloning protocol (“tamI_pET28H8T”). The 1 L cultures were grown overnight at 37°C until an OD₆₀₀ of 0.8-1.0 was reached, cooled in an ice-water bath (10-20 min), induced with 0.4 mM IPTG and 0.4 mM δ -aminolevulinic acid, and expressed for 18-20 hours at 18°C. The cells were harvested by centrifugation and stored at -80°C until used for protein purification.

Protein purification for enzymatic assays and large-scale reactions: All subsequent steps were performed at 4°C. The cells were thawed and resuspended in 70 mL of lysis buffer (50 mM Tris-HCl, 50 mM NaCl, 10% (v/v) glycerol, pH 7.4 at room temperature) supplemented with 1 mM PMSF, 0.5 mg/mL lysozyme, 2 mM MgCl₂, 5-10 U/mL Benzonase nuclease) per 1 L of original overexpression culture. The cell suspension was stirred slowly for 1 hour prior to sonication using a Model 705 Sonic Dismembrator (Thermo Fisher Scientific) and centrifugation at 50,000 x g for 30 min to remove cellular debris. The lysate was filtered through a syringe-operated 0.45 μ m filter (Corning) prior to loading onto a column containing Ni-NTA resin. The resin was then washed with 6 CV of wash buffer 1 (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol, pH 7.4 at RT) followed by 6 CV of wash buffer 2 (50 mM Tris-

HCl, 300 mM NaCl, 50 mM imidazole, 10% (v/v) glycerol, pH 7.4 at RT). The protein was then eluted by 4 CV of elution buffer (50mM Tris-HCl, 300 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol, pH 7.4 at RT). Concentrated protein was desalted by loading onto PD-10 columns and eluting with storage buffer (50 mM Tris- HCl, 10% (v/v) glycerol, pH 7.4 at RT). Protein was flash frozen with liquid N₂ and stored at -80 °C until needed. Protein purity was assessed by SDS-Page.

A.3.2 TamL expression and purification

Expression of TamL: The pET28b_TamL plasmid previously generated² was used to express N-terminally His-tagged TamL. The expression procedure was carried out as described earlier for the P450 TamI except that cells were induced with 0.1mM IPTG only.

Protein purification for enzymatic assays and large-scale reactions: The protein purification procedure was conducted as described earlier for the P450 TamI.

A.4 Protein Crystallization

Structure of Substrate-Free TamI: Initial efforts to determine the TamI crystal structure were conducted using a tamI-pSJ2 expression vector containing an N-terminal His₈ tag.¹² Crystals grown with this construct were used to solve the TamI structure by molecular replacement and revealed a single polypeptide chain in the asymmetric unit; however, the P450 active site was occupied by 16 N-terminal residues of a symmetry-related protomer. The bound N-terminal segment was derived from the linker region between the His₈ tag and the EcoRI restriction site and prompted us to prepare a TEV-cleavable construct (TamI-pET28H8T) where the affinity tag and linker could be removed by digestion. Crystals of substrate-free TamI were grown by vapor diffusion at 20 °C with a well solution composed of 3.5 M sodium formate, pH 7.0, and 0.1 M strontium chloride. Crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. Substrate-free TamI crystallized in the space group I422 with unit cell dimensions of a = 154.5, b = 154.5, c = 91.8 and one chain in the asymmetric unit. X-ray data were collected at 100 K on the Advanced Light Source 8.3.1 beamline in Berkeley CA, USA. Diffraction data were integrated and scaled using MOSFLM³³ and SCALA.³⁴ The structure of TamI was solved by molecular replacement with Phaser-MR³⁵ using PikC (PDBid: 2X5W, chain B) as a search model. Three rounds of autobuild using Buccaneer³⁴ were used to provide initial model. Iterative

rounds of manual building in Coot36 and refinement with Refmac37 and Phenix.refine38 were used to furnish the final model.

Structure of Tirandamycin C-Bound TamI: TEVcleaved TamI in the presence of tirandamycin C gave a different crystal form with eight chains in the asymmetric unit and substrate bound in each active site. Crystals of tirandamycin C-bound TamI were grown by vapor diffusion by mixing 2 μ L of 7 mg/mL tirandamycin C-bound TamI with 2 μ L of a well solution composed of 1.6 M ammonium sulfate, 0.9 M sodium chloride, 0.1 M bis-Tris, pH 6.5, and 2.5% ethylene glycol. Narrow, rod-shaped crystals were cryoprotected by the addition of well solution containing 9% ethylene glycol, 0.2 mM tirandamycin C, 2% dimethyl sulfoxide (DMSO) directly to the sitting droplets. Crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. TamI tirandamycin C complex crystallized in the space group C2 with unit cell dimensions of $a = 224.6$, $b = 57.2$, $c = 282.7$, $\alpha = 90^\circ$, $\beta = 90.9^\circ$, $\gamma = 90^\circ$ and eight chains in the asymmetric unit. X-ray data were collected at 100 K on beamline 23ID-B at the General Medical Sciences and Cancer Institute Structural Biology Facility at the Advanced Photon Source in Argonne, IL, USA using helical data collection along the long axis of the rod-shaped crystal. Diffraction data were integrated and scaled using X-ray diffuse scattering.³⁹ The structure was solved by molecular replacement with Phaser-MR35 using substrate-free TamI as a search model. To generate the initial model, Phenix.autobuild38 was performed using an eightfold-NCS averaged map. Iterative rounds of manual building in Coot36 and refinement with Phenix.refine38 were used to furnish the final model.

A.5 CO-Bound Reduced Difference Spectra of TamI enzymes

Following an established protocol,³ the CO-bound reduced difference spectra were obtained using a SpectraMax M5 UV-visible spectrophotometer (Molecular Devices). TamI in storage buffer was reduced through the addition of a few milligrams of sodium dithionite ($\text{Na}_2\text{S}_2\text{O}_4$) and a spectrum was recorded from 400 to 500 nm (5 nm steps). After CO was bubbled into the solution for 35 s, the spectrum of CO-bound reduced P450 species was recorded and the previous reduced spectrum was used to reference. This assay was used to determine the functional P450 concentration with the extinction coefficient of $91,000 \text{ M}^{-1} \text{ cm}^{-1}$.

A.6 Site-Directed Mutagenesis

All single mutants were prepared with TamI wild-type as the template and the target mutation as the primer using the Quikchange Lightning Site-Directed Mutagenesis Kit and protocol. The reactions included 50 ng template, 0.2 μM of primer (forward only), 0.2 mM dNTPs, 2.5 μL of 10x Pfu buffer, 0.5 μL DMSO and 0.5 μL Pfu Turbo from Agilent in a total of 25 μL . The PCR cycle was 1) 95°C for 3 min, 2) 95°C for 30 sec, 3) 68°C for 11 min, 4) 68°C for 5 min, and steps 2-3 were repeated for 30 cycles. When required, reaction conditions were adjusted by lowering or increasing DMSO (0-1 μL). DpnI digestion contained 0.5 μL of 2 U/ μL DpnI and 25 μL PCR reaction solution for 2 h at 37°C. Plasmids were introduced into competent *E. coli* DH5 α strain and plasmid DNA isolation was performed by alkaline lysis (Purelink Quick Plasmid Miniprep Kit from Invitrogen) following manufacturer's protocols. The mutations were confirmed using DNA Sanger sequencing and verified using sequence alignment (LaserGene). Double substituents were prepared sequentially starting with a single mutation in the same manner as described above.

A.7 Enzymatic Assays

A.7.1 TamI enzymatic assays

Standard conditions: The analytical *in vitro* enzymatic reactions of tirandamycin were performed in a final volume of 100 μL of reaction buffer (50 mM NaH_2PO_4 , 0.2 mM DTT, 10% (v/v) glycerol, pH 7.4 at RT) containing 200 μM substrate, 2 μM P450, 40 μM spinach ferredoxin, 6 μM spinach ferredoxin-NADP⁺ reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP⁺. As negative controls, enzymes were omitted. Reactions were run for 2 h at 30°C, quenched by extraction using 2 x 200 μL of CHCl_3 and dried under N_2 . The resulting material was redissolved in 100 μL of methanol, and analyzed by LC-MS using 354 nm UV detection, positive/negative ion MS detection and the following conditions (A = H_2O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-100% B over 15 min, 100% B for 2 min, 10% B for 2 min; flow rate 0.2 mL/min; injection volume 2 μL . All reactions were performed and analyzed in duplicate. Authentic standards were run simultaneously following the same conditions.

A.7.2 TamL enzymatic activity assays

The analytical *in vitro* enzymatic reactions of tirandamycin with TamL were performed in a final volume of 100 μL of reaction buffer containing 200 μM substrate and 2 μM TamL. Reactions were run and extracted as described above for assays with P450 TamI using standard conditions.

A.7.3 Total turnover number (TTN)

For determination of TTN, the analytical enzymatic reactions included: 100 μL of reaction buffer containing 200 μM substrate, 0.5 μM P450, 10 μM ferredoxin, 1.5 μM ferredoxin reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP^+ . Reactions were run for 2 h at 30°C, quenched by addition of 100 μL of methanol and analyzed by a Shimadzu analytical HPLC using 354 nm UV detection. All reactions were performed and analyzed in triplicate. TTN values were calculated using Equation 1 below. TTN was determined by evaluating the areas of HPLC peaks corresponding to starting material. Standard curves of substrates were run simultaneously in triplicate and fitted to a linear curve with R values of 0.98-0.99.

Equation 1

$$\text{TTN} = \text{mol substrate consumed} / \text{mol P450 used}$$

HPLC conditions were as follows (A = H_2O + 0.1% FA, B = MeCN + 0.1% FA):

Reactions with tirandamycin C and A substrates: 10% B for 2 min, 10-85% B over 15 min, 100% B for 4 min, 10% B for 2 min; flow rate 0.4 mL/min; injection volume 50 μL .

Reactions with tirandamycin D substrate: 35% B for 2 min, 35-60% B over 15 min, 100% B for 8 min, 35% B for 2 min; flow rate 0.4 mL/min; injection volume 50 μL .

Reactions with tirandamycin E substrate: 29% B for 2 min, 29-59% B over 15 min, 100% B for 8 min, 29% B for 2 min; flow rate 0.4 mL/min; injection volume 50 μL .

Reactions with tirandamycin L substrate: 33% B for 2 min, 33-60% B over 15 min, 100% B for 8 min, 33% B for 2 min; flow rate 0.4 mL/min; injection volume 50 μL .

A.7.4 Equilibrium substrate binding assays

Assays with P450s: The spectral shifts induced by tirandamycin binding were performed at room temperature under aerobic conditions using a SpectraMax M5 UV-visible spectrophotometer (Molecular Devices). Compound parent stocks were prepared in DMSO and serial dilutions were made to obtain stock solutions of varying concentrations of substrate for binding assays. Purified P450s were diluted in storage buffer (50 mM Tris-HCl, 10% (v/v) glycerol, pH 7.4 at RT) to a final concentration of 1 μ M. 0.7 mL of diluted enzyme was transferred to a 1 cm quartz cuvette (Beckman) and varying concentrations of substrate were independently titrated into the solution in small aliquots (<2 μ L) to achieve the desired final concentrations. DMSO comprised less than 2% of the solution volume at the conclusion of the experiment. Scans were recorded from 365 nm to 429 nm (2 nm steps) until the P450s became saturated. The starting spectrum was subtracted from subsequent experimental spectra and average absorbance differences ($\Delta A = A_{\text{peak}} - A_{\text{trough}}$) were plotted against concentration of substrate. The dissociation constant (K_d) for the P450s was calculated by fitting data points to either a non-linear tight-binding quadratic equation⁴ (Equation 2) for high-affinity ligands ($K_d \leq 5$ μ M) or a rectangular hyperbolic function (Equation 3) for low-affinity ligands ($K_d \geq 5$ μ M) using the GraphPad Prism 8 software. All titrations were performed in triplicates and the K_d values reported represent the average of the three sets of data.

Assays with TamL: Isothermal Titration Calorimetry (ITC) is used to measure the thermodynamics of a reaction by measuring the change in power required to maintain the system at a constant temperature when the compound is injected into CDK2. ITC experiments were performed in Nano-ITC Low volume (TA Instruments) at the Center for Structural Biology at University of Michigan. 400 μ l of purified protein in storage buffer (50 mM Tris-HCl, 10% (v/v) glycerol, pH 7.4 at RT) was added to the cell and 50 μ l of substrate in storage buffer (concentration of protein and compound varies depending upon the binding affinity) was taken in the syringe. The experiments were performed at 25 °C with 25x2 μ l of substrate injected into protein for every 250 seconds with stirring speed of 250 rpm. Buffer A was injected into the protein and it was used as blank. The K_a and ΔH of the reactions were calculated using the Launch NanoAnalyze software (TA Instruments). Concentrations were as follows: 0.015 mM protein was used with 0.6 mM tirandamycin M and 0.3 mM protein was used with 0.8 mM tirandamycin E. The K_a and ΔH of the reactions were calculated using the Launch NanoAnalyze software (TA Instruments).

Equation 2

$$\Delta A = (A_{\max} / 2[E]) (K_d + [E] + [S]) - (K_d + [E] + [S])^2 - 4 [E] [S]^{1/2}$$

Equation 3

$$\Delta A = A_{\max} [S] / (K_d + [S])$$

ΔA = Average absorbance difference (peak-to-trough) at each titration point

A_{\max} = Maximum absorbance difference at substrate saturation

[E] = Total enzyme concentration

[S] = Substrate concentration

A.7.5 Michaelis-Menten model kinetics

Reactions were pre-incubated at 30°C for 5 min and initiated by addition of 2.5 mM NADPH. Reactions were quenched with methanol by removing 50 μ L at 4 different time points and dispensing the aliquot into 50 μ L of methanol. Quenched reactions were centrifuged at 17,000 x g for 10 min (4°C) and the supernatant was analyzed by Shimadzu HPLC using 354 nm UV detection. The conditions used for HPLC separation were identical to those used for determination of TTN. GraphPad Prism software was used to plot the initial velocities (V_0) against the product concentration and to determine the kinetic constants k_{cat} and k_M by fitting the data to the standard Michaelis-Menten equation (Equation 4). All reactions were performed and analyzed in triplicate. Standard curves of products were run simultaneously in triplicate and fitted to a linear curve with R values of 0.98-0.99.

Substrate tirandamycin C: Reactions were set up in a total volume of 200 μ L and varying substrate concentrations ranging from 2 μ M to 400 μ M, with the following components:

For TamI wild-type (0.018 μ M P450, 0.36 μ M ferredoxin, 0.054 μ M ferredoxin reductase)

For TamI L244A_L295I (0.025 μ M P450, 0.5 μ M ferredoxin, 0.075 μ M ferredoxin reductase)

For TamI L101A_L295I (0.035 μ M P450, 0.7 μ M ferredoxin, 0.1 μ M ferredoxin reductase).

Substrate tirandamycin E: Reactions were set up in a total volume of 200 μ L and varying substrate concentrations ranging from 12.5 μ M to 200 μ M, with the following components:

For TamI wild-type (2.5 μM P450, 50 μM ferredoxin, 7.5 μM ferredoxin reductase)

For TamI L244A_L295I (0.15 μM P450, 3 μM ferredoxin, 0.45 μM ferredoxin reductase)

For TamI L101A_L295I (0.85 μM P450, 1.7 μM ferredoxin, 1.26 μM ferredoxin reductase).

Substrate tirandamycin D: Reactions were set up in a total volume of 200 μL and varying substrate concentrations ranging from 5 μM to 120 μM , with the following components:

For TamI wild-type (0.045 μM P450, 0.9 μM ferredoxin, 0.136 μM ferredoxin reductase)

For TamI L244A_L295I (0.05 μM P450, 1 μM ferredoxin, 0.5 μM ferredoxin reductase)

For TamI L101A_L295I (0.05 μM P450, 1 μM ferredoxin, 0.5 μM ferredoxin reductase).

Substrate tirandamycin A: Reactions were set up in a total volume of 200 μL and varying substrate concentrations ranging from 6.25 μM to 200 μM , with the following components:

For TamI wild-type (0.9 μM P450, 18 μM ferredoxin, 2.7 μM ferredoxin reductase)

For TamI L244A_L295I (0.25 μM P450, 5 μM ferredoxin, 0.75 μM ferredoxin reductase)

For TamI L101A_L295I (0.5 μM P450, 10 μM ferredoxin, 1.5 μM ferredoxin reductase).

Substrate tirandamycin L: Reactions were set up in a total volume of 200 μL and varying substrate concentrations ranging from 10 μM to 400 μM , with the following components:

For TamI wild-type (4 μM P450, 80 μM ferredoxin, 12 μM ferredoxin reductase)

For TamI L244A_L295I (0.3 μM P450, 6 μM ferredoxin, 0.9 μM ferredoxin reductase)

For TamI L101A_L295I (0.18 μM P450, 3.6 μM ferredoxin, 0.54 μM ferredoxin reductase).

Equation 4

$$V_0 / [E] = K_{\text{cat}} [P] / K_M + [P]$$

[E] = Total enzyme concentration

[S] = Product concentration

A.8 TamI Large-Scale Enzymatic Reactions and Isolation of Products

All preparative scale biotransformation reactions were run in reaction buffer containing 50 mM NaH₂PO₄, 0.2 mM DTT, 10% (v/v) glycerol, pH 7.4 at RT. The purity of compounds was determined by LC-MS analysis and ¹H-NMR.

Tirandamycin E (2): Purified tirandamycin E was isolated for use in analytical scale enzymatic reactions and as an authentic standard. Crude lysate of TamI-RhFRED fusion protein² was employed as the biocatalyst. The enzyme was expressed as previously described.² The cells were lysed in reaction buffer (1 L culture in 70 mL) followed by sonication as described earlier. After centrifugation, the supernatant was filtered through a syringe-operated 0.45 μm filter (Corning) and a small aliquot of the protein was used for CO binding assays to quantify active P450. Reactions were run in a 10 mL scale using a 250 mL flask with the following component: 500 μM tirandamycin C substrate, 2 μM TamI-RhFRED, 5 mM G6P, 2 U/mL G6PDH and 1 mM NADP⁺. Reactions were run at 30°C at 200 rpm and product formation was monitored via LC-MS. Once reactions went to completion, they were quenched by adding 3X volume of acetone, followed by filtration using celite and a vacuum system. The filtrate was concentrated under reduced pressure and extracted using chloroform. The crude product was dissolved in methanol and purified by RP-HPLC on a 45 min gradient of 35-90% MeCN and H₂O (no additives) using 354 nm UV detection. Luna Phenyl-Hexyl (100 Å pore size LC column 5 μm, 250 x 10 mm; Phenomenex Inc) was used.

Tirandamycin D (3): Purified tirandamycin E was isolated for use in analytical scale enzymatic reactions and as an authentic standard. Purified TamL was used as the biocatalyst following previously described methods.² The crude product was dissolved in methanol and purified by RP-HPLC on a 45 min gradient of 35-90% MeCN and H₂O (no additives) using 354 nm UV detection.

Tirandamycin L (6): Purified TamI L295A mutant was utilized as the biocatalyst to produce tirandamycin L. Reactions were run in a 34 mL scale using a 250 mL flask with the following components: 500 μM tirandamycin C substrate (approximately 7 mg), 2 μM P450, 40 μM ferredoxin, 12 μM ferredoxin reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP⁺. Reactions were run at 30°C at 200 rpm and product formation was monitored via LC-MS. Once reactions went to completion, they were quenched by adding 3X volume of acetone, followed by

filtration using celite and a vacuum system. The filtrate was concentrated under reduced pressure, extracted using ethyl acetate and washed with brine. The extract was dried with NaSO₂, filtered and concentrated under reduced pressure. The crude product was dissolved in methanol and purified by RP-HPLC using a Luna Phenyl-Hexyl (100 Å pore size LC column 5 μm, 250 x 10 mm; Phenomenex Inc) column on a 30 min isocratic gradient of 42% MeCN and H₂O (no additives) using 354 nm UV detection.

Tirandamycin M (7): Purified TamI L295V mutant was utilized as the biocatalyst to produce tirandamycin M. Reactions were run in a 34 mL scale using a 250 mL flask with the following components: 500 μM tirandamycin C substrate (approximately 7 mg), 2 μM P450, 40 μM ferredoxin, 6 μM ferredoxin reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP⁺. Reactions were run at 30°C at 200 rpm and product formation was monitored via LC-MS. Once reactions went to completion, they were quenched by adding 3X volume of acetone, followed by filtration using celite and a vacuum system. The filtrate was concentrated under reduced pressure, extracted using ethyl acetate and washed with brine. The extract was dried with NaSO₂, filtered and concentrated under reduced pressure. The crude product was dissolved in methanol and purified by RP-HPLC using a Luna Phenyl-Hexyl (100 Å pore size LC column 5 μm, 250 x 10 mm; Phenomenex Inc) column on a 35 min isocratic gradient of 42% MeCN and H₂O (no additives) using 354 nm UV detection.

Tirandamycin N (8): Purified TamI L101A_L295I mutant was utilized as the biocatalyst to produce tirandamycin N. Multiple reactions were run in a 5 mL scale using 50 mL falcon tubes with the following components: 300 μM tirandamycin C substrate (approximately 1 mg), 2 μM P450, 10 μM ferredoxin, 3 μM ferredoxin reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP⁺. Reactions were run at 30°C at 200 rpm and product formation was monitored via LC-MS. Once reactions went to completion, they were quenched by adding 3X volume of acetone, followed by filtration using celite and a vacuum system. The filtrate was concentrated under reduced pressure and extracted using chloroform. The extract was concentrated under reduced pressure, dissolved in methanol and the pH was adjusted to 3-4 using 0.1M HCl. The crude product was purified by RP-HPLC using a Luna Phenyl-Hexyl (100 Å pore size LC column 5 μm, 250 x 10 mm; Phenomenex Inc) column and a step isocratic method with 354 nm UV detection and the following conditions (A = H₂O, B = MeCN): 32% B for 30 min, 42% B for 10 min, 100% B for 15 min.

Tirandamycin O and O' (9 and 10): Purified TamI L295A mutant was utilized as the biocatalyst to produce tirandamycin O and O' (isolated as a single peak in the HPLC). Multiple reactions were run in a 5 mL scale using 50 mL falcon tubes containing: 250 μ M tirandamycin C substrate (approximately 0.5 mg), 2 μ M P450, 20 μ M ferredoxin, 1 μ M ferredoxin reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP⁺. Reactions were run at 30°C at 200 rpm and product formation was monitored via LC-MS. Once reactions went to completion, they were quenched by adding 3X volume of acetone, followed by filtration using celite and a vacuum system. The filtrate was concentrated under reduced pressure and extracted using chloroform. The extract was concentrated under reduced pressure, dissolved in methanol and the pH was adjusted to 3-4 using 0.1M HCl. The crude product was purified by RP-HPLC using a Luna Phenyl-Hexyl (100 Å pore size LC column 5 μ m, 250 x 10 mm; Phenomenex Inc) column and a 35 min isocratic gradient of 40 % MeCN and H₂O (no additives) using 354 nm UV detection.

A.9 Computational Methods

Quantum Mechanics: Conformational searches were performed with the Merck molecular force field in Spartan. All QM calculations were performed with Gaussian 09. Geometry optimizations and frequencies calculations were performed at the B3LYP level with LANL2DZ41 for iron and 6-31G(d) for all other atoms. Transition structures contained one negative frequency. Enthalpies and free energies were computed at 1 atm and 298.15 K. A correction to the entropy was applied in accordance with the work of Zhao and Truhlar. Single point energy computations were performed at the B3LYP-D3(BJ) level with LANL2DZ for iron and 6-311+G(d,p) in the gas phase for all calculations.

MD Simulations: The heme iron(IV)-oxo complex involved in the cytochrome-catalyzed oxidative hydroxylation and epoxidation cycle (compound I) was used to model the active form of the cofactor. Simulations were performed using the GPU code (pmemd)⁴ of the Amber 12 package, The Amber-compatible parameters developed by Cheatham et al. were used for compound I and its axial Cys ligand. Parameters for tirandamycin substrates were generated within the antechamber module using the general AMBER force field (gaff), with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model. The charges were calculated according to the Merz–Singh–Kollman scheme using Gaussian 09. Each protein

was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P water molecules using the tleap module, resulting in the addition of around 11,370 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na^+ and Cl^-). All subsequent calculations were carried out using the widely tested Stony Brook modification of the Amber 99 forced field (ff99sb). The substrate and enzyme were optimized for total 1,000,000 steps, with 750,000 steepest descent steps and 250,000 conjugate gradient steps. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0–300 K) under constant volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between hydrogen atoms were kept fixed. Long-range electrostatic effects were modeled using the particle-meshEwald method. An 8 Å cutoff was applied to the LennardJones and electrostatic interactions. Harmonic restraints of 30 kcal/(mol Å²) were applied to the solute, and the Andersen equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 time step at a constant volume. Production trajectories were then run for an additional 1500 ns under the same simulation conditions.

A.10 Computational Energies and Coordinates of all Optimized Structures

Energies:

Transition states of tirandamycin congeners

ZPVE = zero-point vibrational energy; TCE = thermal correction to energy; TCH = thermal correction to enthalpy; TCG = thermal correction to Gibbs free energy

Ground States

Structure	ZPVE	TCE	TCH	ECG	E	H (E + TCH)	G (E + TCG)
TirE-(R)	0.280413	0.294358	0.295303	0.241596	-655.645751	-655.350449	-655.404156
TirC	0.275838	0.288622	0.289566	0.238497	-580.436806	-580.147240	-580.198309
TirE-(S)	0.280176	0.294165	0.295110	0.241380	-655.639727	-655.344618	-655.398347
TirD	0.256786	0.270271	0.271216	0.218379	-654.452424	-654.181208	-654.234045
Fe(IV)=O	0.317644	0.340369	0.341313	0.266302	-1625.106237	-1624.764924	-1624.839936
TirA	0.261121	0.275303	0.276248	0.221861	-729.644010	-729.367762	-729.422149

Transition States of Tirandamycin C Substrate

Structure	ZPVE	TCE	TCH	ECG	E	H (E + TCH)	G (E + TCG)	Imaginary Frequency
10S hydroxylation	0.587707	0.624672	0.625616	0.517956	-2205.519437	-2204.893820	-2205.001481	-1749.878
10R hydroxylation	0.587704	0.624671	0.625615	0.517936	-2205.519437	-2204.893821	-2205.001500	-1750.162
(11R, 12S) epoxidation	0.593274	0.630036	0.630980	0.525412	-2205.519592	-2204.888613	-2204.994181	-509.239
(11S, 12R) epoxidation	0.593312	0.630052	0.630996	0.526109	-2205.519684	-2204.888688	-2204.993575	-521.792
C18 hydroxylation	0.588071	0.623788	0.624733	0.520565	-2205.515326	-2204.890593	-2204.994761	-1808.538

Transition States of Tirandamycin D Substrate

Structure	ZPVE	TCE	TCH	ECG	E	H (E + TCH)	G (E + TCG)	Imaginary Frequency
(11R, 12S) epoxidation	0.573647	0.611046	0.611990	0.505564	-2279.537491	-2278.925501	-2279.031927	-528.499
(11S, 12R) epoxidation	0.573876	0.611411	0.612355	0.505393	-2279.531122	-2278.918767	-2279.025729	-602.732
C18 hydroxylation	0.568817	0.605309	0.606253	0.501390	-2279.532469	-2278.926216	-2279.031079	-1895.486

Transition State of Tirandamycin A Substrate

Structure	ZPVE	TCE	TCH	ECG	E	H (E + TCH)	G (E + TCG)	Imaginary Frequency
C18 hydroxylation	0.572665	0.610861	0.611805	0.501116	-2354.717111	-2354.105306	-2354.215996	-1702.944

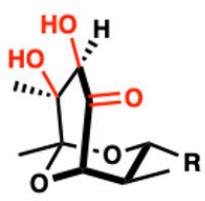
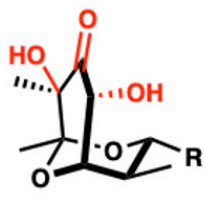
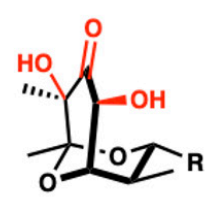
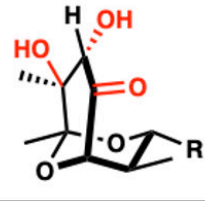
Mechanisms for iterative TamI L244A_L295V mutant

Ground States	E	ZPE	H	G	G-gh
tirandamycin E	-655.90476	-655.62458	-655.60965	-655.66338	-655.6628
Fe(IV)•+ heme model	-1625.5564	-1625.2387	-1625.215	-1625.2901	-1625.2876
3	-654.70639	-654.44961	-654.43518	-654.48801	-654.48765
Fe(III) heme	-1550.3934	-1550.0802	-1550.0572	-1550.1326	-1550.1296
H ₂ O	-76.458939	-76.437799	-76.43402	-76.455465	-76.455465

Route 1/2	E	ZPE	H	G	G-gh
TS-2	-2281.4661	-2280.8736	-2280.8347	-2280.9434	-2280.9357
3a	-2281.5159	-2280.9199	-2280.8808	-2280.9906	-2280.9824
5a	-731.16381	-730.88012	-730.86401	-730.91994	-730.91943
Route 3	E	ZPE	H	G	G-gh
TS-7	-2281.4617	-2280.8695	-2280.8309	-2280.9402	-2280.9313
8a	-2281.4676	-2280.8713	-2280.8317	-2280.9432	-2280.9334
TS-9	-2281.4689	-2280.8769	-2280.8378	-2280.9489	-2280.939

	Imaginary Frequency
TS-2	-1575.201
TS-7	-1211.821
TS-9	-1177.711

Thermodynamic comparison of tirandamycin O tautomers

Ground States	E	ZPE	H	G	G _{qh}
	-1512.3635	-1511.8715	-1511.8394	-1511.9348	-1511.927
	-1512.3628	-1511.871	-1511.8387	-1511.9345	-1511.9265
	-1512.3607	-1511.8686	-1511.8363	-1511.9321	-1511.9242
	-1512.3585	-1511.8674	-1511.835	-1511.9314	-1511.9227

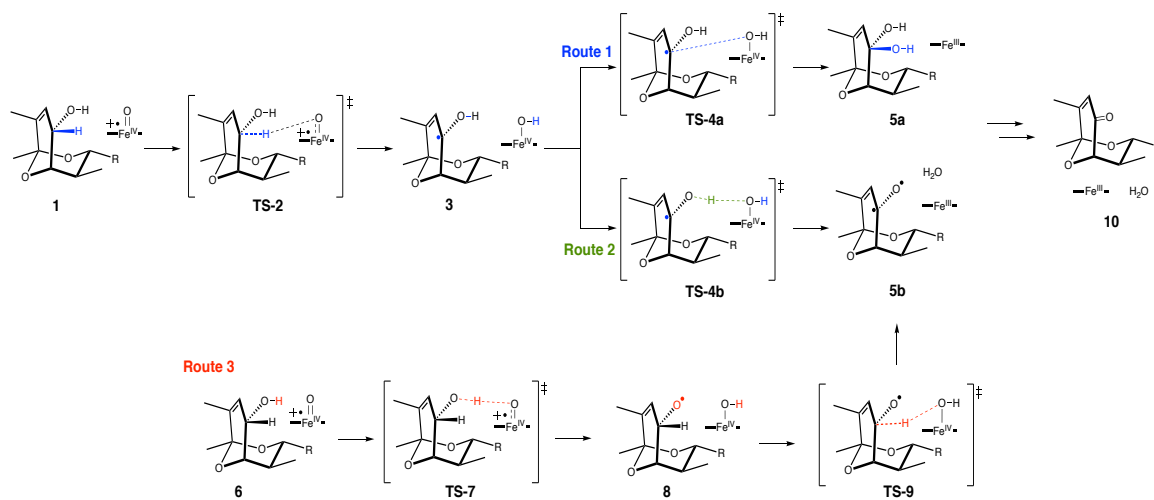
Analysis of TamI L101A_L295I Regioselectivity

Ground States	E	ZPE	H	G	G-qh
tirandamycin L	-655.89197	-655.6113	-655.59693	-655.64943	-655.64912
Fe(IV) ^{•+} heme model	-1625.5564	-1625.2387	-1625.215	-1625.2901	-1625.2876
Transition States	E	ZPE	H	G	G-qh
C18 hydroxylation	-2281.4378	-2280.8455	-2280.8071	-2280.9163	-2280.907
C10 hydroxylation	-2281.4381	-2280.8459	-2280.8073	-2280.916	-2280.9078

	Imaginary Frequency
C18 hydroxylation	-1726.217
C10 hydroxylation	-1745.483

All energy values discussed in the manuscript correspond to the quasi-harmonic corrected Gibbs energies (ΔG -qh). All energies in kcal/mol.

A.11 Supplemental Schemes

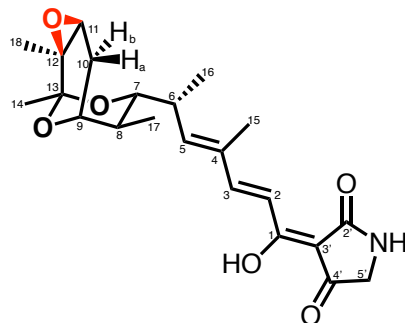


Scheme A.1. Possible mechanisms for iterative TamI L244A_L295V to install the C10 ketone. In Route 1, two consecutive C-H hydroxylation events occur at the C10 site forming a geminal-diol that breaks non-enzymatically into the ketone. Route 2 suggests a double proton abstraction with the H-10 proton being abstracted first followed by the OH-11 proton. In Route 3, a double proton abstraction occurs as well but in an inverted order from route 2 where the OH-11 proton is abstracted first.

A.12 Supplemental Tables

	Substrate-Free TamI	TamI•TirandamycinC
Wavelength (Å)	1.12	1.03
Resolution range	20.9-2.96 (3.07-2.96)	47.9-2.64 (2.73-2.64)
Space group	I 4 2 2	C 1 2 1
Unit cell	154.5×154.5×91.8 Å 90, 90, 90°	224.6×57.2×282.7 90, 90.9, 90°
Total reflections	126589 (18383)	1417501 (123856)
Unique reflections	11820 (1144)	106766 (10575)
Multiplicity	10.7 (10.9)	13.3 (11.7)
Completeness (%)	99.4 (100.00)	99.8 (99.8)
Mean I/sigma(I)	15 (2.9)	10.60 (1.76)
Wilson B-factor	69.4	45.4
R-merge	0.126 (0.954)	0.246 (1.525)
Reflections used in refinement	11818 (1144)	106577 (10565)
R-work	0.255 (0.350)	0.251 (0.346)
R-free	0.314 (0.392)	0.298 (0.393)
Number of non-hydrogen atoms	3176	25091
macromolecules	3114	24302
ligands	43	568
solvent	19	221
Protein residues	410	3185
RMS(bonds)	0.003	0.008
RMS(angles)	0.78	1.11
Ramachandran favored (%)	96.28	97.56
Ramachandran allowed (%)	3.72	2.32
Ramachandran outliers (%)	0.00	0.13
Average B-factor	66.64	51.65
macromolecules	67.01	51.97
ligands	47.68	43.13
solvent	49.12	38.32

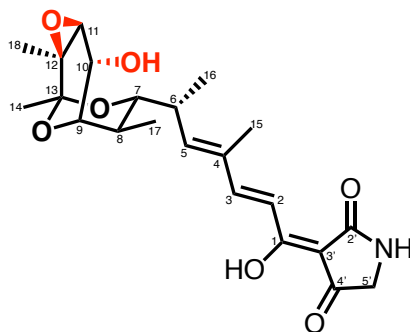
Table A.1. X-ray data collection and refinement statistics.



No.	$\Delta^{13}\text{C}$	$\Delta^1\text{H}$ (J [Hz])	COSY	HMBC	Selective NOE
1	174.9				
2	116.2	7.13, d (15.7)	3	1, 4	
3	149.7	7.60, d (15.7)	2	1, 4, 5, 15	
4	134.3				
5	145.4	6.29, d (10.2)	6	3, 15	
6	34.7	2.85, ddd (9.5, 6.9, 2.3)	5, 16	16	
7	75.8	3.70, m	8	5	6 ^a , 8 ^a , 11 ^a , 16 ^a , 17 ^a , 18 ^a
8	35.1	1.87, m	7, 17	7, 9, 10, 17	9, 17
9	69.3	3.72, m	10a	5, 7, 8, 11, 13	6, 8, 10a, 11, 16, 17, 18
10a	21.7	2.15, m	9, 10b	8, 9, 11, 12	9 ^a , 10b, 11
10b	21.7	1.89, d (5.4)	10a	8, 11	9 ^a , 10a, 11, 17
11	58.5	3.17, d (5.4)	10b	9, 10, 18	7 ^a , 9 ^a , 10a, 10b, 18
12	54.6				
13	95.9				
14	22.5	1.35, s		12, 13	
15	11.9	1.92, d (1.3)		3, 4, 5	
16	16.9	1.13, d (7.0)	6	5, 6, 7	6, 7 ^a , 15
17	12.6	0.70, d (7.0)	8	7, 8, 9	6, 9 ^a , 10b, 15
18	16.6	1.30, bs		11, 12, 13	7 ^a , 11
1' (N)		5.71, bs			
2'	176.3				
3'	99.9				
4'	192.3				
5'	51.6	3.78, s		2', 4'	

Table A.2. NMR characterization for tirandamycin L (6) in CD₂Cl₂.

^aAssigned based upon knowledge of spectra with possible interchanged signals for C9 (3.72, m) and C7 (3.70, m).



No.	$\delta^{13}\text{C}$	$\delta^1\text{H}$ (J [Hz])	COSY	HMBC	Selective NOE
1	175.0				
2	116.0	7.13, d (15.7)	3	1	
3	150.0	7.59, d (15.8)	2	15	
4	134.5				
5	145.3	6.30, d (10.2)	6	3, 15	
6	34.7	2.91, ddd (9.7, 6.9, 2.3)	16		
7 ^a	76.0	3.92, m			8 ^b , 11 ^b , 16 ^b , 17 ^b , 18 ^b
8	36.5	2.0, m	17, 7		5, 9 ^b , 17
9 ^a	71.5	3.93, d (5.7)	8, 10	7, 8, 10, 11, 13	8 ^b , 10 ^b , 17 ^b , 18 ^b
10a	67.3	4.32, d (7.2)	9	8, 9, 11, 12	9 ^b , 11
10b					
11	63.5	3.14, s		9, 10	7 ^b , 9 ^b , 10a, 14, 18
12	55.9				
13	96.2				
14	29.6	1.26			
15	12.0	1.92, bs		3, 4, 5	
16	16.7	1.14, d (7.0)	6	6, 7	5, 6, 7 ^b , 15, 18
17	12.7	0.94, d (7.4)	8	7, 8, 9	
18	16.3	1.36, bs		11, 12, 13	7 ^b , 9 ^b , 11, 16
1' (N)		5.70, bs			
2'	176.4				
3'	99.9				
4'	192.4				
5'	51.5	3.78, s		2', 4'	

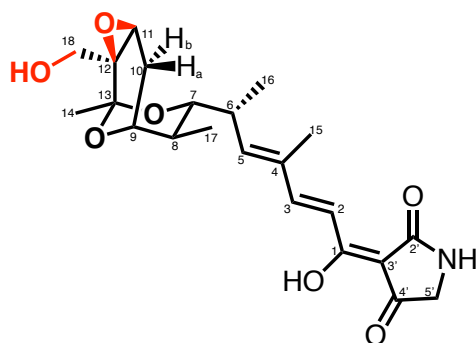
Table A.3. NMR characterization for tirandamycin M (7) in CD₂Cl₂.

^a Assigned based on similarity to the carbon shifts of precursors, tirandamycin E and C.

^b Assigned based upon knowledge of spectra with possible interchanged signals for C9 (3.93, d) and C7 (3.92, m).

	Tirandamycin M (this work)	Previously published spectra, TAM E ¹⁰
No.	δ ¹ H (J Hz)	δ ¹ H (J Hz)
1		
2	7.15, d (15.7)	7.15, d (16.0)
3	7.59, d (15.7)	7.60, d (15.5)
4		
5	6.29, d (10.0)	6.29, d (9.5)
6	2.88, m	2.89, m
7	3.92, d (10.7)	3.92, d (10.5)
8	2.02, m	2.04, m
9	3.98, m	3.98, m
10a	4.41, d (7.4)	4.41, d (7.0)
10b		
11	3.19, s	3.19, s
12		
13		
14	1.42, s	1.42, s
15	1.93, s	1.93, s
16	1.14, d (7.1)	1.14, d (7.0)
17	0.95, d (7.5)	0.96, d (7.5)
18	1.40, s	1.40, s
1' (N)	5.67, bs	
2'		
3'		
4'		
5'	3.82, s	3.82, s

Table A.4. Comparison of ¹H-NMR spectra between tirandamycin M (*in vitro*) and the previously published spectra (*in vivo*, named TAM E).¹⁰ In this paper, this molecule has been renamed as tirandamycin M for clarity. Both spectra were taken in CDCl₃.



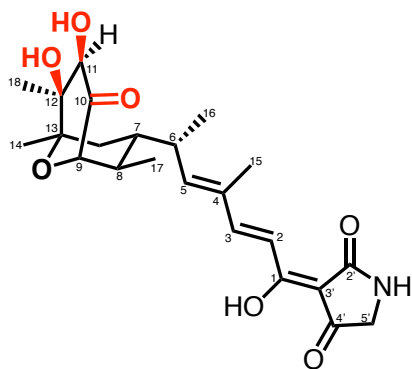
No.	$\delta^{13}\text{C}^c$	$\delta^1\text{H}$ (J [Hz])	COSY	HMBC	Selective NOE
1	175.0				
2	116.2	7.13, d (15.7)	3	1, 4	
3	150.0	7.59, d (15.7)	2	1, 5, 15	
4	134.4				
5	145.2	6.27, d (10.1)	6	3, 15	
6	34.6	2.85, ddd (9.6, 6.8, 2.2)	5, 16		
7	76.2	3.75, m	8, 10a		6 ^b , 8 ^b , 15 ^b , 16 ^b , 17 ^b , 18a ^b
8	35.0	1.90, m	7, 17		
9	69.1	3.77, d (2.5)		7, 11, 13	8 ^b , 10a ^b , 17 ^b
10a	21.3	2.20, dd (16.4, 8.2)	8, 9	8, 9, 11, 12	9, 10b, 11
10b	21.3	1.94, dd (16.4, 5.6)			2, 6, 7, 10a, 11, 17
11	54.0	3.57, d (5.6)	10b	9, 10	7, 9, 10a, 10b, 18a
12	56.9				
13	95.0				
14	23.2	1.37, bs		12, 13	18a, 18b
15	12.1	1.92, bs		3, 4, 5	
16	17.0	1.13, d (7.0)	6	5, 6, 7	5, 6, 7 ^b , 15
17	12.5	0.70, d (7.0)	8	6, 7, 9	6, 7 ^b , 9 ^b , 10b, 15
18a	58.9	3.88, m		11, 12	14
18b	58.9	3.79, s			
1' (N)		5.72, bs			
2'	176.4				
3'	N/A ^a				
4'	192.4				
5'	51.5	3.78, s		2', 4'	

Table A.5. NMR characterization for tirandamycin N (8) in CD_2Cl_2 .

^a Not observed in 2D experiments but observed for tirandamycin C (starting material)

^b Assigned based upon knowledge of spectra with possible interchanged signals for C9 (3.77, d) and C7 (3.75, m).

^c Assigned based on HMBC and HSQC spectra.

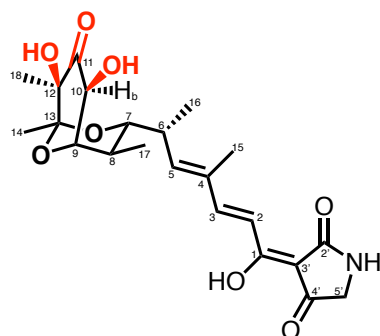


No.	$\delta^{13}\text{C}^{\text{a}}$	$\delta^1\text{H}$ (J [Hz])	COSY	HMBC	Selective NOE
1	175.0				
2	116.0	7.14, d (15.7)	3	1, 4	
3	149.4	7.58, d (15.8)	2	1, 5, 15	
4	134.7				
5	144.3	6.19, d (10.2)	6	3, 15	
6	35.0	2.76, ddd (9.6,6.8,2.3)	16, 5	16	
7	76.0	3.21, d (11.5)	8	5	6, 8, 11, 16, 17, 18
8	32.8	2.02, m		17	5, 7, 9, 17
9	78.0	4.15, d (5.5)	8	7, 8, 10, 11, 13	8, 17
10	204.0				
11a					
11b	72.1	4.11, s		10	7, 17
12	78.3				
13	103.0				
14	21.5	1.50, s		12, 13	18
15	12.0	1.89, bs	5	3, 4, 5	
16	16.3	1.05, d (7.0)	6	5, 6, 7	
17	11.8	0.79, d (7.1)	8	7, 8	6, 7, 8, 9, 11, 15
18	16.0	1.31, s		10, 12, 13	7, 14
1' (N)		5.70, bs			
2'	176.3				
3'	N/A ^b				
4'	192.4				
5'	51.7	3.78, s		2', 4'	

Table A.6. NMR characterization for tirandamycin O (9) in CD_2Cl_2 .

^a Assigned based on HMBC and HSQC spectra

^b Not observed in 2D experiments but observed for tirandamycin C (starting material)



No.	$\delta^{13}\text{C}^{\text{a}}$	$\delta^1\text{H}$ (J [Hz])	COSY	HMBC	Selective NOE
1	175.0				
2	116.6	7.14, d (15.7)	3	1, 4	
3	150.1	7.57, d (15.8)	2	1, 5, 15	
4	135.0				
5	144.3	6.15, d (10.1)	6	3, 15	
6	35.5	2.73, td (7.1, 3.5)	16		
7	76.3	3.44, d (11.4, 2.1)	8		6, 16, 17, 18
8	32.5	2.07, m	7, 17	7	5, 6, 7, 9, 17
9	78.0	4.20, dd (6.1, 2.5)	8	7, 11, 13	8, 10, 17
10a					
10b	74.2	4.28, d (2.7)		11	9, 17
11	205.0				
12	79.9				
13	105.0				
14	21.0	1.47, s		12, 13	18
15	12.2	1.88, s		4, 5	
16	16.7	1.05, d (7.01)	6	5, 6, 7	
17	11.2	0.83, d (7.1)	8	7	6, 7, 8, 9, 10
18	14.8	1.33, s		11, 12, 13	7, 14
1' (N)		5.70, bs			
2'	176.0				
3'	N/A ^b				
4'	192.3				
5'	51.5	3.78, s		2', 4'	

Table A.7. NMR characterization for tirandamycin O' (10) in CD_2Cl_2 .

^a Assigned based on HMBC and HSQC spectra

^b Not observed in 2D experiments but observed for tirandamycin C (starting material)

	Biocatalyst	Tirandamycin C	Tirandamycin E	Tirandamycin D	Tirandamycin A	Tirandamycin L
	WT	368.4 ± 2.5	55.7 ± 0.2	340.1 ± 4.0	143.0 ± 1.1	109.9 ± 2.8
Selective	L295A	116.0 ± 15.1	81.8 ± 6.0	338.8 ± 4.6	89.3 ± 3.9	251.7 ± 15.4
	L295V	368.7 ± 1.4	170.0 ± 3.6	335.6 ± 1.1	47.3 ± 8.8	214.1 ± 16.9
	L101A_L295I	369.9 ± 0.6	324.4 ± 2.0	334.8 ± 3.9	69.1 ± 3.5	160.8 ± 1.3
	L244A_L295I	354.5 ± 1.8	88.0 ± 8.2	313.2 ± 4.4	164.3 ± 6.4	128.6 ± 7.9
Iterative	L244A_L295V	270.8 ± 10.6	89.6 ± 15.6	329.0 ± 4.3	166.7 ± 1.3	187.2 ± 7.3

Table A.8. Total turnover numbers (TTN) for enzymatic reactions involving TamI WT and variants and different tirandamycin substrates. TTN = mol substrate consumed/mol P450. Reported errors are standard deviations calculated from experiments performed in triplicate.

Substrate	P450	K_M (μ M)	K_{cat} (min^{-1})	K_{cat}/K_M ($\mu\text{M}^{-1}\text{min}^{-1}$)	Products observed and tracked	Step(s) catalyzed	Lane
1	WT	16.0 ± 2.6	29.8 ± 1.9	1.87	2	1	1
	L244A_L295V	8.9 ± 1.5	7.1 ± 0.4	0.81	2	1	2
	L101A_L295I	3.8 ± 0.6	6.6 ± 0.3	1.71	6 + 7	1, 3	3
2	WT	108.5 ± 25.9	0.04 ± 0.01	0.0003	4 + traces of 7	1, 2, 3	4
	L244A_L295V	115.6 ± 22.4	1.27 ± 0.10	0.0110	3 + 4 + 5 + traces of 7	2, 3, 4	5
	L101A_L295I	88.6 ± 19.0	1.81 ± 0.19	0.0204	3 + 4 + 7	2, 3	6
3	WT	36.5 ± 7.1	6.8 ± 0.5	0.19	4	3	7
	L244A_L295V	99.0 ± 23.8	6.5 ± 0.9	0.0652	4 + 5	3, 4	8
	L101A_L295I	120.1 ± 33.5	56.4 ± 9.6	0.4693	4	3	9
4	WT	8.2 ± 1.0	0.036 ± 0.001	0.0044	5	4	10
	L244A_L295V	7.0 ± 0.9	0.716 ± 0.023	0.1022	5	4	11
	L101A_L295I	7.4 ± 1.2	0.276 ± 0.011	0.0375	5	4	12
6	WT	34.9 ± 5.5	0.005 ± 0.001	0.0001	8	4	13
	L244A_L295V	55.6 ± 9.9	1.092 ± 0.066	0.0196	8 + 9 + 10	1, 4, new steps	14
	L101A_L295I	51.6 ± 8.6	0.227 ± 0.016	0.0044	8	4	15

Table A.9. Kinetic parameters for TamI WT, iterative L244A_L295V and selective L101A_L295I. K_{cat}/K_M values of TamI variants that are higher than the values calculated for the parent enzyme are highlighted in red.

	Biocatalyst	Tirandamycin C	Tirandamycin E	Tirandamycin D	Tirandamycin A	Tirandamycin L
	WT	0.17 ± 0.02	NS	6.98 ± 0.45	NS	NS
Selective	L295A	NS	NS	NS	NS	NS
	L295V	0.10 ± 0.02	NS	7.04 ± 0.35	12.71 ± 0.56	NS
	L101A_L295I	0.09 ± 0.02	4.75 ± 0.50	4.09 ± 0.54	2.28 ± 0.25	11.98 ± 1.27
	L244A_L295I	4.33 ± 0.17	NS	8.30 ± 0.48	12.72 ± 0.57	NS
Iterative	L244A_L295V	3.18 ± 0.06	NS	8.32 ± 0.27	7.09 ± 0.56	NS
	I247A	1.13 ± 0.06				
	L101Y	NS				
No conversion	H102V	10.27 ± 0.38				

Table A.10. Substrate binding analysis of Tami WT and variants with different tirandamycin substrates. Equilibrium dissociation constant values (K_d) represented in μM . Reported errors are standard deviations calculated from experiments performed in triplicate. NS refers to no apparent shift observed in differential spectrum upon substrate titration to P450, even at substrate saturation levels.

Bacterial strain	Media	Ciprofloxacin	Erythromycin	Streptolydigin	5	4	1	3	2	9	8	7	6
		MIC (μM)											
Gram Positive													
Vancomycin-resistant <i>Enterococcus</i>	2	>100	3.125	>100	12.5	6.25	25	>100	>100	>100	>100	>100	>100
<i>B. anthracis</i> 34f2	1	0.781	3.125	>100	6.25	12.5	3.125	6.25	25	>100	6.25	50	25
<i>B. cereus</i>	1	1.56	3.125	>100	3.125	6.25	3.125	25	>100	>100	50	>100	>100
<i>S. pneumonia</i> ATCC 49619	3	1.56	0.1953	>100	3.125	25	12.5	12.5	>100	>100	25	>100	>100
<i>B. subtilis</i> DHS 5333	1	0.048	0.781	>100	25	25	12.5						
<i>B. subtilis</i> 168	1	0.781	0.781	>100	50	25	25						
<i>S. aureus</i> ATCC 6538P	1	0.781	1.56	>100	>100	>100	50						
<i>S. aureus</i> NorA	1	0.781	1.56	>100	50	>100	50						
Methicillin-resistant <i>S. aureus</i> ATCC 43300	1	3.125	>100	>100	>100	>100	>100						
<i>E. faecalis</i> ATCC 29212	2	3.125	1.56	>100	25	25	>100						
<i>L. monocytogenes</i> ATCC 19115	2	6.25	1.56	>100	50	>100	>100						
<i>K. rhizophilia</i> ATCC 9341	2	3.125	0.391	>100	50	>100	>100						
Gram Negative													
<i>E. coli</i> ToIC	1	0.048	1.56	>100	>100	>100	25	>100	>100	>100	>100	>100	>100
<i>A. baumannii</i> B. ATCC 17978	1	0.781	>100	>100	>100	>100	>100	>100	>100	>100	>100	>100	>100
<i>E. coli</i> MC1061	1	0.048	6.25	>100	12.5	>100	12.5						
<i>S. enterica</i> ATCC 14028	1	0.0976	>100	>100	>100	>100	>100						
<i>S. flexneri</i> BS103	1	0.1953	>100	>100	>100	>100	>100						
<i>K. pneumoniae</i> ATCC 29665	1	0.781	>100	>100	>100	>100	>100						

Table A.11. Antimicrobial activity of tirandamycin congeners against select strains. Media composition, solid, seed culture, MIC culture: 1 – LB agar, Mueller-Hintin broth, Mueller-Hintin broth. 2 – TSA, BHI broth, BHI broth. 3 – Blood agar, 5% blood broth, 0.5% blood broth. Tirandamycin 9 contained a minor presence of 10 in the sample in a 5:1 ratio.

A.13 Supplemental Figures

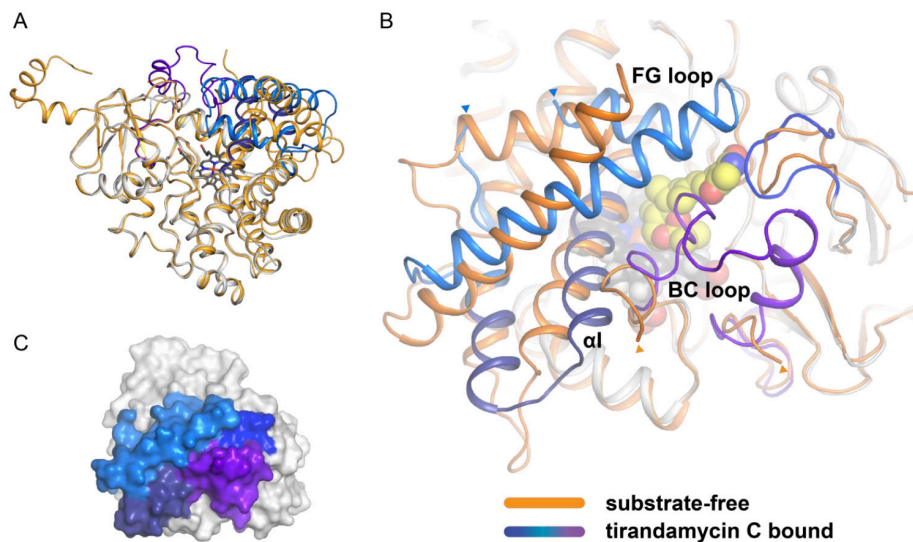


Figure A.1. Comparing substrate-bound and substrate-free TamI. (A) Cartoon representation of the superposition of substrate-free TamI (orange) and substrate-bound TamI (grey and blue). The r.m.s.d between the structures is 1.64 Å over 340 atoms. (B) Close-up view of active site of superposed structures. The FG loop, BC loop, and I helix undergo conformational rearrangements upon substrate binding. (C) Surface representation of the substrate-bound structure. Protein regions that undergo conformational change are colored.

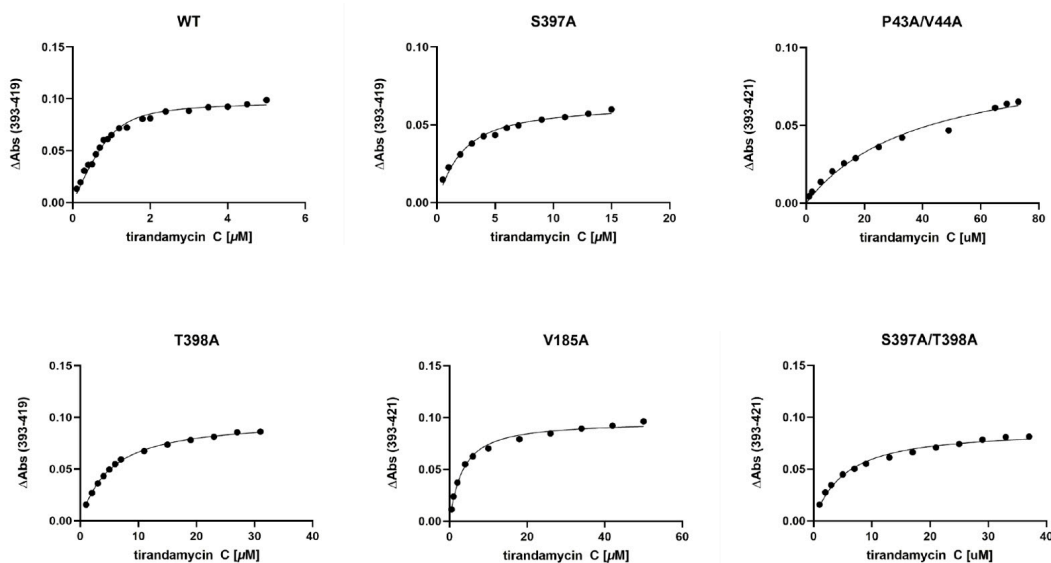


Figure A.2. Equilibrium dissociation constant binding curves for TamI WT and mutants with tirandamycin C (1) substrate.

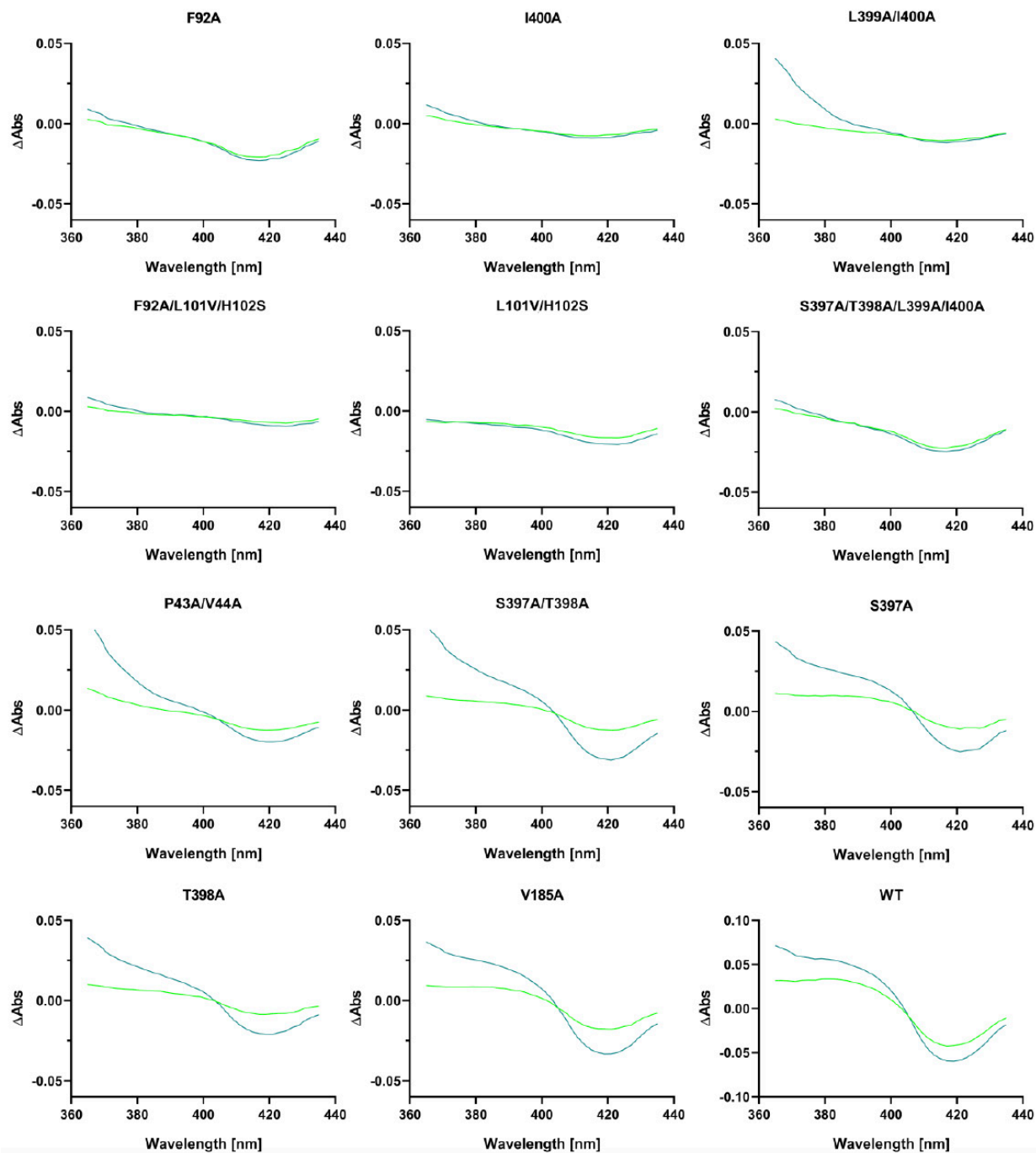


Figure A.3. Qualitative spectral shift assay. For qualitative assessment of spectral shift 1 μM of P450 TamI was incubated with 1 μM tirandamycin C (green trace) and 5 μM tirandamycin C (blue trace). TamI mutants F92A, I400A, L399A_I400A, F92A_L101V_H102S, L101V_H102S, and S397A_T398A_L399A_I400A showed no shift and were not utilized for quantitative measurements.

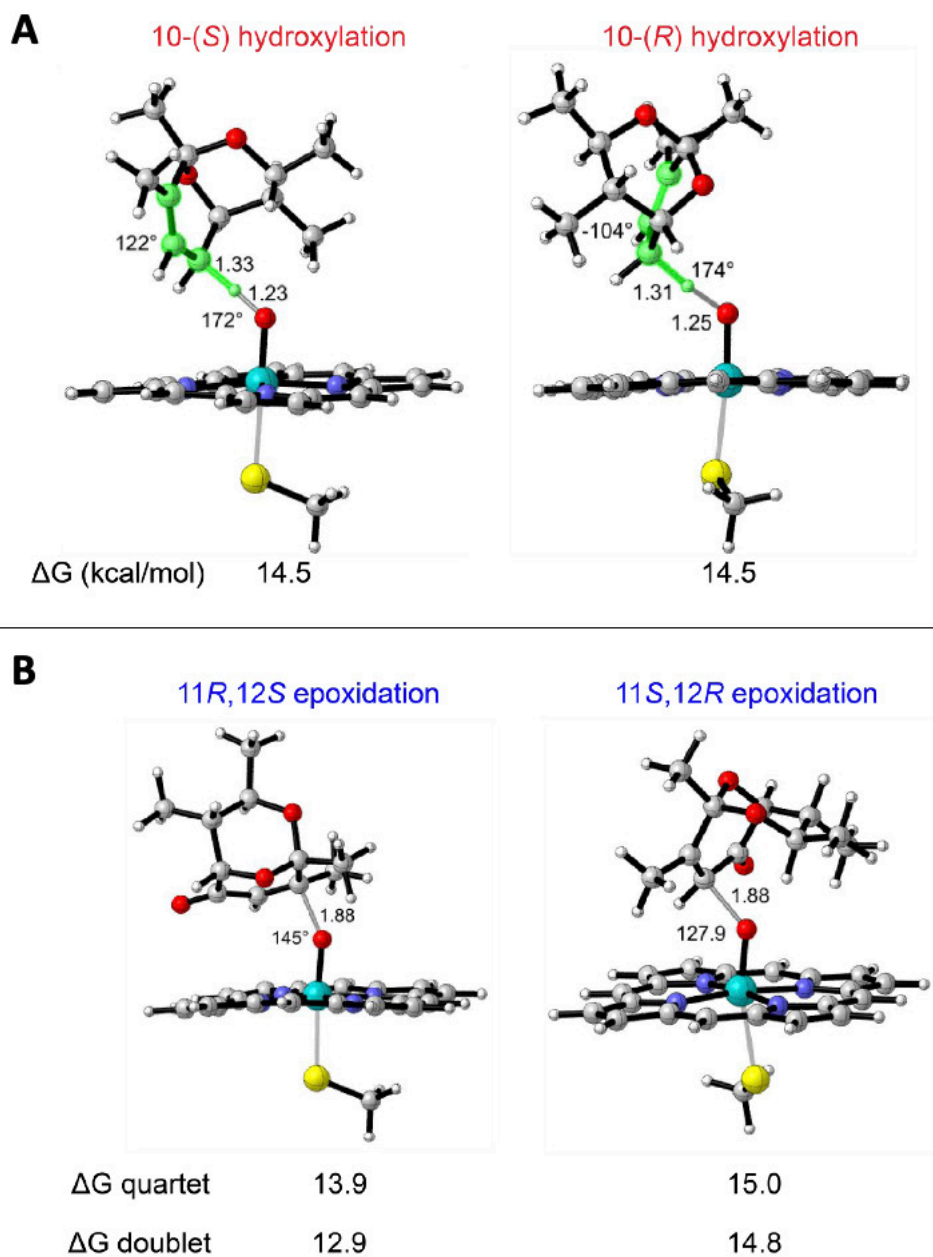


Figure A.4. QM calculations for stereoselectivity in TamI chemistry. (A) No inherent preference in *R* versus *S* selectivity for C10 hydroxylation with tirandamycin C. (B) Innate preference for the 11*R*/12*S* epoxidation with tirandamycin D. B3LYP/6-311+G(d,p)/LanL2DZ^{Fe} // B3LYP/6-31G(d)/LanL2DZ^{Fe}

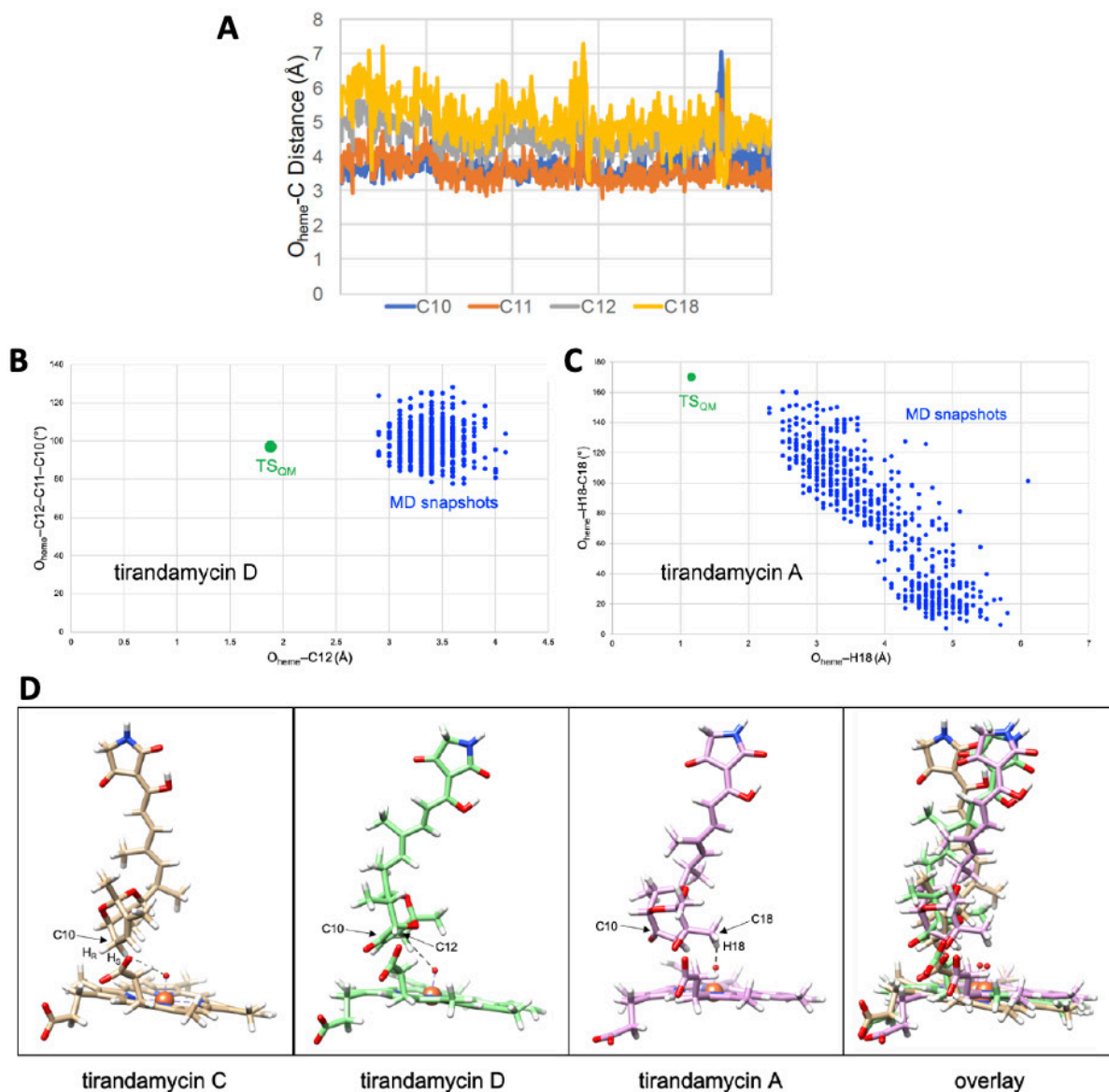


Figure A.5. MD simulations of iterative oxidation cascade with TamI. (A) Bicyclic ketal distance from Fe(IV)-oxo radical cation. (B) The geometric orientation of tirandamycin D is similar to that of the QM transition state. (C) The geometric orientation of tirandamycin A shows a range of snapshots with similar orientation to the QM transition state geometry. (D) The MD simulations show changes in substrate orientation as the oxidation level increases. All simulations were performed for 500 ns using Amber12/ff99SB-ILDN/GAFF.

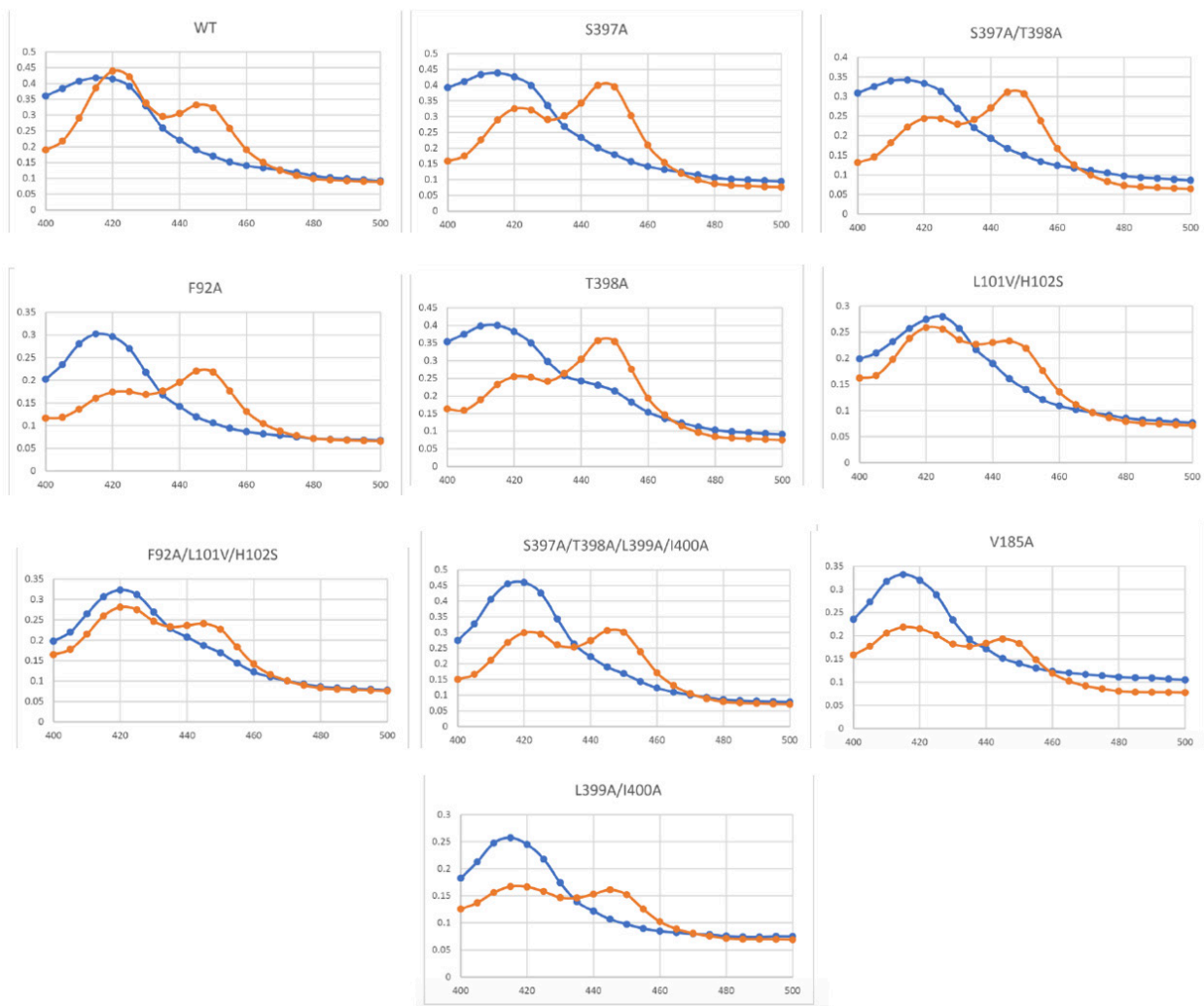


Figure A.6. CO difference spectroscopy for TamI mutants designed to explore the structural basis of TamI. Spectra of sodium dithionite reduced protein (blue) and CO-bound reduced enzyme (orange).

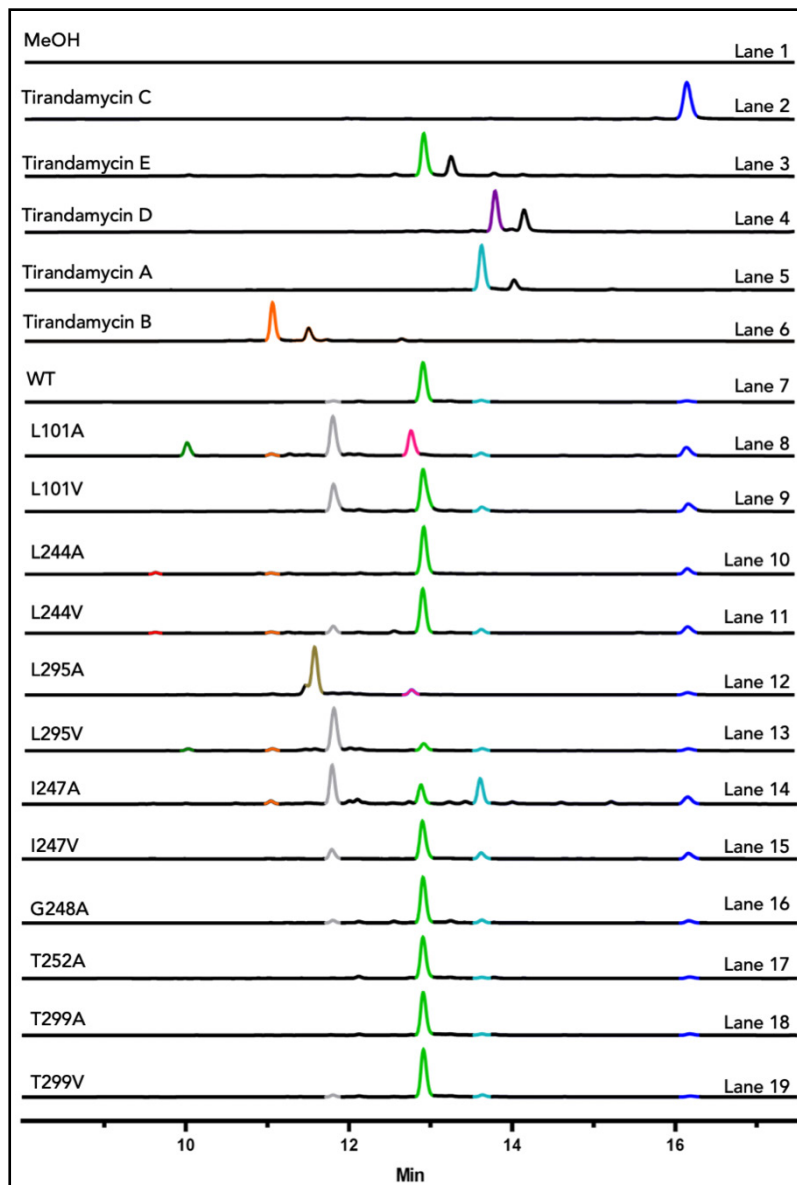


Figure A.7. Analytical enzymatic reactions with TamI WT, TamI variants and tirandamycin C substrate. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A and B (lanes 2-6). Compared to WT activity (lane 7), mutants TamI L101A (lane 8) and L295A (lane 12) showed a distinct product profile generating new tirandamycin congeners (RT 12.7, 11.8, 11.6 and 10min). Although in nominal yields, TamI L244A (lane 10) and L244V (lane 11) also generated a new tirandamycin congener (RT 9.6min). L244V overexpressed poorly in *E. coli* compared to L244A, hence L244A was selected for further studies. TamI I247A (lane 15) led to significant formation of tirandamycin A compared to WT, suggesting an enhanced ability for iterative oxidation catalysis. Moreover, conserved residues G248 and T252, which are involved in the proton delivery mechanism and dioxygen activation of P450 catalysis, led to WT activity when mutated to alanine. This suggests that TamI can tolerate significant deviations from the WT active structure without loss of oxidative activity. It is important to note that the neighboring residues G249 and T253 may compensate for the G249A and T252A mutations.

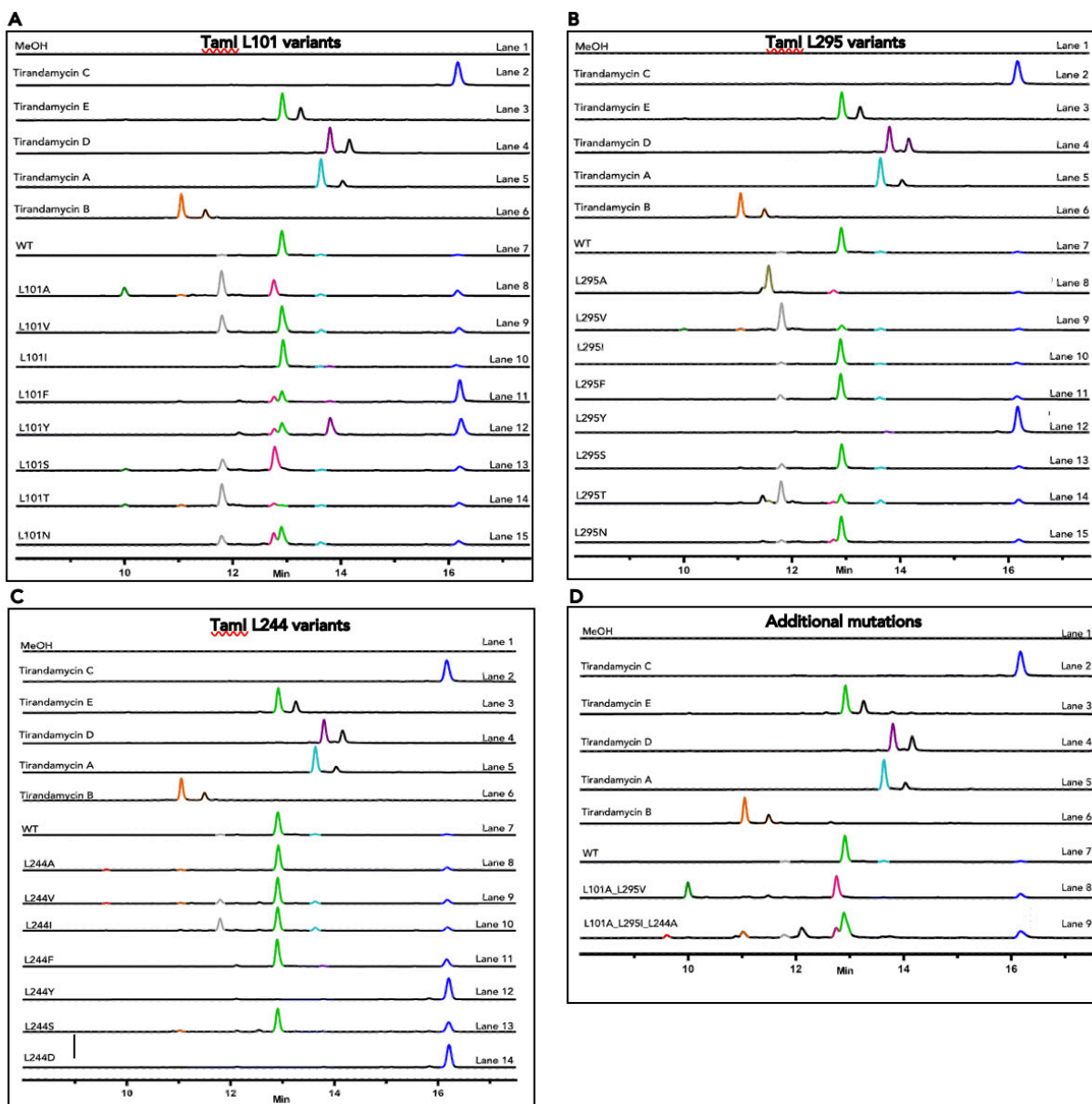


Figure A.8. Analytical enzymatic reactions of TamI mutants and tirandamycin C substrate. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A and B (lanes 2-6). (A) The majority of L101 mutants led to non-selective catalysis. TamI L101V (lane 9) and L101I (lane 10) showed a similar product profile to WT (lane 7). TamI L101S (lane 13) had a similar product profile to L101A but overexpressed poorly in *E. coli*. TamI L101Y (lane 12) accumulated tirandamycin D as the major product. (B) Most variations at the L295 site resulted in WT activity. TamI L295Y (lane 12) led to no product formation. TamI L295T (lane 14) showed a similar product profile as L295V (lane 9) but it overexpressed poorly in *E. coli* compared to L295V. (C) Mutations at the L244 site led to WT activity with the exception of L244Y (lane 12) and L244D (lane 14) which abolished enzymatic activity. (D) Double mutant TamI L101A_L295V (lane 8) was designed to increase formation of the new double oxidation congener with RT 10min. Triple mutant L101A_L295I_L244A (lane 9) was engineered to explore the effect of simultaneously introducing mutations at these three key leucine sites and it displayed a non-selective product profile.

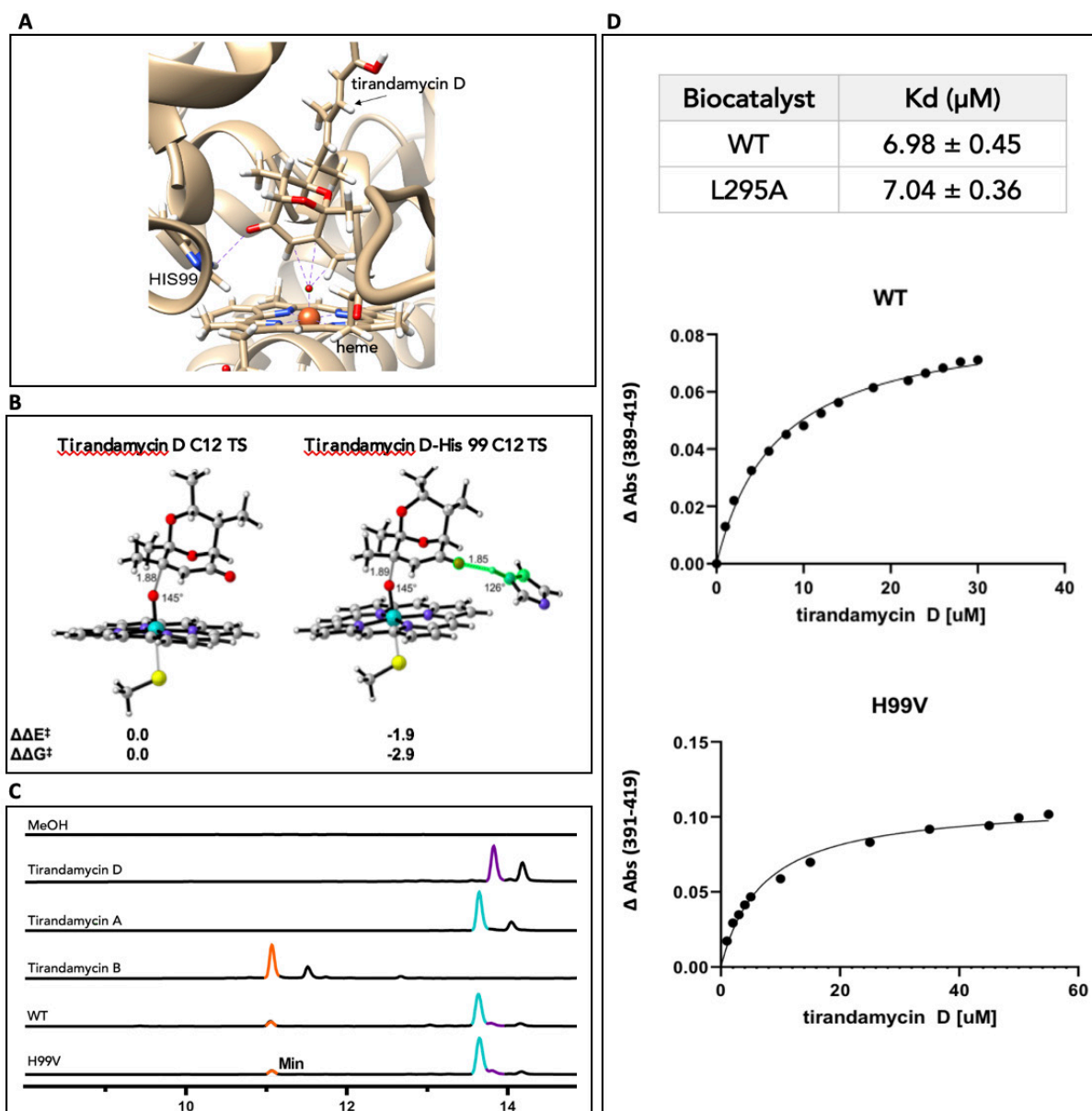


Figure A.9. Suspected H-bonding interaction between His99 and tirandamycin D ketone. (A) MD simulations of TamI WT with tirandamycin D substrate. A predicted hydrogen bonding interaction between Histidine 99 and the C-10 allylic ketone of tirandamycin D is shown in a dashed arrow. (B) QM calculations of the tirandamycin D-His 99 interaction. Energies are shown in kcal/mol. $\Delta\Delta E^\ddagger$ quantifies the electronic energy and $\Delta\Delta G^\ddagger$ is Gibb's free energy. The C=O-His interaction lowers the electronic energy of the rate limiting transition state. The same is true for the free energy of the transition state, therefore it seems favorable to have this His interaction for the epoxidation. (C) Analytical enzymatic reaction of TamI WT and H99V variant with tirandamycin D substrate. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Comparable product conversion to TamI WT is exhibited by H99V mutant. (D) Equilibrium substrate binding analysis of TamI WT and H99V with tirandamycin D. Reported errors are standard deviations calculated from experiments performed in triplicate.

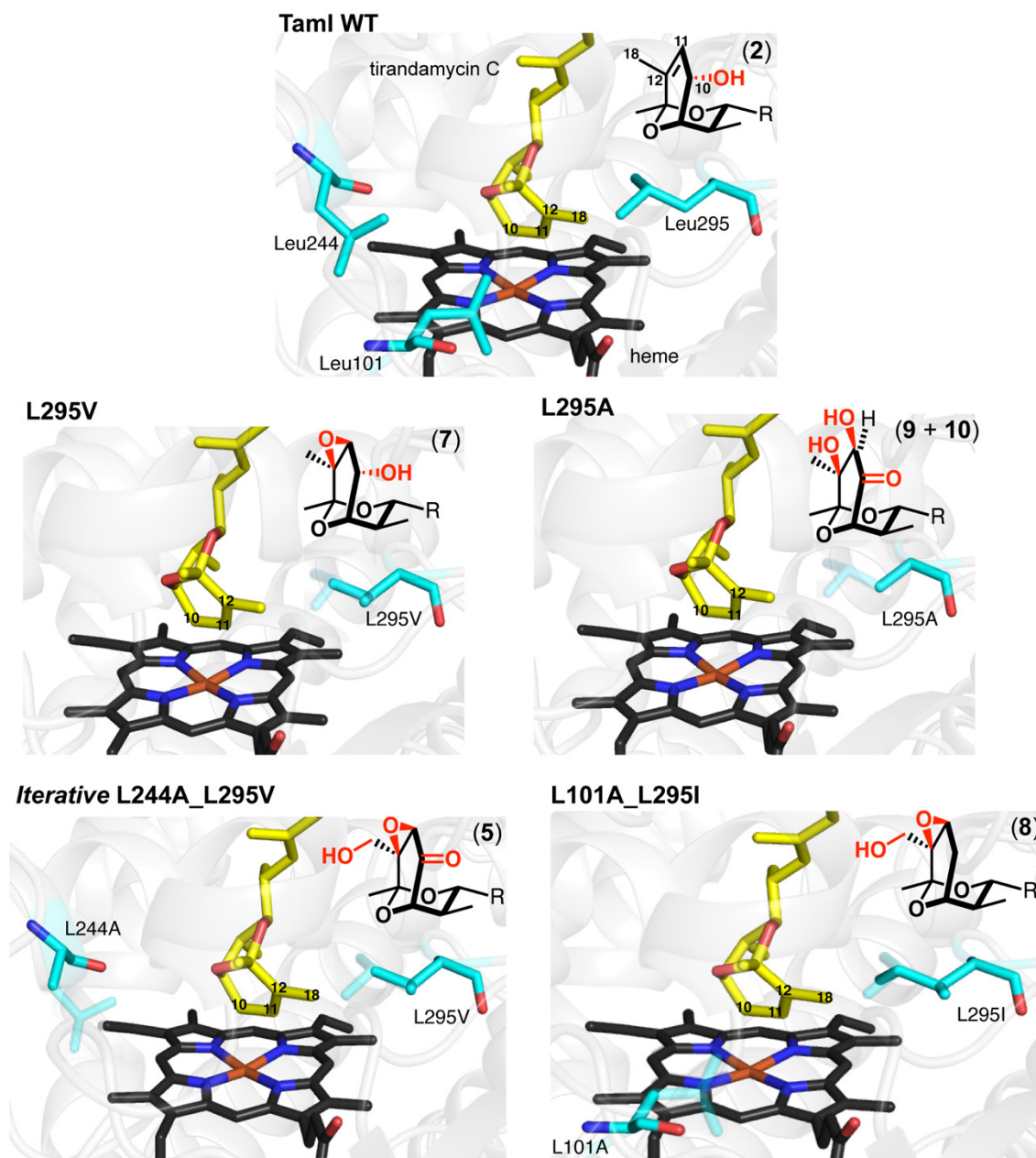


Figure A.10. PyMOL simulations of TamI mutants. Substitutions were modeled in PyMOL from TamI WT in complex with 1. For mutants TamI L295A and L295V: the major products isolated from preparative-scale reactions with the P450 enzyme and 1 as substrate are depicted on the top right corner of each model. For mutants TamI L244A_L295V and L101A_L295I: The more highly oxidized products from analytical enzymatic reactions with the P450 and 1 as substrate are depicted on the top right corner of each model.

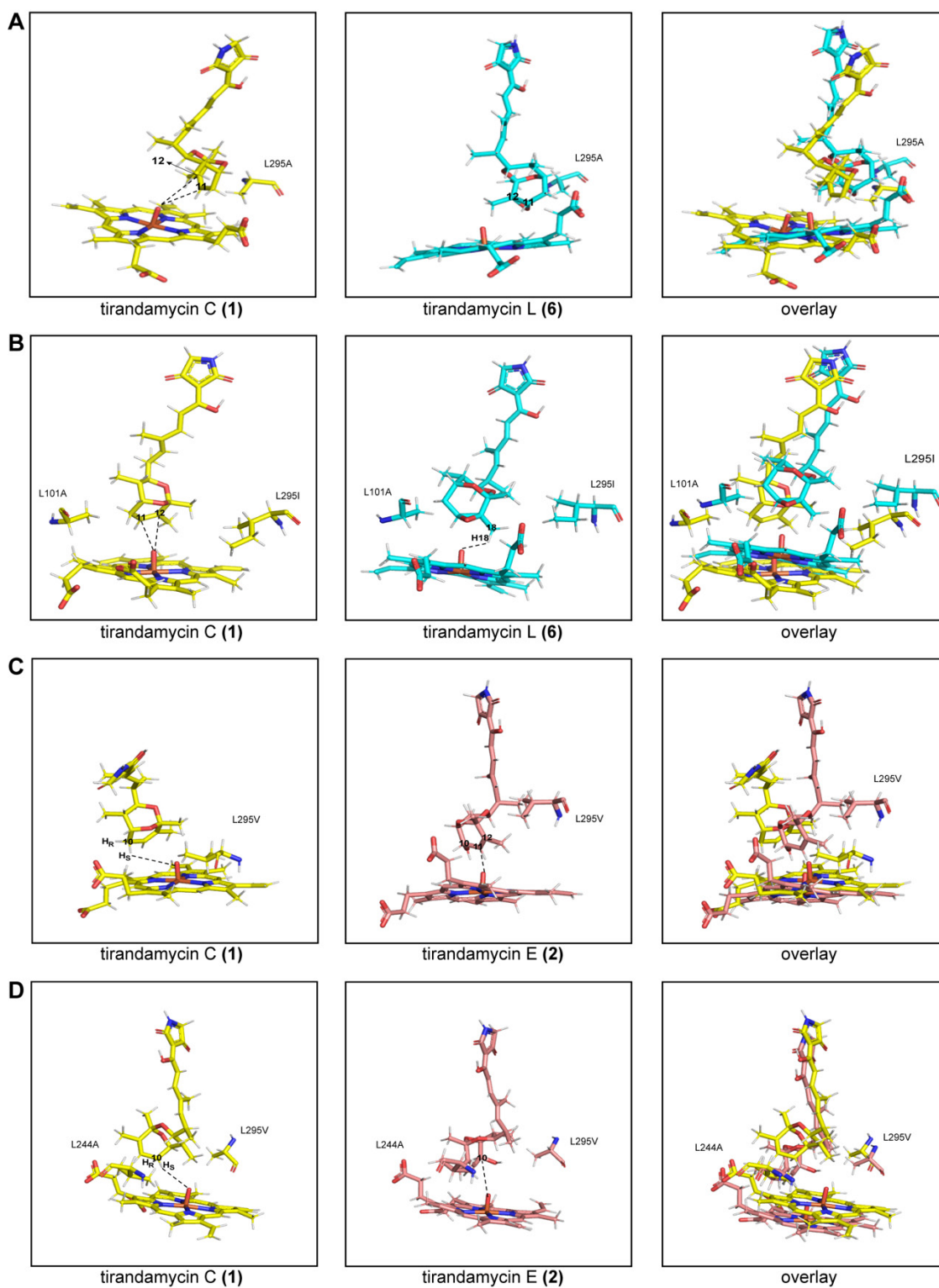


Figure A.11. MD simulations of TamI mutants. The simulations show changes in the orientation between the various tirandamycin substrates that match the geometry for the expected reactivity and selectivity. Simulations were performed using Amber12/ff99SB-ILDN/GAFF for 1500 ns.

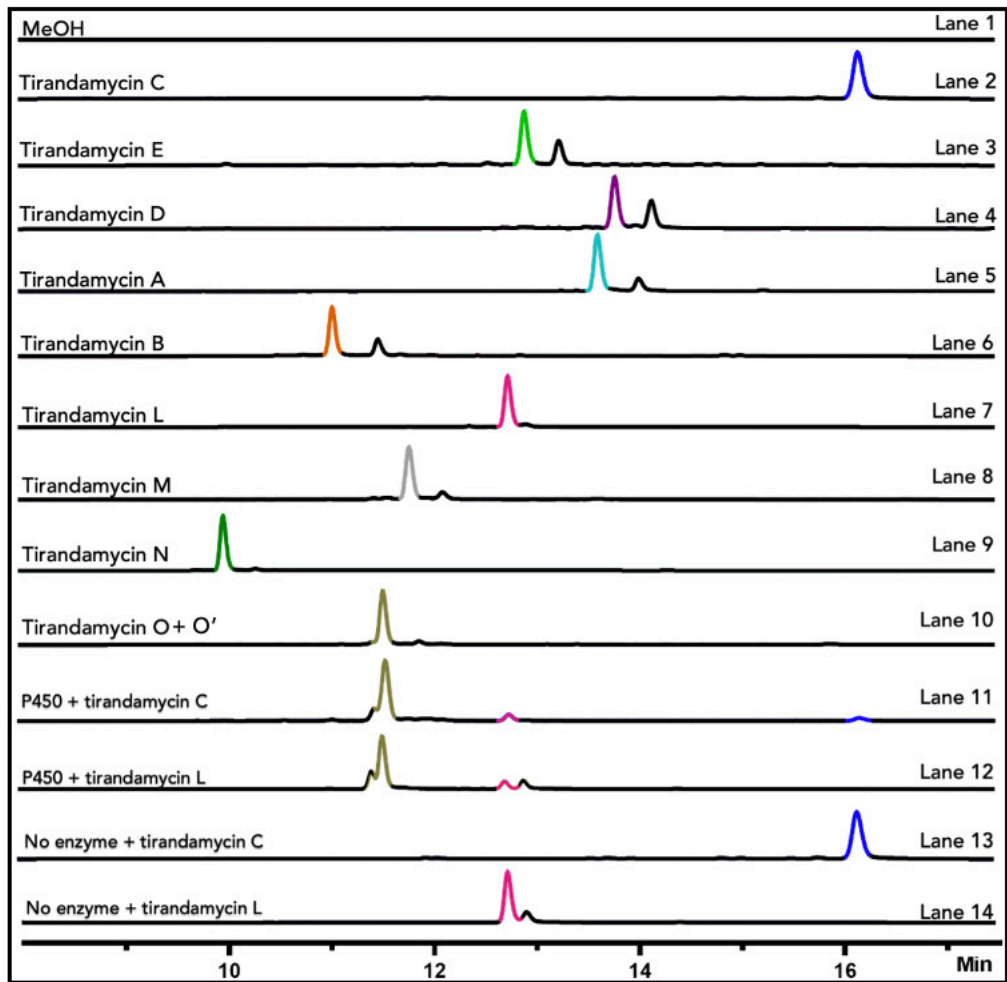


Figure A.12. Endpoint assays with TamI L295A. Analytical enzymatic reactions with TamI L295A and tirandamycin substrates are shown in lanes 11 and 12. Reactions were run using standard conditions. No enzyme control reactions are shown in lanes 13 and 14. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).

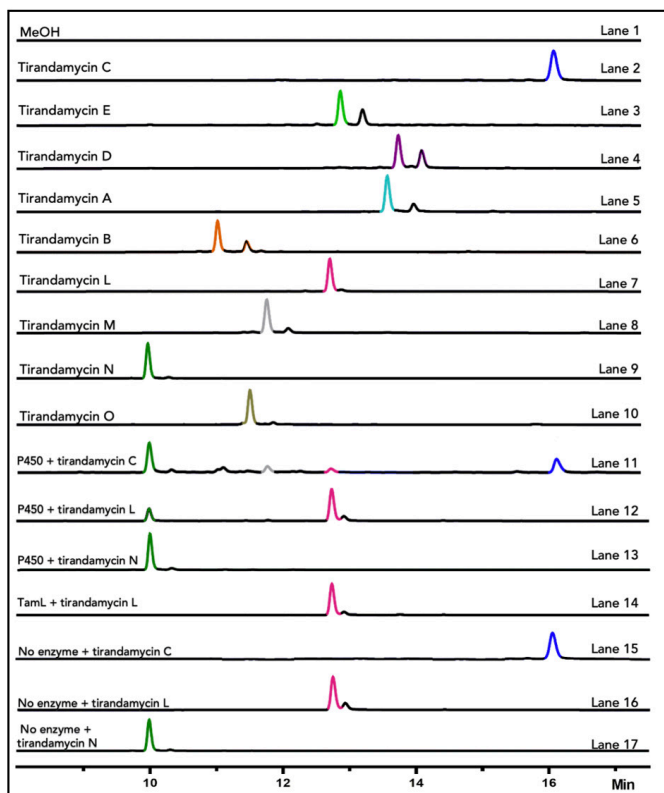
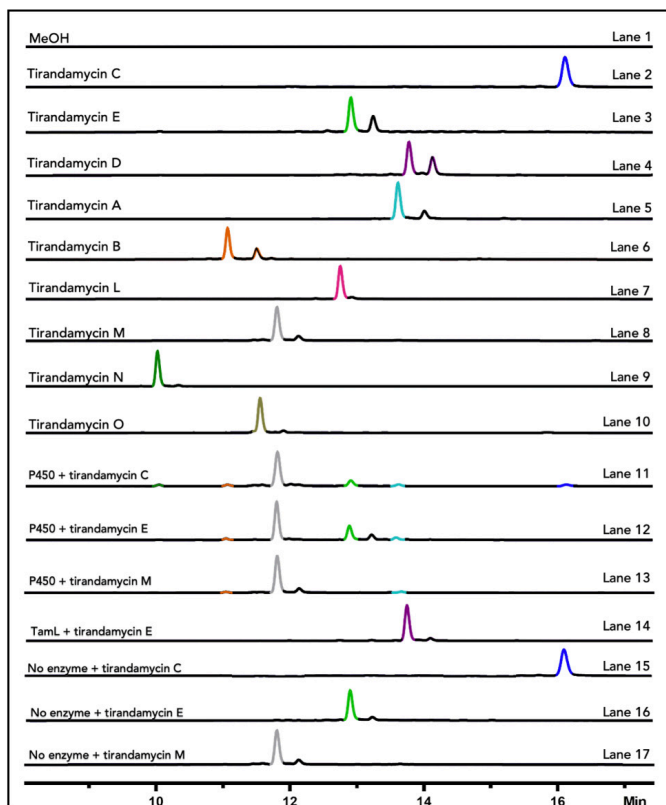


Figure A.13. Endpoint assays with TamI L101A_L295I. Analytical enzymatic reactions with TamI L101A_L295I and various tirandamycin substrates are shown in lanes 11-13. Control reactions are shown in lanes 14-17. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).

Figure A.14. Endpoint assays with TamI L295V. Analytical enzymatic reactions with TamI L295V and various tirandamycin substrates are shown in lanes 11-13. Control reactions are shown in lanes 14-17 including no enzyme controls. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).



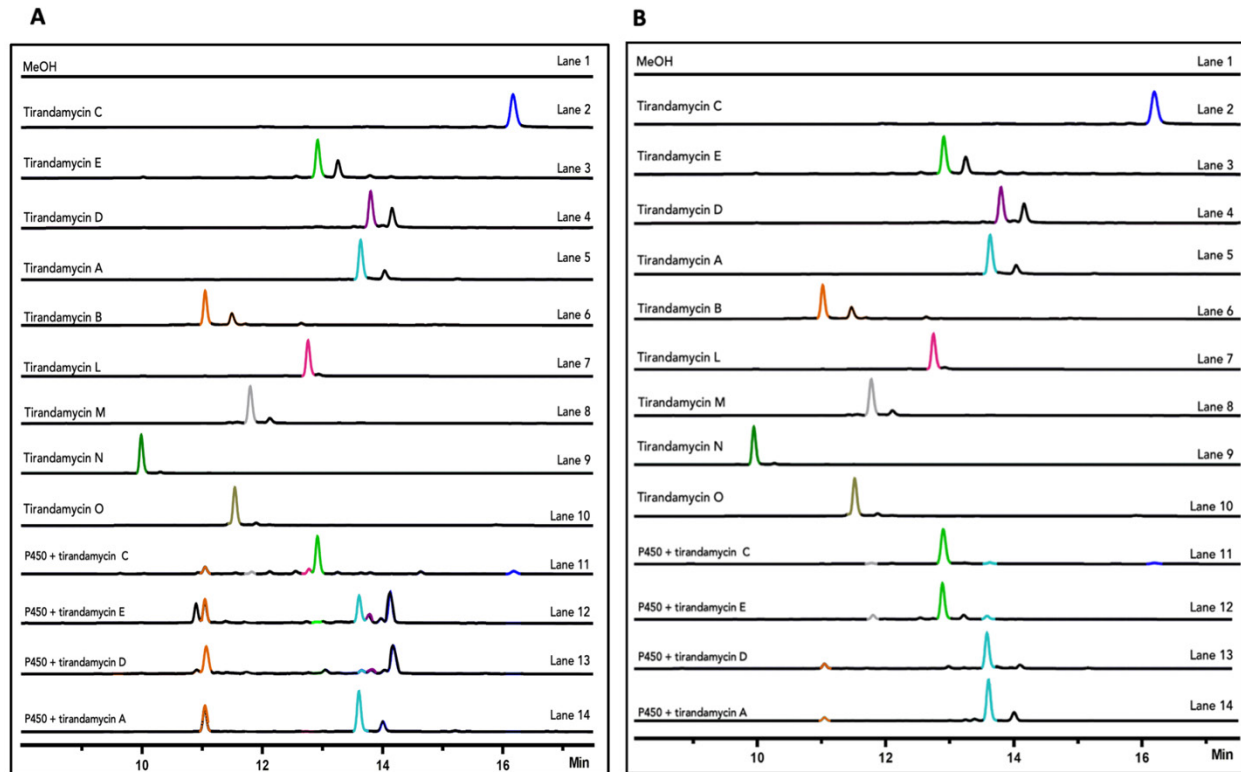


Figure A.15. The stepwise oxidative cascade of iterative TamI L244A_L295V and TamI WT. (A) Analytical enzymatic reactions with TamI L244A_L295V and various tirandamycin substrates. (B) Control reactions with TamI WT and various tirandamycin substrates. Minimal to no formation of tirandamycin B is observed. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).

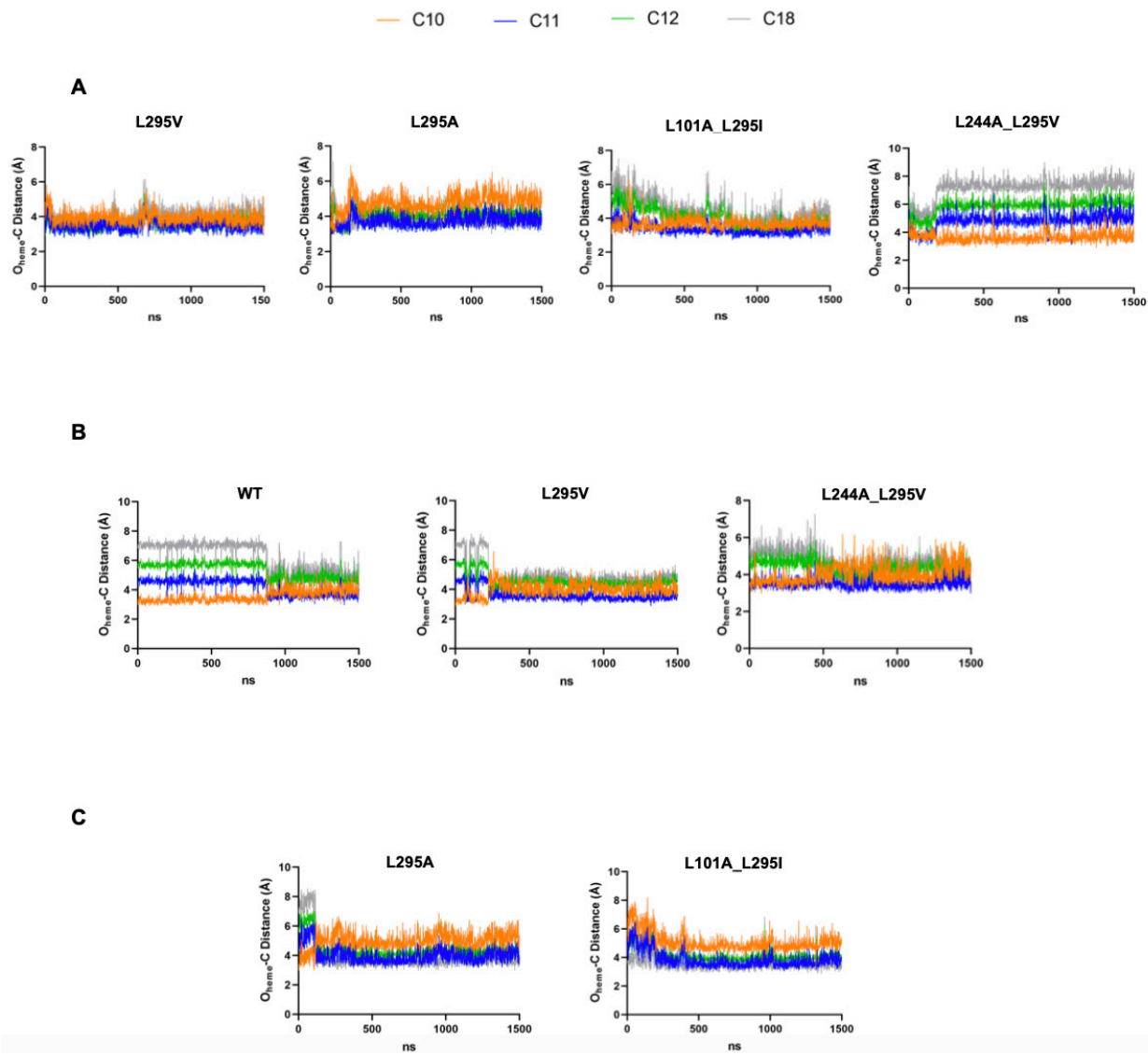


Figure A.16. Bicyclic ketal distance from Fe(IV)-oxo radical cation. (A) MD simulations of TamI mutants with 1. (B) MD simulations with 2. (C) MD simulations with 6. All simulations were run for 1500 ns.

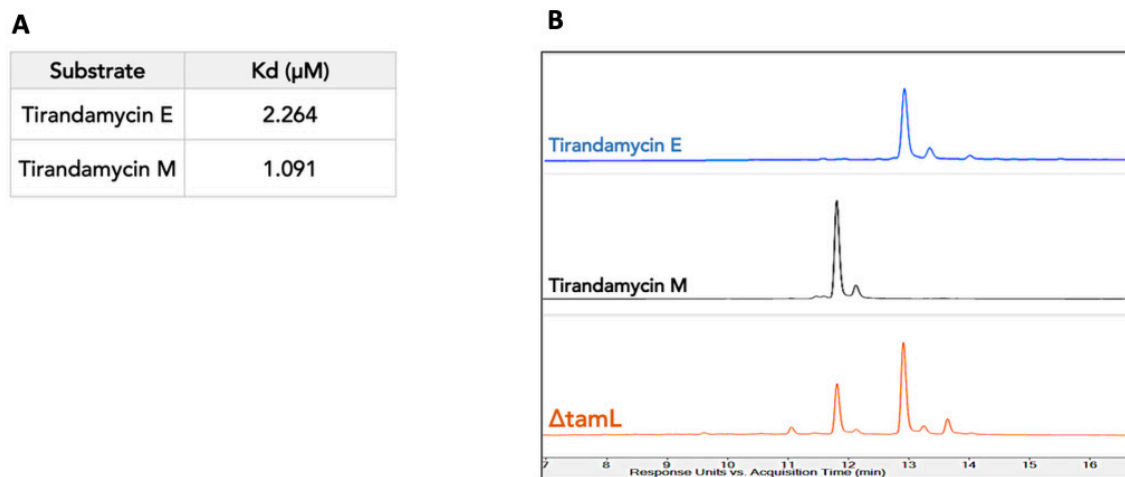


Figure A.17. Assays with the flavoprotein TamL. (A) Binding affinities of tirandamycin E and M to the TamL enzyme using ITC. (B) HPLC analysis of the culture broth of a *Streptomyces* sp. 307-9 ΔtamL flavoprotein mutant strain after 4 days of growth. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin E and M.

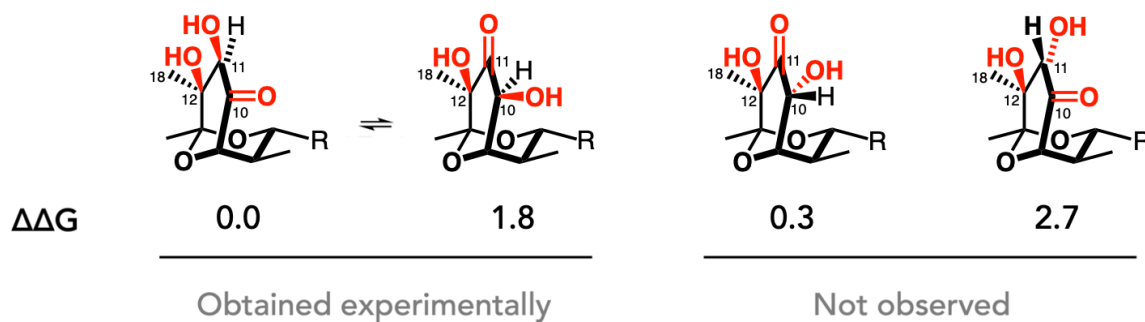


Figure A.18. Stereochemical configuration of tirandamycins O and O'. The $\Delta\Delta G$ of each theoretically possible stereoisomer was calculated.

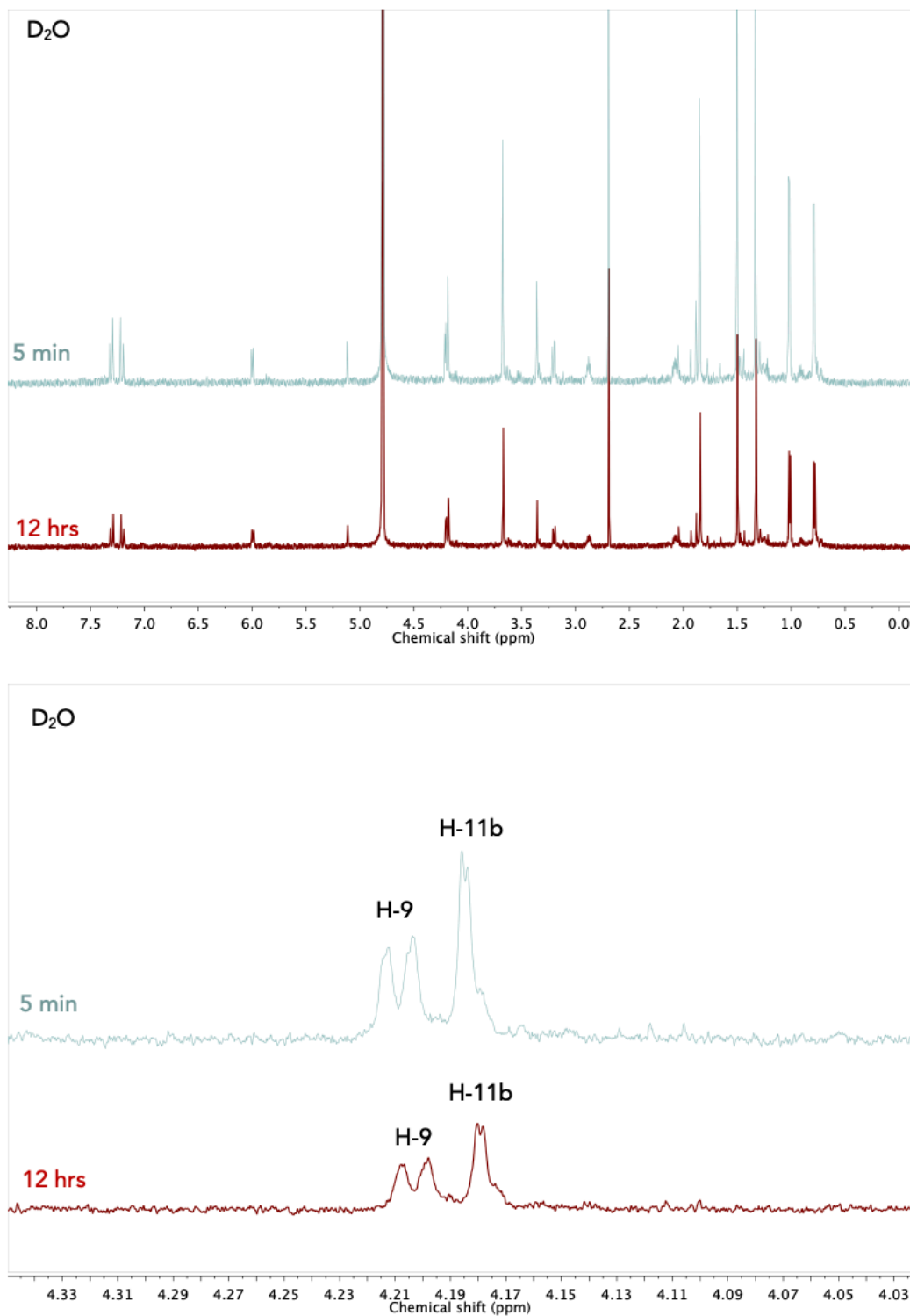


Figure A.19. $^1\text{H-NMR}$ experiments with tirandamycin O (9). C10-keto-tirandamycin O (5:1) was incubated with deuterated water and $^1\text{H-NMR}$ spectra were taken after 5 min and 12 hours. No change in the NMR spectra was detected. Assignments for H-9 and H-11b protons were made based on COSY correlations between H-9 and H-8.

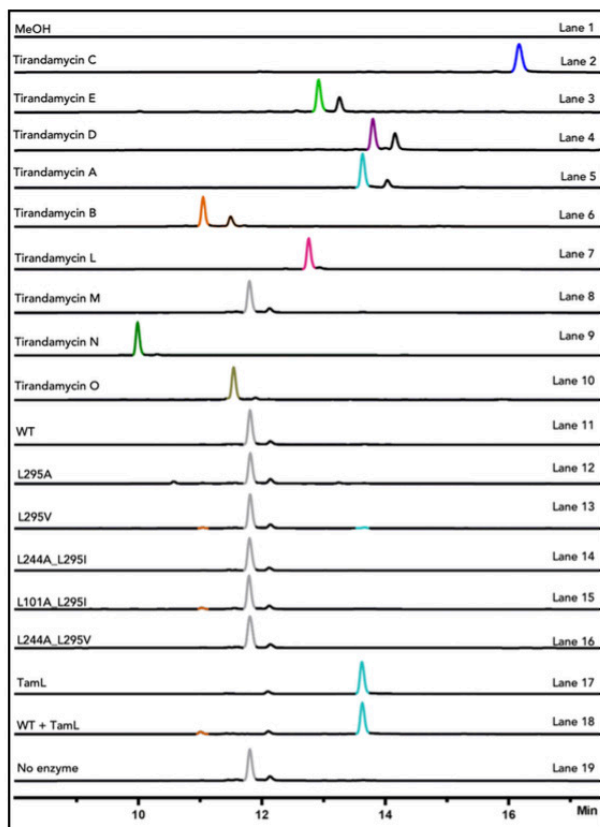
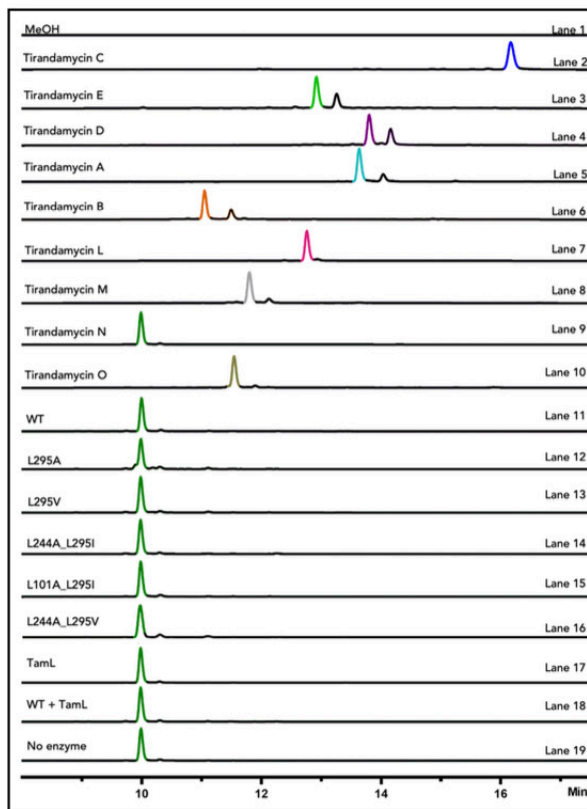


Figure A.20. Endpoint assays with TamI WT and variants using tirandamycin M substrate. Analytical enzymatic reactions using standard conditions are shown in lanes 11-16. Control reactions are shown in lanes 17-19. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).

Figure A.21. Endpoint assays with TamI WT and variants using tirandamycin N substrate. Analytical enzymatic reactions using standard conditions are in lanes 11-16. Control reactions are shown in lanes 17-19. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).



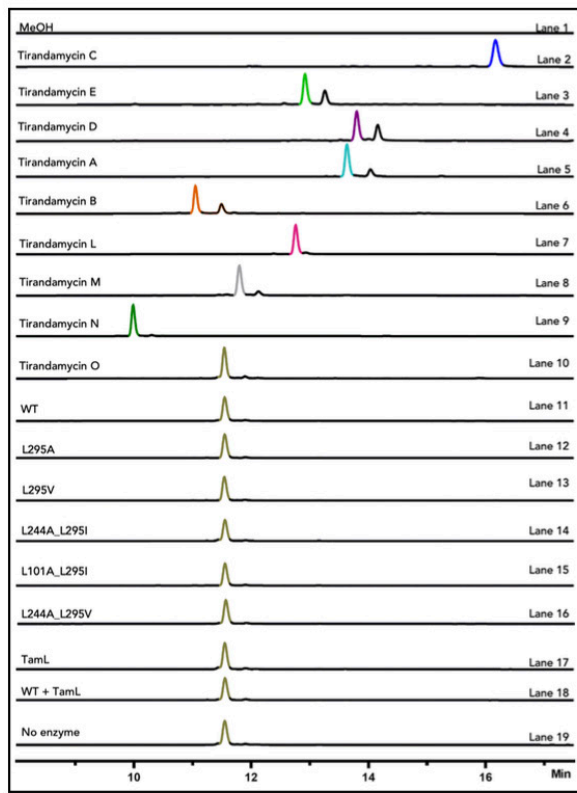


Figure A.22. Endpoint assays with TamI WT and variants using tirandamycin O and O' substrates (5:1 ratio of tirandamycin O to O'). Analytical enzymatic reactions using standard conditions are in lanes 11-16. Control reactions are shown in lanes 17-19. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and a mix of O and O' (in a 5:1 ratio) (lanes 2-10).

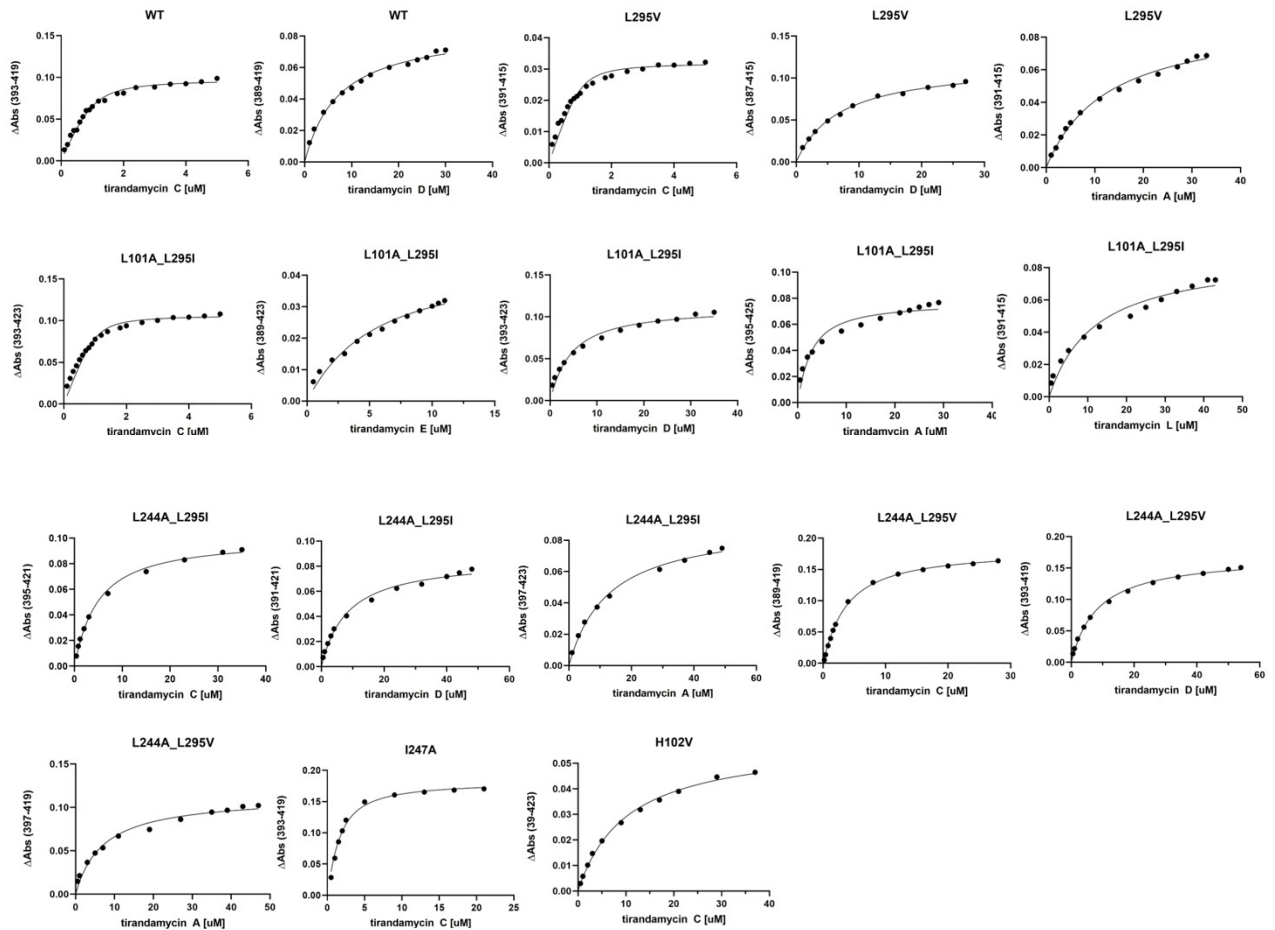


Figure A.23. Equilibrium dissociation constant binding curves for TamI WT and variants with different tirandamycin substrates. Titrations were stopped when no further spectral shift occurred even upon further addition of substrate. All tirandamycin substrates were soluble well above the concentrations used in the titration assays.

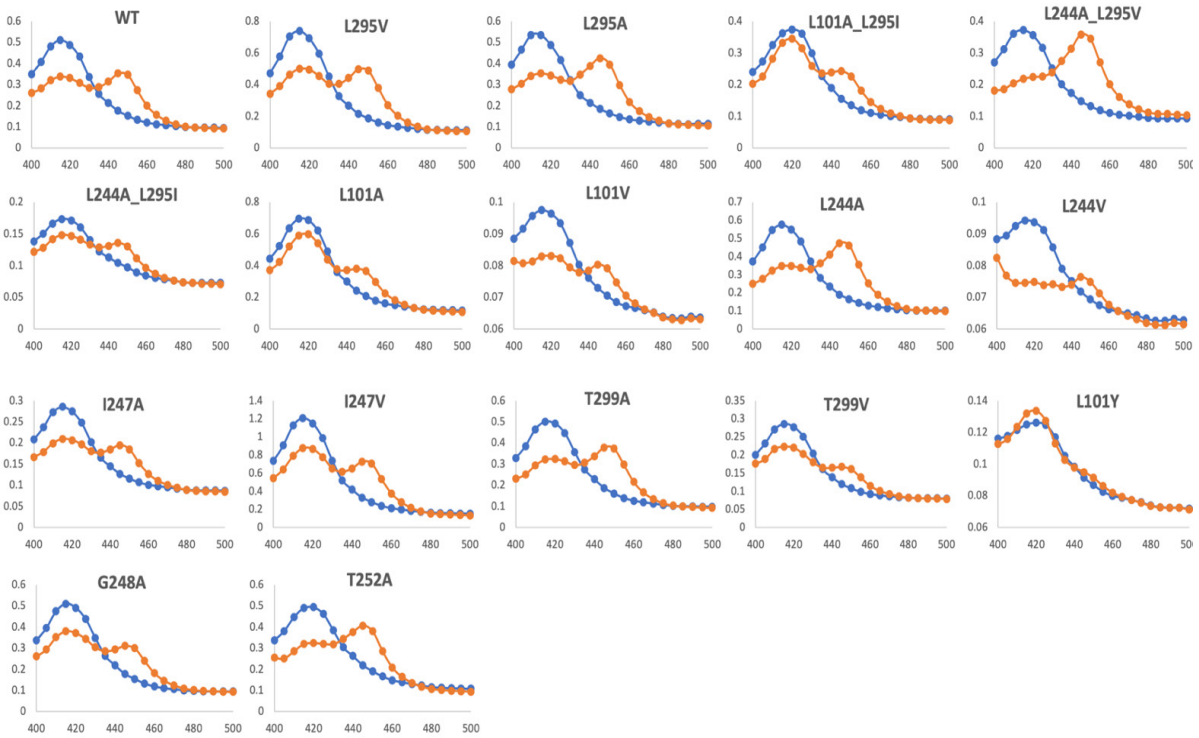


Figure A.24. CO difference spectroscopy for TamI mutants designed to alter the selectivity of TamI.

Blue = spectra of sodium dithionite reduced proteins. Orange = CO-bound reduced forms.

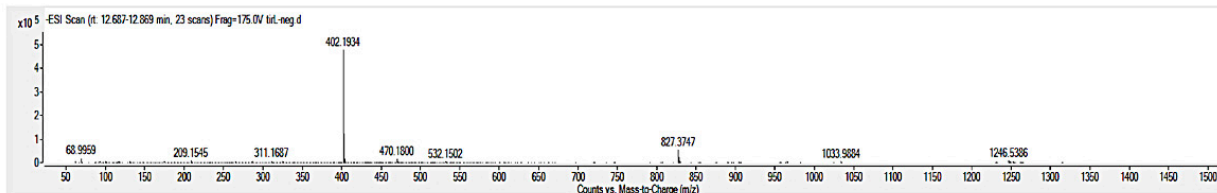
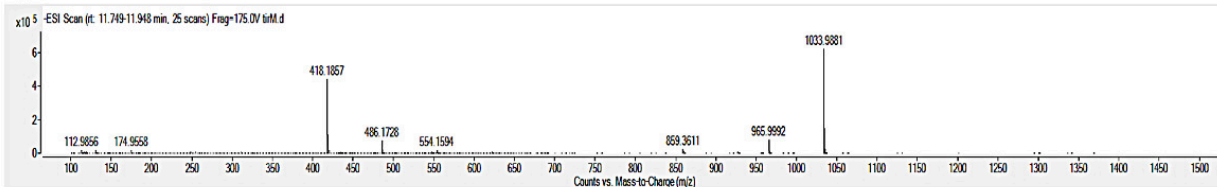
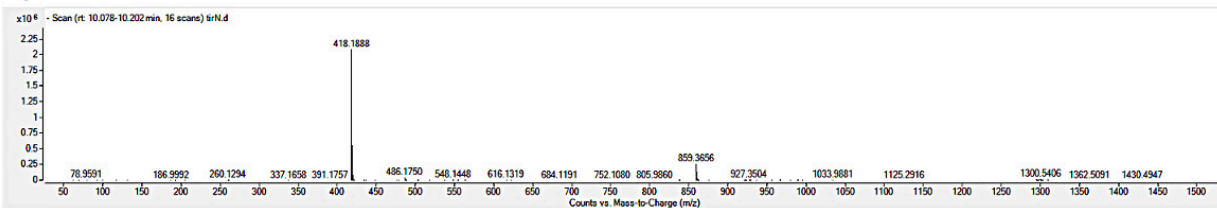
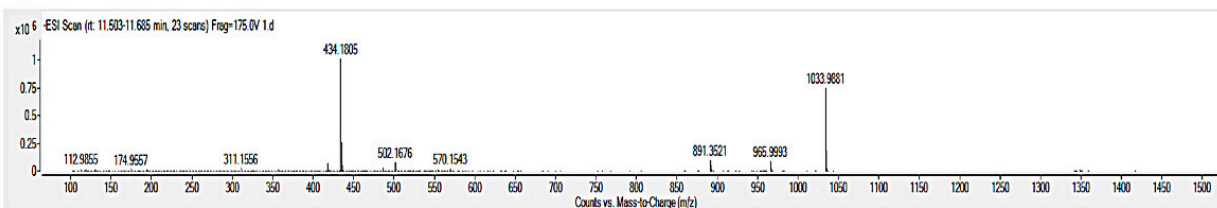
A**B****C****D**

Figure A.25. New tirandamycin compound characterization - HRMS.

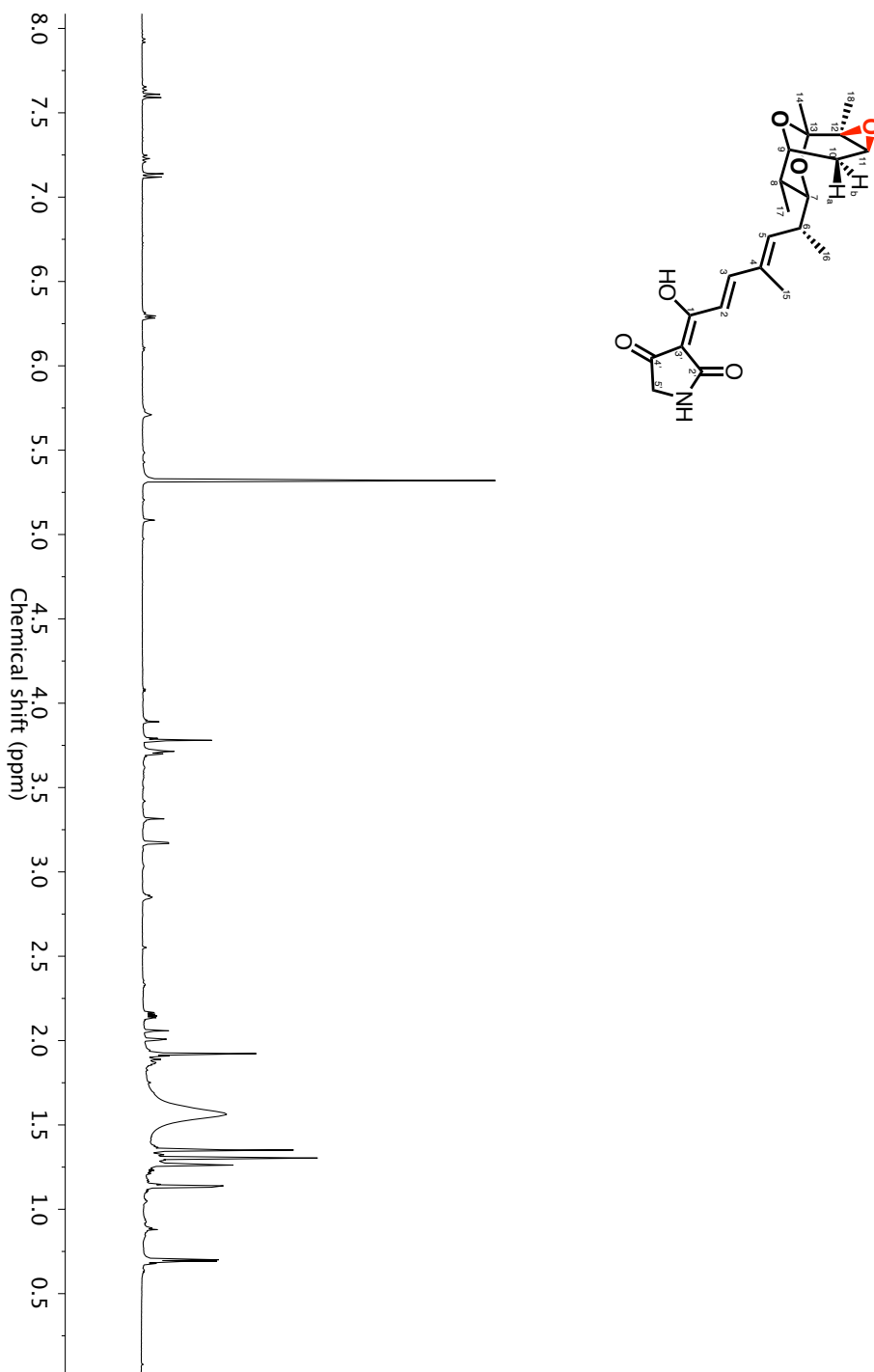
(A) Tirandamycin L (*in vitro* assay), calculated mass $[M-H]^+ = 402.1917$, experimental mass $[M-H]^+ = 402.1934$

(B) Tirandamycin M (*in vitro* assay), calculated mass $[M-H]^+ = 418.1866$, experimental mass $[M-H]^+ = 418.1857$

(C) Tirandamycin N (*in vitro* assay), calculated mass $[M-H]^+ = 418.1866$, experimental mass $[M-H]^+ = 418.1888$

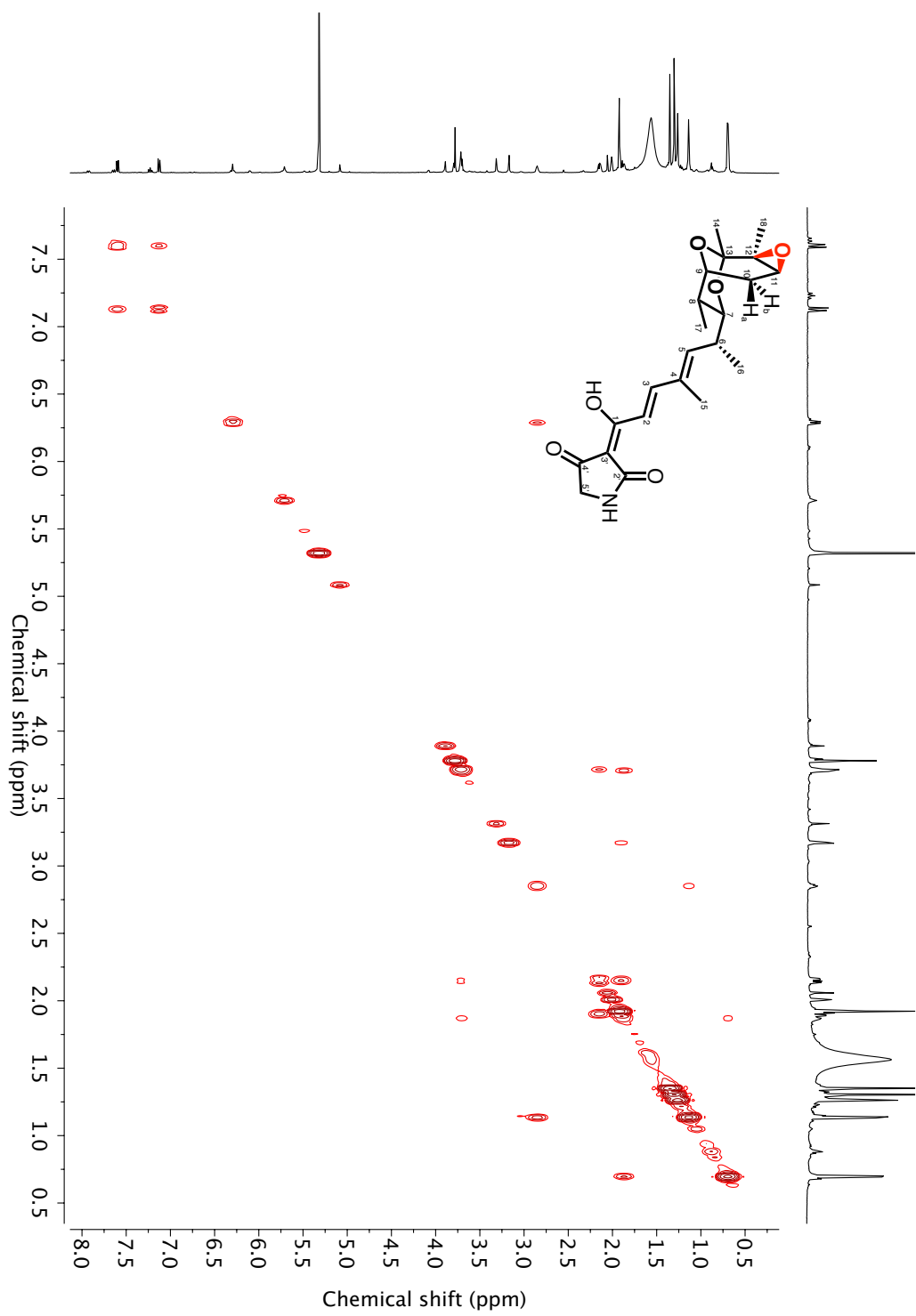
(D) Tirandamycins O and O' (*in vitro* assay), calculated mass $[M-H]^+ = 434.18149$, experimental mass $[M-H]^+ = 434.1805$

Figure A.26 Compound characterization – tirandamycin L (6)

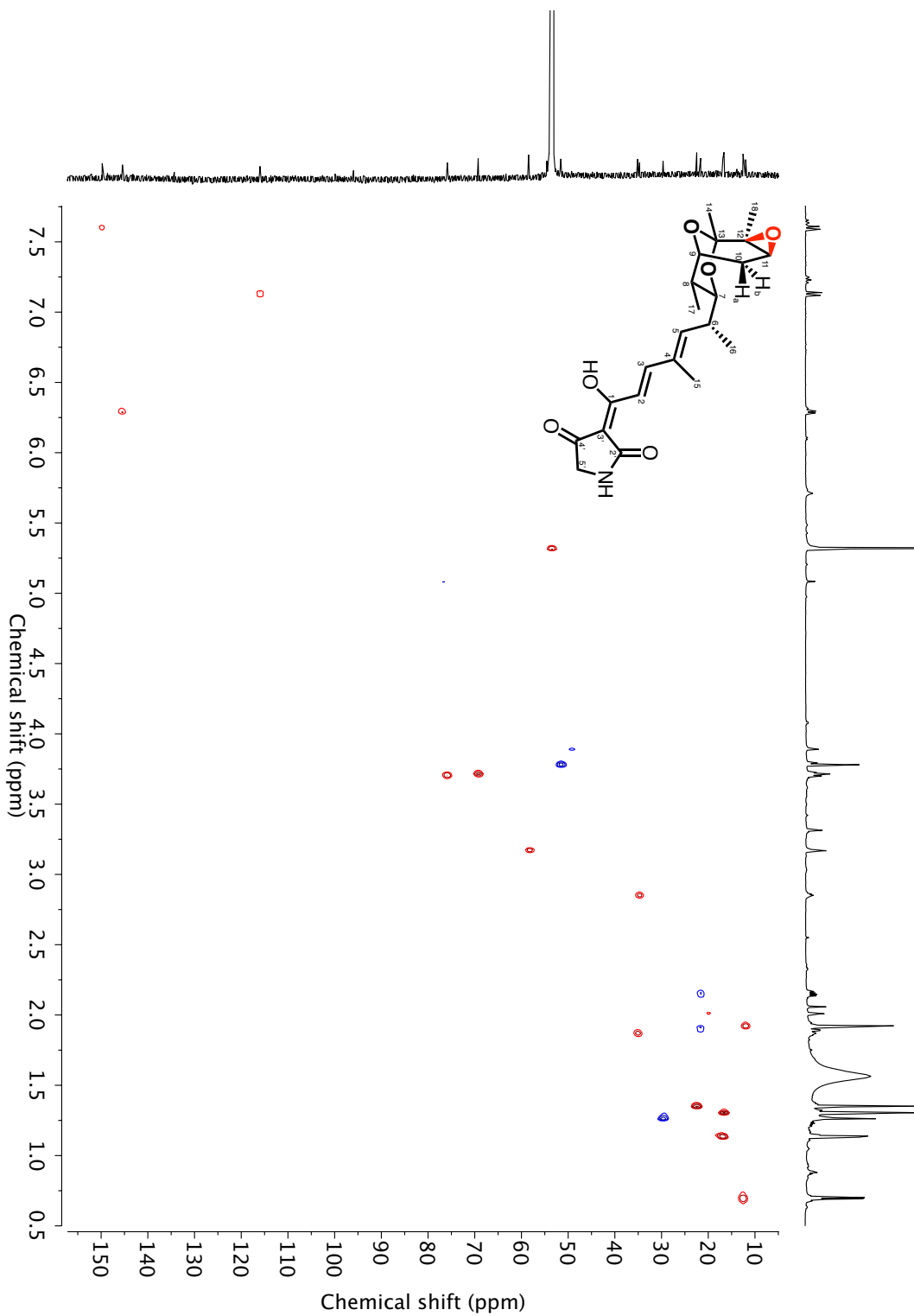


^1H spectrum of tirandamycin L (6) (800 MHz, CD_2Cl_2)

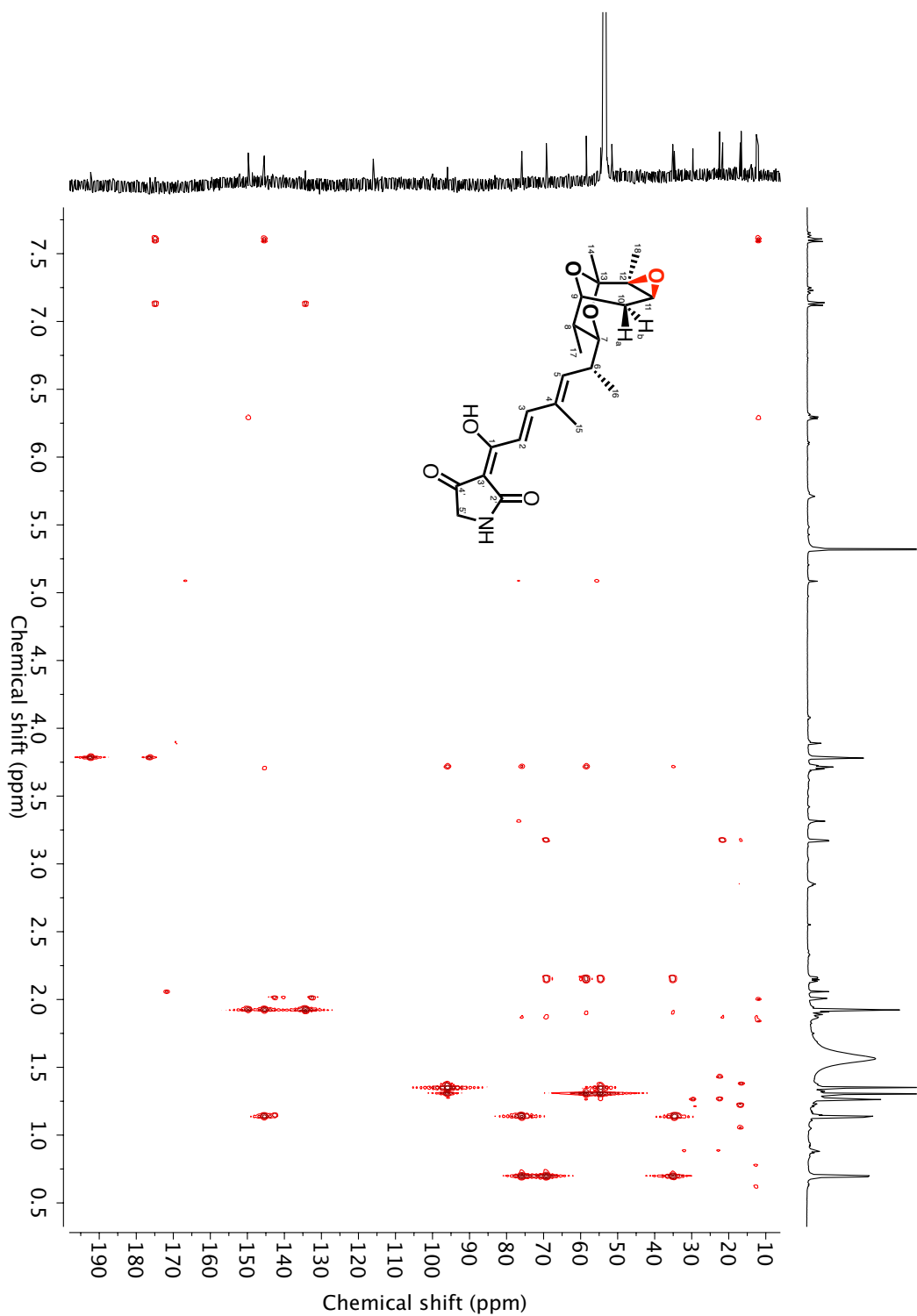
The additional set of doublets in the 7.00 - 8.00 ppm region correspond to the tetramic acid isomer of 6.



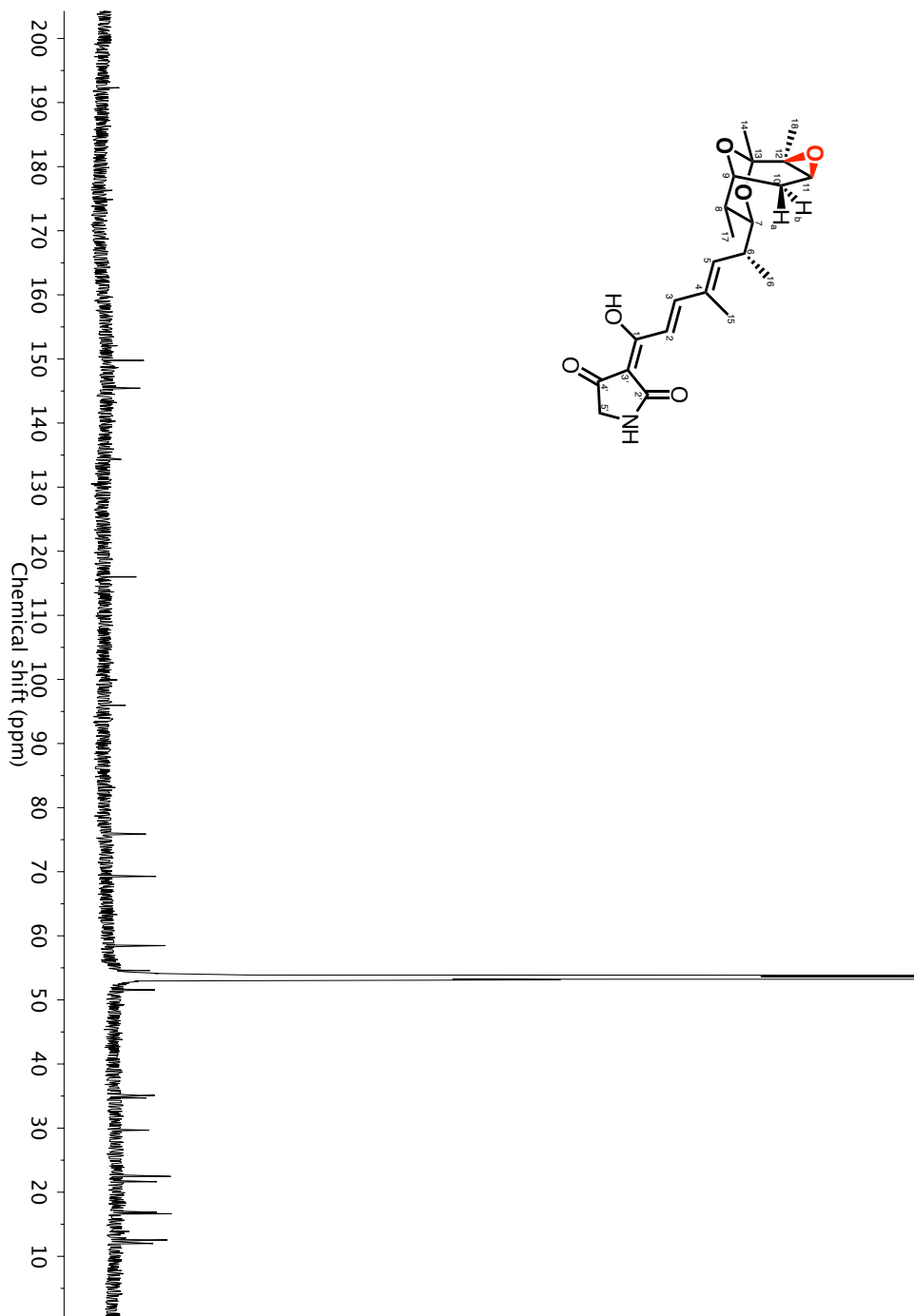
COSY spectrum of tirandamycin L (6) (800 MHz, CD₂Cl₂)



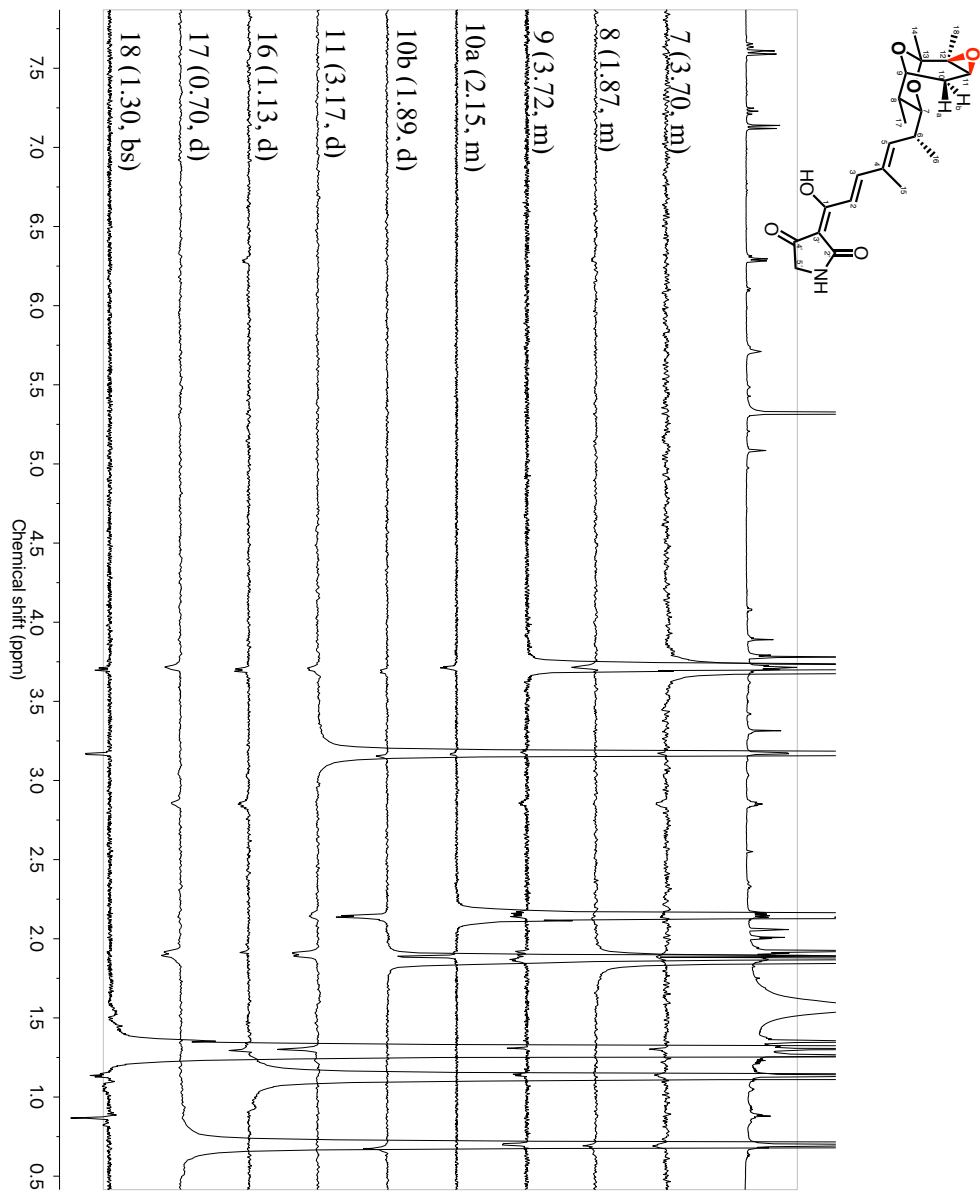
HSQC spectrum of tirandamycin L (6) (800 MHz, CD₂Cl₂)



HMBC spectrum of tirandamycin L (6) (800 MHz, CD₂Cl₂)

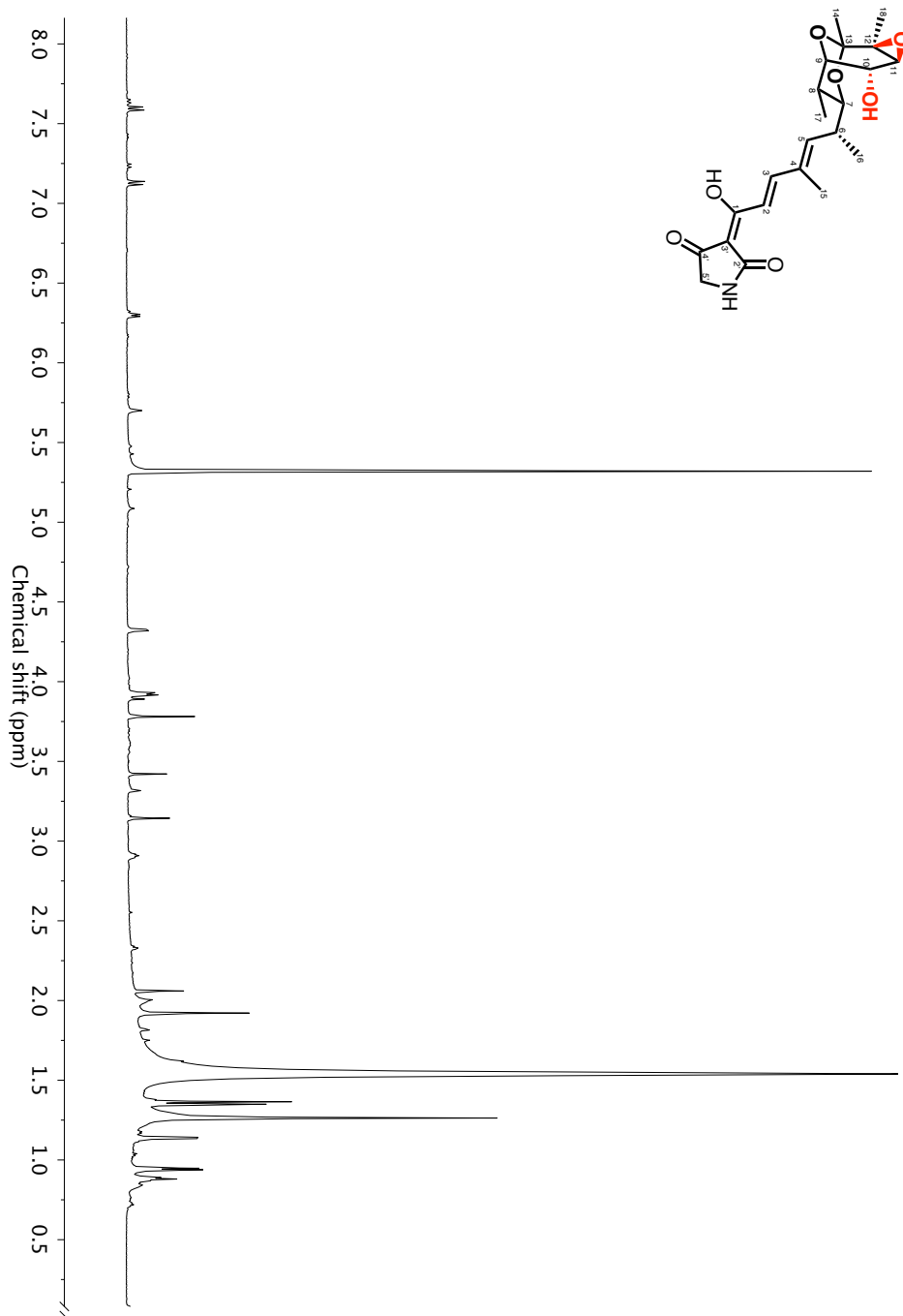


^{13}C spectrum of tirandamycin L (6) (800 MHz, CD_2Cl_2)



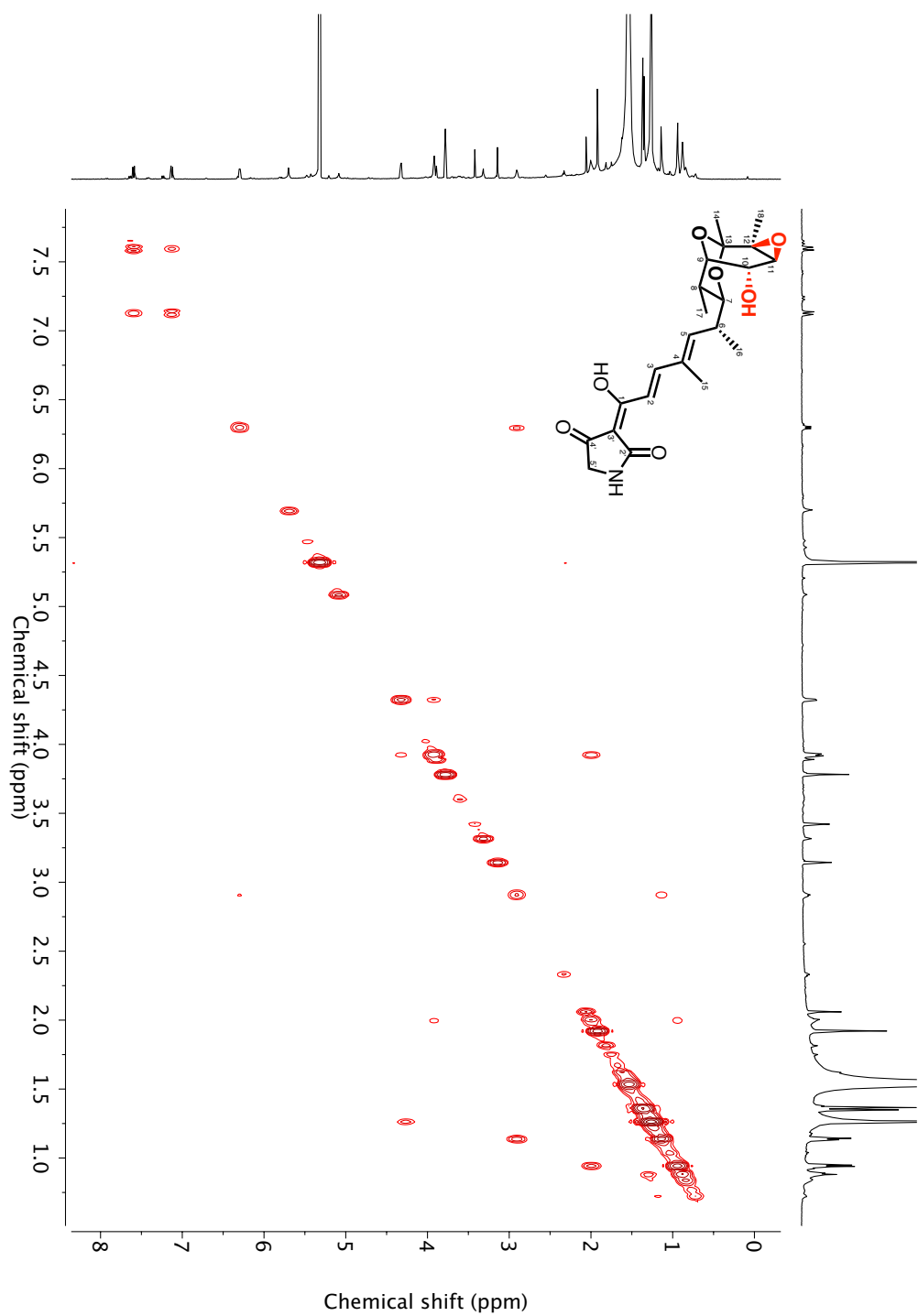
Selective NOE spectra of tirandamycin L (6) (800 MHz, CD₂Cl₂)

Figure A.27 Compound characterization – tirandamycin M (7, also known as TAM E)

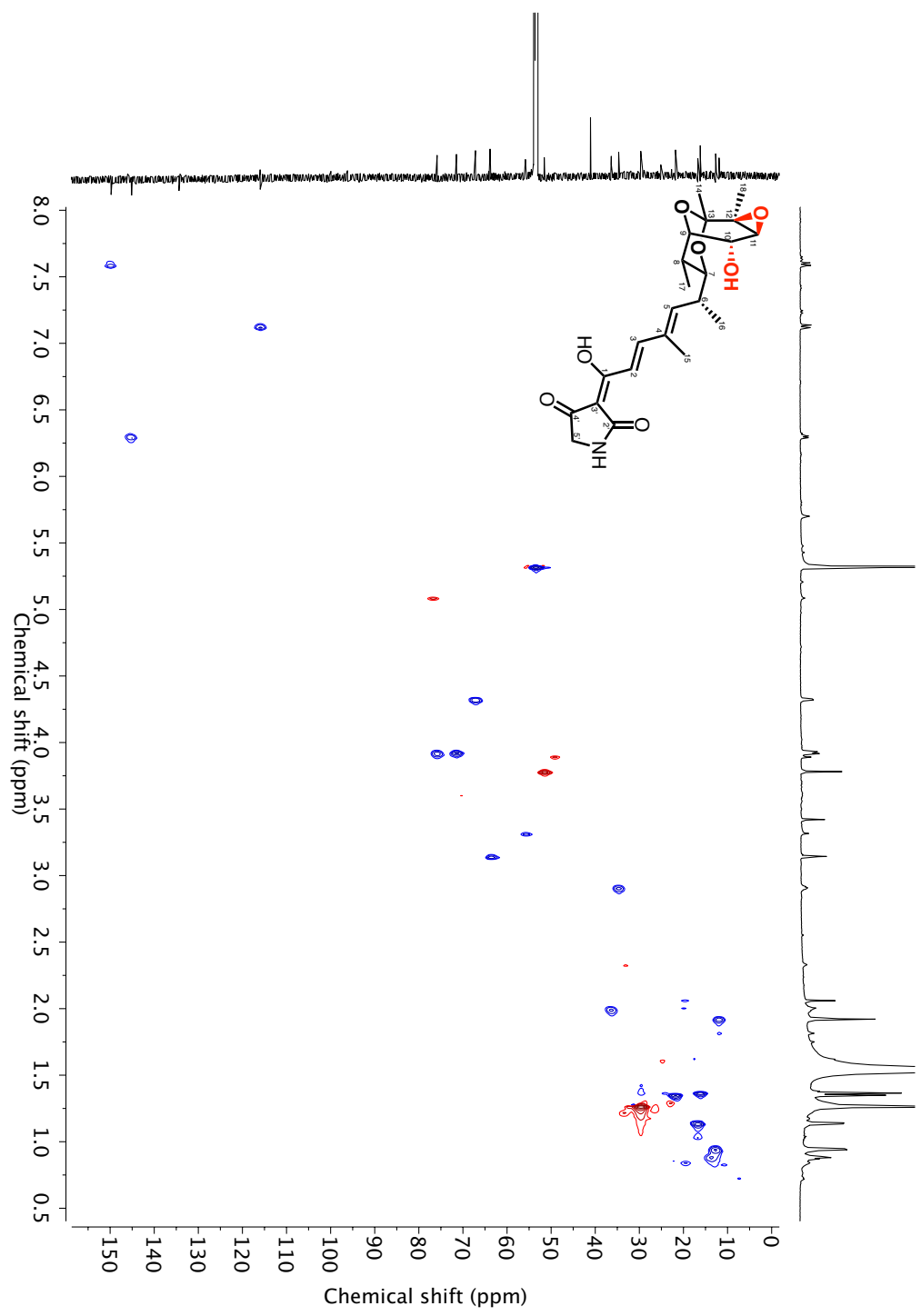


¹H spectrum of tirandamycin M (7) (800 MHz, CD₂Cl₂)

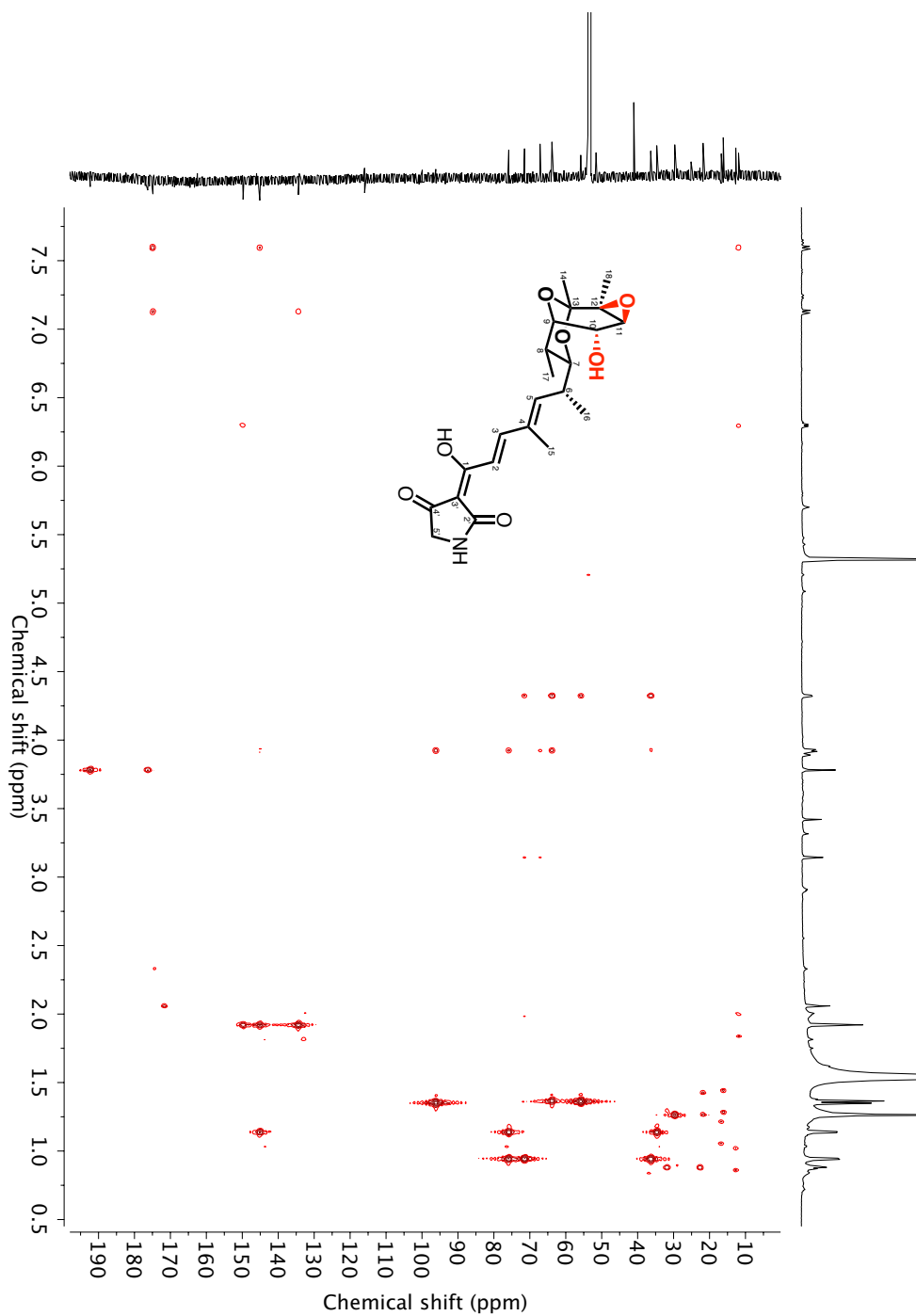
The additional set of doublets in the 7.00 - 8.00 ppm region correspond to the tetramic acid isomer of 7.



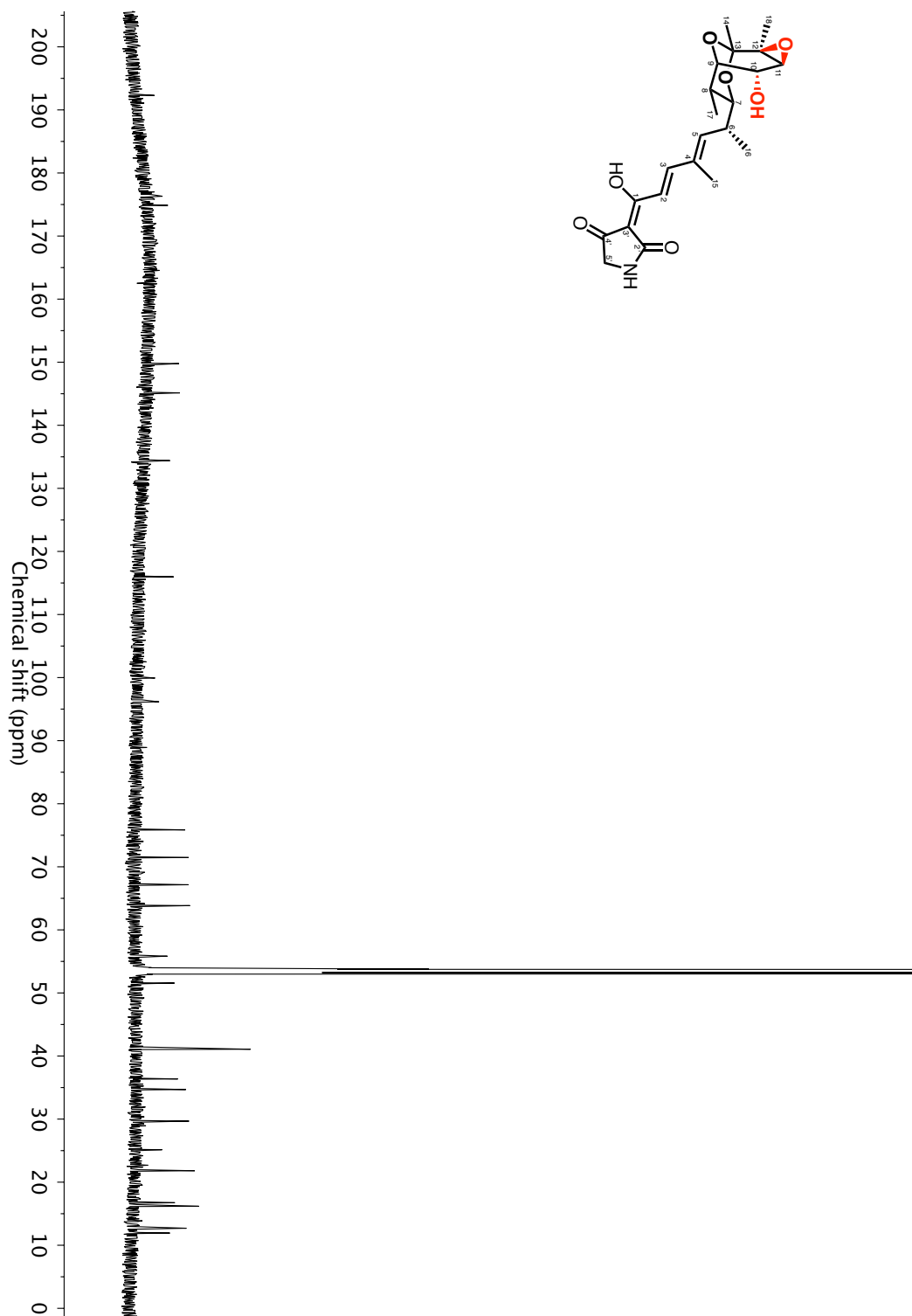
COSY spectrum of tirandamycin M (7) (800 MHz, CD₂Cl₂)



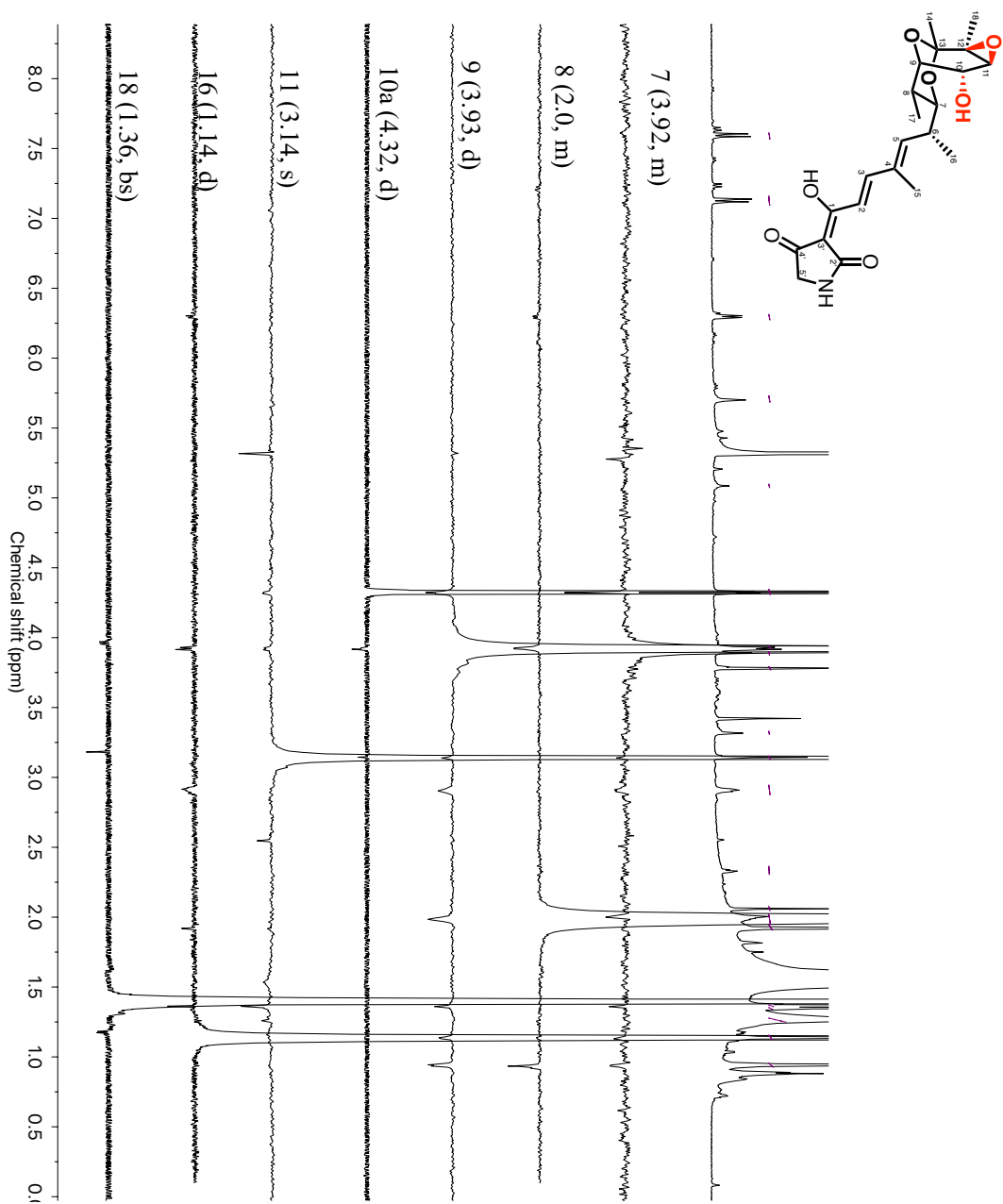
HSQC spectrum of tirandamycin M (7) (800 MHz, CD₂Cl₂)



HMBC spectrum of tirandamycin M (7) (800 MHz, CD₂Cl₂)

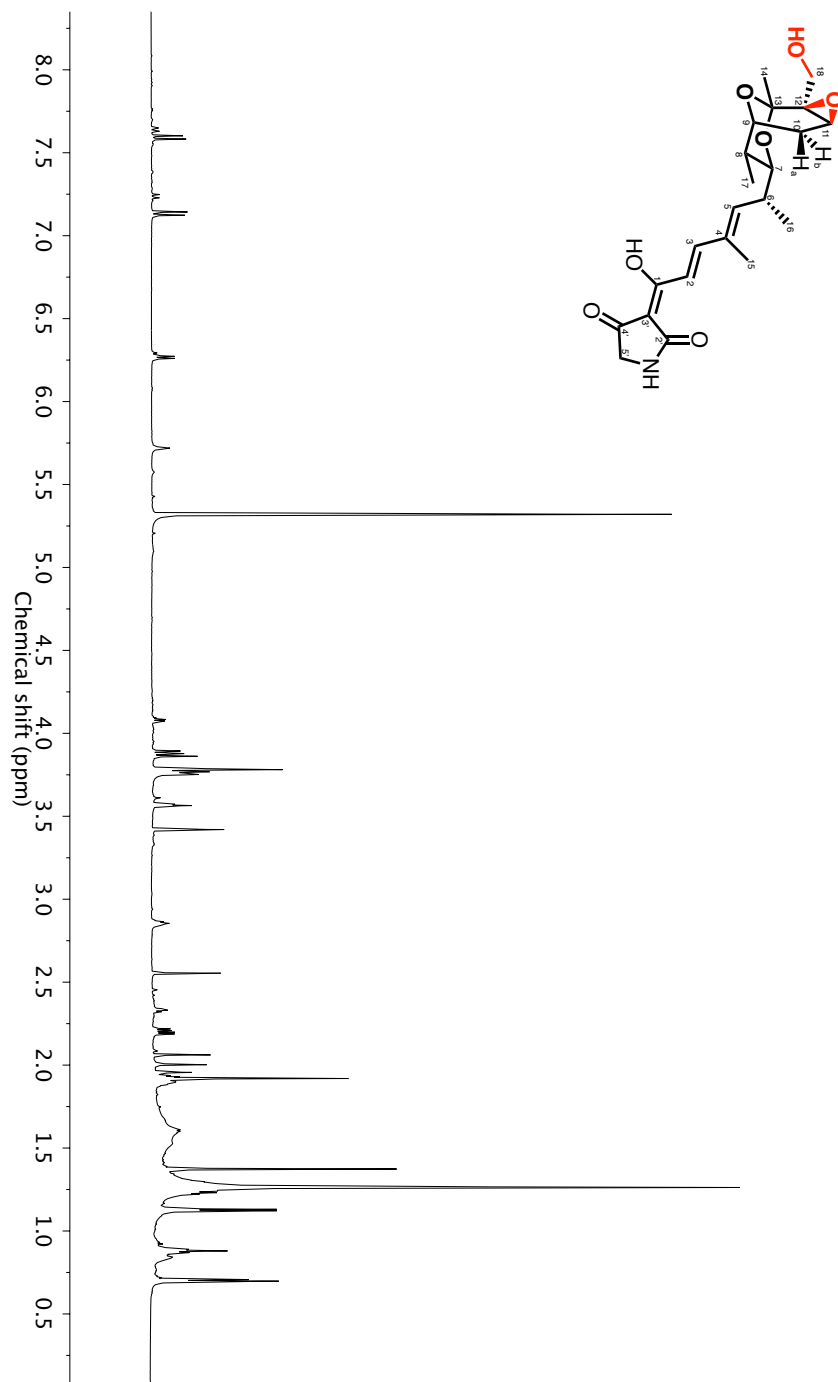


^{13}C spectrum of tirandamycin M (7) (800 MHz, CD_2Cl_2)



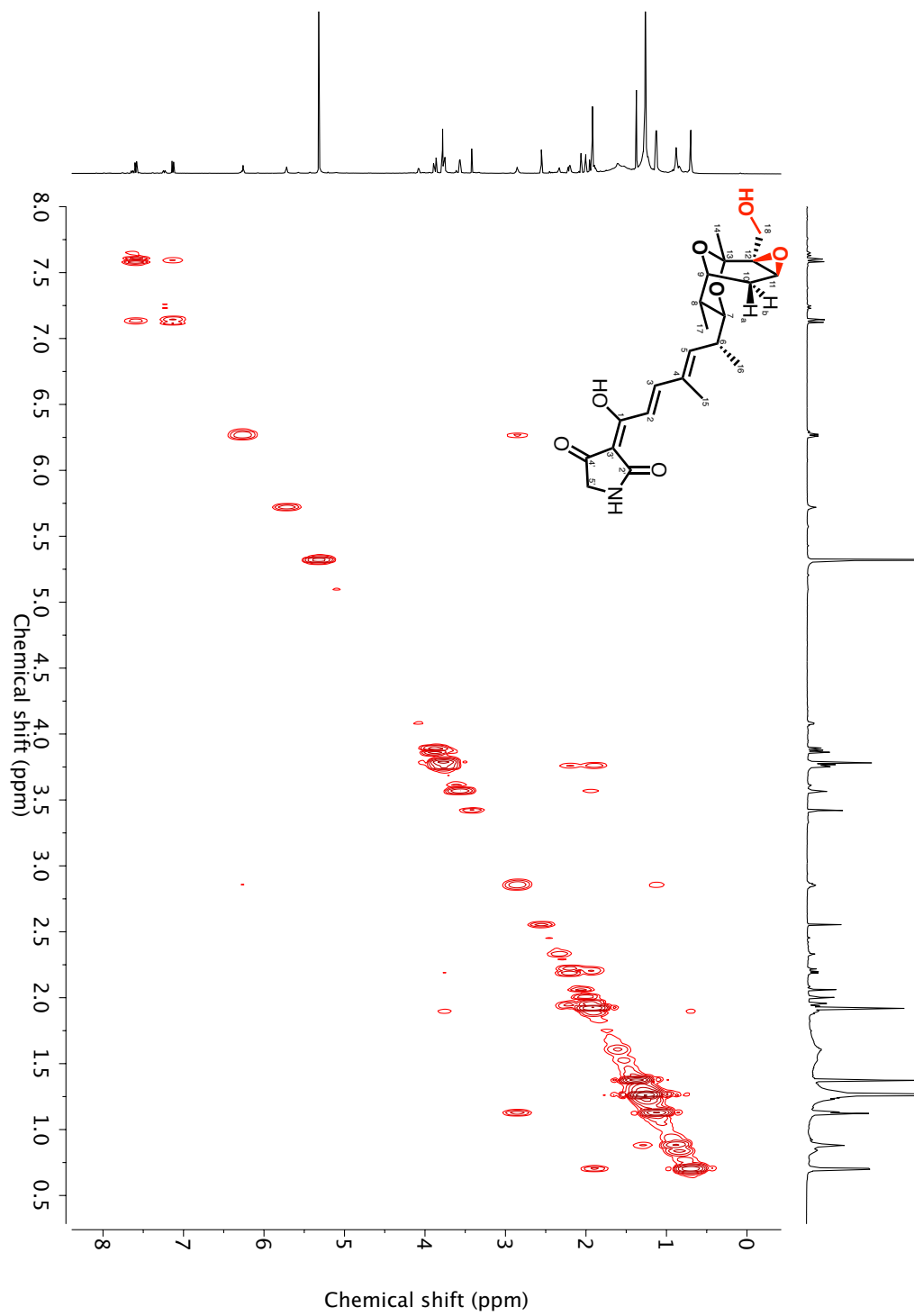
Selective NOE spectra of tirandamycin M (7) (800 MHz, CD₂Cl₂)

Figure A.28 Compound characterization – tirandamycin N (8)

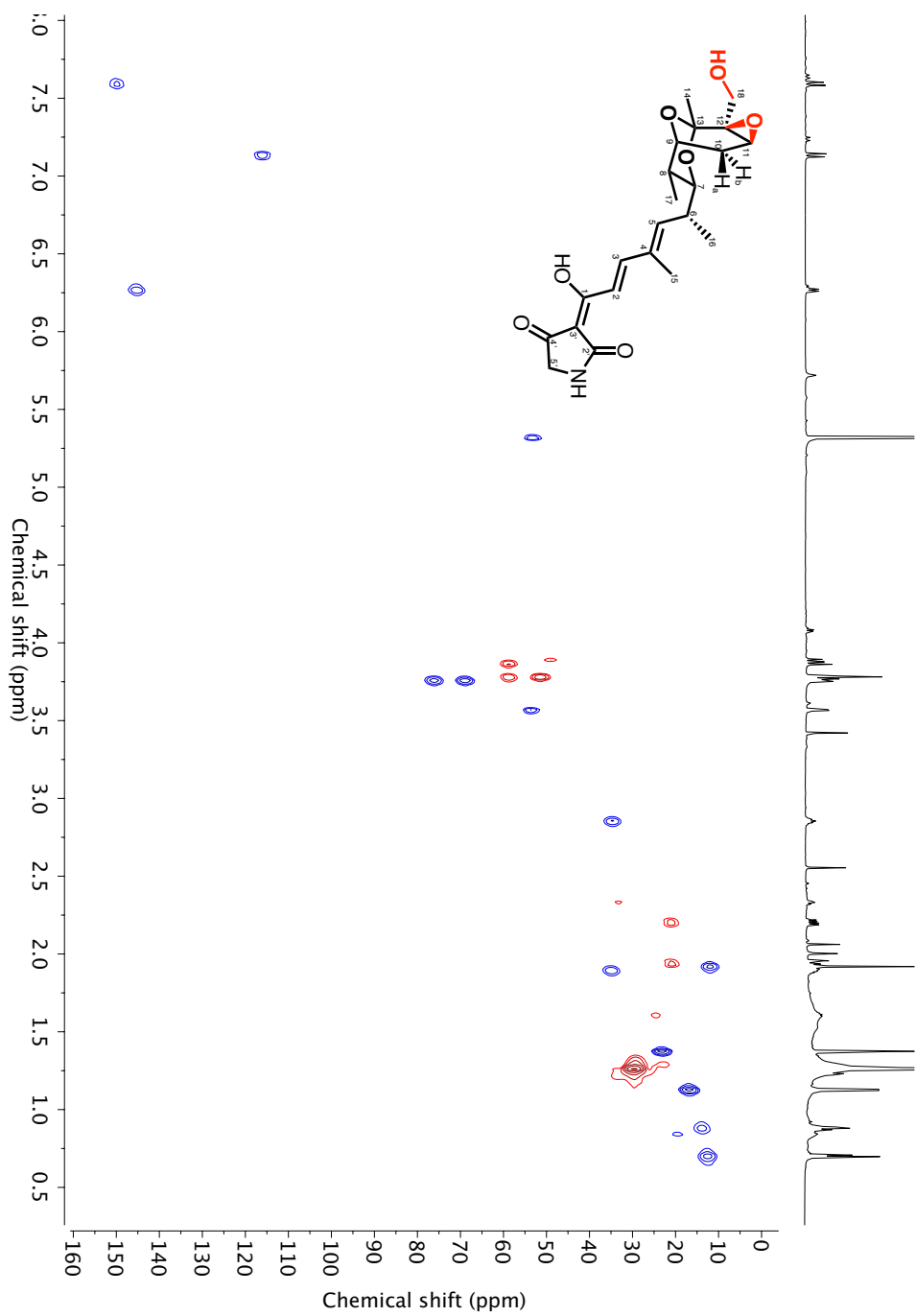


^1H spectrum of tirandamycin N (8) (800 MHz, CD_2Cl_2)

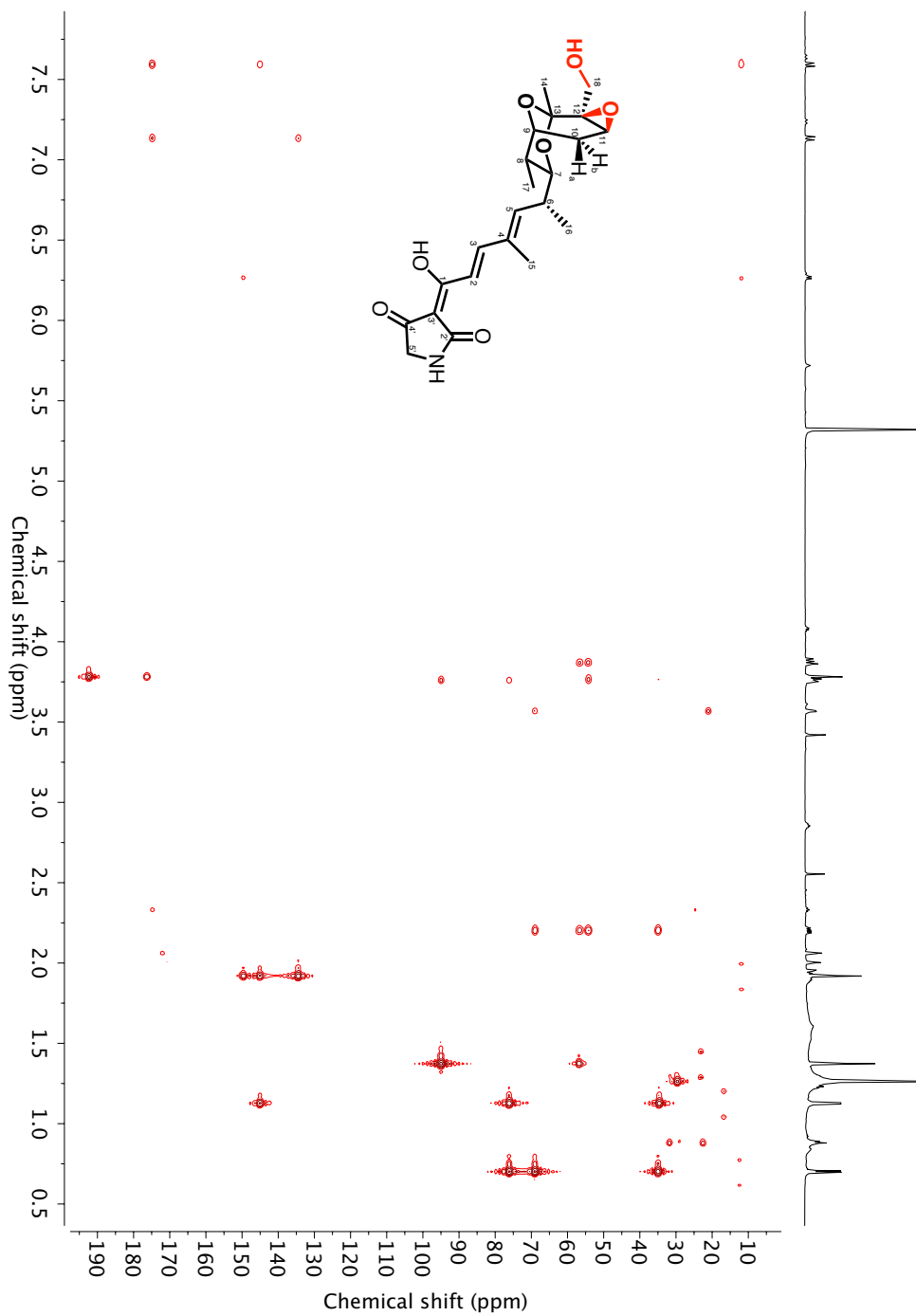
The additional set of doublets in the 7.00 - 8.00 ppm region correspond to the tetramic acid isomer of 8.



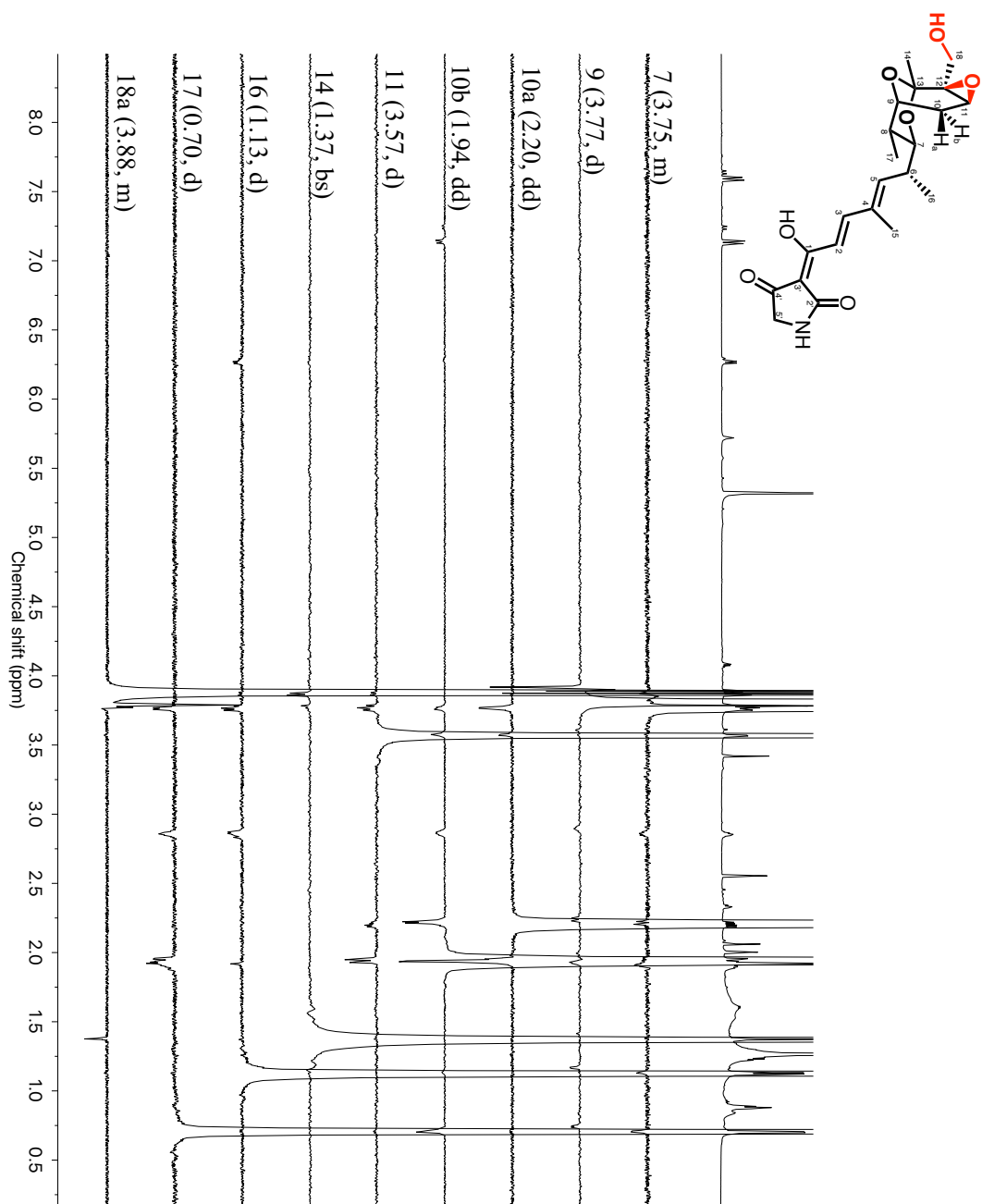
COSY spectrum of tirandamycin N (8) (800 MHz, CD₂Cl₂)



HSQC spectrum of tirandamycin N (8) (800 MHz, CD₂Cl₂)

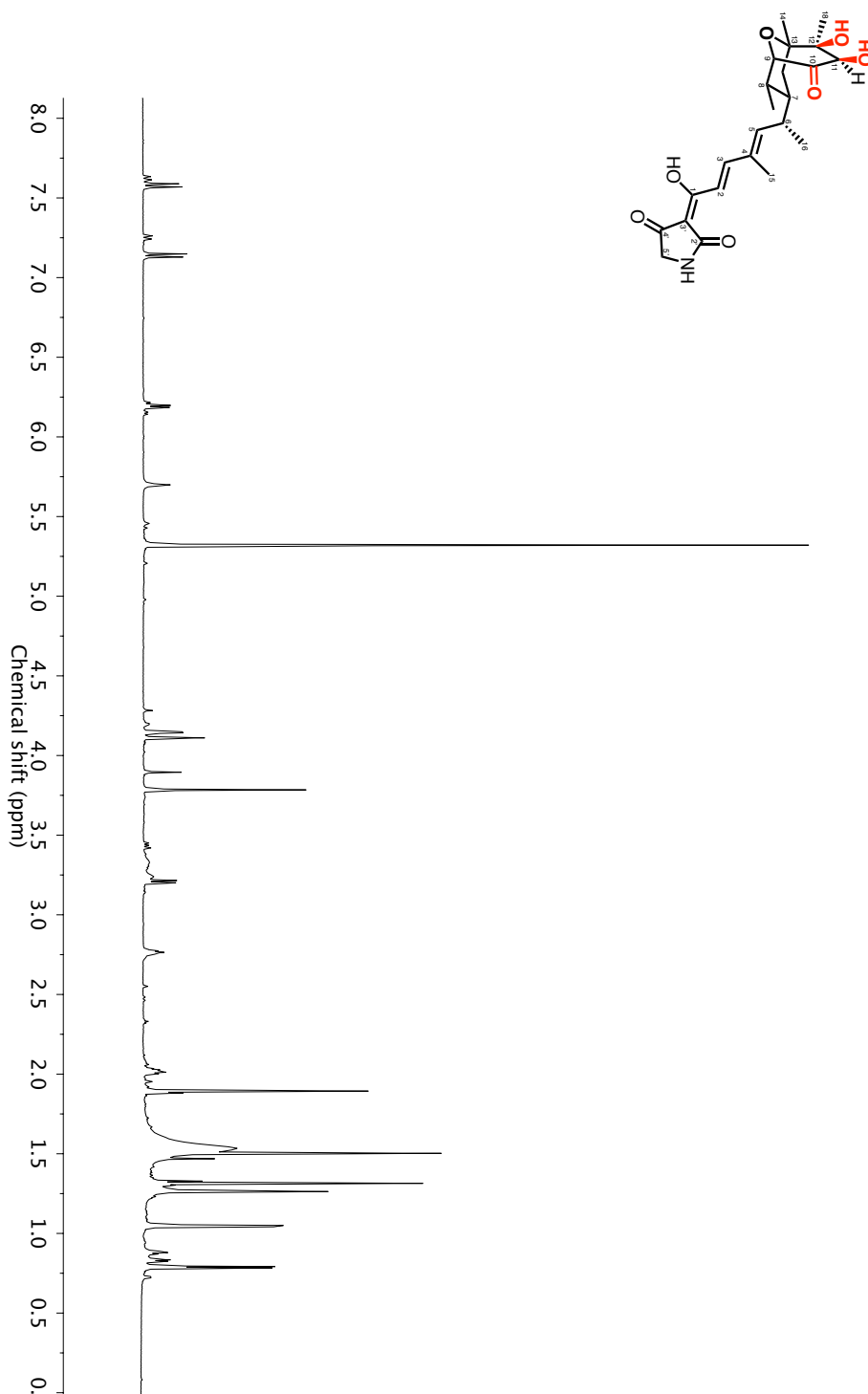


HMBC spectrum of tirandamycin N (8) (800 MHz, CD₂Cl₂)



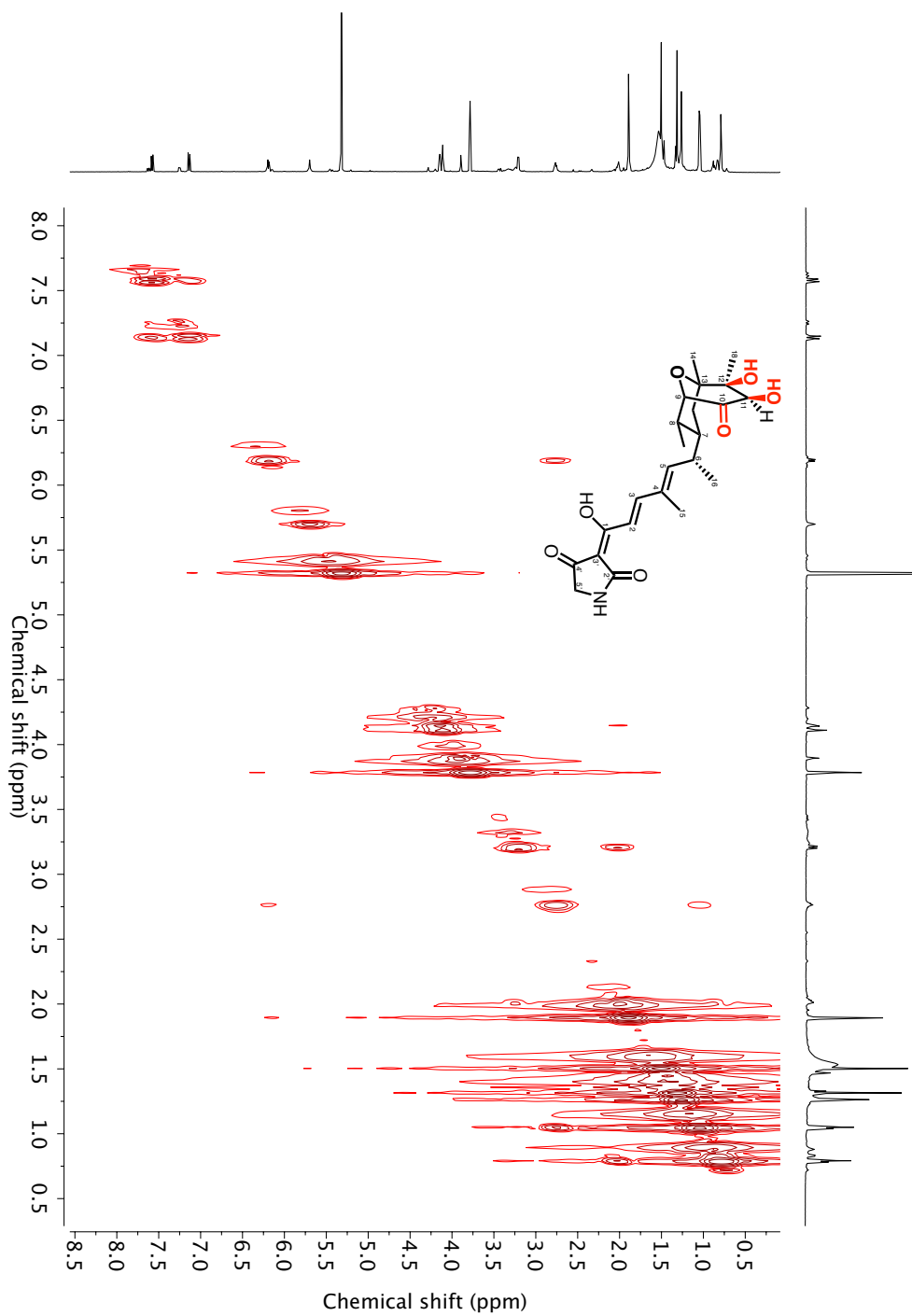
Selective NOE spectra of tirandamycin N (8) (800 MHz, CD₂Cl₂)

Figure A.29 Compound characterization – tirandamycin O (9)

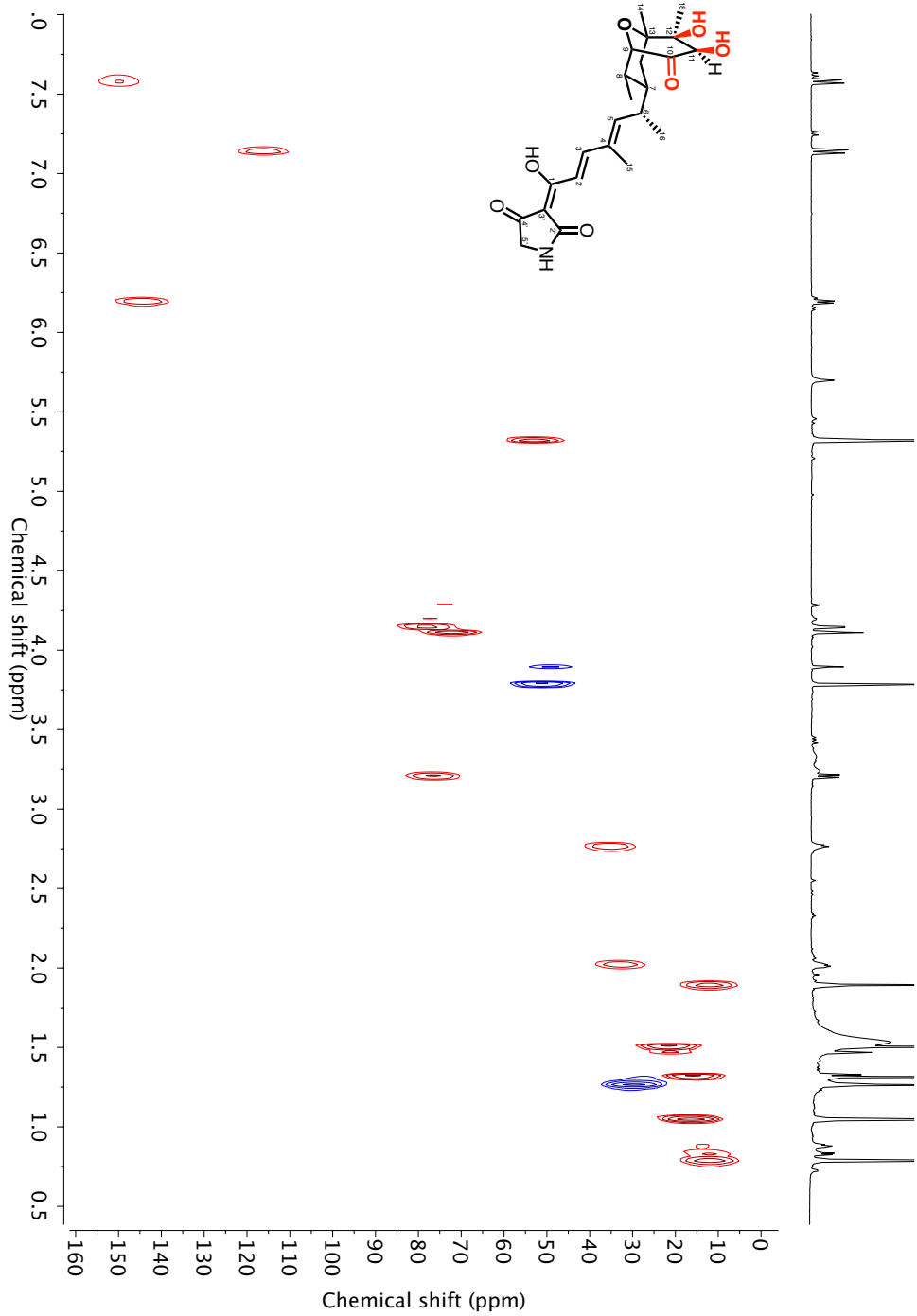


^1H spectrum of tirandamycin O (9) (800 MHz, CD_2Cl_2)

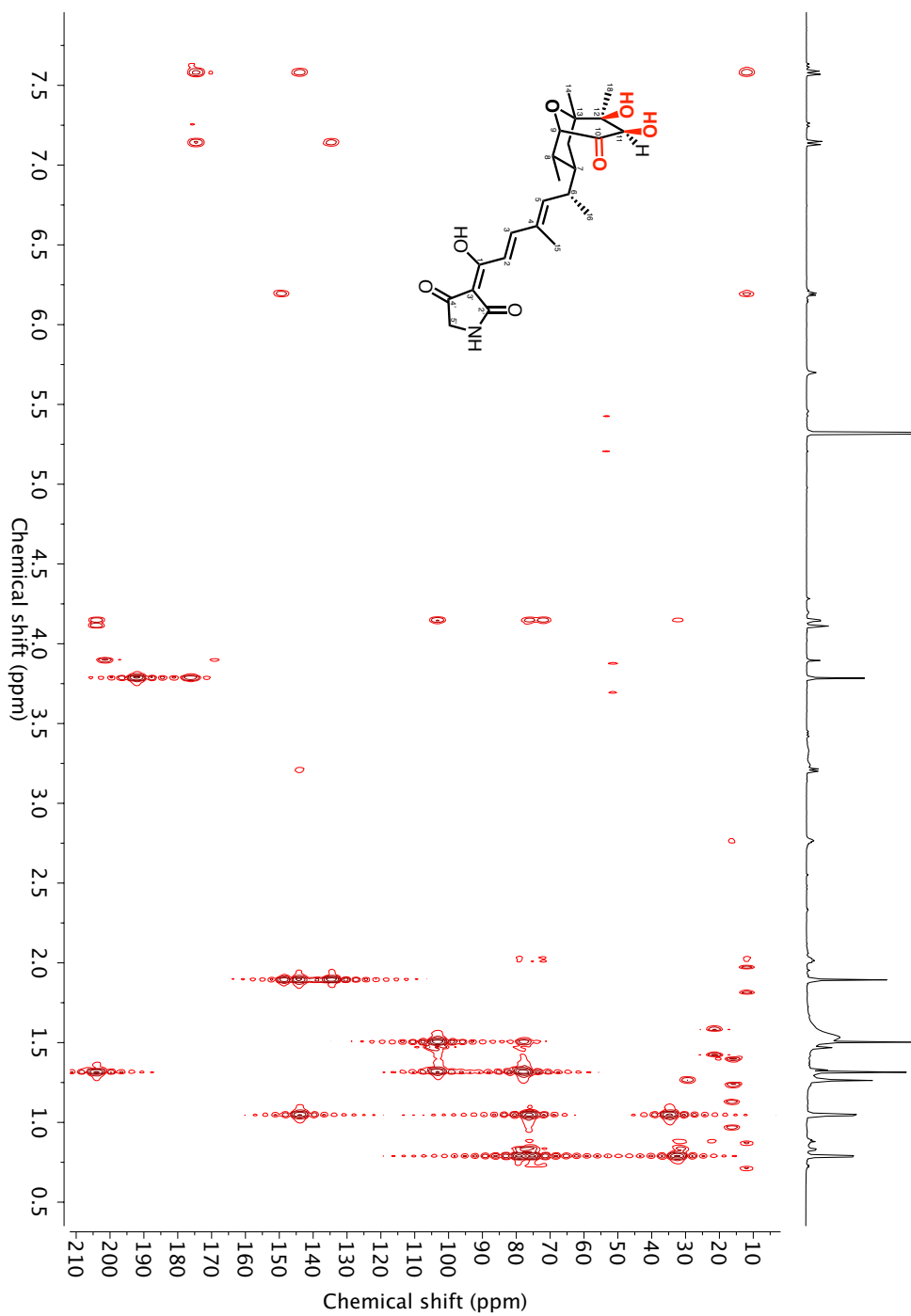
The additional set of doublets in the 7.00 - 8.00 ppm region correspond to the tetramic acid isomer of 9.



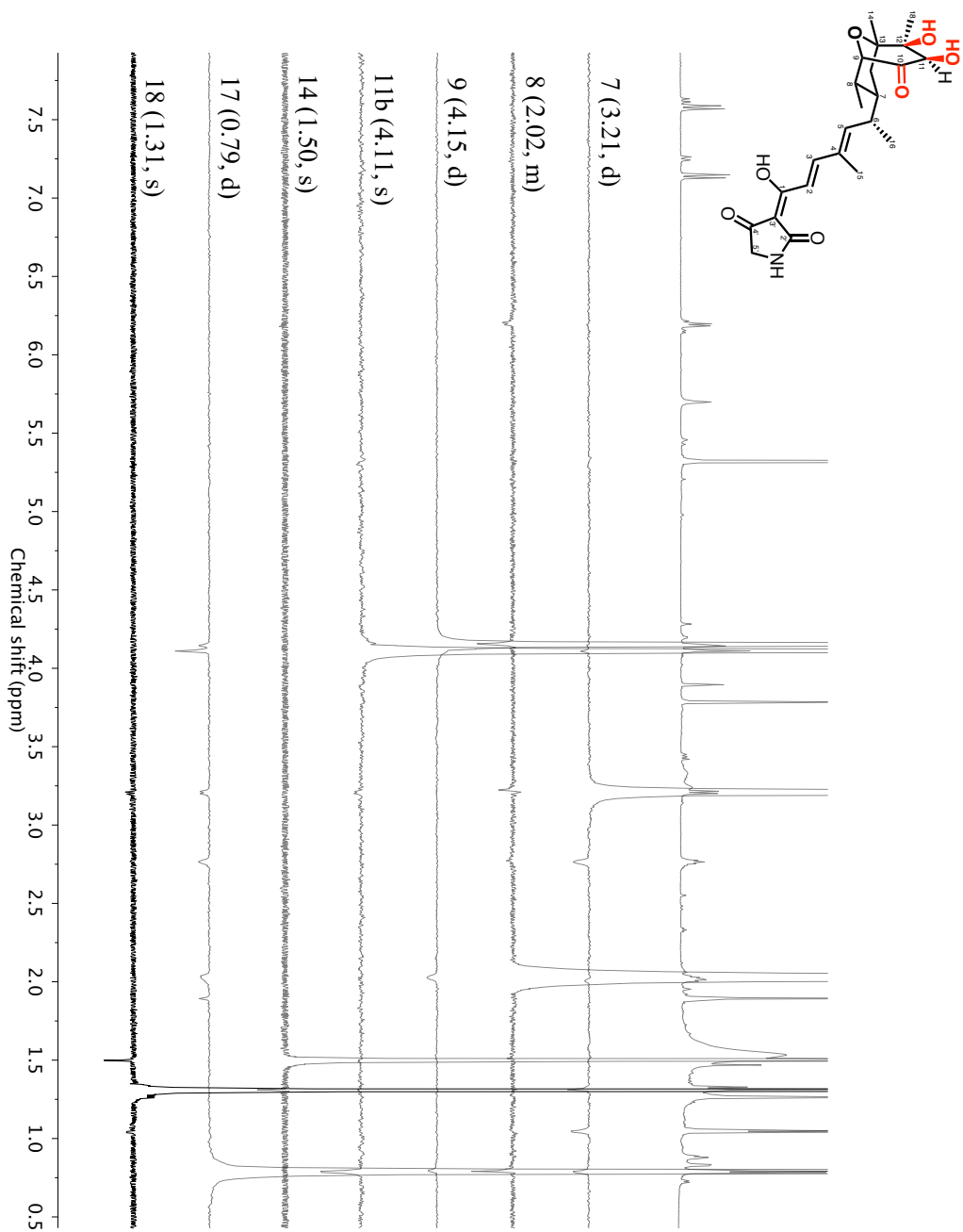
COSY spectrum of tirandamycin O (9) (800 MHz, CD₂Cl₂)



HSQC spectrum of tirandamycin O (9) (800 MHz, CD₂Cl₂)

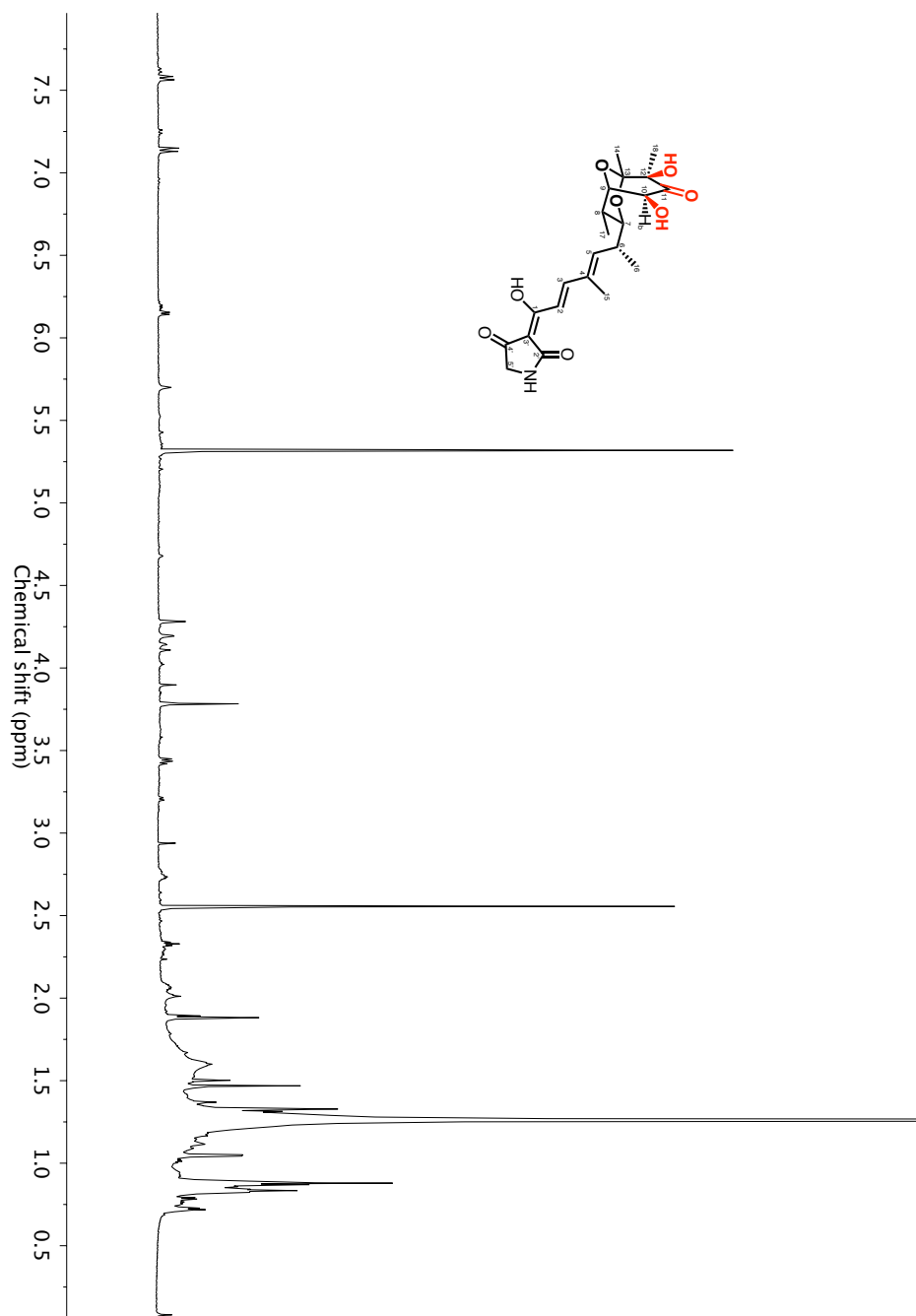


HMBC spectrum of tirandamycin O (9) (800 MHz, CD₂Cl₂)



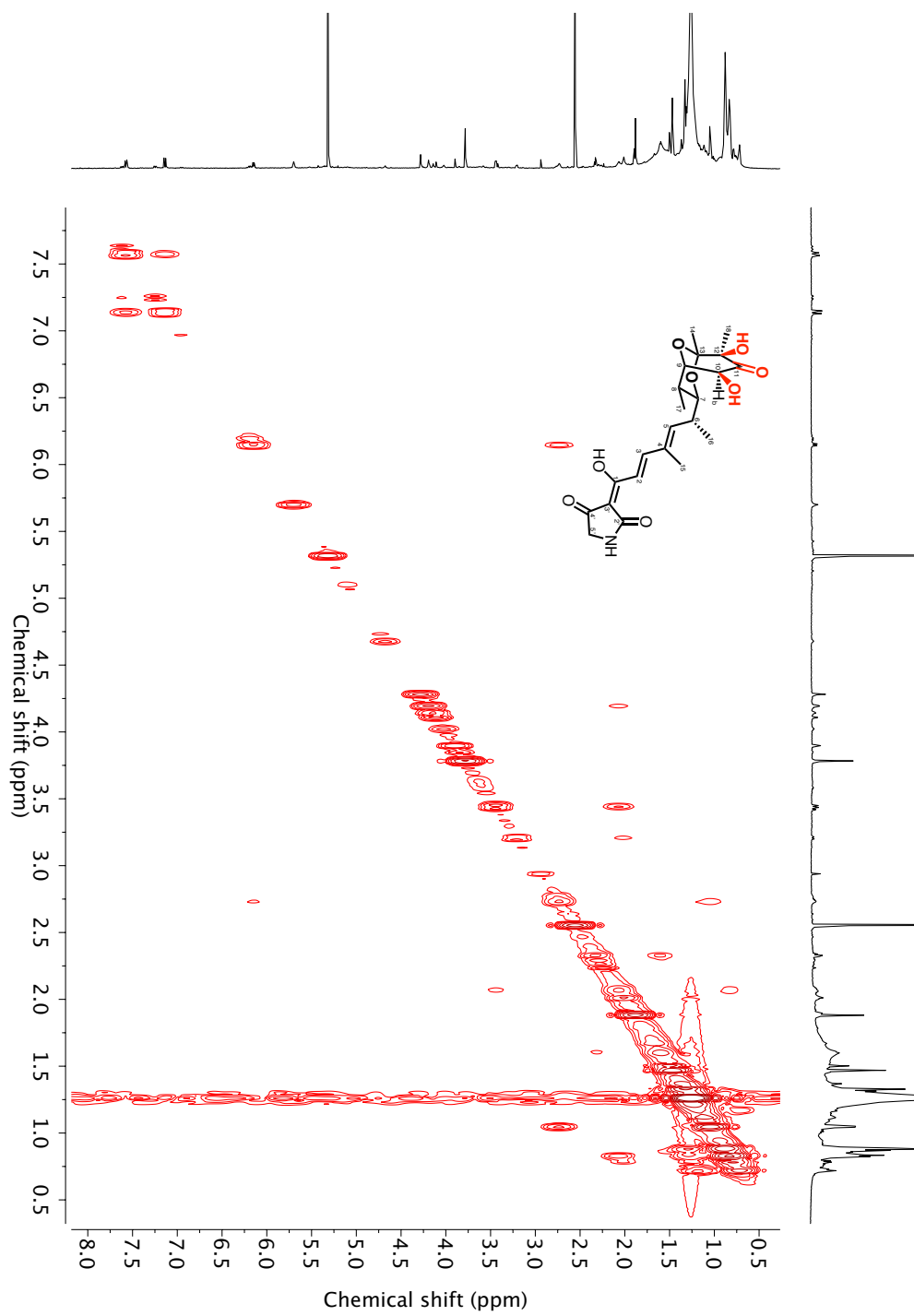
Selective NOE spectra of tirandamycin O (9) (800 MHz, CD₂Cl₂)

Figure A.30 Compound characterization – tirandamycin O' (10)

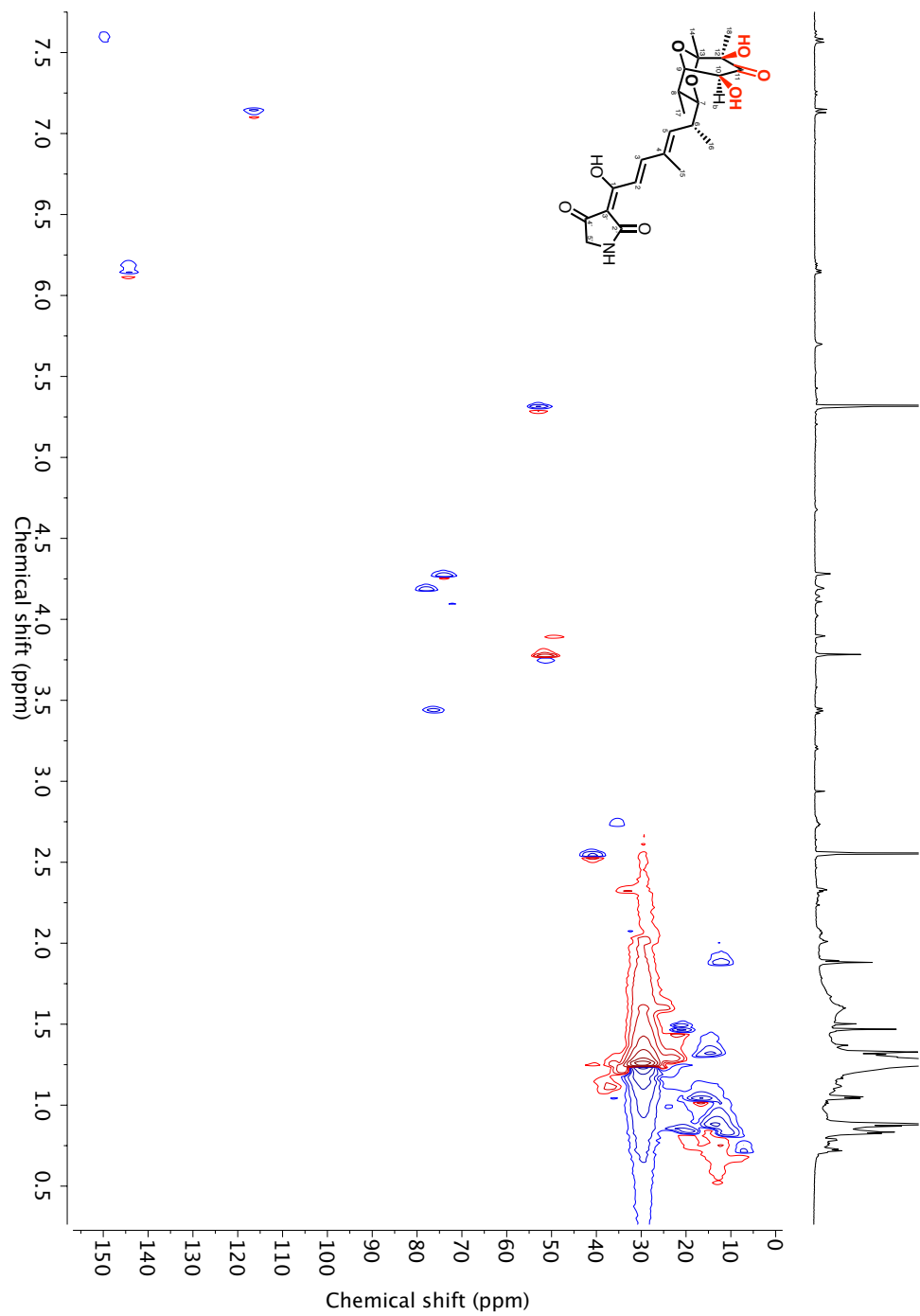


¹H spectrum of tirandamycin O' (10) (800 MHz, CD₂Cl₂)

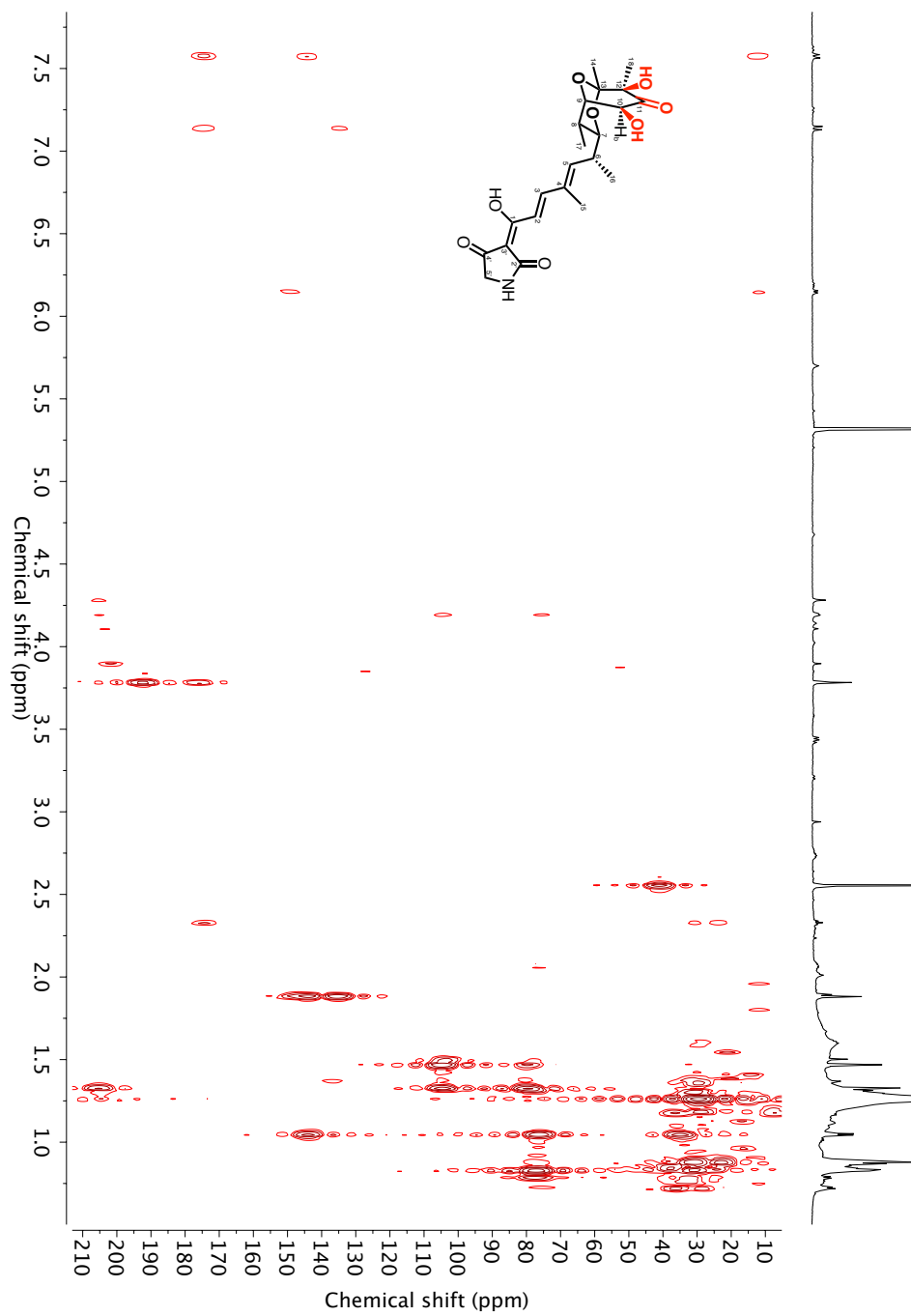
The additional set of doublets in the 7.00 - 8.00 ppm region correspond to the tetramic acid isomer of **10**



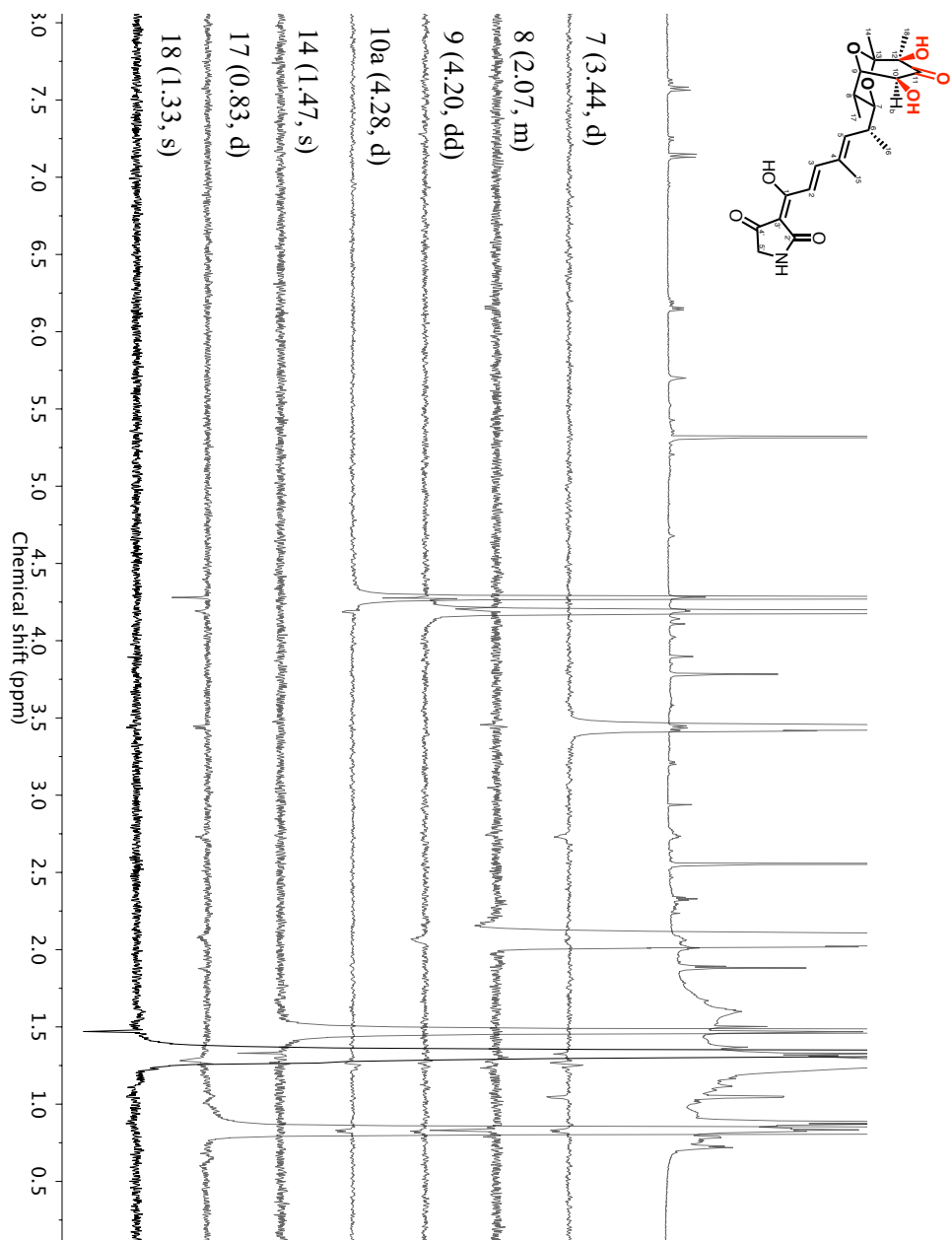
COSY spectrum of tirandamycin O' (10) (800 MHz, CD₂Cl₂)



HSQC spectrum of tirandamycin O' (10) (800 MHz, CD₂Cl₂)



HMBC spectrum of tirandamycin O' (10) (800 MHz, CD₂Cl₂)



Selective NOE spectra of tirandamycin O' (10) (800 MHz, CD₂Cl₂)

A.14 References

1. Carlson, J.C.; Li, S.; Burr, D.A.; Sherman, D.H. Isolation and characterization of tirandamycins from a marine-derived *Streptomyces* sp. *J Nat Prod* **2009**, 72(11):2076-2079.
2. Carlson, J. C.; Li, Shengying; Gunatilleke S. S., Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D.H. Tirandamycin biosynthesis is mediated by co-dependent oxidative enzymes. *Nat Chem* **2011**, 3, 628-633.
3. Omura, T.; Sato, R. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J Biol Chem* **1964**, 239, 2370-2378.
4. Williams, J.W.; Morrison, J.F. The kinetics of reversible tight-binding inhibition. *Methods Enzymol* **1979**, 63, 437-467.
5. Mo, X.; Huang, H.; Ma, J.; Wang, Z.; Wang, B.; Zhang, S.; Zhang, C.; Ju, J. Characterization of Trdl as a 10-hydroxy dehydrogenase and generation of new analogues from a tirandamycin biosynthetic pathway. *Org Lett* **2011**, 13(9), 2212-2215.
6. Hagenmaier, H.; Jaschke, K.H.; Santo, L.; Scheer, M.; Zähler, H. Metabolic products of microorganisms tirandamycin B. *Arch Microbiol* **1976**, 109:65–74.
7. Meyer, C.E. Tirandamycin, a new antibiotic isolation and characterization. *J Antibiot* **1971**, 24:558–560.
8. Mo, X.; Wang, Z.; Wang, B.; Ma, J.; Huang, H.; Tian, X.; Zhang, S.; Zhang, C.; Ju, J. Cloning and characterization of the biosynthetic gene cluster of the bacterial RNA polymerase inhibitor tirandamycin from marine-derived *Streptomyces* sp. SCSIO1666. *Biochem Biophys Res Commun* **2011**, 406(3):341-347.
9. Mo, X.; Ma, J.; Huang, H.; Wang, B.; Song, Y.; Zhang, S.; Zhang, C.; Ju, J. $\Delta 11, 12$ Double bond formation in tirandamycin biosynthesis is atypically catalyzed by TrdE, a glycoside hydrolase family enzyme. *J Am Chem Soc* **2012**, 134(6), pp.2844-2847.
10. Yu, Z.; Vodanovic-Jankovic, S.; Ledebor, N.; Huang, S.X.; Rajska, S.R.; Kron, M.; Shen, B. Tirandamycins from *Streptomyces* sp. 17944 inhibiting the parasite *Brugia Malayi* asparagine tRNA synthetase. *Org Lett* **2011**, 13(8):2034-7.
11. Rateb, M.E.; Yu, Z.; Yan, Y.; Yang, D.; Huang, T.; Vodanovic-Jankovic, S.; Kron, M.A.; Shen, B. Medium optimization of *Streptomyces* sp. 17944 for tirandamycin B production and isolation and structural elucidation of tirandamycins H, I and J. *J Antibiotic Res* **2014**, 67(1):127-32.
12. Cong, Z.; Huang, X.; Liu, Y.; Liu, Y.; Wang, P.; Liao, S.; Yang, B.; Zhou, X.; Huang, D.; Wang, J. Cytotoxic anthracycline and antibacterial tirandamycin analogues from a marine-derived *Streptomyces* sp. SCSIO 41399. *J Antibiot* **2019**, 72(1):45-49.
13. Zhang, X.; Li, Z.; Du, L.; Chlipala, G.E.; Lopez, P.C.; Zhang, W.; Sherman, D.H.; Li, S. Identification of an unexpected shunt pathway product provides new insights into tirandamycin biosynthesis. *Tetrahedron Lett* **2016**, 57(52):5919-5923.
14. Harris, D. L.; Loew, G. H. A role for Thr 252 in cytochrome P450cam oxygen activation. *J Am Chem Soc* **1994**, 11671-11674.
15. Newmister, S. A.; Srivastava, K. R.; Espinoza, R. V.; Caddell Haatveit, K.; Khatri, Y.; Martini, R. M.; Garcia-Borràs, M.; Podust, L. M.; Houk, K. N.; Sherman, D. H. Molecular basis of iterative C–H oxidation by TamI, a multifunctional P450 monooxygenase from the tirandamycin biosynthetic pathway. *ACS Cat* **2020**, 10, 13445-13454.

APPENDIX B:

Experimental Procedures and Supplemental Information for Chapter 3

B.1 Materials and General Methods

Media components, chemical reagents and solvents were purchased from Thermo Fisher Scientific and Sigma-Aldrich unless otherwise stated. Lysozyme was purchased from dot Scientific Inc. Imidazole was purchased from AK Scientific. Kanamycin sulfate and isopropyl- β -D-thiogalactopyranoside (IPTG) were obtained from Gold Biotechnology. Nicotinamide adenine dinucleotide phosphate reduced form (NADPH) was purchased from Chem-Impex Int'l INC. δ -aminolevulinic acid was purchased from Oakwood Chemical. Amicon Ultra centrifugal filters used for protein concentration were from EMD Millipore. PD-10 columns were purchased from GE Healthcare. Deionized water was obtained from a Milli-Q system (EMD Millipore) using QGard 2/Quantum Ex Ultrapure organex cartridges. For media, buffers, and other solutions, pH and pD were measured using a VWR sympHony SB70P pH meter calibrated following manufacturer's specifications. Media solutions were autoclaved before use.

PCR was completed using a Bio-Rad iCycler thermal cycler system and DNA primers for site-directed mutagenesis were obtained from Integrated DNA Technologies. Invitrogen plasmid miniprep kits were purchased from Thermo Fisher Scientific and manufacturer's protocols were followed for all DNA extractions and purifications. A NanoDrop ND-1000 spectrophotometer was used to quantify DNA concentrations. DNA sequencing was performed at Eurofins Genomics. Optical density (OD₆₀₀) was measured using an Eppendorf BioPhotometer. Centrifugations were done using a Beckman Coulter Avanti J-20 XP centrifuge, and sonication was performed using a Fisherbrand Model 705 Sonic Dismembrator.

Reversed-phase high- performance liquid chromatography (RP-HPLC) purification was performed using a Waters XBridge 5 mm C18 column and a solvent system of acetonitrile and H₂O supplemented with 0.1% formic acid (FA) unless stated otherwise. High-resolution ESI-MS spectra was acquired and performed on a quadrupole time-of-flight spectrometer (Agilent Q-TOF 6500 series) using a XBridge™ C18 3.5 μ m 50 mm column with a MeCN and H₂O solvent

system supplemented with formic acid (0.1% v/v). A Shimadzu analytical HPLC was used for analysis of analytical enzymatic reactions using the same column and solvent system.

B.2 Isolation and Purification of Tirandamycin

B.2.1 Streptomyces sp. 307-9 $\Delta tamI$ P450 growth and extraction of tirandamycin C

The culturing and extraction procedures for tirandamycin C were followed as previously described.¹ The mutant strain was maintained on ISP2 agar plates supplemented with 50 $\mu\text{g}/\text{mL}$ apramycin and as spore stocks. Approximately 10-20 mg of pure tirandamycin C were consistently isolated from 1 L of culture.

B.2.2 TamI large-scale enzymatic reactions and isolation of tirandamycin L, O and O'

The enzymatic reactions and isolations of tirandamycin were carried out as previously described.¹ Purified TamI L295A was used as the biocatalyst to produce tirandamycin L, O and O'. When performing the reaction in a large scale (34-mL scale reaction using 500 μM tirandamycin C substrate and 2 μM P450), tirandamycin L accumulated as the main product. When running the reaction at a smaller scale (5-mL scale using 250 μM tirandamycin C substrate and 2 μM P450), tirandamycin O and O' formed as the main products and were isolated as a single peak via HPLC.

B.3 Protein Expression and Purification

B.3.1 TamI protein expression and purification

Expression of TamI and TamI mutants: The pET28b_TamI plasmid previously generated¹ was used to express N-terminally His-tagged TamI. *E. coli* BL21 (DE3) was transformed with the plasmid. Kanamycin (50 $\mu\text{g}/\text{mL}$) was added to 1 L of Terrific Broth (TB) media, which was then inoculated with the transformed *E. coli* cells. The 1 L cultures were grown overnight at 37°C until an OD600 of 0.8-1.0 was reached, cooled in an ice-water bath (10-20 min), induced with 0.4 mM IPTG and 0.4 mM δ -aminolevulinic acid, and expressed for 18-20 hours at 18°C. The cells were harvested by centrifugation and stored at -80°C until used for protein purification.

Protein purification for enzymatic assays and large-scale reactions: Purification was carried out as previously described.¹ Protein purity was assessed by SDS-Page and quantified as previously described.¹

B.4 Enzymatic Assays

B.4.1 TamI variants enzymatic assays

Standard conditions: Reactions were performed *in vitro* in an analytical scale to a final volume of 100 μ L of reaction buffer (50 mM NaH₂PO₄, pH 7.4 at RT) containing 200 μ M substrate, 2 μ M P450, 40 μ M spinach ferredoxin, 6 μ M spinach ferredoxin-NADP⁺ reductase and 2.5 mM NADPH. As negative controls, enzymes were omitted. Reactions were run for 2 h at 30°C, quenched by addition of 100 μ L MeOH and centrifuged for 10 min at 4°C. The supernatant was analyzed by HPLC and LC-MS using 354 nm UV detection, positive/negative ion MS detection when applicable, and the following conditions (A = H₂O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-100% B over 15 min, 100% B for 2 min, 10% B for 2 min; flow rate 0.2 mL/min; injection volume 10 μ L for HPLC and 2 μ L for LC-MS. All reactions were performed and analyzed in duplicate.

B.4.2 TamI enzymatic assays initiated with hydrogen peroxide

Reactions were set up to a final volume of 100 μ L of reaction buffer (50 mM NaH₂PO₄, pH 7.4 at RT) containing 200 μ M substrate, 2 μ M P450 and varying concentrations of H₂O₂ (10, 100 or 200 mM). As negative controls, P450 enzyme was omitted. Reactions were run for 30 min at 30°C (unless otherwise stated), quenched by addition of 100 μ L MeOH and centrifuged for 10 min at 4°C. The supernatant was analyzed by HPLC and LC-MS as described above for TamI mutants' enzymatic assays. Control reactions were set up as described in paragraph B.7.1, using redox partners and NADPH.

B.4.3 TamI enzymatic assays initiated with iodosobenzene

Reactions were set up to a final volume of 100 μ L of reaction buffer (50 mM NaH₂PO₄, pH 7.4 at RT) containing 100 μ M substrate, 10 μ M P450 and 2 mM iodosobenzene (0.1M working solution freshly prepared in MeOH). As negative controls, P450 enzyme was omitted. Reactions were run for 30 min at 30°C (unless otherwise stated), quenched by addition of 100 μ L MeOH and

centrifuged for 10 min at 4°C. The supernatant was analyzed by HPLC and LC-MS as described above for TamI mutants' enzymatic assays. Control reactions were set up as described in paragraph B.4.1, using redox partners and NADPH.

Preparation of iodosobenzene: To commercially available iodosobenzene diacetate (1 g), 3M NaOH (4.65 mL) was added over a 5 min period with stirring. Any solid lumps formed were triturated manually for 15 min. The reaction was stirred for an additional 45 min prior adding 100 mL distilled H₂O with stirring. The resulting solid (crude iodosobenzene) was collected on a Büchner funnel, washed with H₂O, dried extensively by maintaining suction and dissolved in CHCl₃. The pure iodosobenzene was separated by filtration and air-dried prior use.

B.4.4 Measurement of hydrogen peroxide concentration in TamI reactions

TamI reactions were set up as described above for TamI mutants' enzymatic assays, from which 10 µL aliquots of reaction mixture were taken and mixed with 50 µL milliQ water prior quenching with 40 µL chloroform. The mixture was centrifuged for 1 min at 4°C and the aqueous supernatant used for the Amplex® Red Hydrogen Peroxide/Peroxidase Assay (Life Technologies). The manufacturer's protocol was followed to build a standard curve for H₂O₂ using concentrations ranging from 0 to 32 µM in a 96-well plate (Corning Costar). For test reactions, 10 µL TamI reaction aqueous supernatant was added to 40 µL 1X Reaction Buffer and 50 µL of 100 µM Amplex® Red reagent. Next, 0.2 U/mL horseradish peroxxygenase (HRP) was added to each well to initiate reactions. Reactions were incubated at RT for 30 min, protected from the light. Afterwards, the absorbance at 560 nm of each sample well was recorded and used to calculate the concentration of H₂O₂ by fitting into a standard curve. All reactions were performed and analyzed in duplicate.

B.4.5 The effect of Reactive Oxygen Species (ROS) scavengers on TamI reactions

To a TamI reaction containing 1 µM P450, 20 µM ferredoxin, 6 µM ferredoxin reductase and 200 µM tirandamycin substrate, 20 U bovine liver catalase, 2 U SOD from bovine erythrocytes, 20 mM ascorbate and the combination of these three ROS scavengers were added in a 50 µL reaction mixture. Reactions were initiated with 2.5 mM NADPH and incubated for 2 h at 30 °C, prior quenching with 50 µL methanol. The reaction supernatants were analyzed by HPLC using 354 nm UV detection. Dose responses were evaluated using 0, 1, 5, 10 and 100 U catalase; 0, 0.05, 0.5, 1 and 10 U SOD; 0, 1, 5, 10 and 100 mM ascorbate in parallel reactions. The total

product formation was determined by obtaining the areas under the curve of HPLC peaks corresponding to product formation. Standard curves were run simultaneously in duplicate. HPLC conditions were as follows (A = H₂O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-85% B over 15 min, 100% B for 4 min, 10% B for 2 min; flow rate 0.2 mL/min; injection volume 10 μ L. The dose response curves were drawn with the overall conversion to product under the different concentration of ROS scavengers. All reactions were performed and analyzed in duplicate.

B.4.6 Steady-State Kinetic Solvent Isotope Effects (KSIE)

Reactions were performed in a H₂O-based buffer consisting of 50 mM NaH₂PO₄, 9% glycerol (to replicate the viscosity of D₂O) with pH 7.4 at RT or a D₂O-based buffer comprising the same components minus glycerol with pD 7.4 at RT. The D₂O-based buffer was prepared as follows: NaH₂PO₄ was dissolved in 99.9% D₂O to a final 50 mM concentration and incubated overnight. The solution was evaporated to dryness using a Biotage® V-10 Touch. This process was repeated a total of three times to obtain a reaction buffer with 99.9% D₂O. The buffer was adjusted to pD 7.4 (pD = pH meter reading + 0.4 to correct for the acidity of the pH electrode) using sodium deuterioxide.³ All enzymes were exchanged into the D₂O-based buffer a total of five times using centrifugal filters (Amicon Ultra).

Reactions were set up in a final volume of 200 μ L following the conditions described below. Reactions were pre-incubated at 30°C for 5 min and initiated by addition of 2.5 mM NADPH. Fixed substrate concentrations were used. Reactions were quenched with methanol by removing 25 μ L at 6 different time points and dispensing the aliquot into 25 μ L of methanol. Quenched reactions were centrifuged for 10 min at 4°C and the supernatant analyzed by HPLC using 354 nm UV detection and the following conditions: (A = H₂O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-85% B over 15 min, 100% B for 4 min, 10% B for 2 min; flow rate 0.2 mL/min; injection volume 10 μ L. GraphPad Prism software was used to plot the product concentration against time to obtain the initial velocities (V_0). The K_h (in H₂O-based buffer) and K_d (in D₂O-based buffer) values were obtained by dividing V_0 by enzyme concentration. The SKIE values were calculated using Equation 2. All reactions were performed and analyzed in duplicate. Standard curves of products were run simultaneously and fitted to a linear curve.

Equation 2

$$SKIE: \frac{K_H (\mu\text{M prod}/\mu\text{M P450}/\text{min})}{K_D (\mu\text{M prod}/\mu\text{M P450}/\text{min})}$$

For reactions with TamI WT and tirandamycin C substrate:

0.05 μM P450, 1 μM ferredoxin, 0.15 μM ferredoxin reductase, 32 μM substrate, 2.5 mM NADP

For reactions with TamI T252A_T253A and tirandamycin C substrate:

0.1 μM P450, 2 μM ferredoxin, 0.3 μM ferredoxin reductase, 64 μM substrate, 2.5 mM NADPH

For reactions with TamI L295A and tirandamycin C substrate:

0.5 μM P450, 10 μM ferredoxin, 1.5 μM ferredoxin reductase, 96 μM substrate, 2.5 mM NADP

For reactions with TamI T252A_T253A_L295A and tirandamycin C substrate:

0.5 μM P450, 10 μM ferredoxin, 1.5 μM ferredoxin reductase, 160 μM substrate, 2.5 mM NADPH

For reactions with TamI L295A and tirandamycin L substrate:

0.4 μM P450, 8 μM ferredoxin, 1.2 μM ferredoxin reductase, 60 μM substrate, 2.5 mM NADPH

For reactions with TamI T252A_T253A_L295A and tirandamycin L substrate:

0.4 μM P450, 8 μM ferredoxin, 1.2 μM ferredoxin reductase, 60 μM substrate, 2.5 mM NADPH

B.4.7 Determination of NADPH consumption rate, turnover frequency (TOF) and coupling efficiency (%)

Analytical enzymatic reactions were carried out in a clear Greiner 384-deep well plate following the conditions described below. All components (except for NADPH) were mixed together in a total volume of 50 μL . Reactions were incubated for 5 min at RT prior initiation with 500 μM NADPH using a multichannel pipette. The absorbance at 340 nm was monitored over 10 min (15 s intervals) using a SpectraMax M5 Microplate Reader (Molecular Devices). The plate-reading chamber was kept at 30 $^{\circ}\text{C}$ for the duration of the experiment. Raw absorbance values were converted to NADPH concentration using $\epsilon_{340} = 6.22 \text{ mM}^{-1}\text{cm}^{-1}$ and pathlength = 0.3542 cm. Initial velocity rates of NADPH consumption were obtained by plotting NADPH concentration as a function of time within the linear range (5-10 min). Normalized rates were

calculated by subtracting the measured value of the reaction buffer and dividing the resulting numbers by the enzyme concentration. The data was fitted to a linear equation from which the initial velocity rates (V_o) for NADPH consumption per minute were obtained. All reactions were performed and analyzed in duplicate. After the absorbance values were recorded, reactions were immediately quenched with 50 μ L methanol and centrifuged for 10 min at 4°C. The supernatant was analyzed via HPLC using the following conditions (A = H₂O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-85% B over 15 min, 100% B for 4 min, 10% B for 2 min; flow rate 0.2 mL/min; injection volume 10 μ L. TOF values were determined by evaluating the areas of HPLC peaks corresponding to products formed. Standard curves were run simultaneously and fitted to a linear curve. The resulting values were divided by enzyme concentration and time of reaction in minutes. The coupling efficiency for each enzyme and substrate combination was calculated using Equation 1.

Equation 1

$$\text{coupling efficiency (\%)}: \left(\frac{\text{TOF } (\mu\text{M product formation}/\mu\text{M P450}/\text{min})}{\text{NADPH consumption rate } (\mu\text{M NADPH consumed}/\mu\text{M P450}/\text{min})} \right) * 100$$

For reactions with TamI WT and tirandamycin C substrate:

0.05 μ M P450, 0.5 μ M ferredoxin, 0.1 μ M ferredoxin reductase, 32 μ M substrate, 500 μ M NADPH

For reactions with TamI T252A_T253A and tirandamycin C substrate:

0.1 μ M P450, 1 μ M ferredoxin, 0.2 μ M ferredoxin reductase, 64 μ M substrate, 500 μ M NADPH

For reactions with TamI L295A and tirandamycin C substrate:

0.5 μ M P450, 5 μ M ferredoxin, 1 μ M ferredoxin reductase, 96 μ M substrate, 500 μ M NADPH

For reactions with TamI T252A_T253A_L295A and tirandamycin C substrate:

0.5 μ M P450, 5 μ M ferredoxin, 1 μ M ferredoxin reductase, 160 μ M substrate, 500 μ M NADPH

For reactions with TamI L295A and tirandamycin L substrate:

0.4 μ M P450, 4 μ M ferredoxin, 0.8 μ M ferredoxin reductase, 60 μ M substrate, 500 μ M NADPH

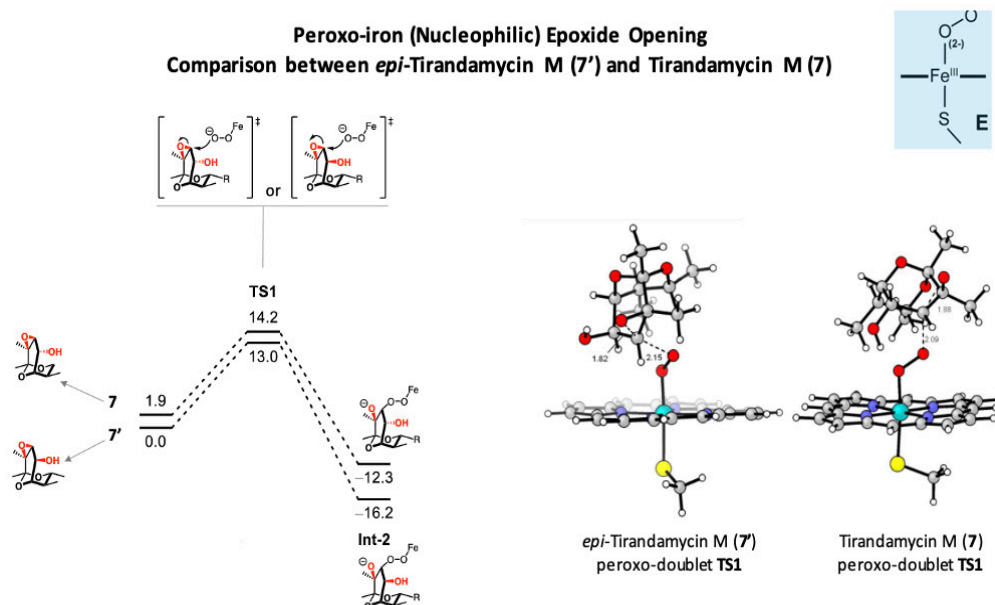
For reactions with TamI T252A_T253A_L295A and tirandamycin L substrate:

0.4 μM P450, 4 μM ferredoxin, 0.8 μM ferredoxin reductase, 60 μM substrate, 500 μM NADPH

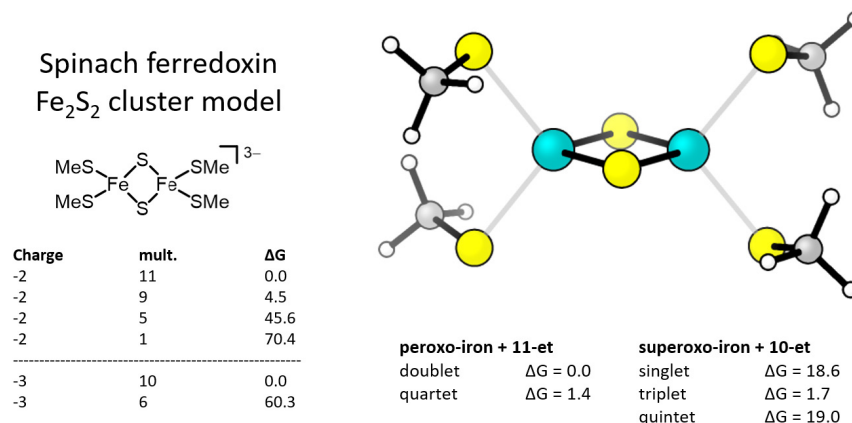
B.5 Computational Methods

The coupling efficiency All geometry optimization calculations were conducted with Gaussian 16⁴ with the uB3LYP-D3(BJ) functional⁵⁻⁸ and def2-SVP basis set⁹ with the SMD solvation model for diethylether.¹⁰ Frequency calculations were carried out in Gaussian 16 at the same level of theory as geometry optimization to verify stationary points. Single point calculations were computed in Gaussian 16 with the uB3LYP-D3(BJ) functional and def2-TZVPP basis set⁹ with the SMD solvation model for diethylether. Conformational searches and preoptimizations were conducted with xTB/CREST.¹¹⁻¹² Quasiharmonic corrections to entropy and enthalpy were computed with GoodVibes at 1 M and 298.15 K.¹³ Visualizations of optimized structures were generated using CylView¹⁴ and PyMOL.¹⁵ Ferredoxin models were constructed per literature precedent from Presti et al.¹⁶

B.6 Supplemental Schemes

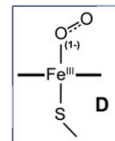
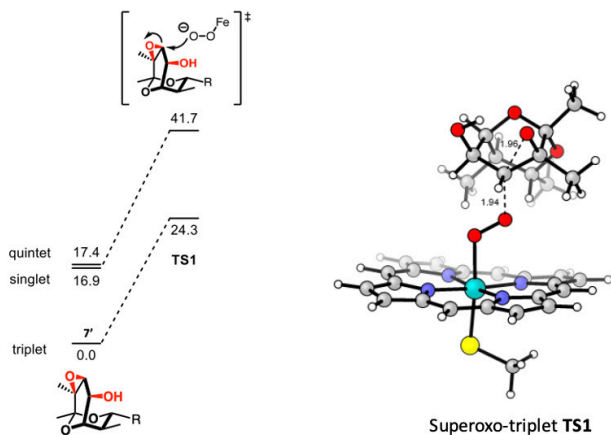


Scheme B.1. DFT calculations on P450 TamI (peroxo-iron, shown as species E in Scheme 3.1) and intermediates 7 and 7'. Energy values are in kcal/mol. Charge of complex for peroxo is -2. SP//OPT: B3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//B3LYP-D3(BJ)/def2-SVP, SMD(diethylether).



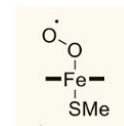
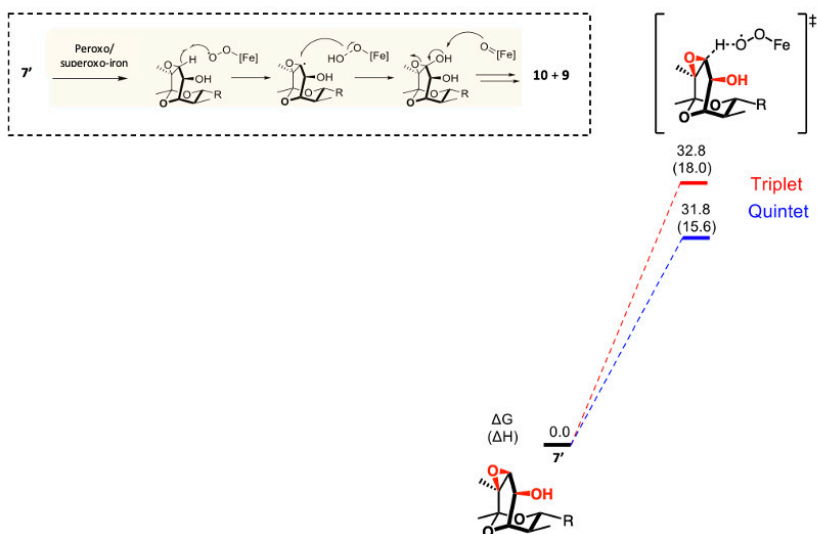
Scheme B.2. DFT calculations on ferredoxin $[\text{Fe}_2\text{S}_2]$ cluster model, showing charge -3/multiplicity 10 as representative structure. Structures of multiplicity 1, 5, 6, 9, 10, 11 were found to converge. Structures of multiplicity 2, 3, 4, 7, 8 did not converge. Qualitative agreement was found with Presti et al. that the 10-et and 11-et were lowest in energy of their respective charge series. Pairing 10-et with superoxo-iron complex and 11-et with peroxo-iron complex permitted quantitative comparison. The lowest superoxo-iron complex (triplet) was found to be 1.7 kcal/mol higher in energy than the lowest peroxo-iron complex (doublet). Free energies (ΔG) are reported in kcal/mol; all -3 complexes are relative to the 10-et, all -2 complexes are relative to the 11-et. SP//OPT: uB3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//uB3LYP-D3(BJ)/def2-SVP, SMD(diethylether).

Superoxo-Iron Epoxide Opening, *epi*-Tirandamycin M (7')



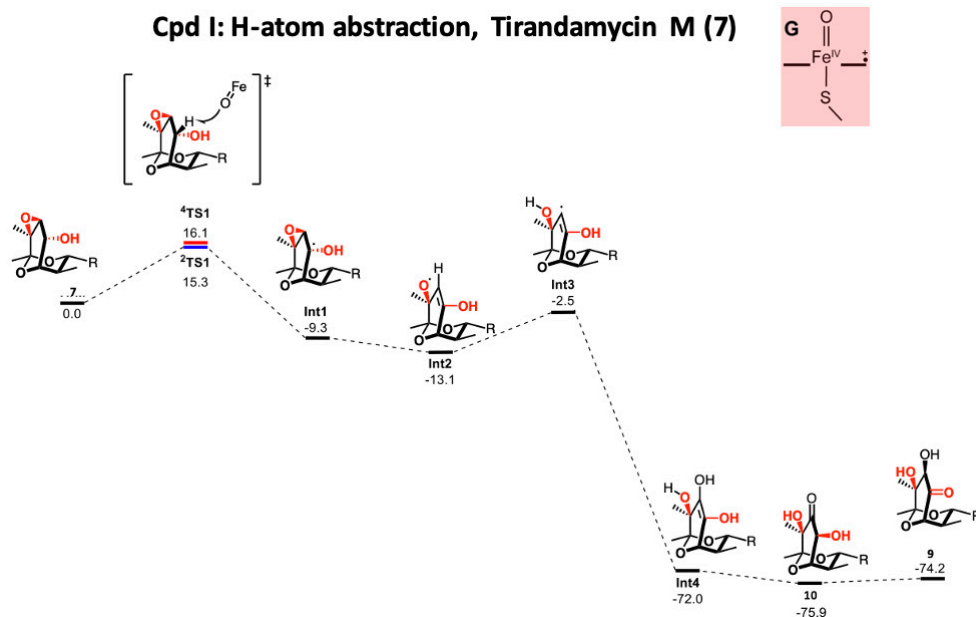
Scheme B.3. DFT calculations on P450 TamI (superoxo-iron, shown as species D in Scheme 3.1) and intermediate 7'. Free energies (ΔG) are reported in kcal/mol. Transition structure searches for the quintet were unsuccessful. Complex has charge -1 . SP//OPT: uB3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//uB3LYP-D3(BJ)/def2-SVP, SMD(diethylether).

Ferric superoxo H-atom abstraction, *epi*-Tirandamycin M (7')



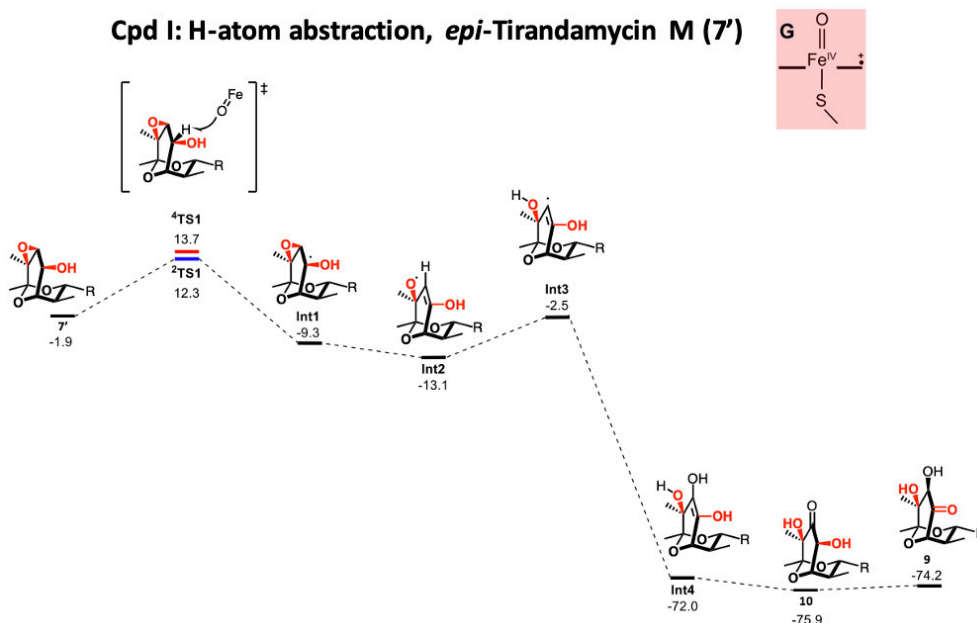
Scheme B.4. DFT calculations on P450 TamI heme model (superoxo-iron, shown as species D in Scheme 3.1) and intermediate 7'. Barriers for initial hydrogen atom abstraction were found to be prohibitively high. Peroxo structures (species E, multiplicities 2/4) were also investigated but no converged transition structure geometries could be found. Free energies (ΔG) are reported in kcal/mol. Complex has charge -1 . SP//OPT: uB3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//uB3LYP-D3(BJ)/def2-SVP, SMD(diethylether).

Cpd I: H-atom abstraction, Tirandamycin M (7)



Scheme B.5. DFT calculations on P450 TamI heme model (Cpd I, shown as species G in Scheme 3.1) and intermediate 7. Energy values are in kcal/mol. Charge of complex for superperoxo is -1. SP//OPT: B3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//B3LYP-D3(BJ)/def2-SVP, SMD(diethylether).

Cpd I: H-atom abstraction, *epi*-Tirandamycin M (7')



Scheme B.6. DFT calculations on P450 TamI heme model (Cpd I, shown as species G in Scheme 3.1) and intermediate 7'. Free energies (ΔG) are reported in kcal/mol. SP//OPT: uB3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//uB3LYP-D3(BJ)/def2-SVP, SMD(diethylether).

B.7 Supplemental Tables

P450 TamI	Tirandamycin C	Tirandamycin L
WT	0.17 ± 0.02	X
T252A	15.74 ± 2.10	X
T252A_T253A	12.55 ± 1.73	X
L295A	NS	NS
T252A_L295A	4.07 ± 0.42	NS
T252A_T253A_L295A	1.28 ± 0.18	NS
E251Q_L295A	NS	NS

Table B.1. Substrate binding analysis of P450 TamI with different tirandamycin substrates. Equilibrium dissociation constant values (K_d) represented in μM . Reported errors are standard deviations calculated from experiments performed in duplicate. NS refers to no apparent shift observed in differential spectrum upon substrate titration to P450, even at substrate saturation levels. The binding affinity of TamI L295A with tirandamycin M (7) was also measured but no spin shift (NS) was observed.

P450 TamI	Tirandamycin L (6)					
	TOF ($\mu\text{M prod}/\text{min}/\mu\text{M P450}$)		Coupling efficiency (%)		NADPH consumption ($\mu\text{M consumed}/\text{min}/\mu\text{M P450}$)	
	H ₂ O	D ₂ O	H ₂ O	D ₂ O	H ₂ O	D ₂ O
L295A	1.55 ± 0.31	1.91 ± 0.22	13.1	20.2	11.8 ± 1.0	9.5 ± 0.4
T252A_T253A_L295A	2.10 ± 0.60	0.85 ± 0.10	9.4	3.5	23.9 ± 1.3	24.3 ± 2.0

P450 TamI	Tirandamycin C (1)					
	TOF ($\mu\text{M prod}/\text{min}/\mu\text{M P450}$)		Coupling efficiency (%)		NADPH consumption ($\mu\text{M consumed}/\text{min}/\mu\text{M P450}$)	
	H ₂ O	D ₂ O	H ₂ O	D ₂ O	H ₂ O	D ₂ O
WT	5.32 ± 0.05	4.68 ± 0.05	14.9	14.0	35.7 ± 1.9	33.4 ± 4.0
T252A_T253A	4.61 ± 0.16	2.62 ± 0.16	4.6	3.2	99.6 ± 7.5	82.2 ± 3.4
L295A	1.52 ± 0.03	1.76 ± 0.03	13.5	19.9	11.3 ± 1.0	8.9 ± 0.8
T252A_T253A_L295A	2.88 ± 0.12	1.79 ± 0.12	8.8	6.5	32.9 ± 3.9	27.6 ± 3.3

Table B.2. Determination of turnover frequency (TOF), coupling efficiency (%) and NADPH consumption rate. Coupling efficiency was calculated by dividing TOF over NADPH consumption rate values as described in Methods. Experiments were run in duplicate.

B.8 Supplemental Figures

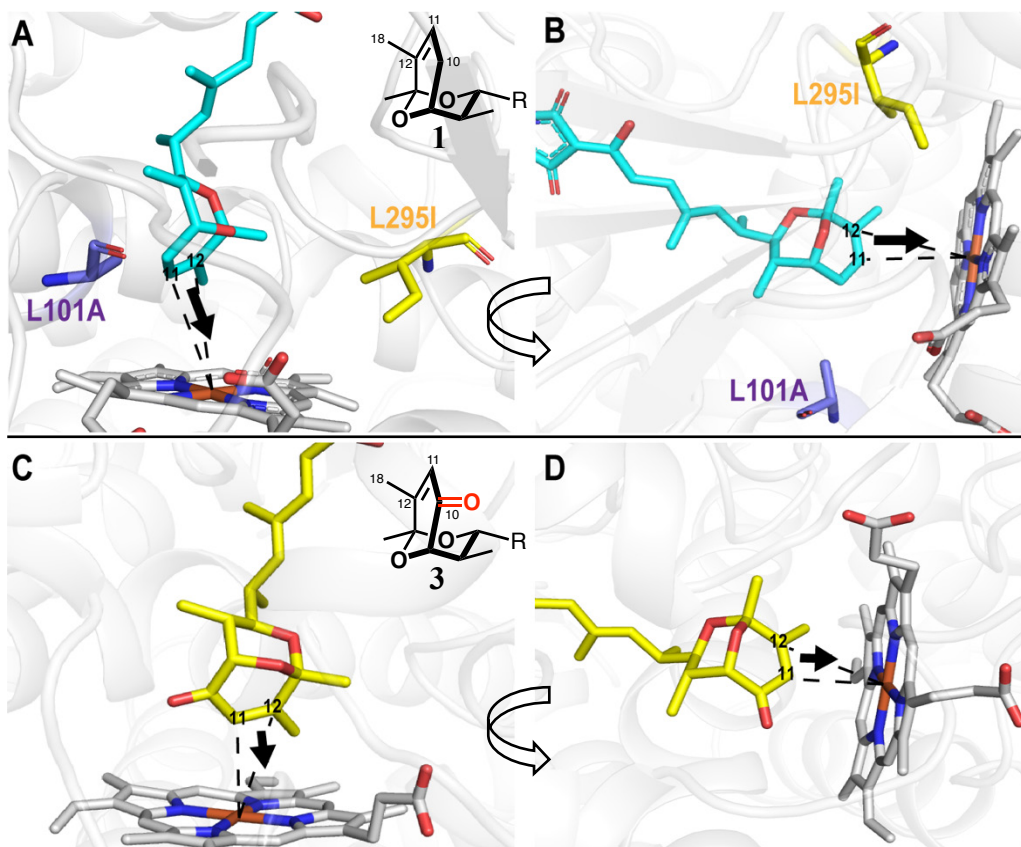
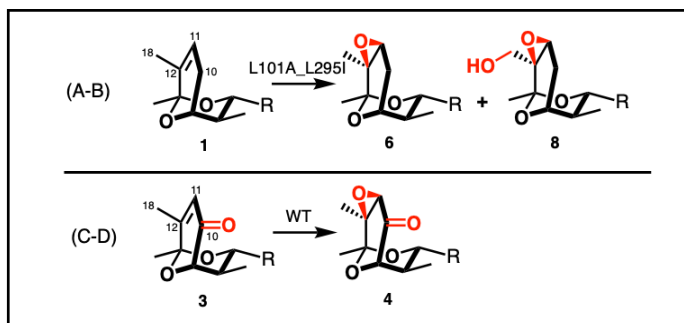


Figure B.1. Molecular Dynamics simulations of TamI variants and tirandamycin congeners where C11/12 alkene epoxidation takes place. (A-B) Two different views of variant TamI L101A_L295I in complex with **1** (light blue) emphasizing the orientation of the C11/12 alkene in the substrate towards the heme iron (light grey). Black arrows show the orientation of the C11/12 π -orbitals. Enzymatic assays show conversion of **1** \rightarrow **6** (epoxide-containing) and **8** with this double mutant. (C-D) Two different views of TamI WT in complex with **3** (yellow). Enzymatic assays show conversion of **3** \rightarrow **4** (epoxide-containing) with the WT enzyme.

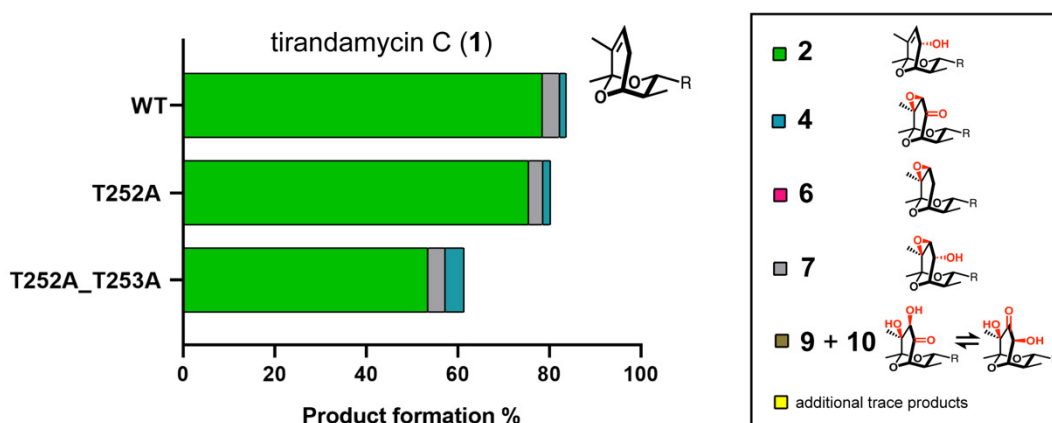


Figure B.2. Additional end-point assays with TamI mutants and 1. Mutants T252A and T252A_T253A were designed to test the effect of interrupting the conserved alcohol residue in the P450 TamI without the L295A variation present. Ratios of product formation were estimated based on HPLC traces. Compared to WT, similar conversion is observed for T252A while the double mutant T252A_T253A displays diminished yields. The data could suggest that the presumed accumulated Cpd 0 species may be a viable oxidant for the oxidative reactions observed, however, previous literature demonstrates Cpd 0 is a poor electrophilic oxidant for hydroxylation reactions compared to Cpd I. These results indicate that the conserved alcohol is not essential for TamI catalysis related to allylic hydroxylation.

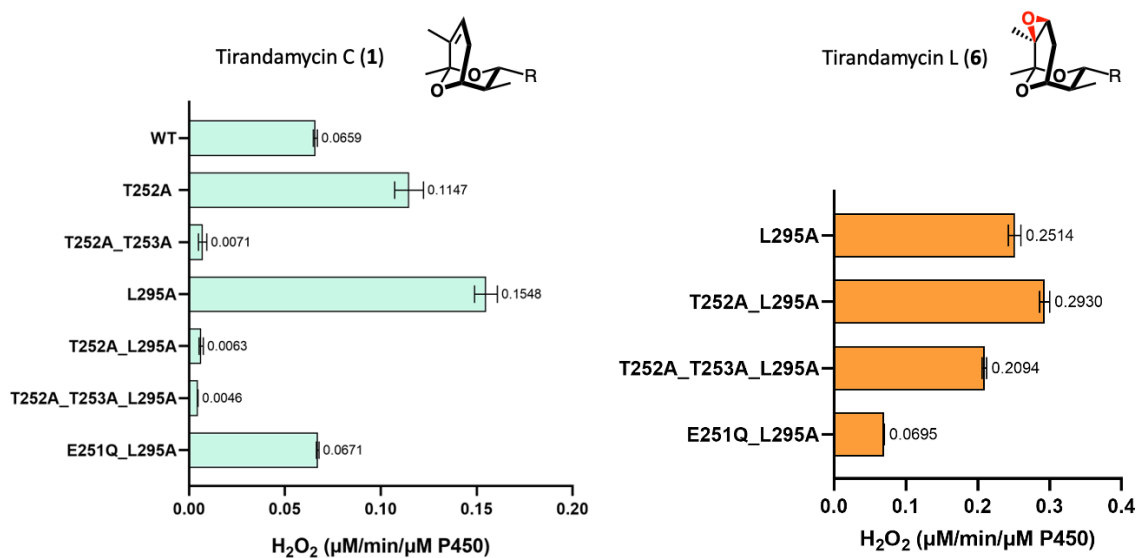
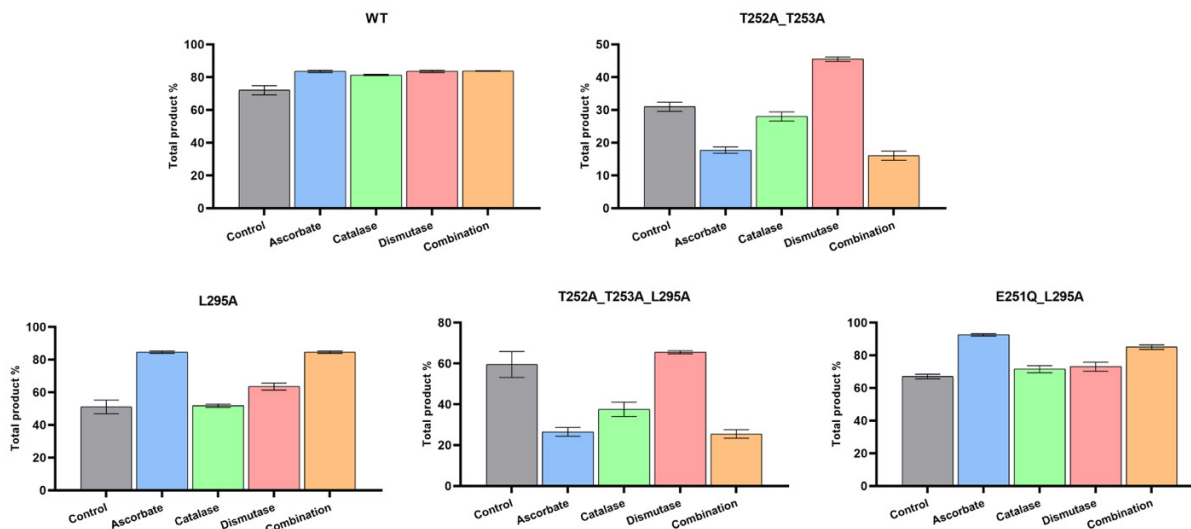


Figure B.3. Hydrogen peroxide formation in TamI-catalyzed reactions. Exact values are shown to the right of bar graphs. Substrates are displayed above each graph. Experiments were performed in duplicate.

Tirandamycin C (1)



Tirandamycin L (6)

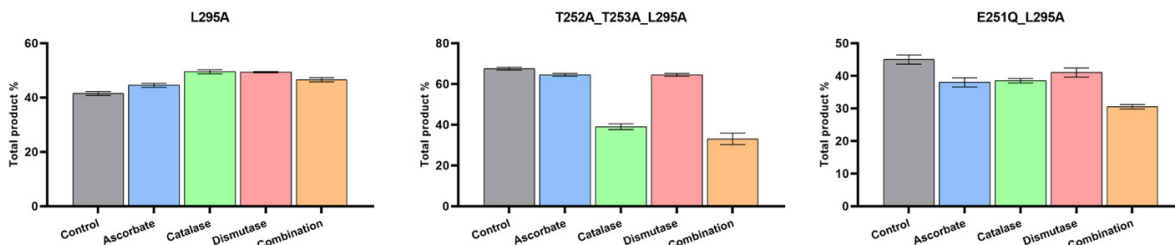


Figure B.4. TamI reactions supplemented with Reactive Oxygen Species (ROS) scavengers. To a typical TamI reaction (see Methods), 20 U bovine liver catalase, 2 U SOD from bovine erythrocytes, 20 mM ascorbate and the combination of these three ROS scavengers were added and the reaction was run for 30 min prior quenching. Control reactions did not include ROS scavengers. Substrates are shown above each bar graph. The error bars represent the standard deviation of experiments performed in duplicate.

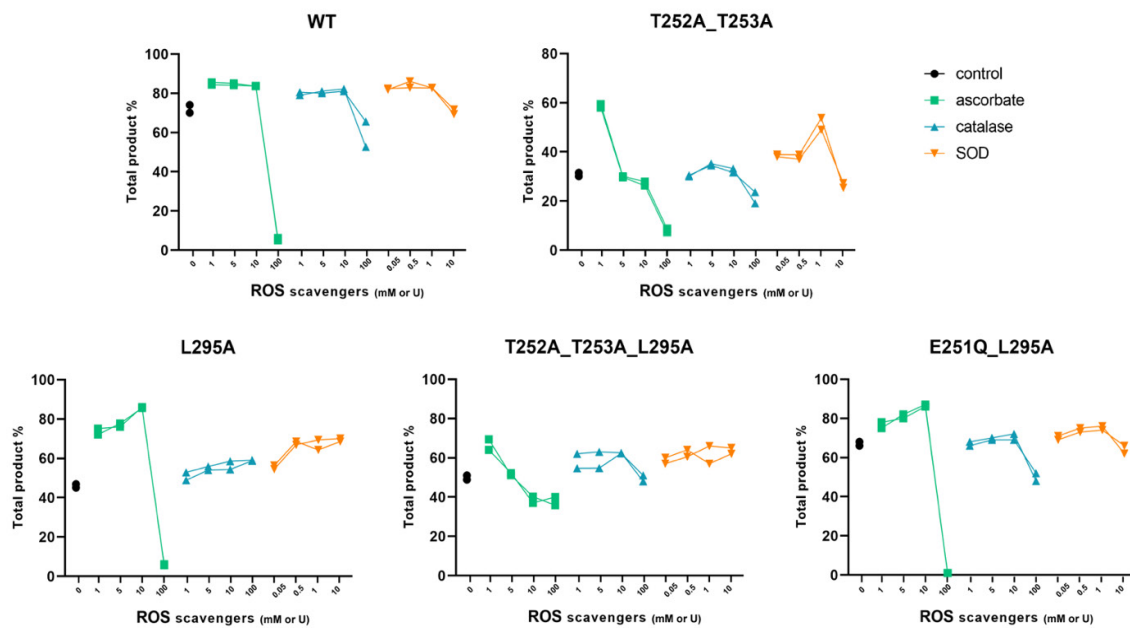


Figure B.5. TamI reactions with tirandamycin C (1) substrate and supplemented with Reactive Oxygen Species (ROS) scavengers in a dose-dependent manner. Experiments were performed in duplicate.

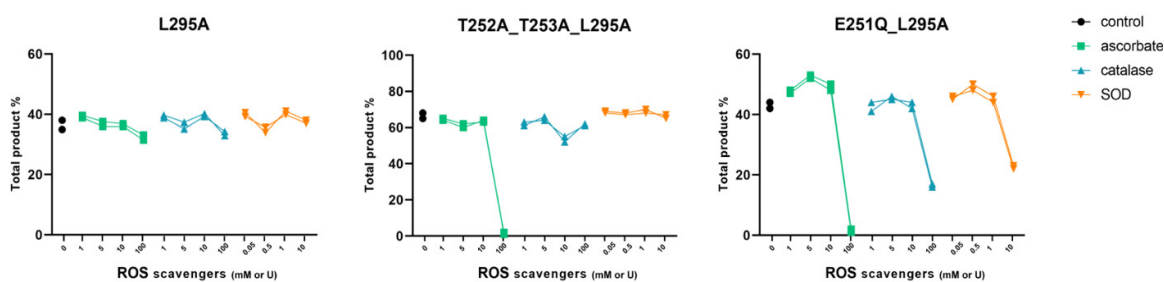


Figure B.6. TamI reactions with tirandamycin L (6) substrate and supplemented with Reactive Oxygen Species (ROS) scavengers in a dose-dependent manner. Experiments were performed in duplicate.

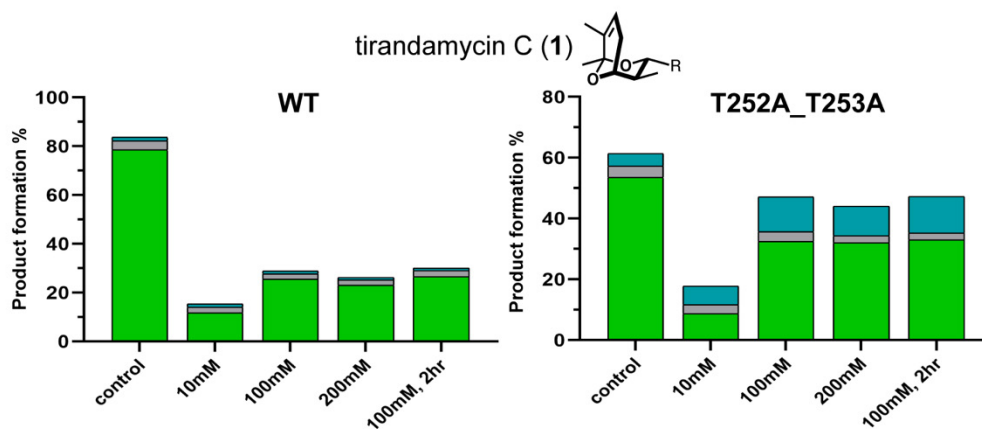
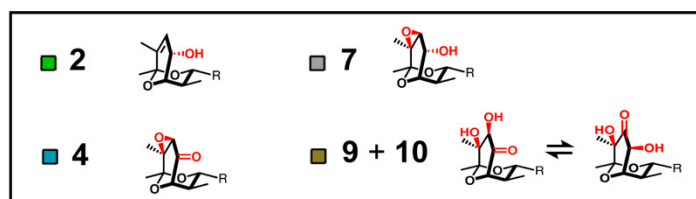


Figure B.7. Additional TamI-catalyzed oxidations via peroxide shunt pathway. Reactions were initiated with increasing concentrations of H_2O_2 and run for 30 min unless otherwise stated. Substrate is shown above bar graphs. These results indicate that TamI WT and T252A_T253A are capable of accepting H_2O_2 as a cosubstrate for oxidative transformations.

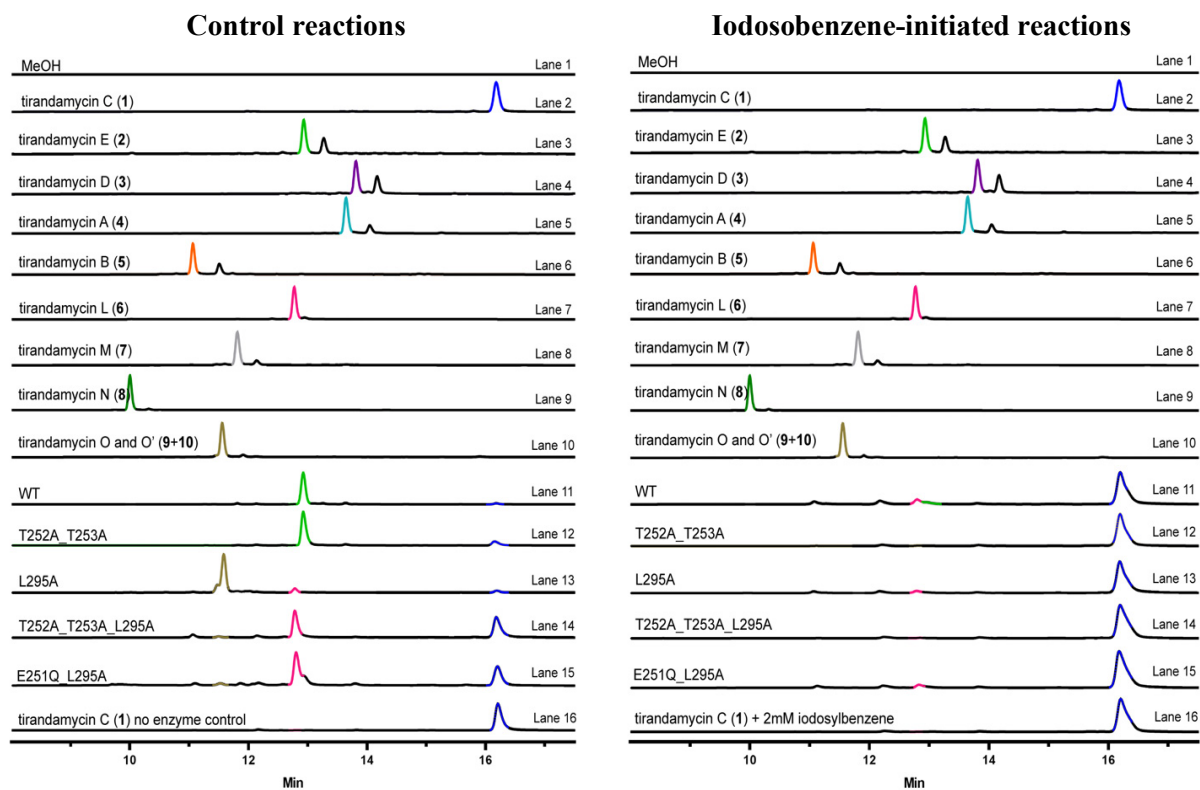


Figure B.8. End-point assays with P450 TamI mutants and tirandamycin C (1) substrate initiated with iodosobenzene. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N, O and O' (lanes 2-10). Left panel: control reactions using P450, ferredoxin, ferredoxin reductase and NADPH as described in Methods (lanes 11-15). Right panel: test reactions using P450 and 2mM iodosobenzene as described in Methods (lanes 11-16); Lane 16 in right panel refers to an additional control reaction incubating the substrate with iodosobenzene only.

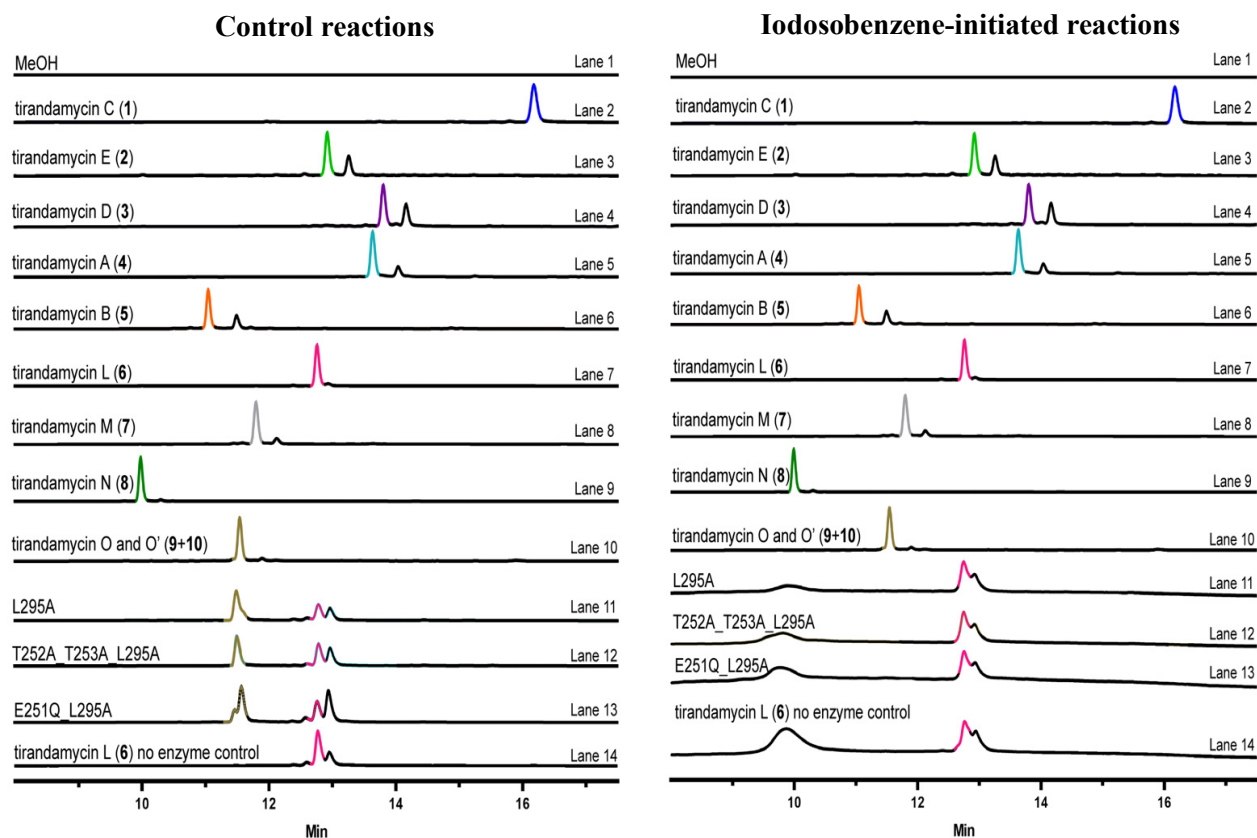


Figure B.9. End-point assays with P450 TamI mutants and tirandamycin L (6) substrate initiated with iodosobenzene. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N, O and O' (lanes 2-10). Left panel: control reactions using P450, ferredoxin, ferredoxin reductase and NADPH as described in Methods (lanes 11-15). Right panel: test reactions using P450 and 2mM iodosobenzene as described in Methods (lanes 11-16); Lane 16 in right panel refers to an additional control reaction incubating the substrate with iodosobenzene only.

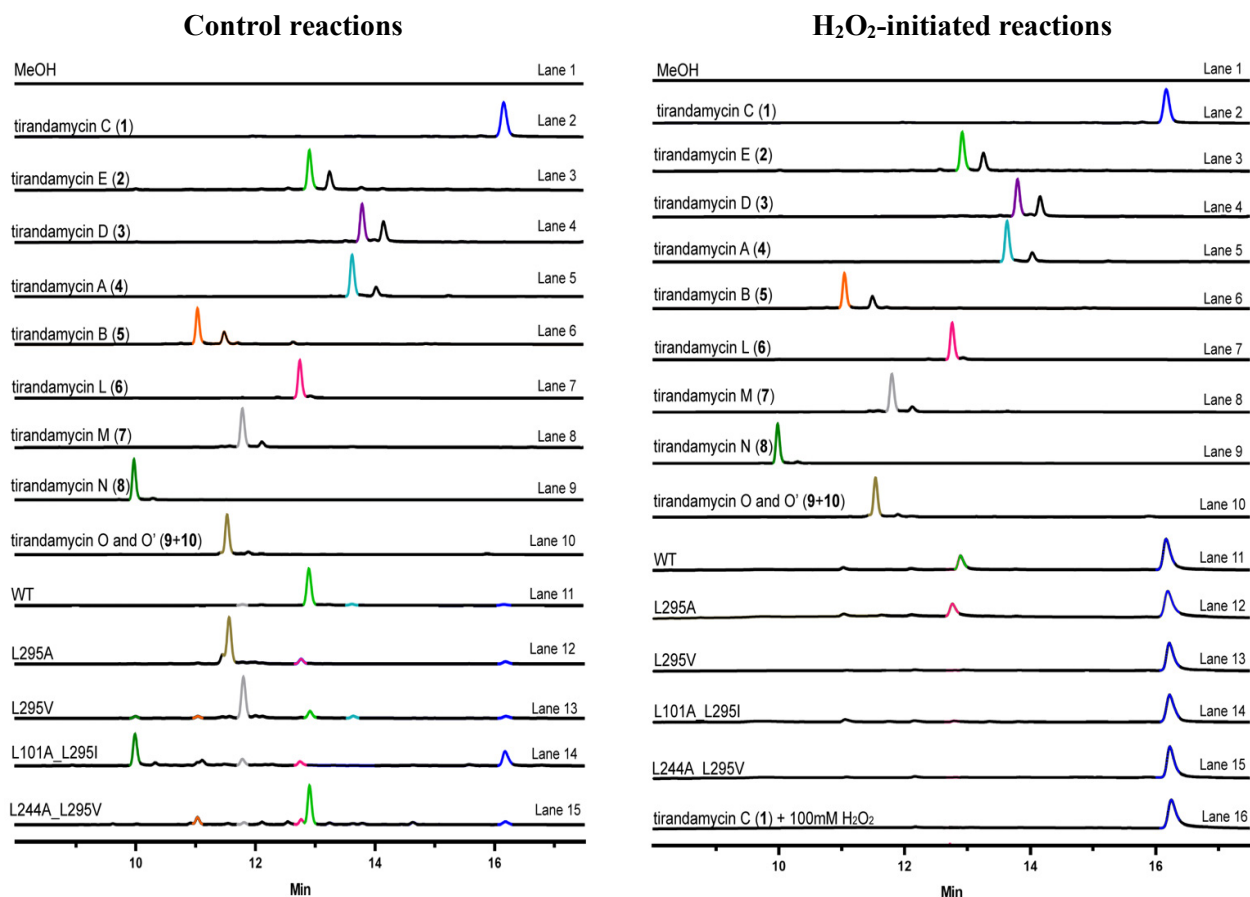


Figure B.10. End-point assays with P450 TamI mutants and tirandamycin C (1) substrate initiated with NADPH versus H₂O₂. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N, O and O' (lanes 2-10). Left panel: control reactions using P450, ferredoxin, ferredoxin reductase and NADPH as described in Methods (lanes 11-15). Right panel: test reactions using P450 and 100mM H₂O₂ as described in Methods (lanes 11-16); Lane 16 in right panel refers to an additional control reaction incubating the substrate with H₂O₂ only.

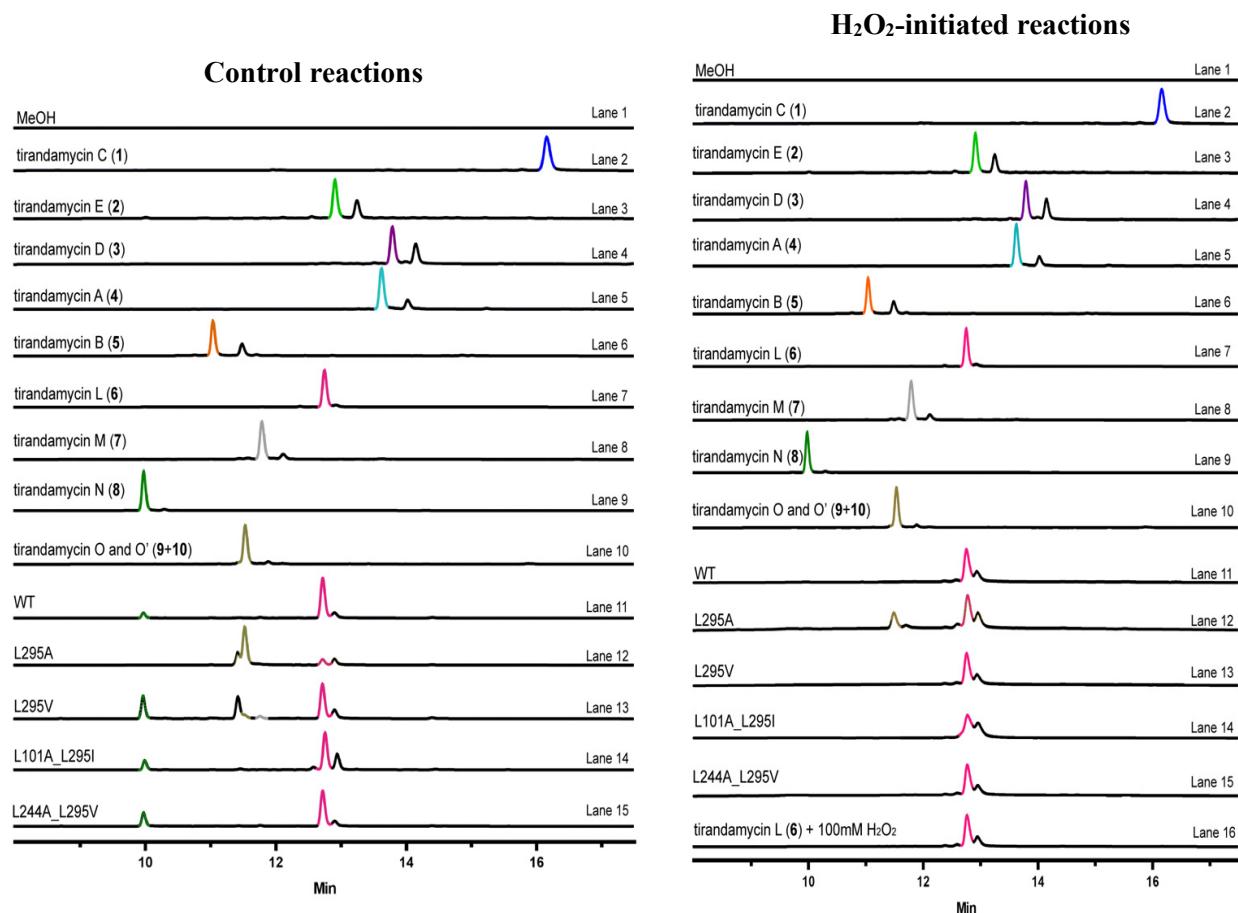


Figure B.11. End-point assays with P450 TamI mutants and tirandamycin L (6) substrate initiated with NADPH versus H₂O₂. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N, O and O' (lanes 2-10). Left panel: control reactions using P450, ferredoxin, ferredoxin reductase and NADPH as described in Methods (lanes 11-15). Right panel: test reactions using P450 and 100mM H₂O₂ as described in Methods (lanes 11-16); Lane 16 in right panel refers to an additional control reaction incubating the substrate with H₂O₂ only.

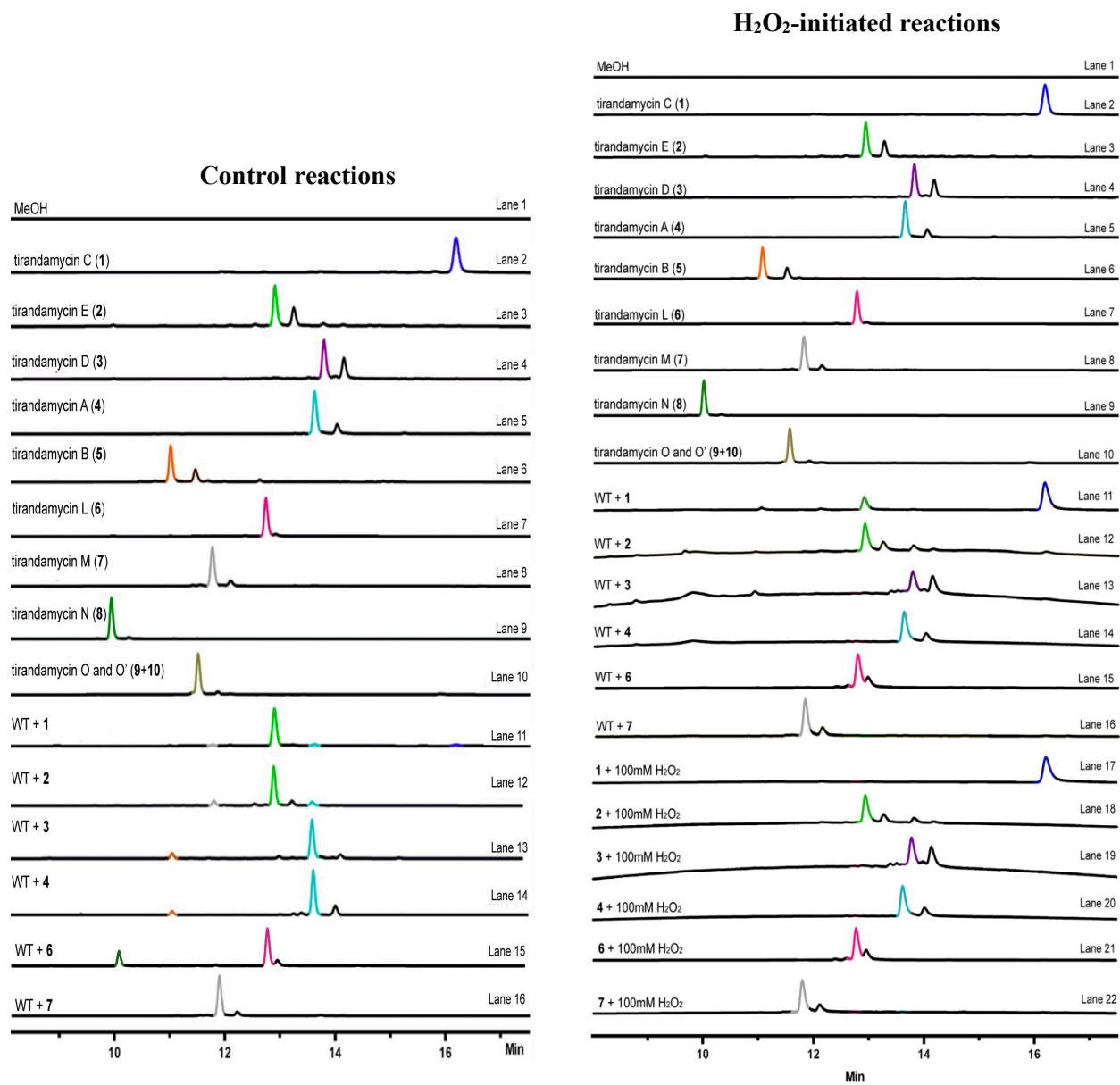


Figure B.12. End-point assays with P450 Tam1 WT and other various tirandamycin substrates initiated with NADPH versus H₂O₂. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N, O and O' (lanes 2-10). Left panel: control reactions using P450, ferredoxin, ferredoxin reductase and NADPH as described in Methods (lanes 11-16). Right panel: test reactions using P450 and 100mM H₂O₂ as described in Methods (lanes 11-16); Lanes 17-22 in right panel refer to additional control reactions incubating the substrate with H₂O₂ only.

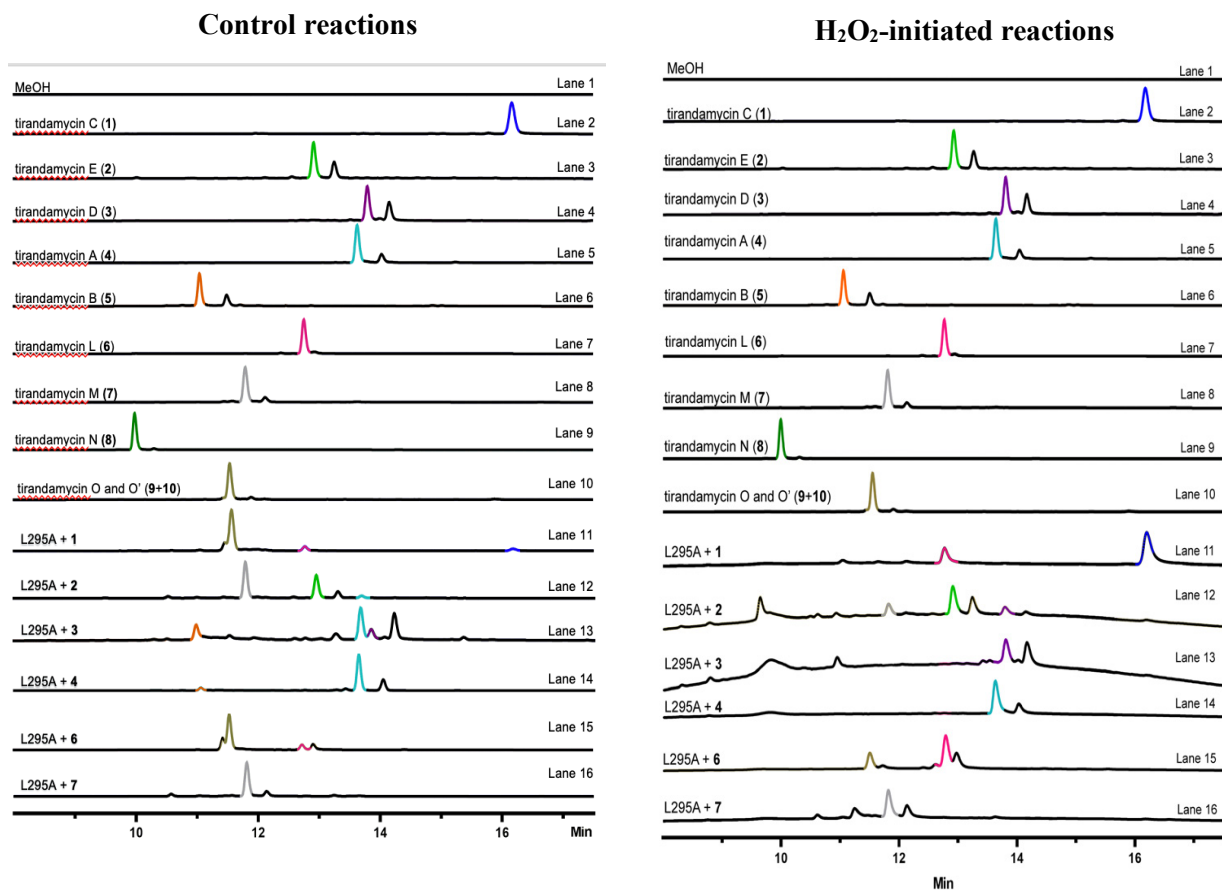


Figure B.13. End-point assays with P450 TamI L295A and various tirandamycin substrates initiated with NADPH versus H₂O₂. LC-MS traces are extracted at 354nm, signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N, O and O' (lanes 2-10). Left panel: control reactions using P450, ferredoxin, ferredoxin reductase and NADPH as described in Methods (lanes 11-16). Right panel: test reactions using P450 and 100mM H₂O₂ as described in Methods (lanes 11-16).

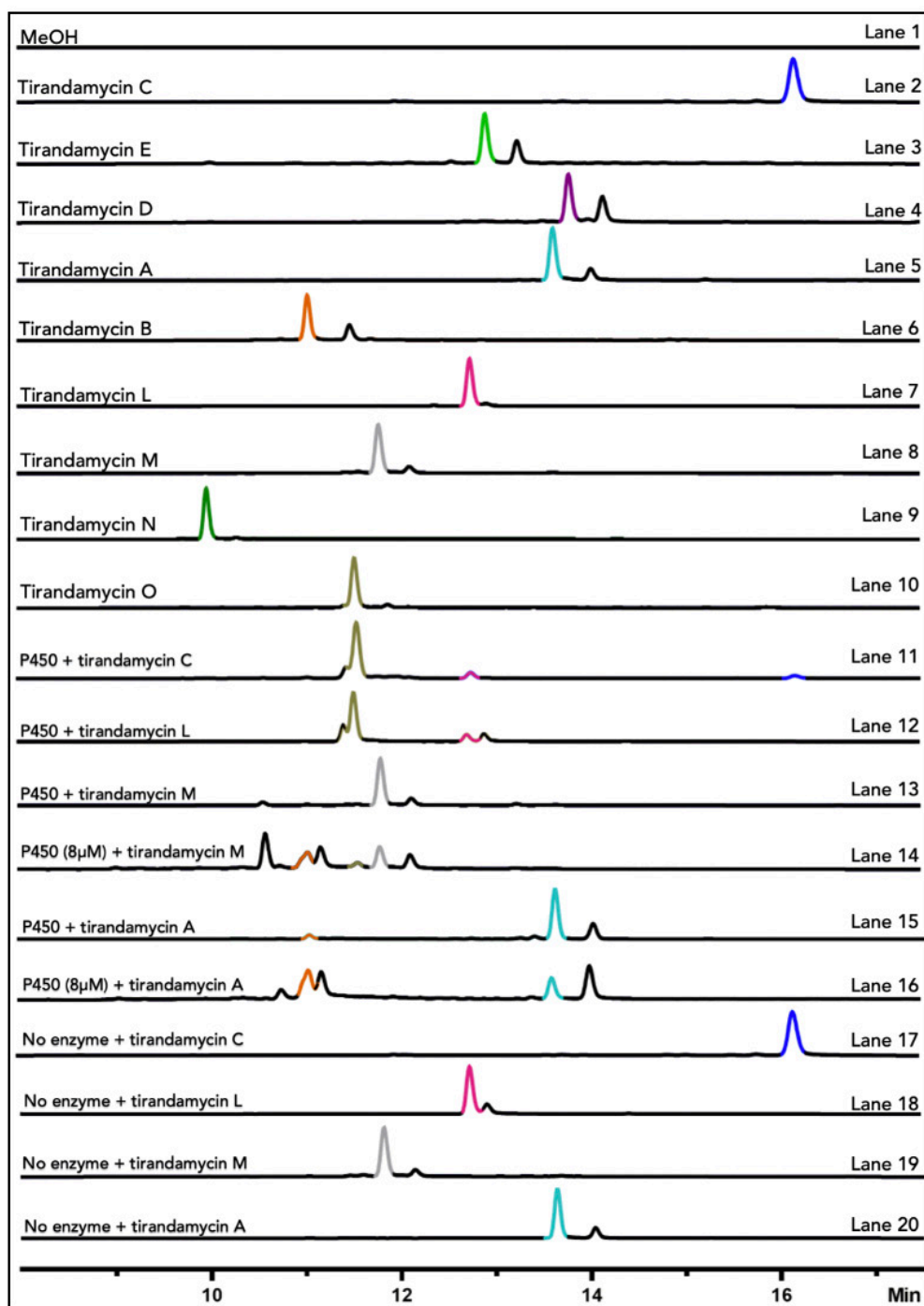


Figure B.14. Endpoint assays with TamI L295A. Analytical enzymatic reactions with TamI L295A and various tirandamycin substrates are shown in lanes 11-16. Reactions were run using standard conditions, with the exception of lanes 14 and 16 where 8 μ M P450 was used instead to assess the effect of increased P450 catalyst loading in product formation. No enzyme control reactions are shown in lanes 17-20. LC-MS traces are extracted at 354nm, the signature UV of tirandamycin. Authentic standards include tirandamycin C, E, D, A, B, L, M, N and O (lanes 2-10).

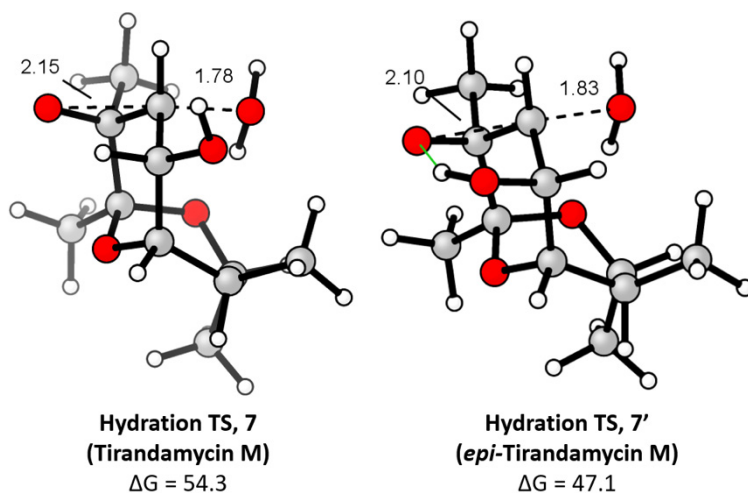


Figure B.15. Uncatalyzed hydration of congeners 7 and 7' epoxides. Congener *epi*-tirandamycin M (7') exhibits internal hydrogen bond from hydroxyl to epoxide, activating the epoxide for opening. SP//OPT: uB3LYP-D3(BJ)/def2-TZVPP, SMD(diethylether)//uB3LYP-D3(BJ)/def2-SVP, SMD(diethylether).

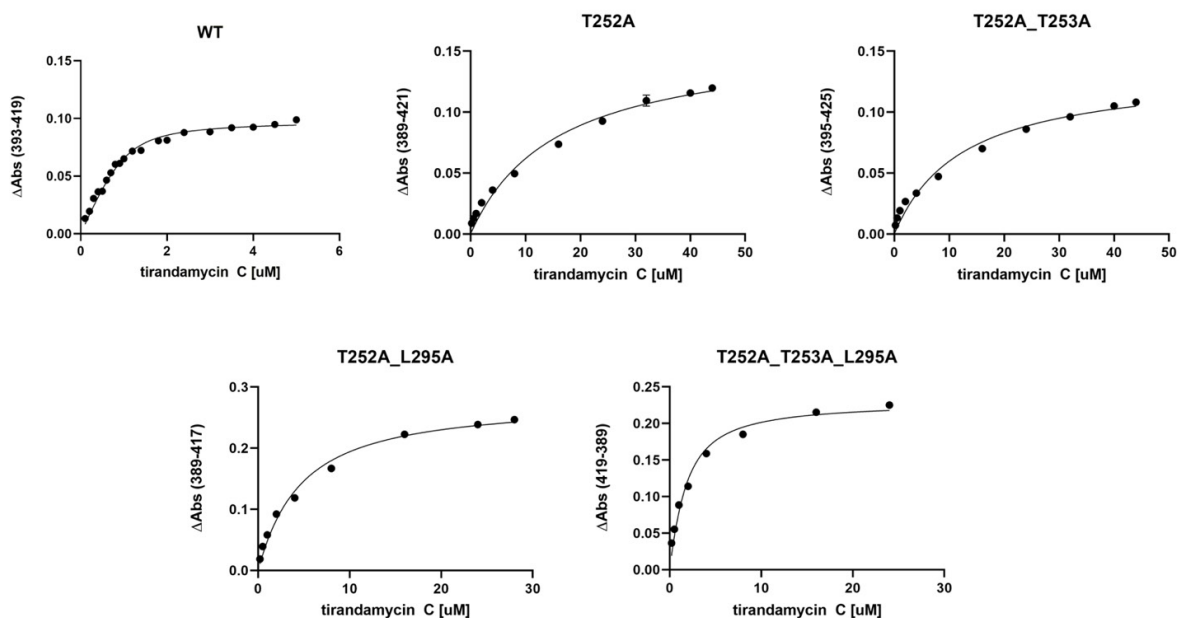


Figure B.16. Equilibrium dissociation constant binding curves for TamI WT and mutants with tirandamycin C (1). Titrations were stopped when no further spectral shift occurred even upon further addition of substrate. For K_d values refer to table S1.

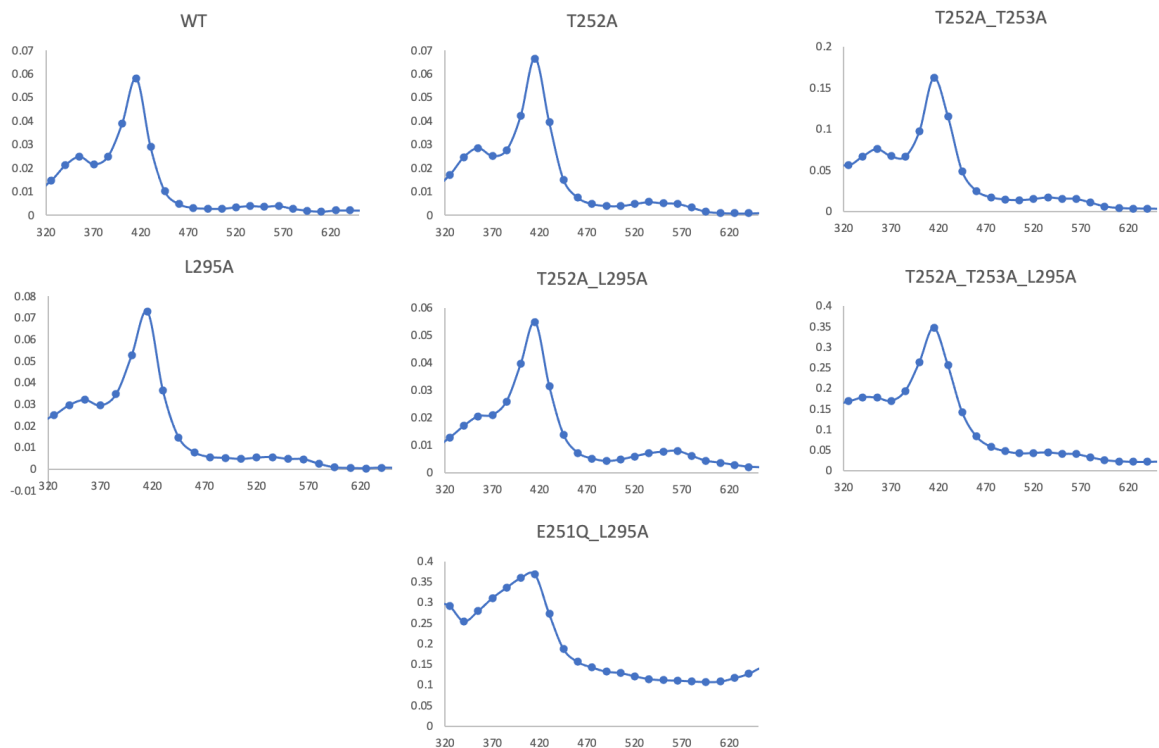


Figure B.17. Oxidized spectra for P450 TamI and variants. Measurements were taken using 1 μ M P450 in storage buffer.

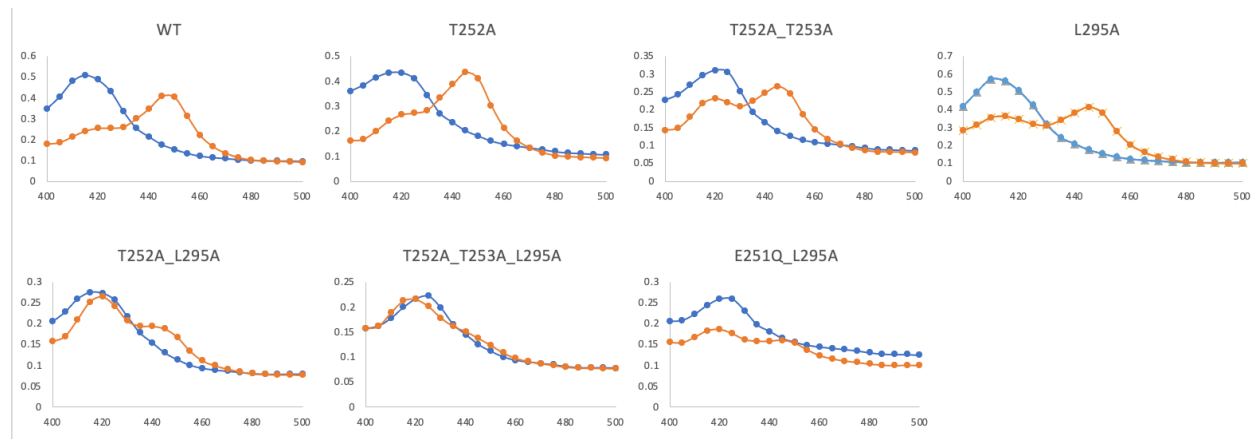


Figure B.18. CO difference spectroscopy for TamI mutants. Spectra of sodium dithionite reduced proteins are shown in blue, and CO-bound reduced forms are shown in orange.

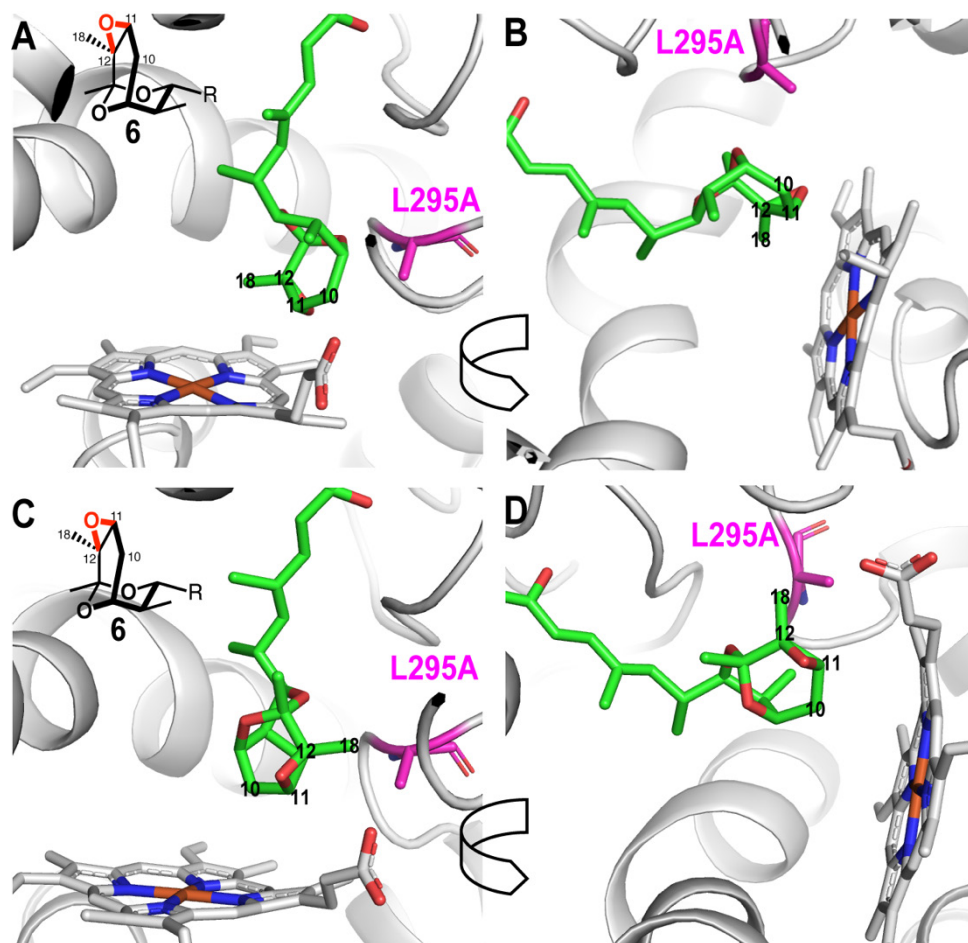


Figure B.19. MD simulations of TamI L295A and tirandamycin L (6).¹ (A) The 4th top occupied clusters show C18 and C11 are close to the heme iron within a 5Å distance. (B) The 5th top occupied cluster shows the C10 atom is the closest to the iron within a 5Å distance.

B.9 References

1. Espinoza, R.V.; Haatveit, K.C.; Grossman, S.W.; Tan, J.Y.; McGlade, C.A.; Khatri, Y.; Newmister, S.A.; Schmidt, J.J.; Garcia-Borràs, M.; Montgomery, J.; Houk, K.N.; Sherman, D.H., Engineering P450 TamI as an iterative biocatalyst for selective late-stage C–H functionalization and epoxidation of tirandamycin antibiotics. *ACS Catal* **2021**, 11 (13), 8304–8316.
2. Omura, T.; Sato, R., The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for its hemoprotein nature. *J Biol Chem* **1964**, 239, 2370-2378.
3. Pitsawong, W.; Haynes, C.A.; Koder Jr, R.L.; Rodgers, D.W.; Miller, A.F., Mechanism-informed refinement reveals altered substrate-binding mode for catalytically competent nitroreductase. *Structure* **2017**, 25(7), 978-987.
4. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.;

Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Rev. A.03, Wallingford, CT, 2016.

5. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* **1993**, 98 (7), 5648–5652.

6. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* **1988**, 37 (2), 785–789.

7. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **2011**, 32 (7), 1456–1465.

8. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J Phys Chem* **1994**, 98 (45), 11623–11627.

9. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys Chem Chem Phys* **2005**, 7 (18), 3297–3305.

10. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* **2009**, 113 (18), 6378–6396.

11. Pracht, P.; Bohle, F.; Grimme, S., Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys Chem Chem Phys* **2020**, 22 (14), 7169–7192.

12. Bannwarth, C.; Ehlert, S.; Grimme, S., GFN2-xTB-An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion Contributions. *J Chem Theory Comput* **2019**, 15 (3), 1652–1671.

13. Luchini, G.; Alegre-Requena, J. V.; Funes-Ardoiz, I.; Paton, R. S., GoodVibes: automated thermochemistry for heterogeneous computational chemistry data. *FI000Research* **2020**, 9.

14. CYLview20; Legault, C. Y., Université de Sherbrooke, 2020 (<http://www.cylview.org>)

15. The PyMOL molecular graphics system, Version 2. *Schrödinger LLC*.

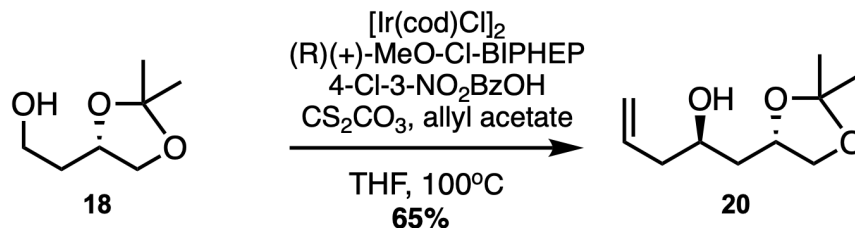
16. Presti, D.; Stoneburner, S. J.; Truhlar, D. G.; Gagliardi, L., Full correlation in a multiconfigurational study of bimetallic clusters: Restricted active space pair-density functional theory study of [2Fe–2S] Systems. *J Phys Chem* **2019**, 123 (18), 11899–11907.

APPENDIX C: Experimental Procedures and Supplemental Information for Chapter 4

C.1 Materials and General Methods

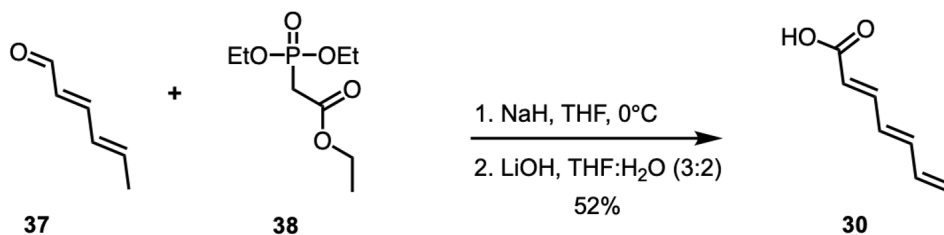
Media components, chemical reagents and solvents were purchased from Thermo Fisher Scientific, Sigma-Aldrich and Chem-Impex International unless otherwise stated. Chemical reactions were conducted in flame-fried glassware with magnetic stirring unless otherwise stated. Reaction temperature conditions were adjusted and controlled using an IKA RCT stir plate with ETS-D5 electronic contact thermometer. Solvents were purified under an atmosphere of dry nitrogen using a solvent purification system (Innovative Technology, Inc., Model #SPS- 400- 3 and PS- 400- 3). Analytical-scale thin layer chromatography (TLC) was completed using SiliaPlate TLC 60Å F-254 (250 µm silica gel). Ceric ammonium molybdate or aqueous KMnO₄ stain were used for compound visualization. SiliaFlash® P60 (230- 400 mesh) silica gel was used to perform flash column chromatography. High-resolution ESI-MS spectra was acquired and performed on a quadrupole time-of-flight spectrometer (Agilent Q-TOF 6500 series) using a XBridge™ C18 3.5 µm 50 mm column with a MeCN and H₂O solvent system supplemented with formic acid (0.1% v/v). A Shimadzu analytical HPLC was used for analysis of analytical enzymatic reactions using the same column and solvent system.

C.2 Synthesis of Tirandamycin Analogues



Scheme C.1, compound 20:

Catalyst [Ir(cod)]Cl₂ (5 mol%, 3 mg, 0.00447 mmol, 0.025 equiv), chiral ligand (10 mol%), and CS₂CO₃ (11.65 mg, 0.03576 mmol, 0.2 equiv) were mixed in a flame-dried pressure tube in the glovebox.¹ (4S)-(+)-4-(2-Hydroxyethyl)-2,2-dimethyl-1,3-dioxolane (**18**, 25.4 uL, 0.1784 mmol, 1.0 equiv), THF (0.45 mL) and 4-chloro-3-nitrobenzoic acid (3.6 mg, 0.01788 mmol, 0.1 equiv) were pre-mixed and added under nitrogen, followed by allyl acetate (96.4 uL, 0.8935 mmol, 5 equiv). Reaction tube was sealed and heated at 100°C in a digital dry bath heater and run for 72 hours. Reaction mixture was allowed to cool to room temperature and diluted with DCM (0.66 mL) prior filtering through celite. The crude mixture was concentrated *in vacuo* to yield a brown oil that was subjected to flash chromatography (EtOAc:hexanes 1:4) to give desired product. The spectra matched that of previous reports.² Refer to Table C.1. below for optimization conditions and a list of chiral ligands used.

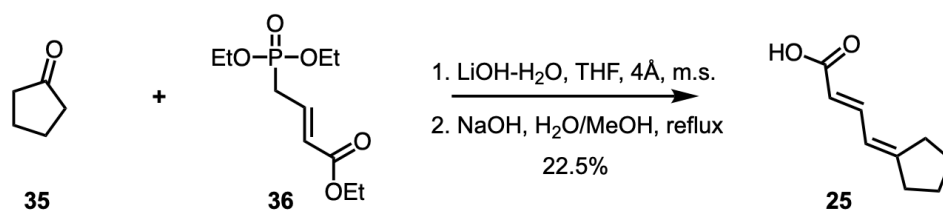


Scheme C.2, compound 30:

Reaction 1: Sodium hydride (60% dispersion in mineral oil, 0.6523 g, 27.1819 mmol, 1.2 equiv) was suspended in THF (141.62 mL) in a flame-dried glass vial at 0°C under nitrogen. Triethyl phosphonoacetate (5.39 mL, 27.1819 mmol, 1.2 equiv) was added slowly with a syringe over 5-10 min at 0°C. Reaction was stirred for 30 min at 0°C (suspension turns clear). Separately, freshly distilled sorbaldehyde (distilled at 69°C, 20 torr; 2.5 mL, 22.6516 mmol, 1 equiv) was suspended in THF (42.47 mL) in a flame-dried glass round bottom flask at 0°C under nitrogen, and added dropwise at 0°C to the reaction mixture. The reaction was allowed to warm to room temperature and stirred for 16 hours. Upon completion, the reaction was quenched by dropwise addition of NH₄Cl (saturated aq.) until bubbling subsided. The resulting solution was mildly acidified to pH 4-5, diluted with water and extracted with Et₂O. The organic extracts were combined, washed with brine, dried over MgSO₄ and concentrated *in vacuo*. The crude mixture was subjected to flash

chromatography (EtOAc:hexanes 1:5) to give the ester version of the product as a colourless liquid.³

Reaction 2 (hydrolysis): The ester (1.3 g, 7.8209 mmol, 1 equiv) was suspended in a mixture of THF:H₂O (3:2) in a flame-dried glass round bottom flask prior adding lithium hydroxide (anhydrous; 2.847g, 118.877 mmol, 15.2 equiv). The reaction mixture was stirred for 72 hours at room temperature. The reaction was acidified with HCl (aq) to reach a pH of 2-3, diluted with H₂O and extracted with CHCl₃.⁴ The organic layers were combined, dried over MgSO₄ and concentrated *in vacuo*. The crude reaction mixture (orange solid) was crystallized with CHCl₃ and minimal EtOAc, followed by sublimation at 150°C to give the desired carboxylic acid product (20:1 E:Z) as a solid white powder. The spectra matched that of previous reports.³



Scheme C.3, compound 25:

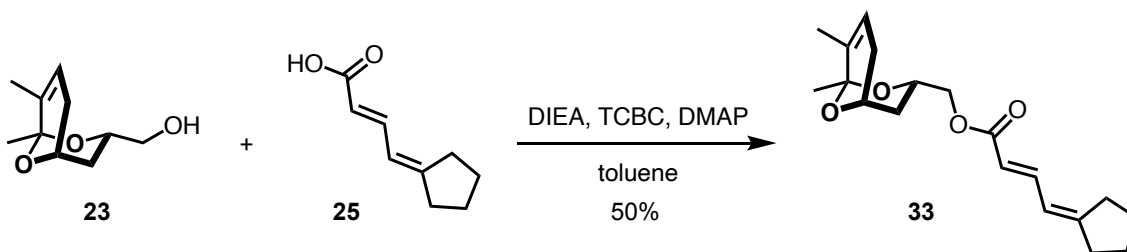
Reaction 1: Cyclopentanone (**35**, 1.28 mL, 14.5341 mmol, 1 equiv) was suspended in a solution of THF (122.24 mL) containing triethyl 4-phosphonocrotonate (**36**, 7.09 mL, 31.9706 mmol, 2.2 equiv), lithium hydroxide (anhydrous; 1.3415g, 31.9706 mmol, 2.2 equiv) and molecular sieves (4Å, 4-8 mesh) in a flame-dried glass round bottom flask under nitrogen. The reaction was heated at reflux (66 °C) for 13 hours. Upon completion, the reaction was filtered through a short plug of silica, eluted with Et₂O and concentrated *in vacuo*.⁵ The crude reaction mixture (dark yellow oil) was purified using flash chromatography (EtOAc:hexanes 1:9) to give the desired ester version of the product (only *E* isomer obtained as observed by H-NMR).

Reaction 2 (hydrolysis): The ester (74 mg, 0.4105 mmol, 1 equiv) was suspended in a solution of H₂O/MeOH (2:1; 7.27 mL:3.65 mL) containing sodium hydroxide (1.1196 g, 27.9972 mmol, 68 equiv) in a flame-dried glass round bottom flask. The reaction was refluxed (85°C) for 45 min under nitrogen. Upon completion, the reaction was cooled to room temperature, diluted with brine and extracted with Et₂O. The aqueous layer was acidified with 10% HCl and extracted with

EtOAc.⁶ The organic layers were combined, washed with brine, dried over MgSO₄ and concentrated *in vacuo*. The crude reaction mixture was purified using flash chromatography (EtOAc:hexanes 1:2) to yield the desired product as a white solid.

¹H NMR (500 MHz, CDCl₃): δ 7.53 (dd, J = 15.2, 11.7 Hz, 1H), 6.13 (d, J = 11.8 Hz, 1H), 5.71 (d, J = 15.3 Hz, 1H), 2.50 (t, J = 7.6 Hz, 2H), 2.41 (t, J = 7.8 Hz, 2H), 1.75 (q, J = 6.9 Hz, 2H), 1.70 (q, J = 6.9 Hz, 2H).

¹³C NMR (500 MHz, CDCl₃): δ 172.7, 160.8, 144.8, 119, 116.6, 34.8, 30.2, 26.1, 25.8



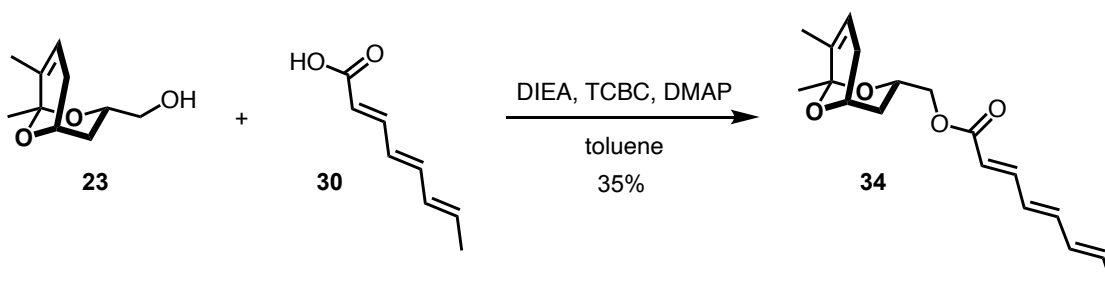
Scheme C.4, compound 33:

Dry carboxylic acid (**25**, 86.4 mg, 0.7703 mmol, 1.70 equiv) was added to a flame-dried round bottom flask containing toluene (25.1 mL), N, N-diisopropylethylamine (DIEA; 641.1 uL, 3.6807 mmol, 8.10 equiv), 2, 4, 6-trichlorobenzoic acid (TCBC; 290.1 uL, 1.8565 mmol, 4.09 equiv) and 4-dimethylaminopyridine (DMAP; 227 mg, 1.8565 mmol, 4.09 equiv). The reaction was stirred for 20 min. Alcohol (**23**, 50 mg, 0.4539 mmol, 1 equiv) was added as a solution in toluene (22.6 mL) and the reaction was stirred overnight (15 hours) at room temperature and protected from light. The reaction was quenched by adding saturated NaHCO₃ until bubbling stopped and diluted with EtOAc.⁷ The organic layers were combined, washed with brine, dried with Na₂SO₄ and concentrated *in vacuo*. The crude reaction product was subjected to flash chromatography (EtOAc:hexanes 1:3) to yield the desired product.

¹H NMR (500 MHz, CDCl₃): δ 7.45 (dd, J = 15.2, 11.6 Hz, 1H), 6.10 (dt, J = 11.7, 2.3 Hz, 1H), 5.80 – 5.75 (m, 1H), 5.72 (s, 0H), 4.35 (t, J = 6.6 Hz, 1H), 4.31 – 4.24 (m, 1H), 4.23 – 4.15 (m, 1H), 4.12 – 4.06 (m, 1H), 2.71 – 2.62 (m, 1H), 2.51 – 2.45 (t, 2H), 2.39 (t, J = 7.0 Hz, 2H), 2.00 – 1.93 (m, 1H), 1.92 (s, 1H), 1.76 – 1.64 (m, 4H), 1.59 (s, 3H), 1.44 (s, 3H), 1.18 (d, J = 6.7 Hz, 1H).

^{13}C NMR (500 MHz, CDCl_3): δ 167.9, 159.9, 143.0, 128.4, 124.1, 119.4, 117.5, 95.9, 69.1, 67.2, 66.5, 35.2, 32.7, 32.6, 30.5, 26.6, 26.3, 24.9, 18.9.

HRMS (ESI+) m/z : $[\text{M}+\text{H}]^+$ predicted for $\text{C}_{19}\text{H}_{26}\text{O}_4$, 319.1904; found 319.1925

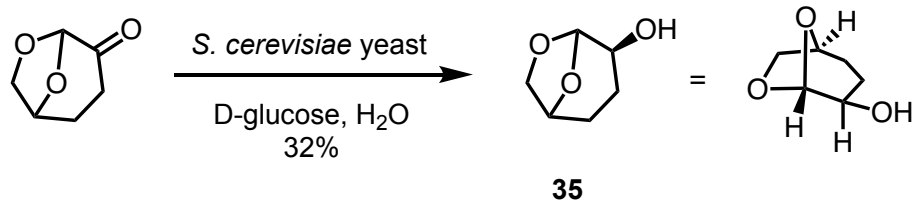


Scheme C.5, compound 34:

Dry carboxylic acid (**30**, 15 mg, 0.1086 mmol, 1.70 equiv) was added to a flame-dried round bottom flask containing toluene (3.5 mL), N, N-diisopropylethylamine (DIEA; 90.4 μL , 0.5188 mmol, 8.10 equiv), 2, 4, 6-trichlorobenzoic acid (TCBC; 40.9 μL , 0.2617 mmol, 4.09 equiv) and 4-dimethylaminopyridine (DMAP; 64 mg, 0.5234 mmol, 8.109 equiv). The reaction was stirred for 20 min. Alcohol (**23**, 11.79 mg, 0.0639 mmol, 1 equiv) was added as a solution in toluene (3.19 mL) and the reaction was stirred overnight (15 hours) at room temperature and protected from light. The reaction was quenched by adding saturated NaHCO_3 until bubbling stopped and diluted with EtOAc.⁷ The organic layers were combined, washed with brine, dried with Na_2SO_4 and concentrated *in vacuo*. The crude reaction product was subjected to flash chromatography (EtOAc:hexanes 1:4) to yield the desired product.

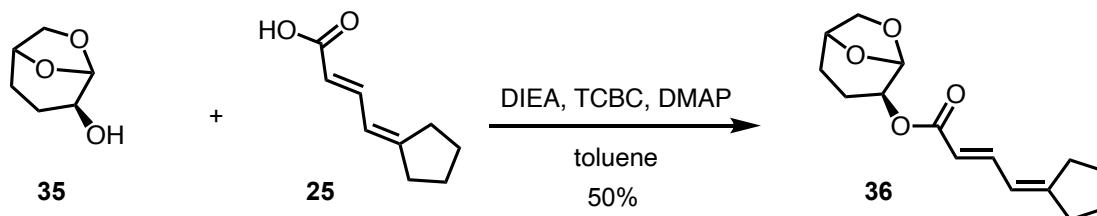
^1H NMR (500 MHz, CDCl_3): δ 7.32-7.24 (m, 1H), 6.52 (dd, $J = 14.9, 10.7$ Hz, 1H), 6.25-61.7 (m, 1H) 6.17-6.11 (m, 1H), 5.95 (dt, $J = 14.9, 7.0$ Hz, 1H), 5.86 (d, $J = 15.3$ Hz, 1H), 5.77 (d, $J = 4.4$ Hz, 1H), 4.35 (t, $J = 6.6$ Hz, 1H), 4.27 (qd, $J = 8.0, 6.7$ Hz, 1H), 4.16, (m, 1H), 2.66 (d, $J = 19.8$ Hz, 1H), 2.01-1.92 (m, 1H), 1.97 – 1.83 (m, 1H), 1.82 (d, $J = 7.0$ Hz, 3H), 1.68 – 1.56 (m, 3H), 1.43 (d, $J = 1.2$ Hz, 3H), 1.17 – 1.11 (m, 1H)

HRMS (ESI+) m/z : $[\text{M}+\text{H}]^+$ predicted for $\text{C}_{18}\text{H}_{24}\text{O}_4$, 305.1753; found 305.1775



Scheme C.6, compound 35:

D-glucose (0.96 g, 5.36 mmol, 5.5 equiv) was added to a flame-dried round bottom flask containing H₂O (9.75 mL) and *Saccharomyces cerevisiae* yeast (1.25 g). The reaction was stirred at 30 °C for 1 hr. Cyrene was added and the reaction was left stirring at room temperature for 24 hrs. Dichloromethane was added to stop the reaction and the mixture was filtered and concentrated *in vacuo* prior extracting with EtOAc at pH 3-4. The crude product was purified using flash chromatography (EtOAc:hexanes 1:2) to yield the desired product as a clear oil. The spectra matched that of previous reports.⁸



Scheme C.7, compound 36:

Dry carboxylic acid (**25**, 15 mg, 0.1086 mmol, 1.70 equiv) was added to a flame-dried round bottom flask containing toluene (3.5 mL), N, N-diisopropylethylamine (DIEA; 90.4 uL, 0.5188 mmol, 8.10 equiv), 2, 4, 6-trichlorobenzoic acid (TCBC; 40.9 uL, 0.2617 mmol, 4.09 equiv) and 4-dimethylaminopyridine (DMAP; 64 mg, 0.5234 mmol, 8.109 equiv). The reaction was stirred for 20 min. Alcohol (**35**, 8.3 mg, 0.0639 mmol, 1 equiv) was added as a solution in toluene (3.19 mL) and the reaction was stirred overnight (15 hours) at room temperature and protected from light. The reaction was quenched by adding saturated NaHCO₃ until bubbling stopped and diluted with EtOAc.⁷ The organic layers were combined, washed with brine, dried with Na₂SO₄ and concentrated *in vacuo*. The crude reaction product was subjected to flash chromatography (EtOAc:hexanes 1:3) to yield the desired product.

¹H NMR (500 MHz, CDCl₃): δ 7.46 (dd, J = 15.2, 11.6 Hz, 1H), 6.09 (d, J = 11.8 Hz, 1H), 5.73 (d, J = 15.0 Hz, 1H), 5.43 (s, 1H), 4.82 (dd, J = 10.9, 5.4 Hz, 1H), 4.54 (bs, 1H), 3.94 (d, J = 7.2 Hz, 1H), 3.85 (t, J = 5.4 Hz, 1H), 2.49 (t, J = 7.5 Hz, 2H), 2.39 (t, J = 7.4 Hz, 2H), 2.03-1.01 (m, 1H), 2.0-1.94 (m, 2H) 1.83-1.80 (m, 1H), 1.75-1.70 (m, 2H), 1.71-1.64 (m, 2H).

¹³C NMR (500 MHz, CDCl₃): δ 159.9, 167.0, 143.3, 119.0, 117.1, 100.6, 73.1, 70.9, 68.5, 34.8, 30.2, 27.8, 25.9, 26.2, 22.0

HRMS (ESI+) m/z: [M+H]⁺ predicted for C₁₅H₂₀O₄, 265.1440; found 265.1438

C.3 Protein Expression and Purification

C.3.1 TamI protein expression and purification

The pET28b_TamI plasmid previously generated⁹ was utilized to express N-terminally His-tagged TamI. *E. coli* BL21 (DE3) was transformed with the plasmid. Overexpression, protein purification and protein quantification were carried out exactly as previously described in Appendix A without any modifications.⁹

C.3.2 Site-directed mutagenesis

All single mutants were prepared with TamI wild-type as the template and the desired mutation as the primer using the Quikchange Lightning Site-Directed Mutagenesis Kit and protocol, and carried out exactly as previously described in Appendix A without any modifications.⁹

C.4 Analytical Enzymatic Assays

Reactions were performed *in vitro* in an analytical scale to a final volume of 100 μL of reaction buffer (50 mM NaH₂PO₄, pH 7.4 at RT) containing 200 μM substrate, 2 μM P450, 40 μM spinach ferredoxin, 6 μM spinach ferredoxin-NADP⁺ reductase, 5 mM G6P, 1 U/mL G6PDH and 1 mM NADP⁺. As negative controls, enzymes were omitted. Reactions were run for 2 h at 30°C, quenched by addition of 100 μL MeOH and centrifuged for 10 min at 4°C. The supernatant was analyzed by HPLC and LC-MS using 280 nm UV detection, positive/negative ion MS detection when applicable, and the following conditions (A = H₂O + 0.1% FA, B = MeCN + 0.1% FA): 10% B for 2 min, 10-100% B over 15 min, 100% B for 4 min, 10% B for 2 min; flow rate 0.2 mL/min;

injection volume 10 μL for HPLC and 2 μL for LC-MS. All reactions were performed and analyzed in triplicate.

C.5 Computational Methods

C.5.1 MD Simulations: The heme iron(IV)-oxo complex involved in the cytochrome-catalyzed oxidative hydroxylation and epoxidation cycle (compound I) was used to model the active form of the cofactor. Simulations were performed using the GPU code (pmemd)⁴ of the Amber 12 package, The Amber-compatible parameters developed by Cheatham et al. were used for compound I and its axial Cys ligand.¹⁰ Parameters for tirandamycin analogues were generated within the antechamber module using the general AMBER force field (gaff), with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model. The charges were calculated according to the Merz–Singh–Kollman scheme using Gaussian 09. Each protein was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P water molecules using the tleap module, resulting in the addition of around 11,370 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na^+ and Cl^-). All subsequent calculations were carried out using the widely tested Stony Brook modification of the Amber 99 forced field (ff99sb). The substrate and enzyme were optimized for total 1,000,000 steps, with 750,000 steepest descent steps and 250,000 conjugate gradient steps. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0–300 K) under constant volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between hydrogen atoms were kept fixed. Long-range electrostatic effects were modeled using the particle-meshEwald method. An 8 Å cutoff was applied to the LennardJones and electrostatic interactions. Harmonic restraints of 30 kcal/(mol Å²) were applied to the solute, and the Andersen equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 time step at a constant volume. Production trajectories were then run for an additional 1500 ns under the same simulation conditions. These experiments were carried out in a similar manner as those described in Appendix A.⁹

C.5.2 Molecular Docking: Docking experiments were performed using the AutoDock Tools (ADT) 4.0.¹¹ Genetic algorithms were utilized to identify the preferred protein-substrate (receptor-ligand) binding conformation.

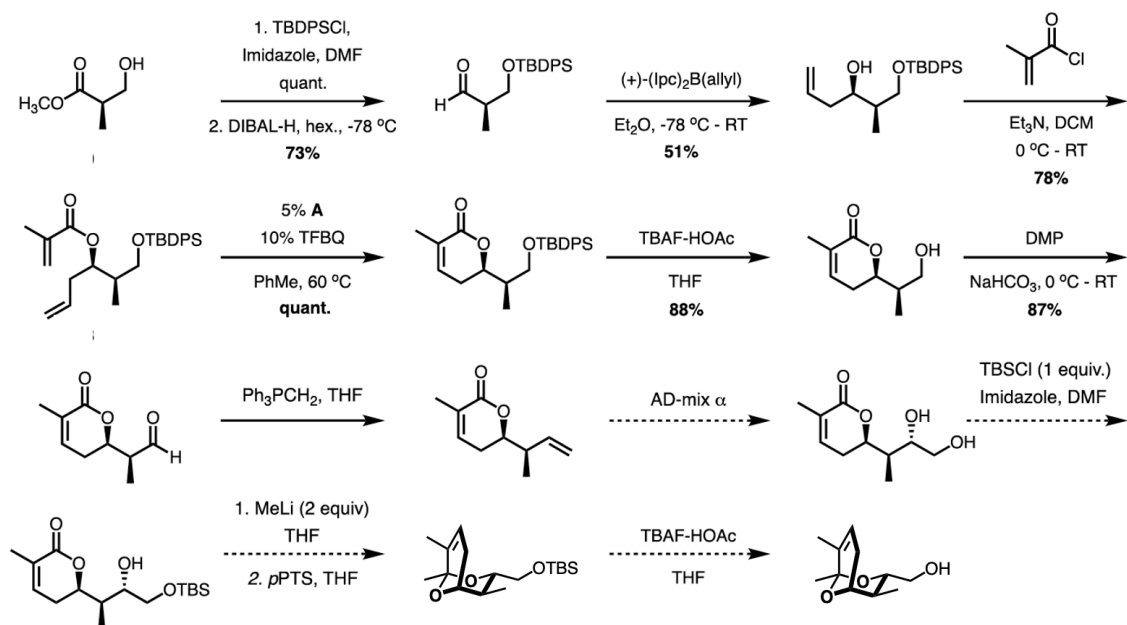
Receptor: The substrate-bound crystal structure of P450 TamI and tirandamycin C was used as the receptor after deleting the tirandamycin substrate from the crystal structure. Hydrogens were added in ADT.

Ligand: The 3D structures were built using Chem3D, their energy minimized, and the files converted into .pdb format. ADT added and merged hydrogens, and computed Gasteiger charges for the entire ligand. ADT identified a root position in the ligand (atom with the smallest largest sub-tree) and chose points of torsion (which bonds could be rotated).

Grid parameters file: The grids were selected as such that they include the conserved active-site regions of the P450 enzyme, while allowing the ligand to freely rotate inside the grid.

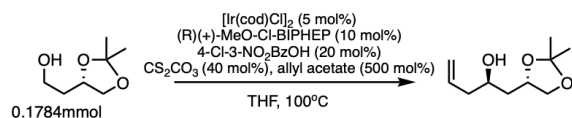
Docking parameters file: The dpf file included docking parameters and instructions for a Lamarckian Genetic Algorithm (LGA) docking. Modeling of P450 TamI and ligands was repeated with 100 generations in each run to improve precision and accuracy in the results obtained. The receptor-ligand interactions were analyzed in relation to their binding energy (kcal/mol) and inhibition constant (μM). The number of hydrogen bonding interactions observed with surrounding active site amino acids was also considered. A picture of the most preferred docked conformations was obtained using PyMOL.

C.6 Supplemental Schemes



Scheme C.8. First-generation total synthesis of a tirandamycin bicyclic analogue. For a description of reaction conditions, refer to Stachowski et al.

C.7 Supplemental Tables



Entry	Ligand	Variations	Yield
1	(<i>R</i>)-BINAP	none	28%
2	(<i>R</i>)-BINAP	1mmol scale	37%
3	(<i>R</i>)-BINAP	120°C	19%
4	(<i>R</i>)-BINAP	1000 mol% allyl acetate	25%
5	(<i>R</i>)-3,5-Xyl-MeO-BIPHEP	none	39%
6	(<i>R</i>)(+)-Cl-MeO-BIPHEP	none	43%
7	(<i>R</i>)-MeO-BIPHEP	none	37%
8	(<i>R</i>)(+)-tol-BINAP	none	8%
9	(<i>R</i>)-C3-TUNEPHOS	none	16%
10	(<i>R</i>)(+)-Cl-MeO-BIPHEP	1mmol scale	40%
11	(<i>R</i>)(+)-Cl-MeO-BIPHEP	200 mol% allyl acetate	23%
12	(<i>R</i>)(+)-Cl-MeO-BIPHEP	10 mol% catalyst	65%
13	(<i>R</i>)(+)-Cl-MeO-BIPHEP	10 mol% ligand	17%
14	(<i>R</i>)(+)-Cl-MeO-BIPHEP	Dioxane, 120°C	32%

Table C.1. Asymetric allylation optimization.

C.8 Supplemental Figures

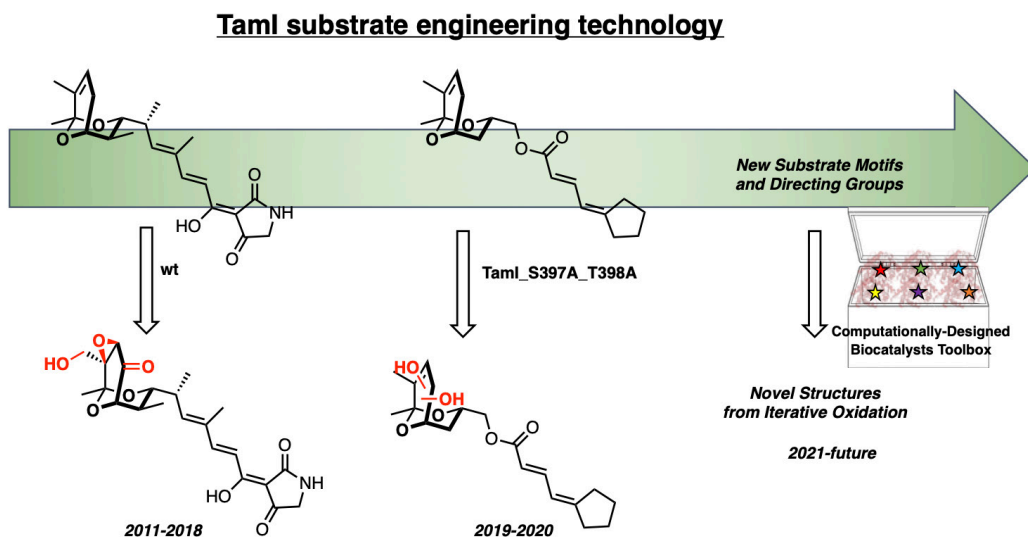


Figure C.1. Substrate engineering strategy to expand the substrate scope of P450 TamI.

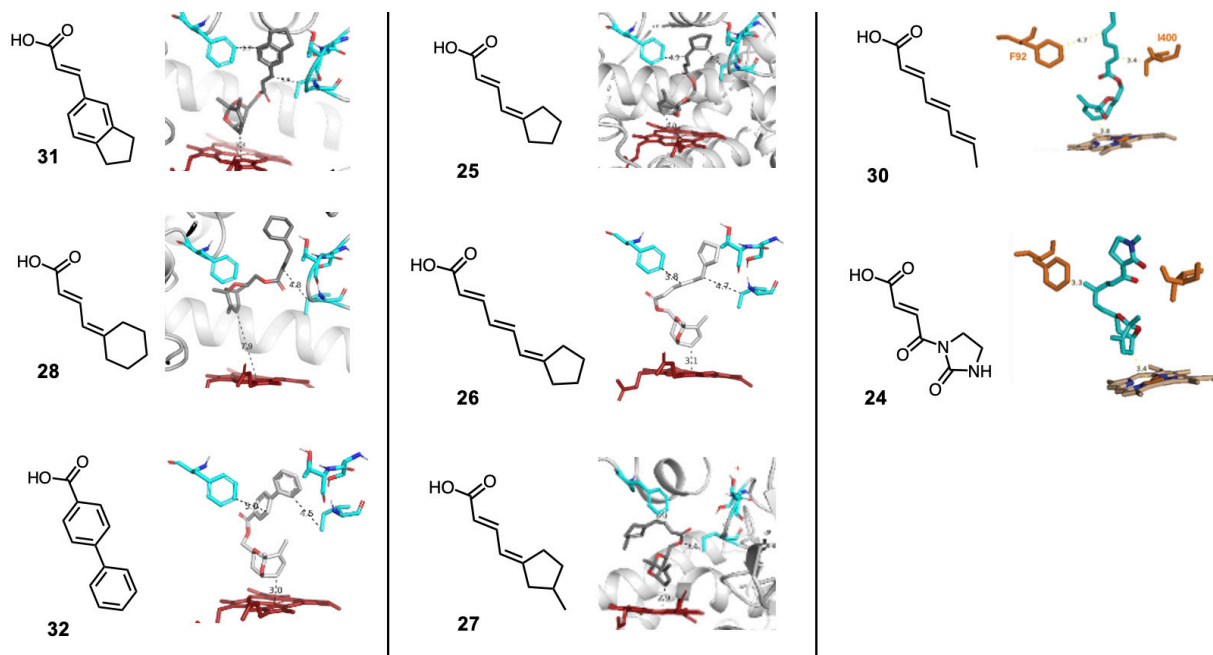


Figure C.2. Computational docking of simplified synthetic anchors. The carboxylic acid-containing anchors coupled to the hydroxyl-containing tirandamycin bicyclic analogue (refer to Scheme 4.3 for an example) were docked into P450 TamI. The most occupied and lowest binding affinity poses are displayed above.

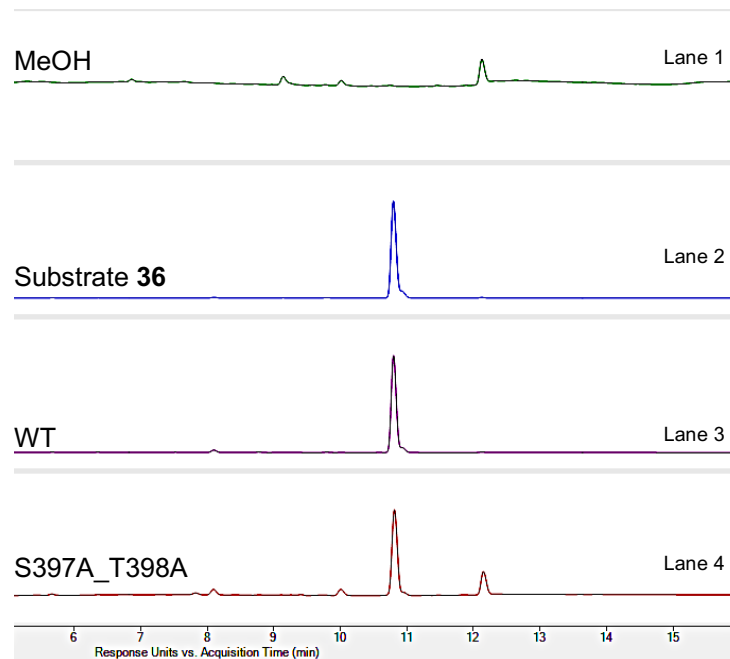
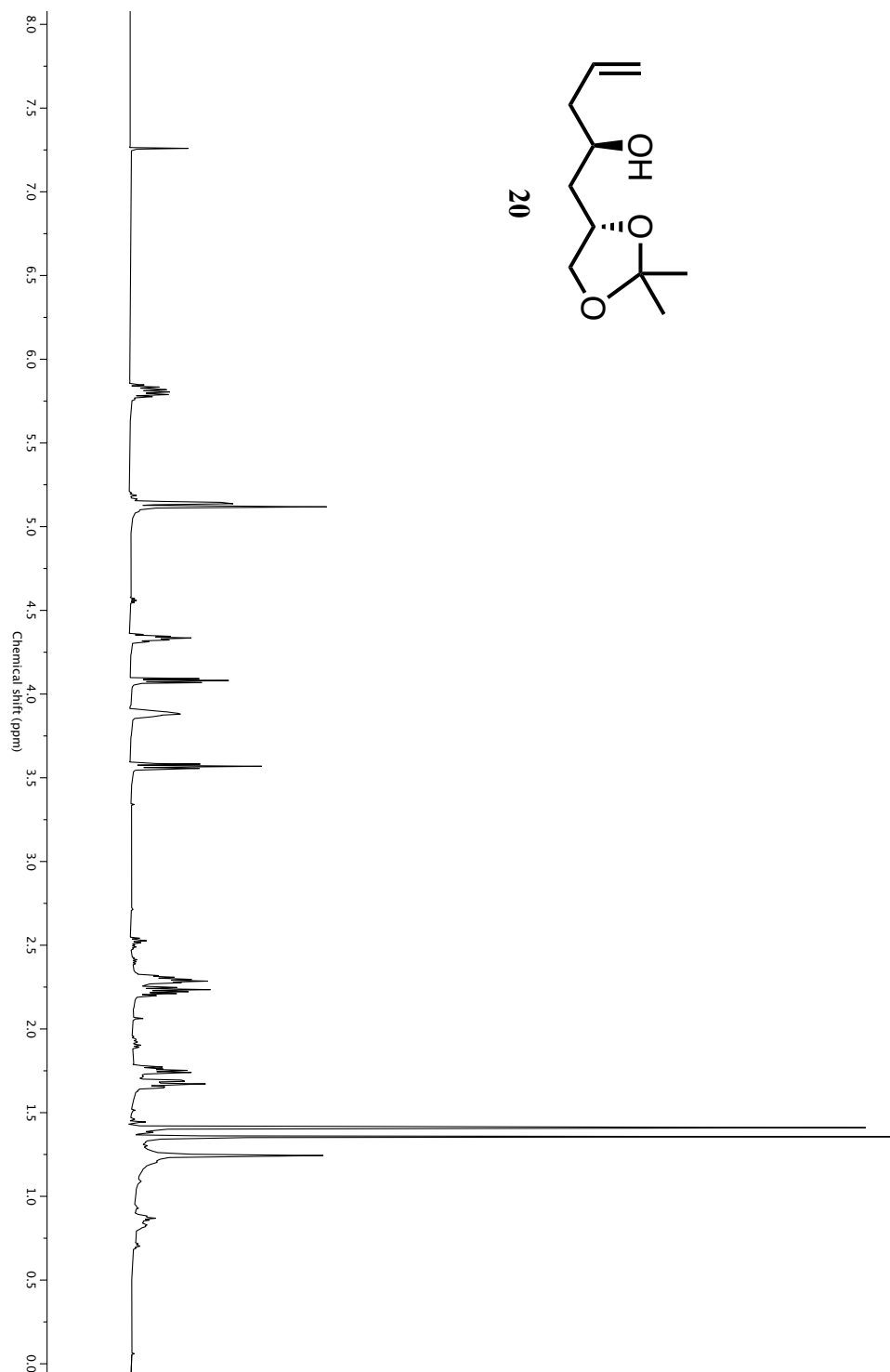
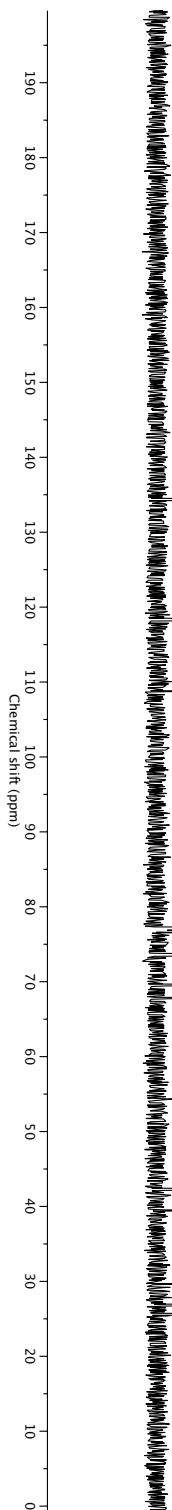
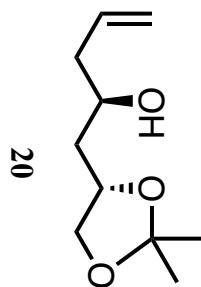


Figure C.3. Enzymatic reactions with synthetic substrate 36. Lanes 1-2 show the LCMS traces (extracted at 280 nm) of the enzymatic reactions with substrate 36 and TamI enzymes. Lane 3 includes the authentic standard for the substrate. No oxidative reactivity was identified with substrate 36 and any of the TamI enzymes.

Figure C.4 Compound characterization – Compound 20

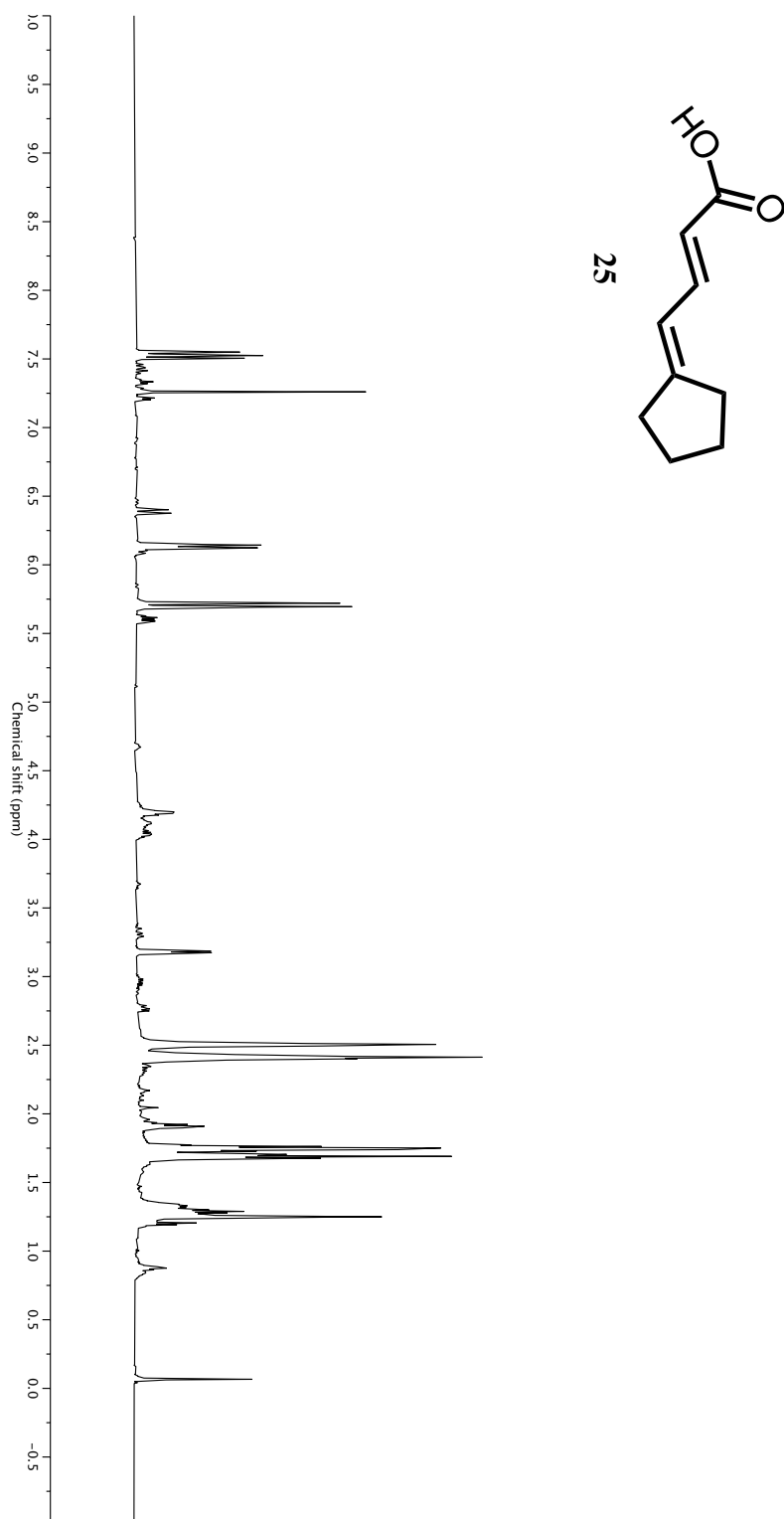


^1H spectrum of Compound 20 (500 MHz, CDCl_3).

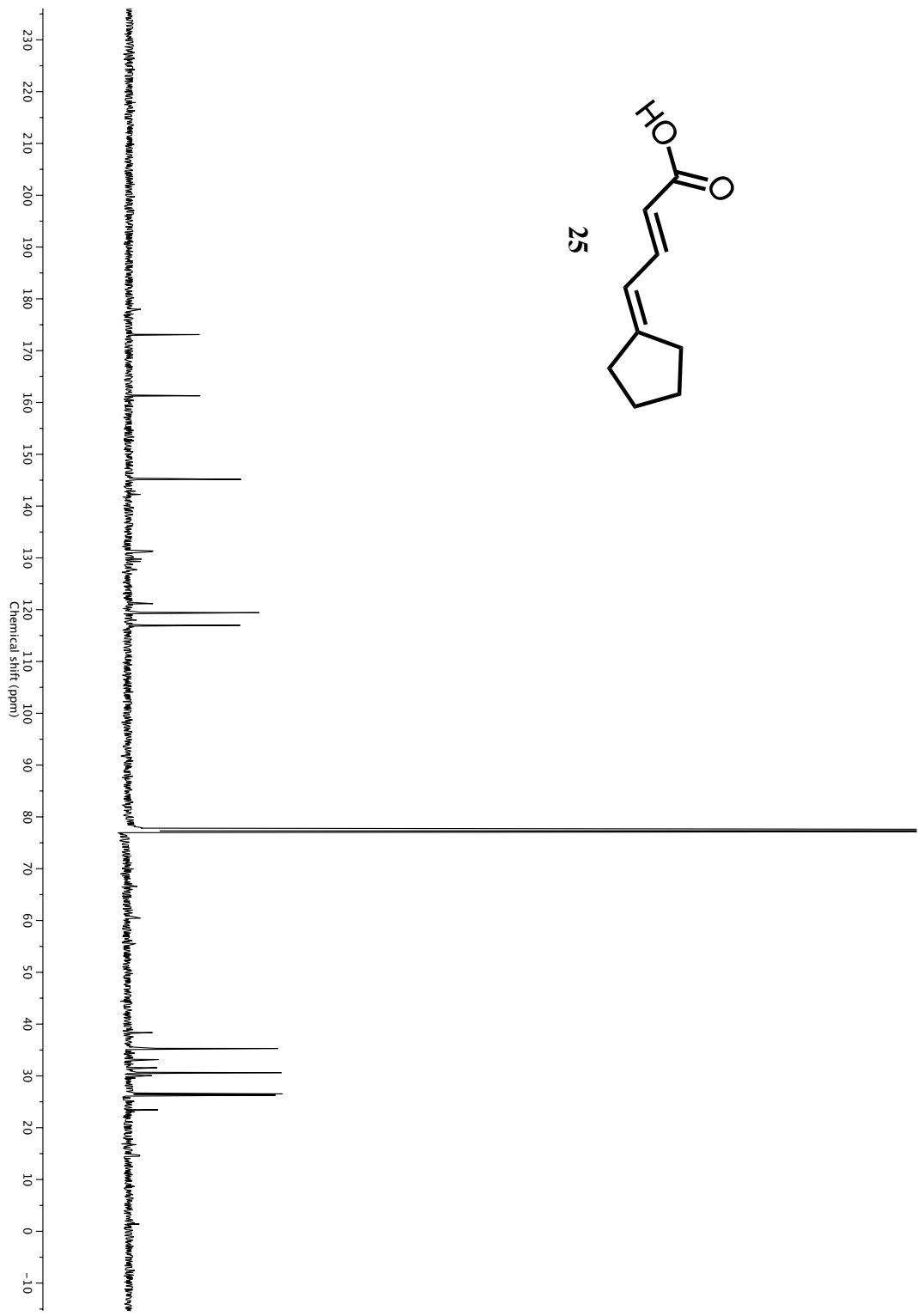
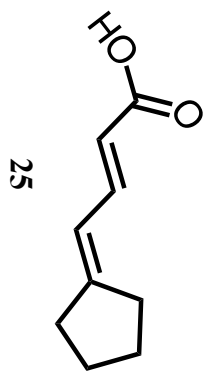


^{13}C spectrum of Compound 20 (500 MHz, CDCl_3).

Figure C.5 Compound characterization – Compound 25

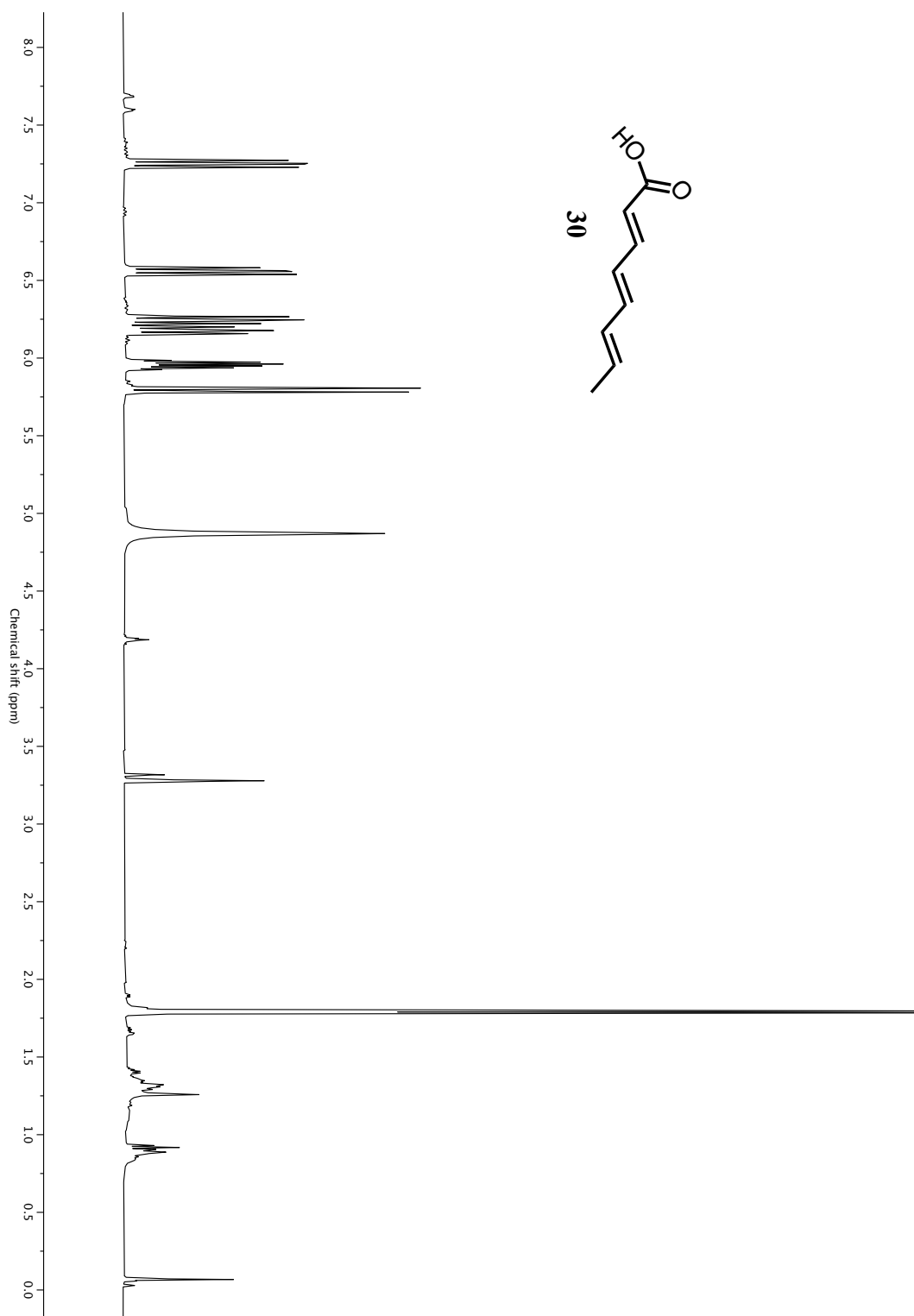


¹H spectrum of Compound 25 (500 MHz, CDCl₃).

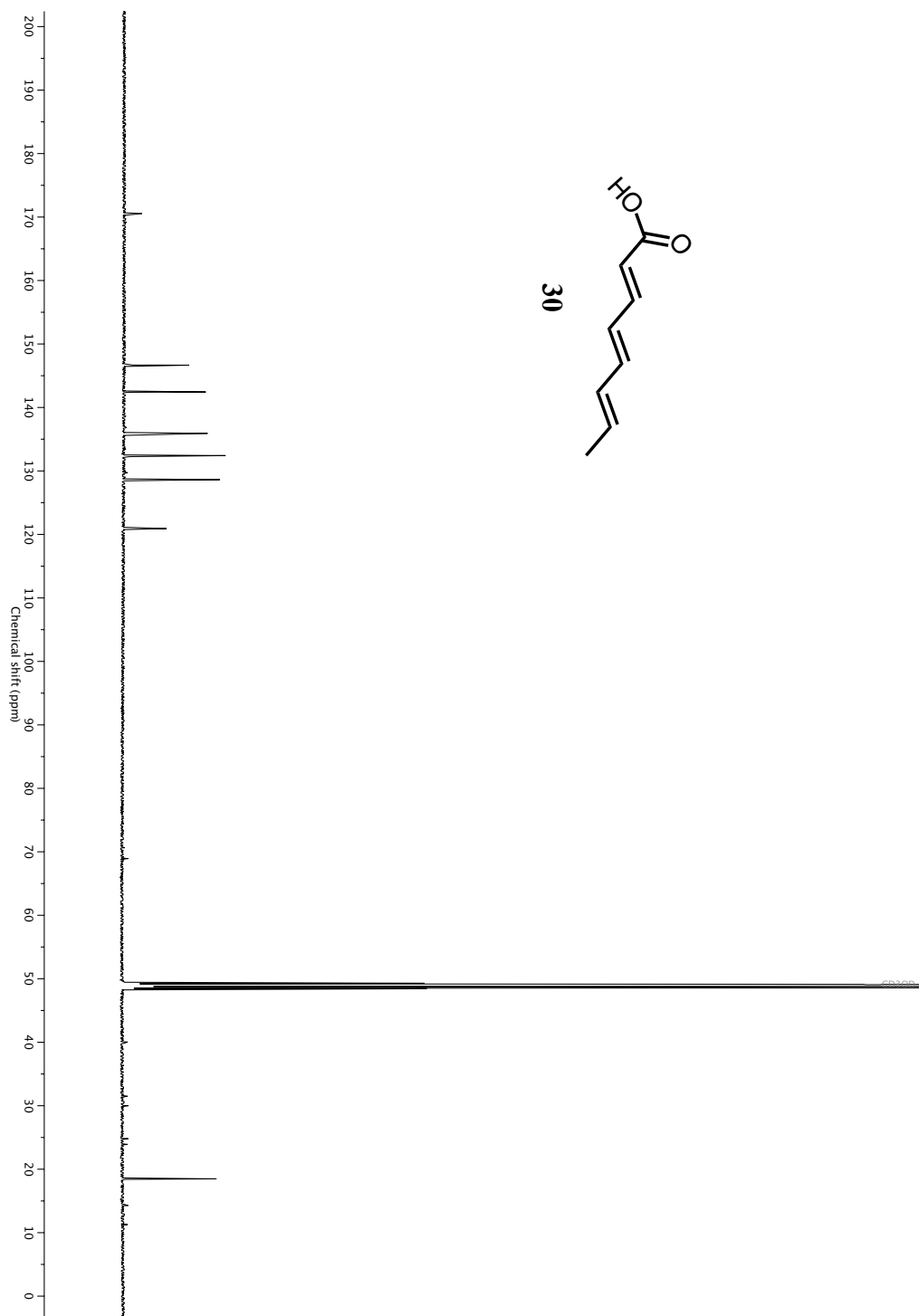


¹³C spectrum of Compound 25 (500 MHz, CDCl₃).

Figure C.6 Compound characterization – Compound 30

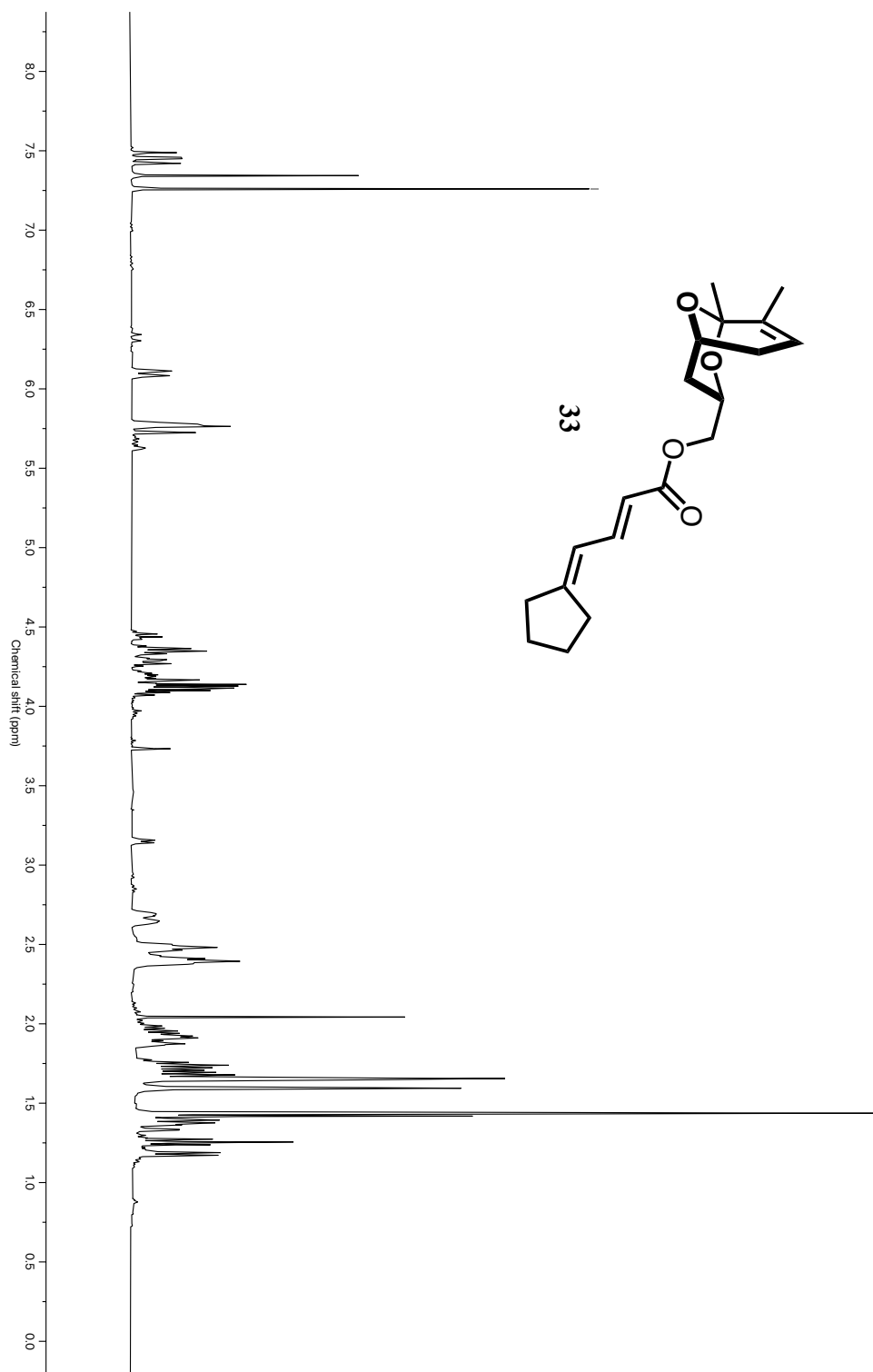


¹H spectrum of Compound 30 (500 MHz, MeOD).

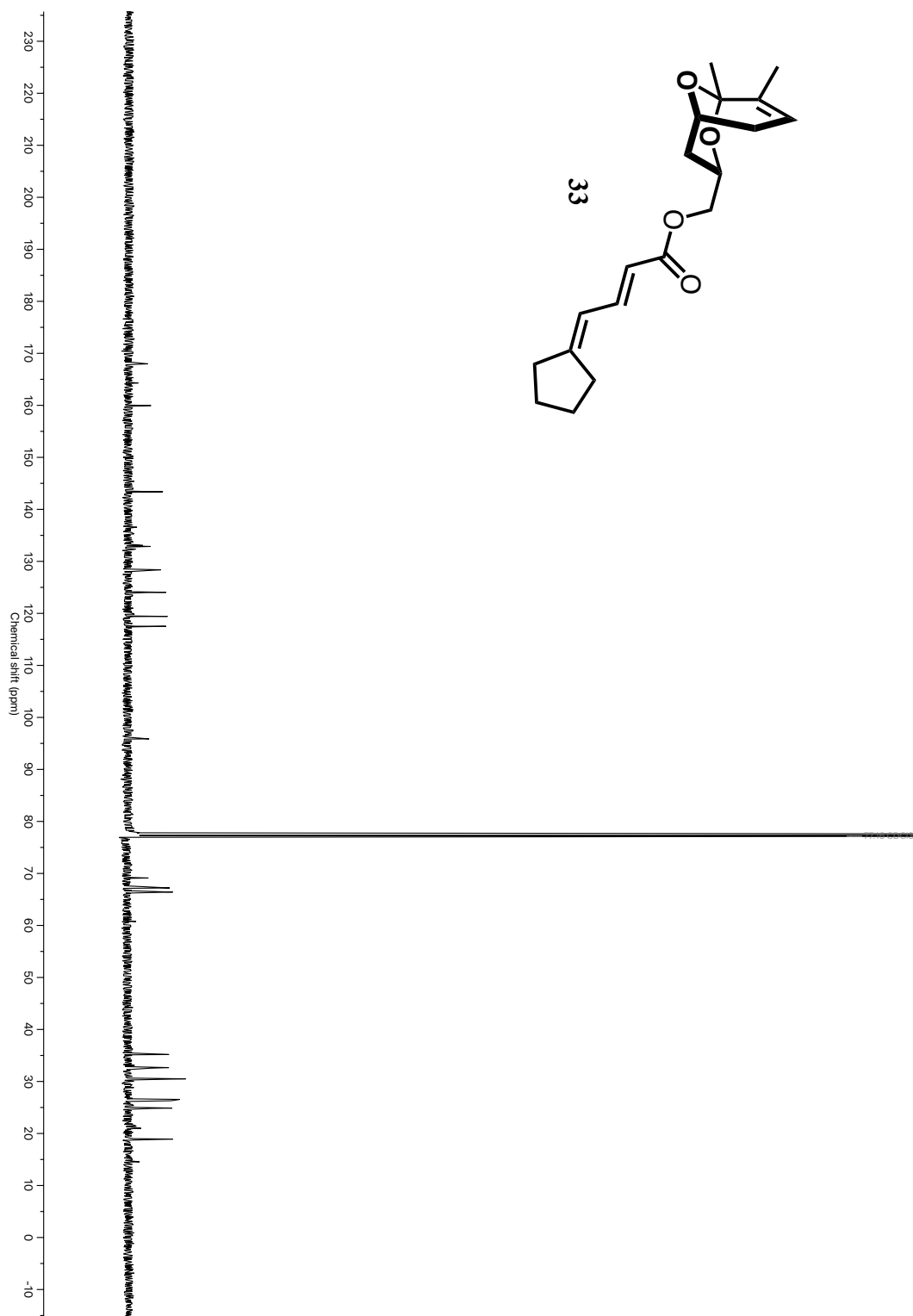


^{13}C spectrum of Compound 30 (500 MHz, MeOD).

Figure C.7 Compound characterization – Compound 33

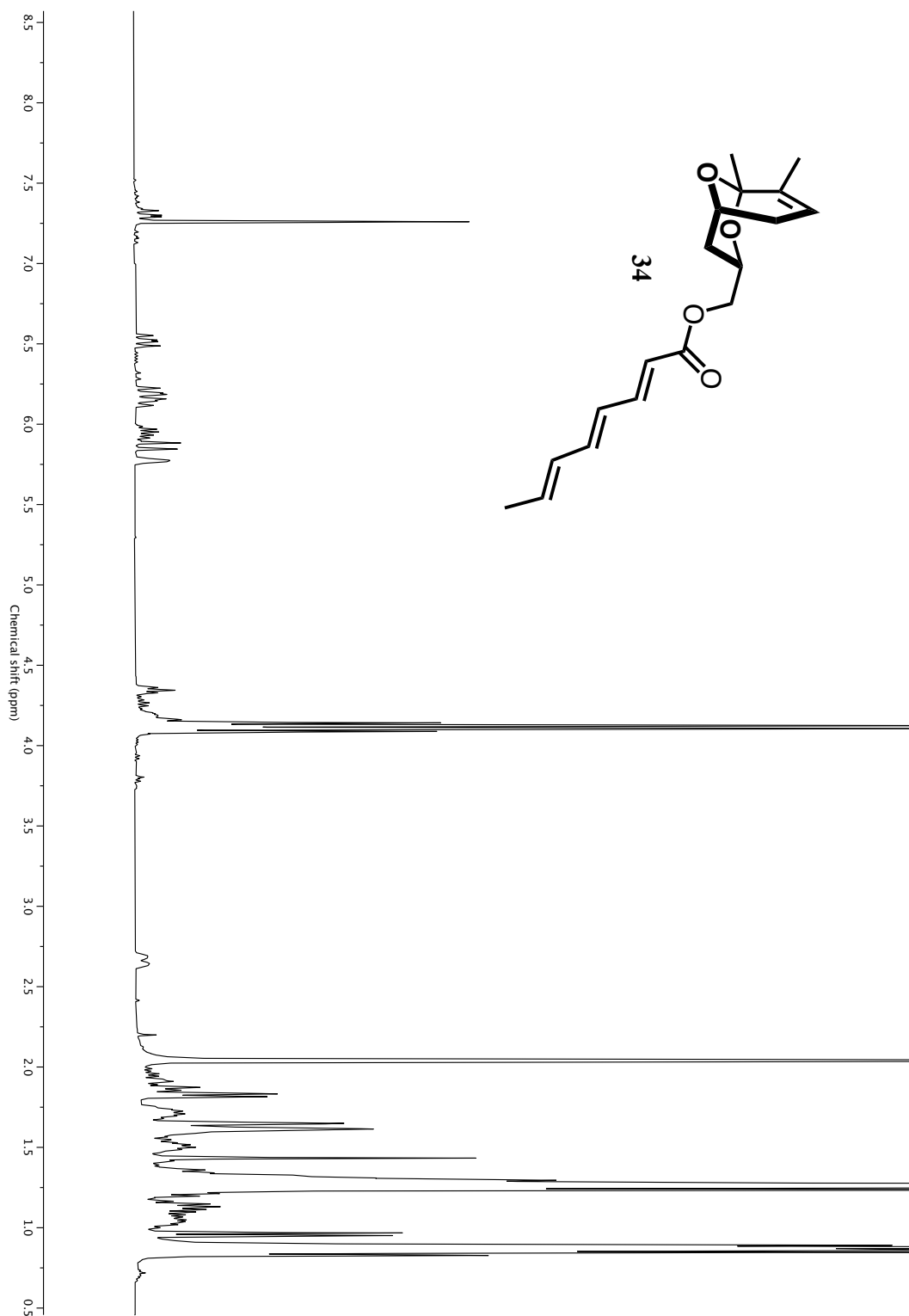


¹H spectrum of Compound 33 (500 MHz, CDCl₃).



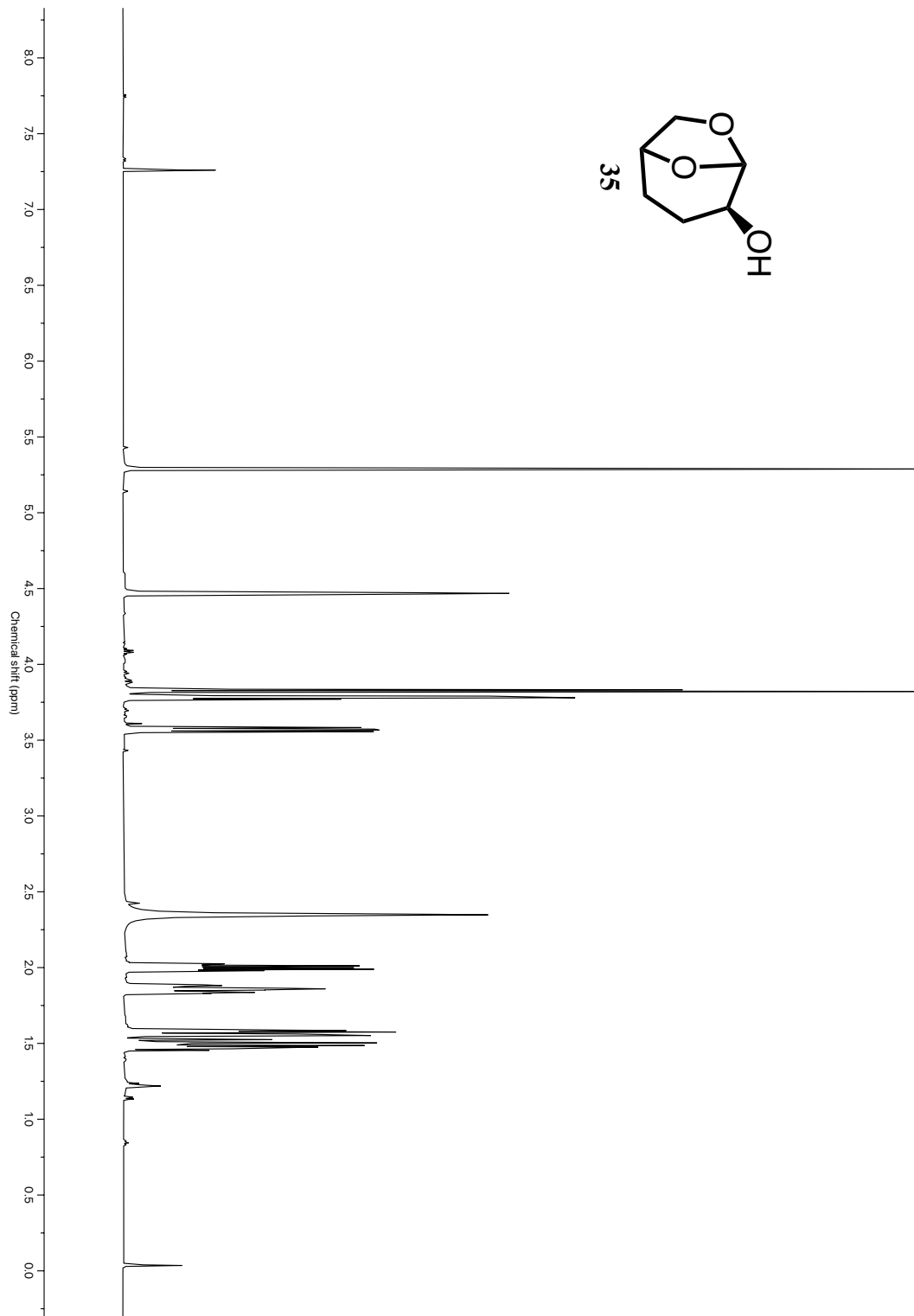
¹³C spectrum of Compound 33 (500 MHz, CDCl₃).

Figure C.8 Compound characterization – Compound 34

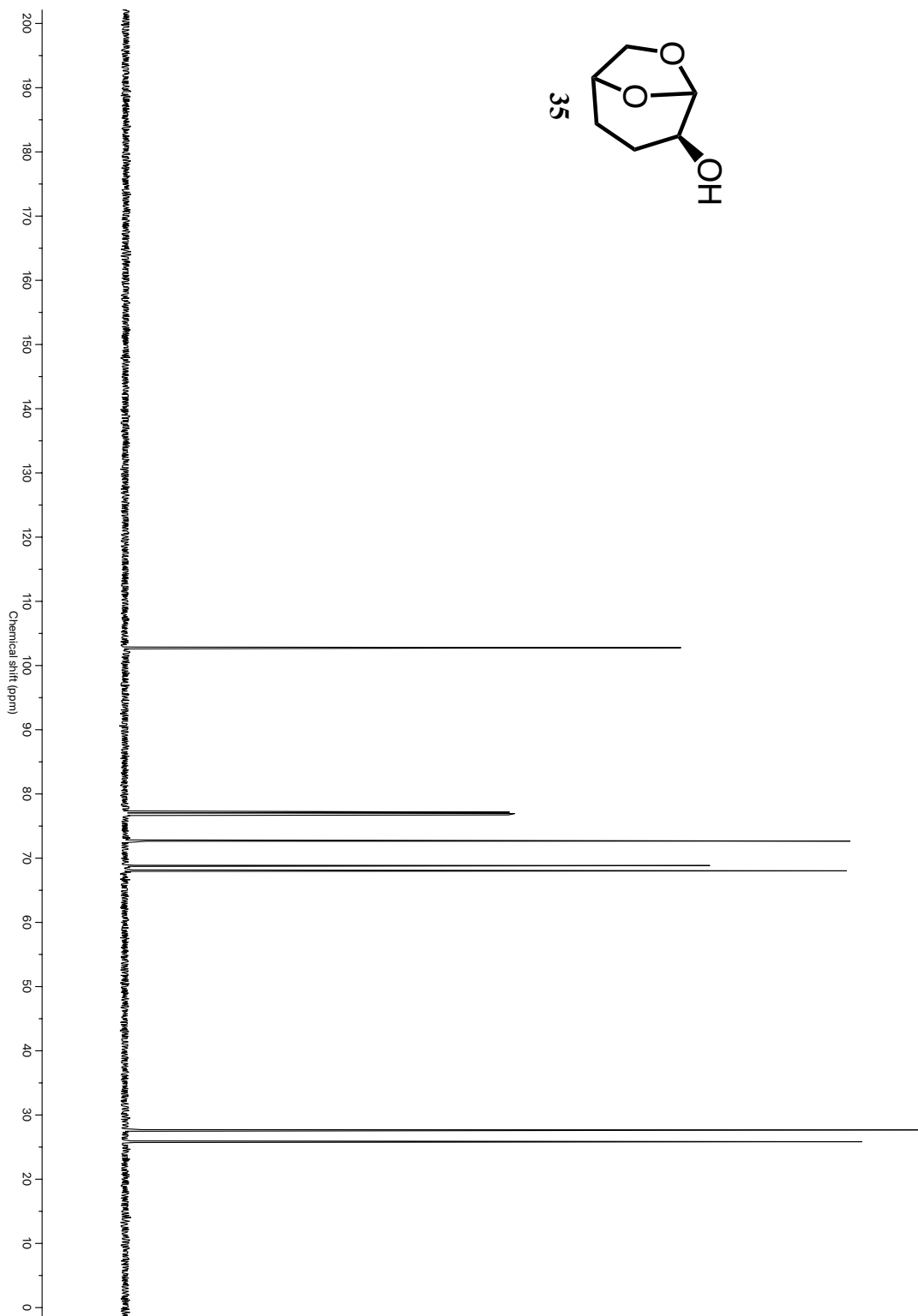
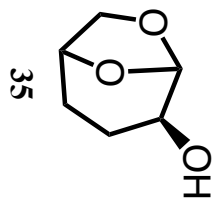


¹H spectrum of Compound 34 (500 MHz, CDCl₃).

Figure C.9 Compound characterization – Compound 35

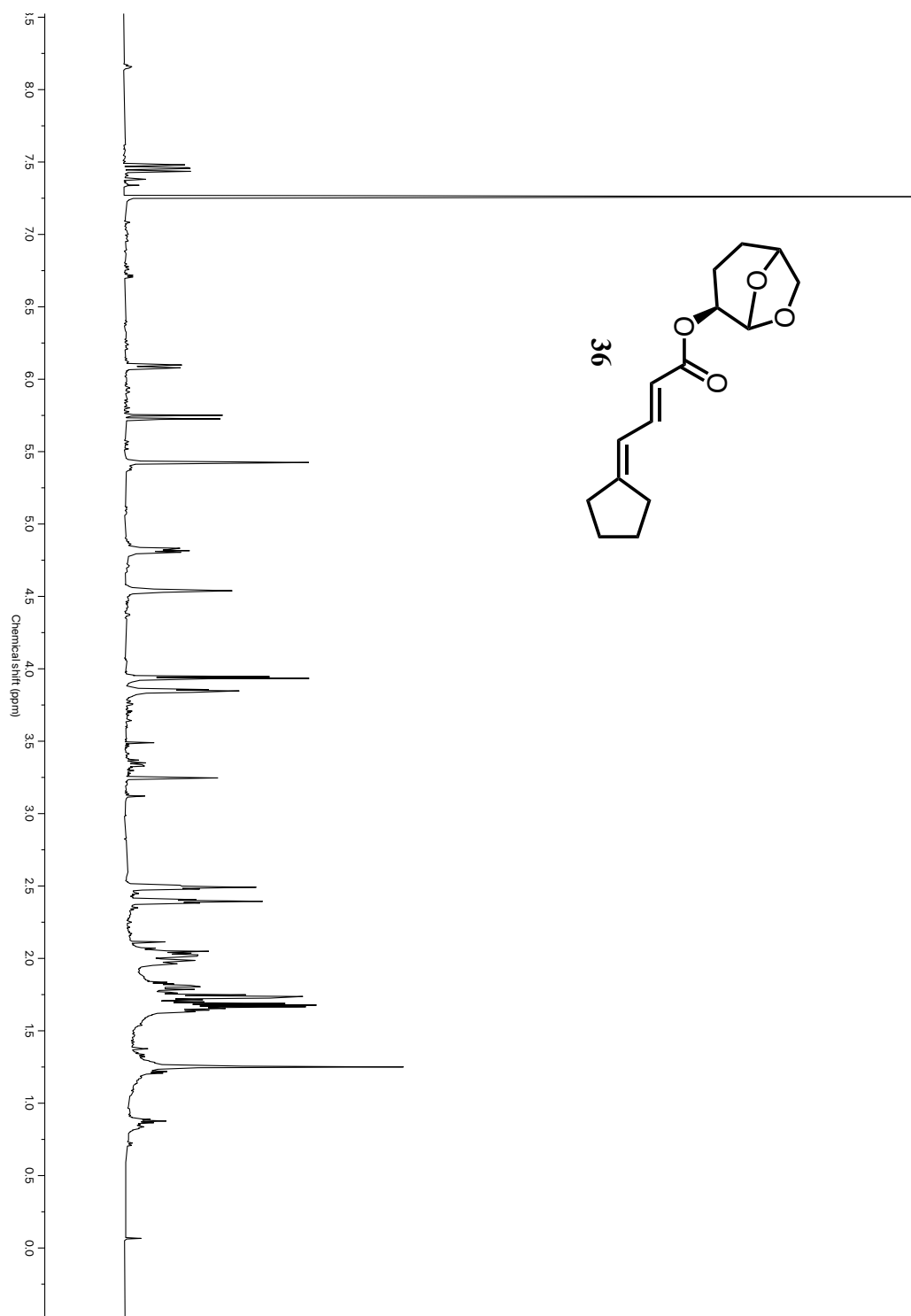


^1H spectrum of Compound 35 (600 MHz, CDCl_3).

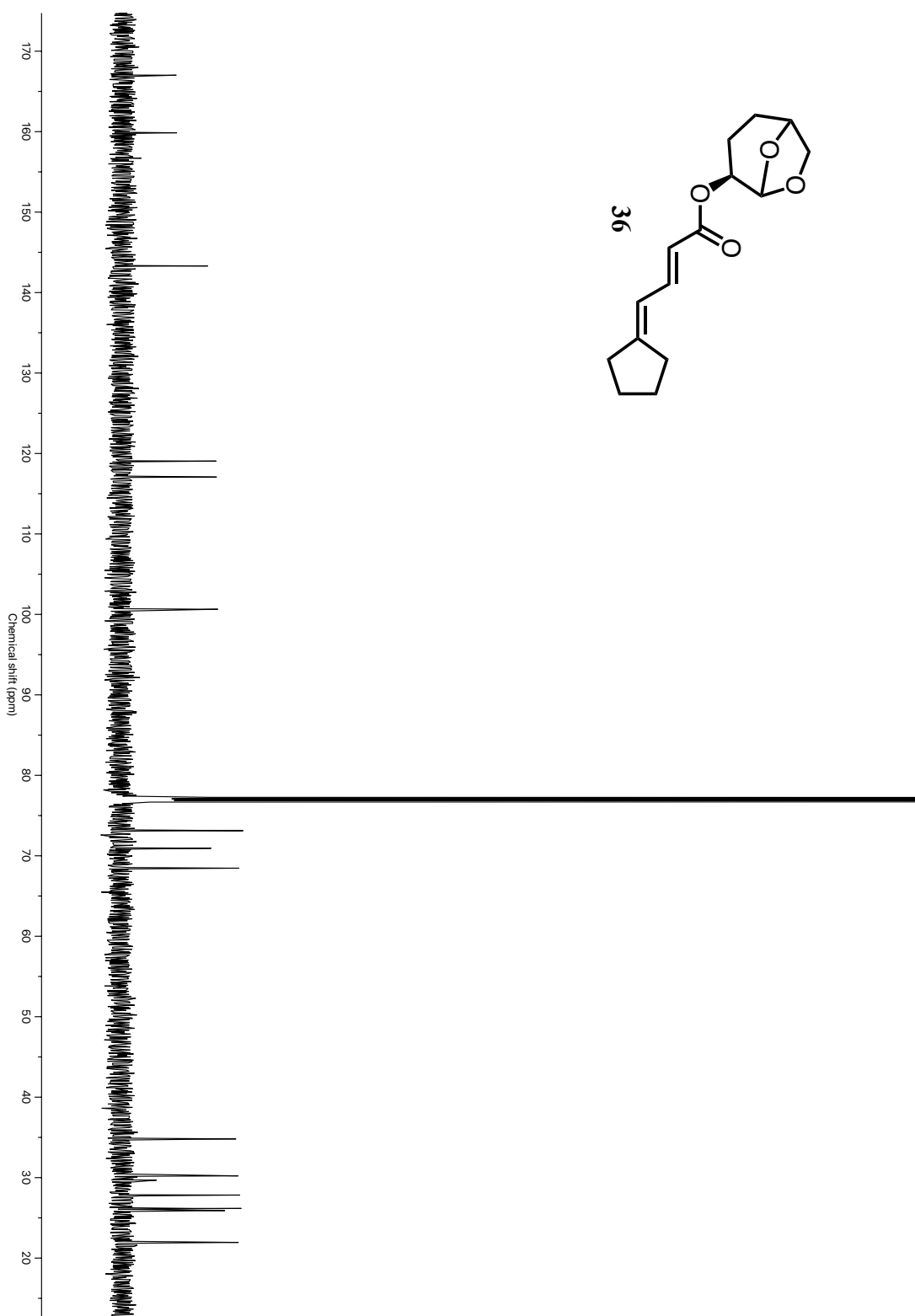


^{13}C spectrum of Compound 35 (600 MHz, CDCl_3).

Figure C.10 Compound characterization – Compound 36



¹H spectrum of Compound 36 (600 MHz, CDCl₃).



^{13}C spectrum of Compound 36 (600 MHz, CDCl_3).

C.9 References

1. Itoh, T.; Montgomery, T. P.; Recio, A., 3rd; Krische, M. J. Asymmetric alcohol C-H allylation and syn-crotylation: C9-C20 of tetrafibricin. *Org Lett* **2014**, 16, 820-823.
2. Stachowski, J. Investigation of cytochrome P450 enzymes as biocatalysts for multifunctional C-H oxidation; and a case study of a graduate/undergraduate laboratory exchange program *Doctoral Dissertation, University of Michigan* **2018**.
3. Sun, H.; Kong, R.; Zhu, D.; Lu, M.; Ji, Q.; Liew, C. W.; Lescar, J.; Zhong, G.; Liang, Z. X. Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis. *Chem Commun (Camb)* **2009**, 7399-7401.
4. Goodreid, J. D.; Wong, K.; Leung, E.; McCaw, S. E.; Gray-Owen, S. D.; Lough, A.; Houry, W. A.; Batey, R. A. Total synthesis and antibacterial testing of the A54556 cyclic acyldepsipeptides isolated from *Streptomyces hawaiiensis*. *J Nat Prod* **2014**, 77, 2170-2181.
5. Bonadies, F., Cardilli, A., Lattanzi, A., Orelli, L.R. and Scettri, A. A new procedure for Horner-Wadsworth-Emmons olefination of carbonyl compounds. *Tetrahedron Lett* **1994**, 35(20), 3383-3386.
6. Stephen A. DiBiase, B. A. L., Anthony Haag, Raymond A. Wolak, and George W. Gokel. Direct synthesis of .alpha.,.beta.-unsaturated nitriles from acetonitrile and carbonyl compounds: survey, crown effects, and experimental conditions. *J Org Chem* **1979**, 1979, 44, 25, 4640-4649
7. Aubry, S.; Lee, R. E.; Mahrous, E. A.; Small, P. L.; Beachboard, D.; Kishi, Y. Synthesis and structure of mycolactone E isolated from frog mycobacterium. *Org Lett* **2008**, 10, 5385-5388.
8. Sharipov, B. T., Davydova, A.N., Faizullina, L.K. and Valeev, F.A. Preparation of the diastereomerically pure 2S-hydroxy derivative of dihydrolevoglucosenone (cyrene). *Mendeleev Comms* **2019**, 29(2), 200-202.
9. V. Espinoza, R., Haatveit, K.C., Grossman, S.W., Tan, J.Y., McGlade, C.A., Khatri, Y., Newmister, S.A., Schmidt, J.J., Garcia-Borràs, M., Montgomery, J., Houk, K.N and D.H., Sherman. Engineering P450 TamI as an iterative biocatalyst for selective late-stage C-H functionalization and epoxidation of tirandamycin antibiotics. *ACS Catal* **2021**, 11, 8304-8316.
10. Shahrokh, K. O., A.; Yost, G. S.; Cheatham, T. E. III Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem* **2012**, 33, 119-133.
11. Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* **1996**, 9, 1-5.

APPENDIX D: Experimental Procedures and Supplemental Information for Chapter 5

D.1 Boiling River Expedition

D.1.1 Sample collection

A total of 114 sediment-containing samples were collected, processed and analyzed for this study. Six samples were collected at each sampling station, with a total of 19 stations included. Samples were collected by Rosa Vásquez Espinoza, Felipe Huanachín Carahuanco (UNALM, Laboratory of Mycology and Biotechnology) and Dr. David H. Sherman in August of 2019 with all the collection permits necessary¹ as described in Chapter 5 section 5.2.2. Refer to Table 5.1 for details on the unique sample ID codes and sampling station names. Refer to Figure D.12 for high-quality aerial photos of each station. Refer to Table D.5 for GPS coordinates for each station.

D.1.2 Physicochemical parameters

See Table 5.1 in Chapter 5 for environmental data most closely associated with each sample including temperature (°C), pH and electrical conductivity. Hach 40d, YSI63 and YSIProPlus multimeters were used to obtain the physicochemical parameters of the local waters. The values included in Table 5.1 refer to data collected between 2014-2019 by Andrés Ruzo (Boiling River Project and Southern Methodist University). Where corresponding chemical data was not available for the 2019 field season, archival water quality data is used as a best-fit reference measurement. In situ water temperatures were measured using a FLIR E95 thermal camera. For associated pH and electrical conductivity (uS/cm), water from each station was allowed to cool down prior measurement to ensure proper functioning of the equipment. All instruments used for measurements were properly calibrated prior to use in the field. An in-depth discussion on these measurement tools, calibrations and year-by-year measurements at the Boiling River site can be found in *Ruzo et al., in progress*.²

D.1.3 GPS coordinates collection

Field methods using GPS coordinates to identify the station locations and unique sample sites proved totally ineffective and unusable due to significant GPS measurement error. In order to correct this, all Boiling River stations and sample sites were site-calibrated to a high-accuracy base map created from multiple drone campaigns by the Boiling River Project. To ensure the accuracy of each location used in this study, each site was individually located and referenced on the Boiling River Project base map, using site descriptions and photographs. The resultant map provides locations accurate to within ± 3 m of their absolute position on the World Geodetic System (WGS 1984), and with < 1 m accuracy in relation to one another. A full description of the base map can be found in *Ruzo et al., in progress*.²

D.2 Metagenomics Methods

D.2.1 Isolation of purified mixed DNA

The DNeasy PowerSoil Kit (Qiagen) was used for total DNA extraction. The manufacturer's instructions were followed with minor modifications. Approximately 250-300 mg of environmental sample were added to the PowerBead Tube provided containing 75 μ L of solution C1. Samples were vortexed for a total of 8 min (1 min on, 30 sec off), centrifuged at 10,000 x g for 30 sec and transferred to a clean collection tube. A total of 300 μ L of solution C2 were added, sample was vortexed for 5 sec and incubated at 4 °C for 5 min. Sample was centrifuged for 1 min at 10,000 x g, and the supernatant was transferred to a clean collection tube containing 250 μ L of solution C3. Tube was vortexed for 5 sec and incubated at 4 °C for 5 min. Sample was centrifuged for 1 min at 10,000 x g, and the supernatant was transferred to a clean collection tube containing 1200 μ L of solution C4. Tube was inverted twice for gentle mixing. 700 μ L of sample were loaded on a MB spin column followed by centrifugation at 10,000 x g for 1 min. The flow-through was discarded, and this process was repeated until all sample was processed. Sample was washed with 550 μ L of solution C5 and centrifuged for 30 sec at 10,000 x g. After discarding the flow-through, sample was centrifuged once more at 10,000 x g for 1 min to ensure solution C5 is completely removed. 40 μ L of autoclaved H₂O were added to the column, sample was incubated at room temperature for 15-20 min (to maximize recovery) and centrifuged for 30 sec at 10,000 x g. This elution step was repeated twice, and the resulting DNA samples were combined for

quantification and quality control using a Nanodrop ND-1000 spectrophotometer. Samples were stored at – 80 °C.

D.2.2 Sequencing library preparation (16S rRNA) and Illumina MiSeq Sequencing

The 16S rRNA gene next generation sequencing library preparations and Illumina MiSeq sequencing were conducted at GENEWIZ, Inc. (South Plainfield, NJ, USA). The sequencing library was prepared using a MetaVx™ 16S rRNA Library Preparation kit (GENEWIZ, Inc., South Plainfield, NJ, USA). Briefly, the DNA was used to generate amplicons that cover V3 and V4 hypervariable regions of bacteria and archaea 16S rRNA. Indexed adapters were added to the ends of the 16S rRNA amplicons by limited cycle PCR. Pooled DNA libraries were validated and quantified before loading. The pooled DNA libraries were loaded on an Illumina MiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). The samples were sequenced using a 2x 250 paired-end (PE) configuration. Image analysis and base calling were conducted by the Illumina Control Software on the Illumina instrument. Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into FASTQ files after removing primers and de-multiplexed using Illumina's bcl2fastq 2.17 software.

D.2.3 Sequence analyses and taxonomic assessment

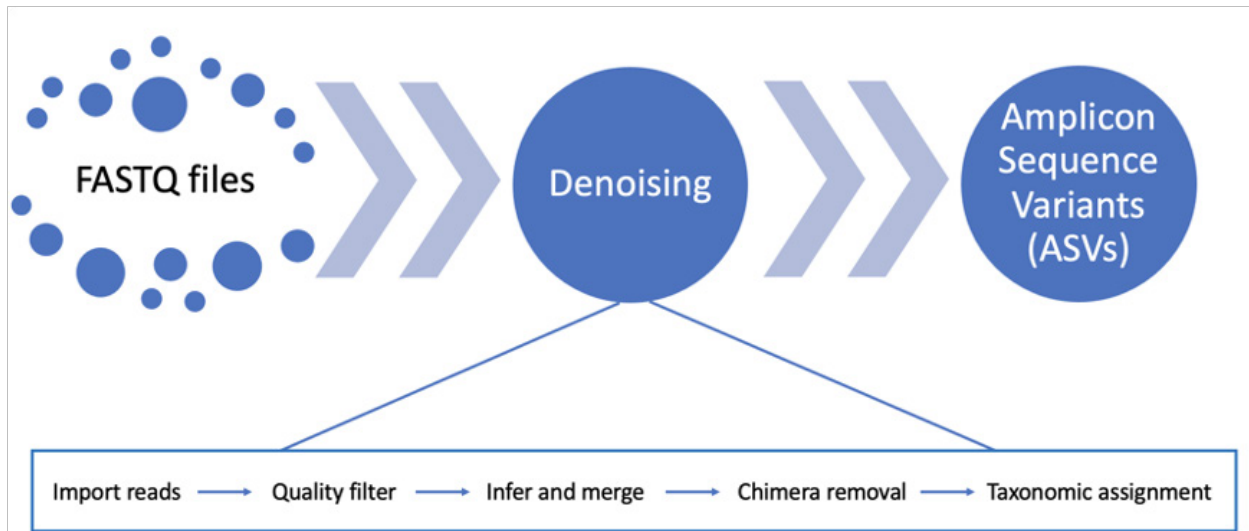
In collaboration with Felipe Huanachín Carahuano in the Laboratory of Mycology and Biotechnology at UNALM (Perú), a DADA2/QIIME2 bioinformatic pipeline was established to quality-filter, denoise, generate Amplicon Sequence Variants (ASVs), and assign taxonomy. Felipe developed a private Google cloud environment consisting of 16 CPU that could support the bioinformatic pipeline for metagenomic analyses. Demultiplexed sequences were imported into QIIME2 in the CASAVA format (1_01_L001_R1_001.fastq) and visualized using view.qiime2.org (.qzv file). We investigated and compared the optimal parameters for sequence denoising using data quality profiles and DADA2 and found that truncating reads at positions 247 (forward) and 238 (reverse) allowed the removal of low-quality sequences while maintaining sufficient length to overlap and merge matching F/R paired-end reads with a high sequence quality score (Q-value: 37). Denoising using DADA2 includes quality-based filtering, sequence inferring and merging, and chimera removal. This allows for computational detection and correction of any error in the Illumina amplicon's nucleotide sequence by correcting sample sequences and resolving differences (down to a single nucleotide) using the error rates estimated by the DADA2 algorithm

based on the input sequences. This limits the number of false or low-quality sequences that gets incorporated in the outcome read. The “classify-sklearn” command (scikit-learn 0.23.1) was used to assign taxonomy to each ASV using the Silva SSU 138.1 bacterial reference database³ and a 99% identity cutoff.

D.2.4 Statistical analyses

Statistical analyses were performed in QIIME2, R (version 4.0.1) and RStudio (version 1.3.959) including alpha diversity metrics (within-sample comparison; observed features and Shannon index) and beta diversity metrics (between-sample analysis; unweighted UniFrac distances). Permutational Multivariate Analysis of Variance (PERMANOVA) were performed with an adjusted p-value and 999 permutations. Rarefaction curves were constructed with QIIME2 at the 99% similarity level. The nonparametric Kruskal-Wallis test was used to measure microbial evenness.

D.3 Supplemental Schemes



Scheme D.1. Denoising procedure using DADA2. The steps required for sequence denoising using DADA2/QIIME 2 are shown.

D.4 Supplemental Tables

Sample ID	Station ID	Station Name	Station Coordinates
P5, P3, P10, P9, P6, P4	1BR	Poza Las Cortinas	8°48'52.25"S, 74°44'9.89"W // 8°48'51.53"S, 74°44'8.63"W // 8°48'52.88"S, 74°44'7.75"W // 8°48'53.60"S, 74°44'9.07"W
P148, P194, P195, P184, P11, P152	2BR	Escaleras del Kakataibo	8°48'55.55"S, 74°44'8.31"W // 8°48'54.89"S, 74°44'7.32"W // 8°48'55.85"S, 74°44'6.40"W // 8°48'56.66"S, 74°44'7.37"W
P108, P128, P138, P139, P166, P140	3BR	Tambos Yacumama	8°48'58.73"S, 74°43'57.95"W // 8°48'57.63"S, 74°43'58.17"W // 8°48'57.57"S, 74°43'56.68"W // 8°48'58.72"S, 74°43'56.50"W
P104, P110, P461, P426, P123, P105	4BR	Yacumama & Paredes Fumonas	8°48'58.71"S, 74°43'56.00"W // 8°48'57.57"S, 74°43'56.22"W // 8°48'57.59"S, 74°43'54.35"W // 8°48'58.69"S, 74°43'54.29"W
P87, P91, P94, P100, P96, P98	5BR	Plaza de las Fallas & KM 6	8°48'58.11"S, 74°43'52.37"W // 8°48'57.27"S, 74°43'52.25"W // 8°48'57.53"S, 74°43'51.23"W // 8°48'58.29"S, 74°43'51.43"W
P72, P74, P462, P75, P79, P81	6BR	Poza Escalera de Raíces	8°48'59.26"S, 74°43'50.60"W // 8°48'58.13"S, 74°43'50.46"W // 8°48'58.32"S, 74°43'49.14"W // 8°48'59.53"S, 74°43'49.28"W
P62, P63, P64, P65, P66, P68	7BR	Poza Dos Cataratas	8°48'59.46"S, 74°43'48.38"W // 8°48'58.38"S, 74°43'48.55"W // 8°48'58.12"S, 74°43'47.38"W // 8°48'59.15"S, 74°43'47.14"W
P42, P47, P50, P48, P45, P46	8BR	Poza del Sheripiari	8°48'55.88"S, 74°43'44.44"W // 8°48'54.86"S, 74°43'44.86"W // 8°48'54.43"S, 74°43'43.78"W // 8°48'55.41"S, 74°43'43.38"W
P30, P36, P38, P39, P477, P29	9BR	Cauce del Sheripiari	8°48'55.25"S, 74°43'43.22"W // 8°48'54.43"S, 74°43'43.55"W // 8°48'53.99"S, 74°43'42.65"W // 8°48'54.74"S, 74°43'42.27"W
P18, P25, P26, P27, P20, P16	10BR	Catarata del Sumiruna	8°48'54.48"S, 74°43'41.57"W // 8°48'53.59"S, 74°43'42.00"W // 8°48'53.00"S, 74°43'40.79"W // 8°48'53.81"S, 74°43'40.35"W
P243, P246, P247, P256, P265, P266	11BR	Cañón del Sumiruna	8°48'52.89"S, 74°43'39.74"W // 8°48'52.54"S, 74°43'40.52"W // 8°48'51.07"S, 74°43'39.83"W // 8°48'51.37"S, 74°43'39.09"W
P438, P458, P476, P465, P467, P405	12BR	Las Aguas Sagradas	8°48'49.46"S, 74°43'38.07"W // 8°48'49.07"S, 74°43'38.78"W // 8°48'47.81"S, 74°43'38.15"W // 8°48'48.28"S, 74°43'37.45"W
P162, P178, P200, P204, P181, P188	13BR	Rápidos de Mayantuyacu	8°48'47.13"S, 74°43'36.49"W // 8°48'46.79"S, 74°43'37.16"W // 8°48'45.92"S, 74°43'36.67"W // 8°48'46.34"S, 74°43'36.02"W
P190, P192, P210, P212, P213, P214	14BR	Quebradita Mayantuyacu	8°48'45.77"S, 74°43'36.21"W // 8°48'45.16"S, 74°43'37.37"W // 8°48'44.51"S, 74°43'37.08"W // 8°48'45.18"S, 74°43'35.87"W
P301, P298, P306, P307, P308, P314	15BR	Poza Salada	8°48'23.10"S, 74°44'42.27"W // 8°48'17.19"S, 74°44'41.12"W // 8°48'18.23"S, 74°44'33.14"W // 8°48'24.20"S, 74°44'33.72"W
P230, P281, P284, P288, P160, P342	16BR	Plaza de Azufre	8°48'49.67"S, 74°43'26.09"W // 8°48'47.29"S, 74°43'25.85"W // 8°48'47.63"S, 74°43'22.83"W // 8°48'49.83"S, 74°43'23.03"W
P221, P443, P259, P355, P348, P223	17BR	Catarata Impicciatori	8°48'40.06"S, 74°43'4.73"W // 8°48'38.76"S, 74°43'3.46"W // 8°48'40.43"S, 74°43'1.73"W // 8°48'41.58"S, 74°43'2.86"W
P341, P345, P352, P357, P359, P326	18BR	Escaleras de Morán	8°48'35.86"S, 74°43'7.41"W // 8°48'34.20"S, 74°43'9.22"W // 8°48'32.11"S, 74°43'6.79"W // 8°48'33.83"S, 74°43'4.97"W

P343, P318, P360, P429, P329, P349	19BR	Boca del Shanay-timpishka	8°48'2.87"S, 74°42'32.40"W // 8°48'1.82"S, 74°42'33.30"W // 8°48'0.82"S, 74°42'31.96"W // 8°48'1.86"S, 74°42'31.13"W
------------------------------------	------	---------------------------	---

Table D.1. Boiling River stations and GPS coordinates. The 114 samples collected are shown with a unique sample ID. The GPS coordinates include four points (each with a unique latitude and longitude) that make up a polygon shape encompassing the area of study per station as seen in Figure 5.4.

Sequence Count	Min Length	Max Length	Mean Length	Range	Standard Deviation		
60296	247	473	443.3	226	30.74		
Percentile:	2%	9%	25%	50%	75%	91%	98%
Length* (nts):	292	432	436	450	460	464	467

Table D.2. Sequence length statistics.

Metric	Sample
Number of samples	114
Number of features	60,296
Total frequency	7,992,826
	Frequency
Minimum frequency	1.0
1st quartile	9.0
Median frequency	54.0
3rd quartile	172.0
Maximum frequency	7,307.0
Mean frequency	132.55980496218655

Table D.3. Frequency per ASV feature.

	forward reads	reverse reads
Minimum	35143	35143
Median	243800	243800
Mean	247918	247918
Maximum	850821	850821
Total	28262668	28262668

Table D.4. Demultiplexed sequence counts summary.

Order taxa	Relative abundance (%)
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales	11.59
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales	8.66
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales	4.7
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales	4.58
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales	4.25
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales	3.44
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales	3.23
d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales	3.22
d__Bacteria;p__Actinobacteriota;c__Thermoleophilia;o__Solirubrobacterales	3.08
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales	2.98
d__Bacteria;p__Cyanobacteria;c__Cyanobacteriia;o__Cyanobacterales	2.76
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales	2.71
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Frankiales	2.5
d__Bacteria;p__Actinobacteriota;c__Thermoleophilia;o__Gaiellales	2.4
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales	2.31
d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Gemmatimonadales	2.25
d__Bacteria;p__Chloroflexi;c__Chloroflexia;o__Chloroflexales	2.17
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacterales	2.16
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Acetobacterales	2
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacterales	1.98
d__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales	1.98
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Propionibacterales	1.95
d__Bacteria;p__Chloroflexi;c__Ktedonobacteria;o__Ktedonobacterales	1.86
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales	1.59
d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Microtrichales	1.57
d__Bacteria;p__Firmicutes;c__Clostridia;o__Peptostreptococcales-Tissierellales	1.55
d__Bacteria;p__Firmicutes;c__Negativicutes;o__Veillonellales-Selenomonadales	1.52
d__Bacteria;p__Firmicutes;c__Bacilli;o__Paenibacillales	1.33
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Pseudonocardiales	1.26
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micromonosporales	1.17
d__Bacteria;p__Acidobacteriota;c__Acidobacteriiae;o__Bryobacterales	1.09
d__Bacteria;p__Myxococcota;c__Polyangia;o__Polyangiales	1.07
d__Bacteria;p__Acidobacteriota;c__Acidobacteriiae;o__Acidobacterales	1.01
d__Bacteria;p__Chloroflexi;c__Anaerolineae;o__SBR1031	1.01
d__Bacteria;p__Chloroflexi;c__Chloroflexia;o__Thermomicrobiales	0.98
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales	0.94
d__Bacteria;p__Firmicutes;c__Clostridia;o__Lachnospirales	0.91
d__Bacteria;p__Firmicutes;c__Clostridia;o__Oscillospirales	0.87
d__Bacteria;p__Acidobacteriota;c__Blastocatellia;o__Pyrinomonadales	0.86
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Tistrellales	0.86
d__Bacteria;p__Myxococcota;c__Polyangia;o__Haliangiales	0.83
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Streptomycetales	0.81

Table D.5. Bacterial abundance at the order level. Cutoff 0.7%

Family taxa	Relative abundance (%)
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae	8.72
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Microscillaceae	5.11
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Propionibacteriales;f__Nocardoidaceae	4.90
d__Bacteria;p__Actinobacteriota;c__Thermoleophilla;o__Solirubrobacterales;f__Solirubrobacteraceae	4.58
d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacteriales;f__Vicinamibacteraceae	4.46
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae	3.88
d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacteriales;f__uncultured	3.84
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Frankiales;f__Acidothermaceae	3.62
d__Bacteria;p__Actinobacteriota;c__Thermoleophilla;o__Gaiellales;f__uncultured	3.33
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Pseudonocardiales;f__Pseudonocardaceae	3.26
d__Bacteria;p__Actinobacteriota;c__Thermoleophilla;o__Solirubrobacterales;f__67-14	3.18
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micromonosporales;f__Micromonosporaceae	3.04
d__Bacteria;p__Chloroflexi;c__Chloroflexia;o__Chloroflexales;f__Chloroflexaceae	2.88
d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Bryobacteriales;f__Bryobacteraceae	2.82
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Microbacteriaceae	2.47
d__Bacteria;p__Actinobacteriota;c__Thermoleophilla;o__Gaiellales;f__Gaiellaceae	2.34
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae	2.32
d__Bacteria;p__Actinobacteriota;c__Blastocatellia;o__Pyrinomonadales;f__Pyrinomonadaceae	2.24
d__Bacteria;p__Acidobacteriota;c__Blastocatellia;o__Blastocatellales;f__Blastocatellaceae	2.17
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Streptomycetales;f__Streptomycetaceae	2.10
d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Microtrichales;f__Illumatobacteraceae	2.05
d__Bacteria;p__Actinobacteriota;c__MB-A2-108;o__MB-A2-108;f__MB-A2-108	1.82
d__Bacteria;p__Chloroflexi;c__Anaerolineae;o__SBR1031;f__A4b	1.65
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Prolixibacteraceae	1.64
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Saprosipiraceae	1.63
d__Bacteria;p__Armatimonadota;c__uncultured;o__uncultured;f__uncultured	1.61
d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__IMCC26256;f__IMCC26256	1.42
d__Bacteria;p__Acidobacteriota;c__Thermoanaerobaculia;o__Thermoanaerobaculales;f__Thermoanaerobaculaceae	1.41
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Micrococcaceae	1.40
d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Solibacteriales;f__Solibacteraceae	1.31
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Intrasporangiaceae	1.29
d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Subgroup_2;f__Subgroup_2	1.27
d__Bacteria;p__Acidobacteriota;c__Holophagae;o__Subgroup_7;f__Subgroup_7	1.18
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Frankiales;f__Frankiaceae	1.18
d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Acidobacteriales;f__uncultured	1.18
d__Bacteria;p__Acidobacteriota;c__Blastocatellia;o__11-24;f__11-24	1.17
d__Bacteria;p__Bacteroidota;c__Kryptonia;o__Kryptoniales;f__BSV26	1.14
d__Bacteria;p__Aquificota;c__Aquificae;o__Hydrogenothermales;f__Hydrogenothermaceae	1.13
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Cyclobacteriaceae	1.12
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidetes_vadinHA17	1.09
d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Acidobacteriales;f__Acidobacteriaceae_(Subgroup_1)	1.08

Table D.6. Bacterial abundance at the family level. Cutoff 0.7%

Genera taxa	Relative abundance (%)
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas	6.12
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Brevundimonas	5.91
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__	4.67
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas	3.80
d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Gemmatimonadales;f__Gemmatimonadaceae;g__uncultured	3.69
d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__uncultured;g__uncultured	3.56
d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__Vicinamibacteraceae;g__Vicinamibacteraceae	3.52
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium	3.50
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Propionibacteriales;f__Nocardioideae;g__Nocardioides	3.41
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Frankiales;f__Acidothermaceae;g__Acidothermus	3.36
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__	3.20
d__Bacteria;p__Actinobacteriota;c__Thermoleophillia;o__Gaiellales;f__uncultured;g__uncultured	3.09
d__Bacteria;p__Actinobacteriota;c__Thermoleophillia;o__Solirubrobacterales;f__67-14;g__67-14	2.95
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Acetobacterales;f__Acetobacteraceae;g__	2.90
d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus	2.88
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Microsillaceae;g__uncultured	2.83
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter	2.65
d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Bryobacterales;f__Bryobacteraceae;g__Bryobacter	2.62
d__Bacteria;p__Methylomirabilota;c__Methylomirabilia;o__Rokubacteriales;f__Rokubacteriales;g__Rokubacteriales	2.53
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__uncultured	2.37
d__Bacteria;p__GAL15;c__GAL15;o__GAL15;f__GAL15;g__GAL15	2.32
d__Bacteria;p__Firmicutes;c__Bacilli;o__Paenibacillales;f__Paenibacillaceae;g__Paenibacillus	2.30
d__Bacteria;p__Actinobacteriota;c__Thermoleophillia;o__Solirubrobacterales;f__Solirubrobacteraceae;g__Conexibacter	2.25
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium	2.20
d__Bacteria;p__Actinobacteriota;c__Thermoleophillia;o__Gaiellales;f__Gaiellaceae;g__Gaiella	2.17
d__Bacteria;p__Acidobacteriota;c__Blastocatellia;o__Pyrinomonadales;f__Pyrinomonadaceae;g__RB41	2.07
d__Bacteria;p__Deinococcota;c__Deinococci;o__Thermales;f__Thermaceae;g__Meiothermus	1.85
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Streptomycetales;f__Streptomycetaceae;g__Streptomyces	1.84
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__	1.75
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Gammaproteobacteria_Incertae_Sedis;f__Unknown_Family;g__Acidibacter	1.74
d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Hydrogenophaga	1.69
d__Bacteria;p__Actinobacteriota;c__MB-A2-108;o__MB-A2-108;f__MB-A2-108;g__MB-A2-108	1.68
d__Bacteria;p__Chloroflexi;c__Chloroflexia;o__Thermomicrobiales;f__JG30-KF-CM45;g__JG30-KF-CM45	1.67
d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__uncultured;f__uncultured;g__uncultured	1.67
d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micromonosporales;f__Micromonosporaceae;g__	1.64
d__Bacteria;p__Chloroflexi;c__KD4-96;o__KD4-96;f__KD4-96;g__KD4-96	1.61

Table D.7. Bacterial abundance at the genera level. Cutoff 0.5%

D.5 Supplemental Figures

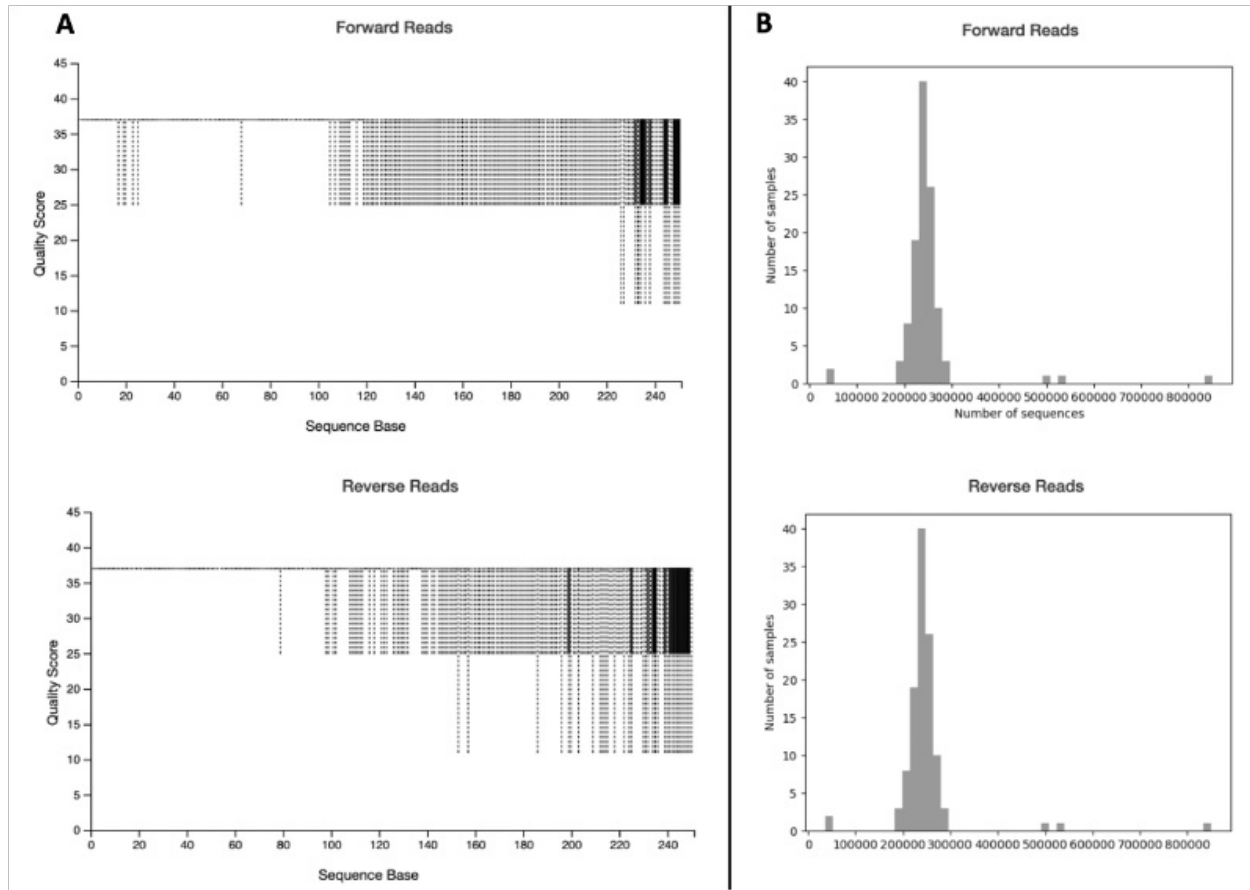


Figure D.1. Sequence quality control. A) Quality plots of the forward and reverse reads visualized with QIIME 2 View shows a high-quality sequence with Q-value of 37. B) Forward and reverse reads frequency histograms.

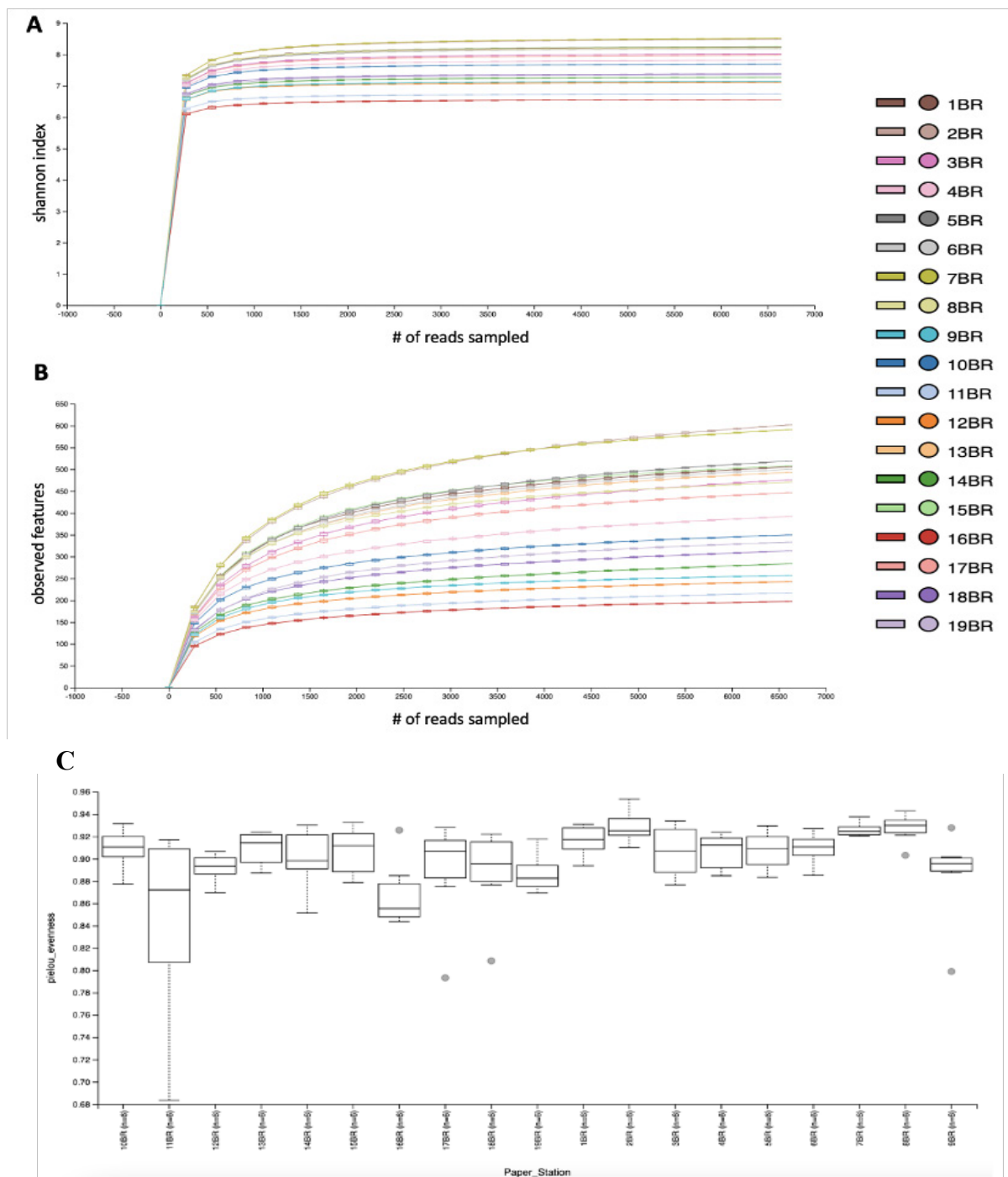
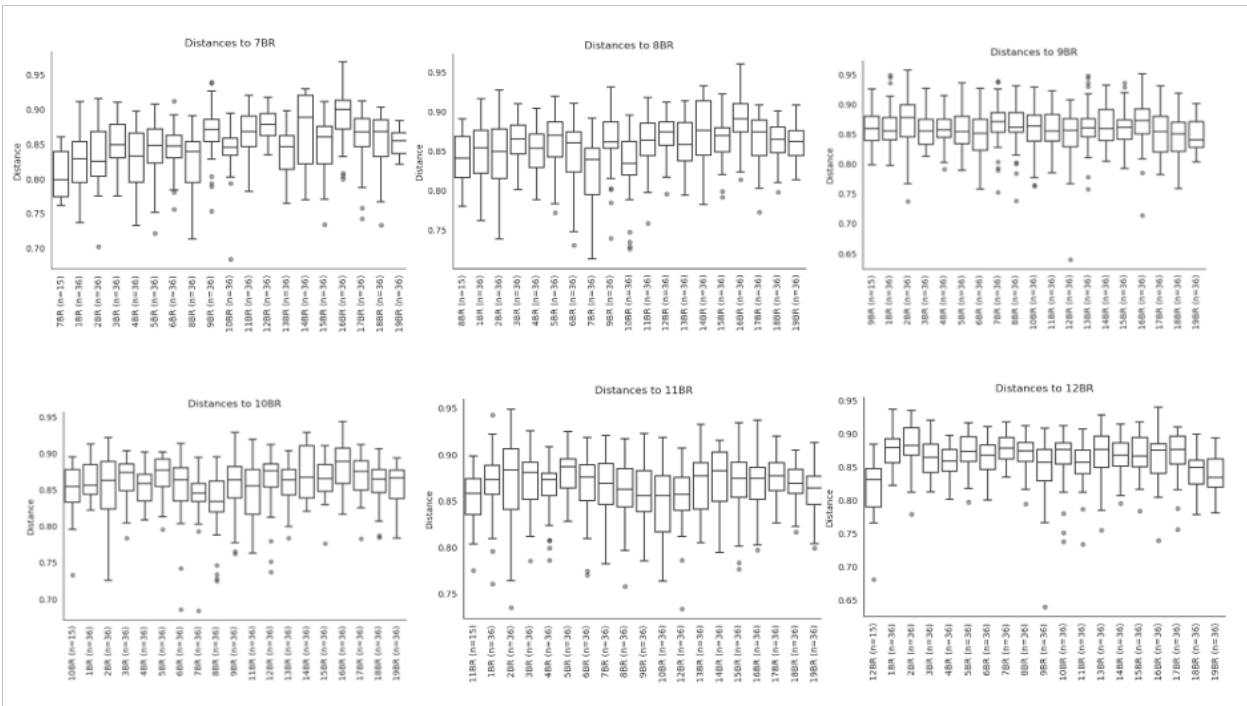
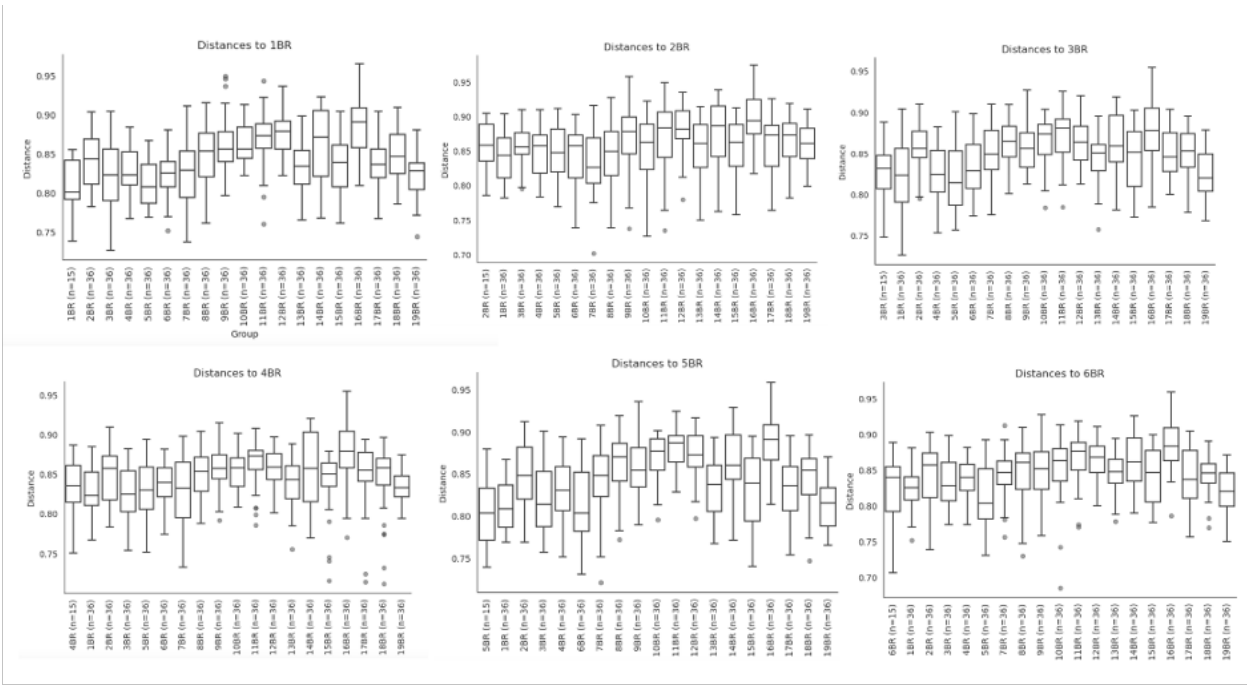


Figure D.2. Alpha diversity metrics. A) Rarefaction curve sequences for Shannon diversity indices of bacteria in Boiling River samples show that the curves almost reached a plateau indicating that the sequencing depth was sufficient to capture and identify the full scope of local microdiversity. The curve was constructed with QIIME2 and calculated at the 99% similarity level. B) The number of observed features (ASVs) is shown. C) Microbial evenness shown as boxplots. The nonparametric Kruskal-Wallis test was used and a p-value of 0.00212 was obtained.



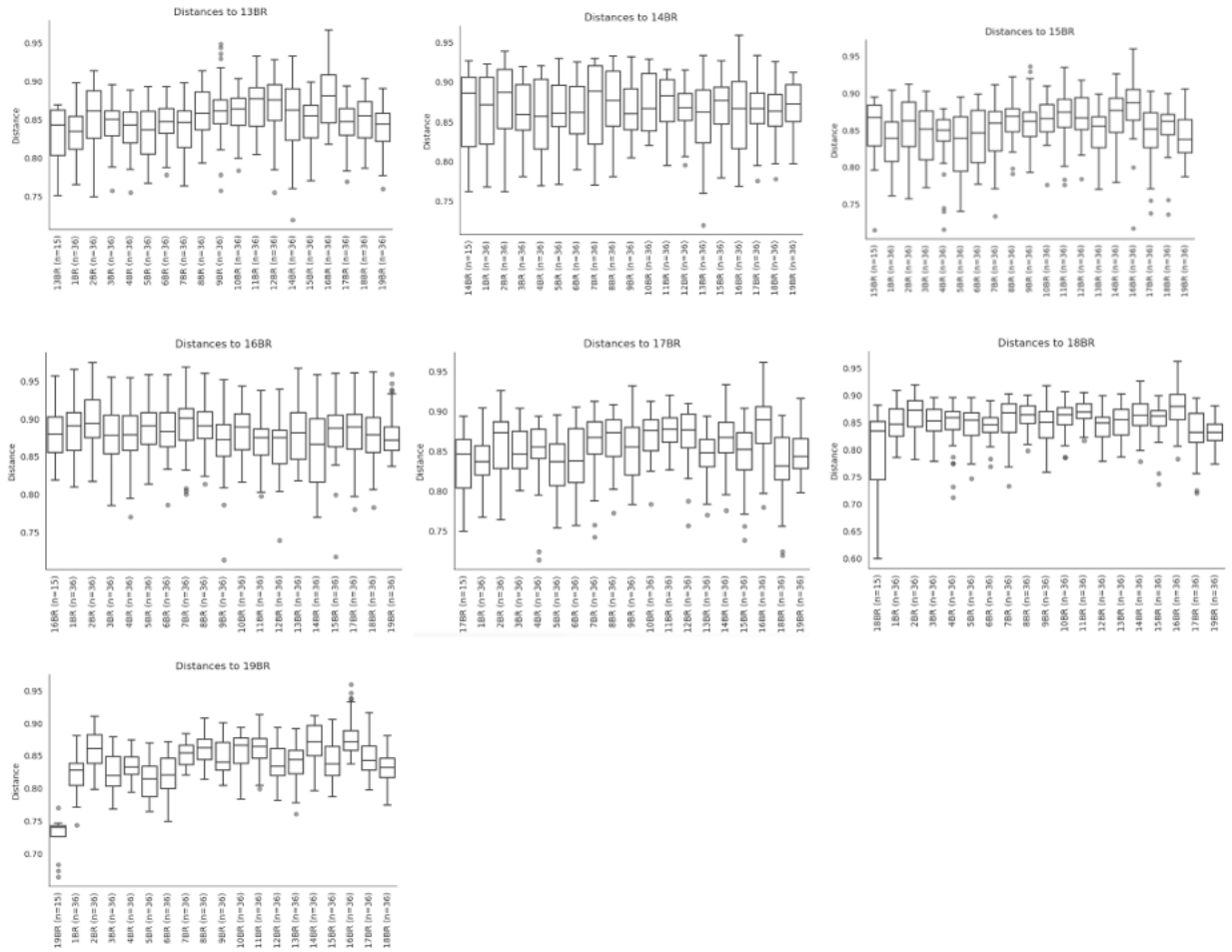


Figure D.3. Beta diversity analyses. These metrics show the unweighted UniFrac distances between the stations' microbiomes. Boxplots summarize the intra- and interstation distances. PERMANOVA analyses indicated that the beta diversity of the microbiome of Boiling River stations did not differ in a statistically significant manner.

pairs	F.Model	R2	p.value	p.adjusted	sig
10 vs 6	0.9511304	0.08685226	0.391	1.000	
10 vs 7	0.6828313	0.06391857	0.649	1.000	
10 vs 8	0.8968962	0.08230749	0.554	1.000	
10 vs 9	2.2896484	0.18630707	0.098	1.000	
10 vs 1	1.6256506	0.13983309	0.142	1.000	
10 vs 14	3.4174443	0.25470158	0.005	0.855	
10 vs 11	1.2129469	0.10817378	0.311	1.000	
10 vs 12	2.2576353	0.18418196	0.090	1.000	
10 vs 16	1.2960710	0.11473644	0.242	1.000	
10 vs 13	1.2286184	0.10941848	0.265	1.000	
10 vs 2	1.8868890	0.15873699	0.109	1.000	
10 vs 15	0.4301050	0.04123688	0.804	1.000	
10 vs 19	2.7525506	0.21584315	0.009	1.000	
10 vs 17	2.6392218	0.20881205	0.055	1.000	

Figure D.4. PERMANOVA analyses between stations at the phyla level. No significant difference was found in the microbial composition of Boiling River stations. Adjusted p-value: 0.001; Number of permutations: 999; Test statistics: 1.36352

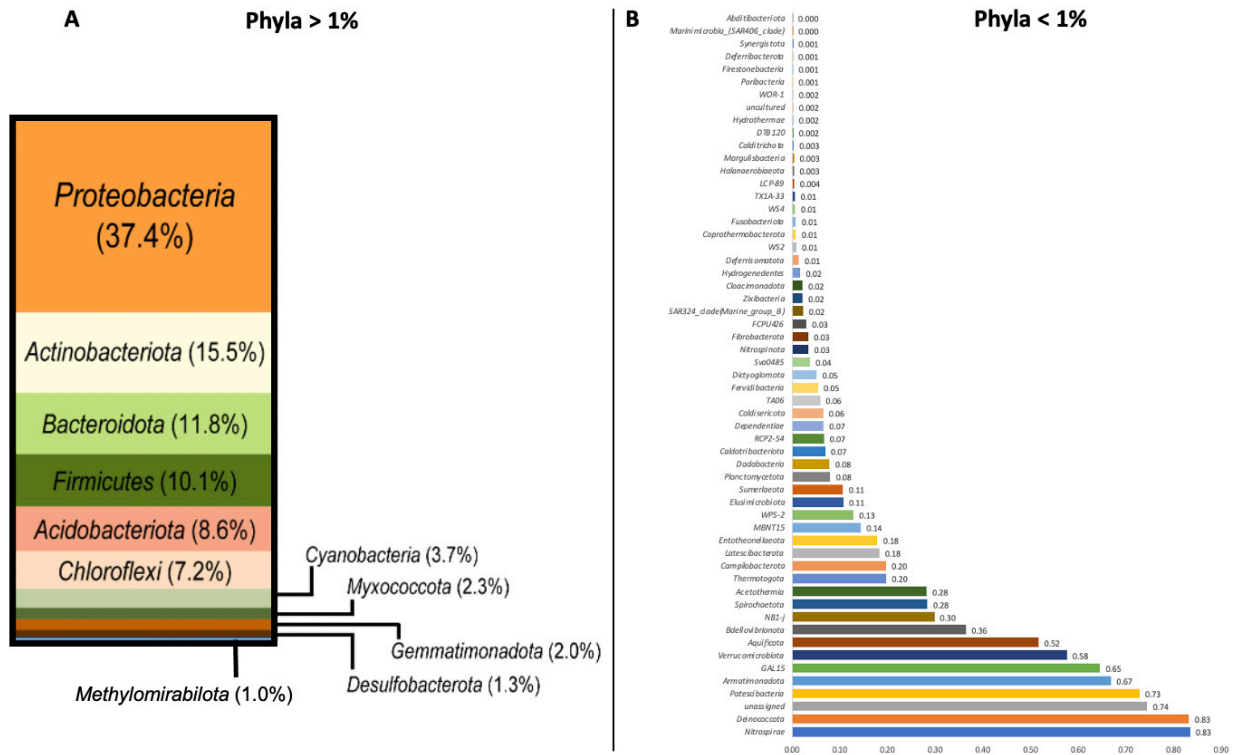


Figure D.6. Total bacterial phyla composition in the Boiling River. A) Phyla groups with a 1% or more abundance. B) Least numerous phyla groups (less than 1% abundance).

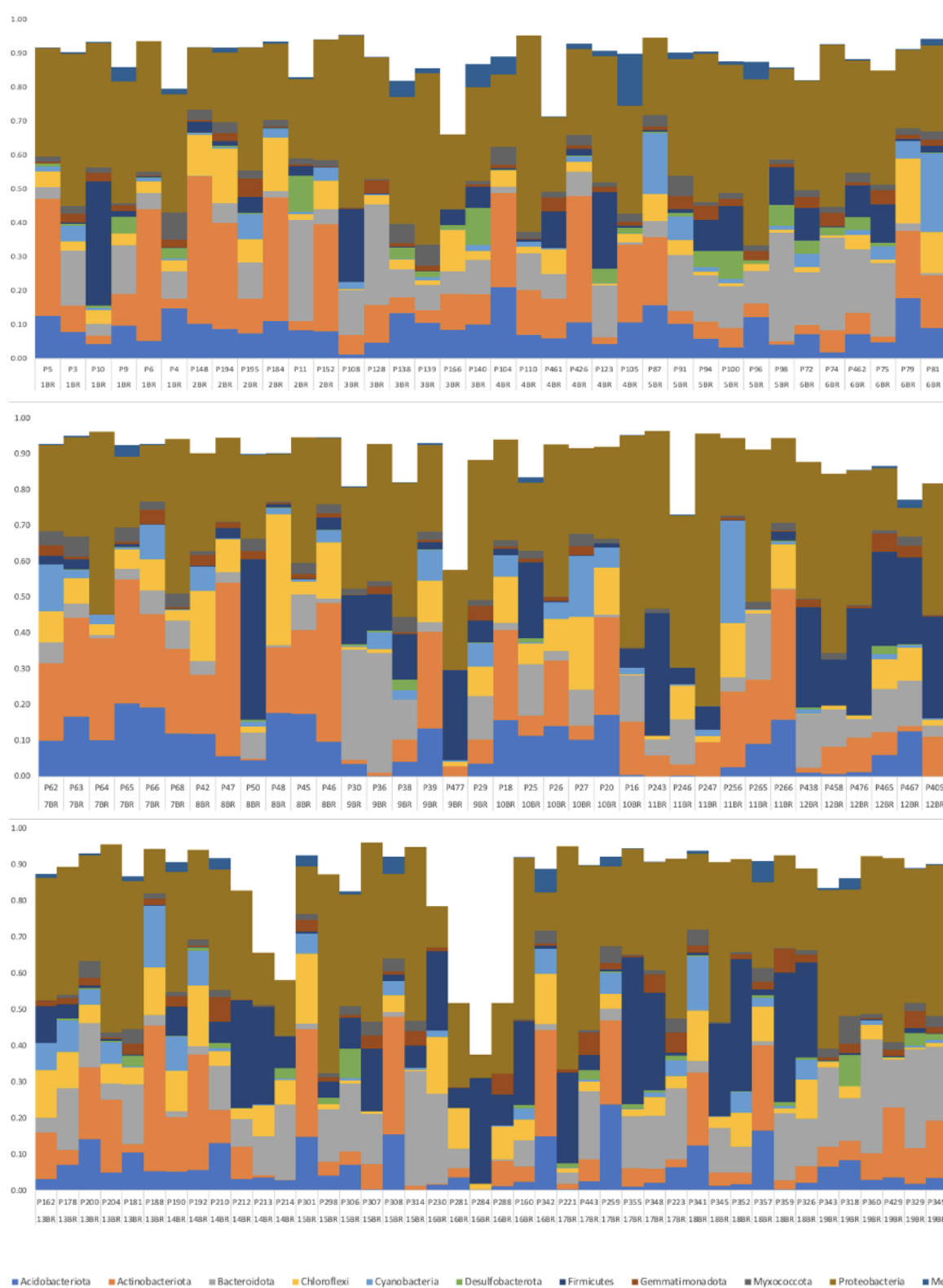
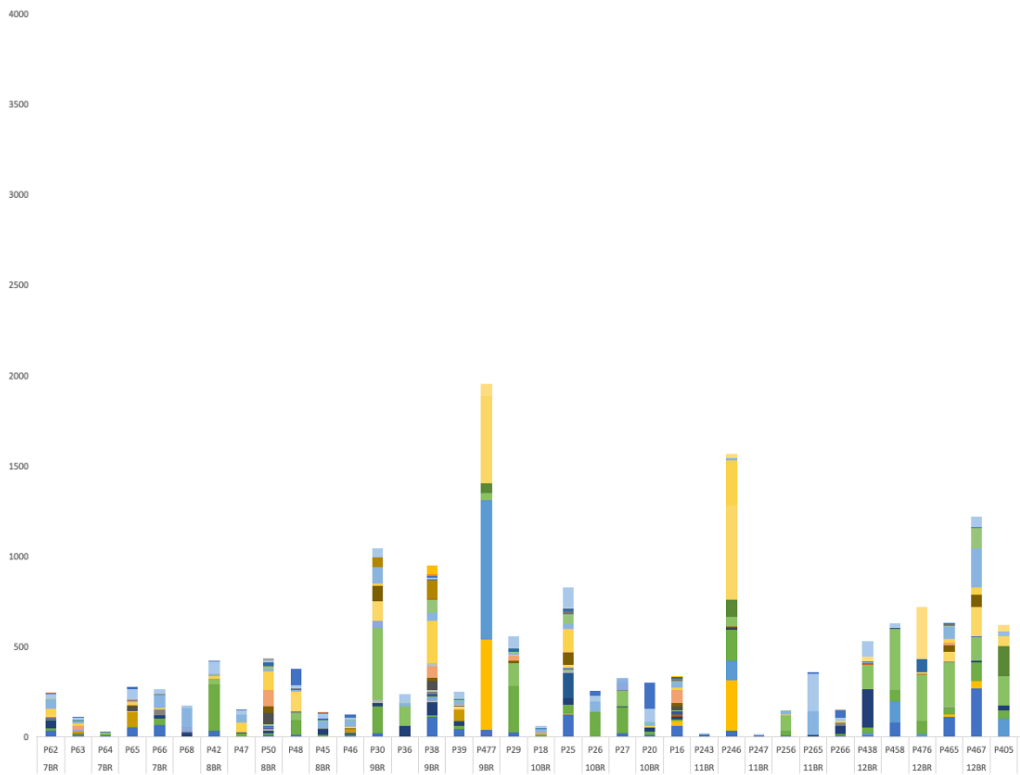
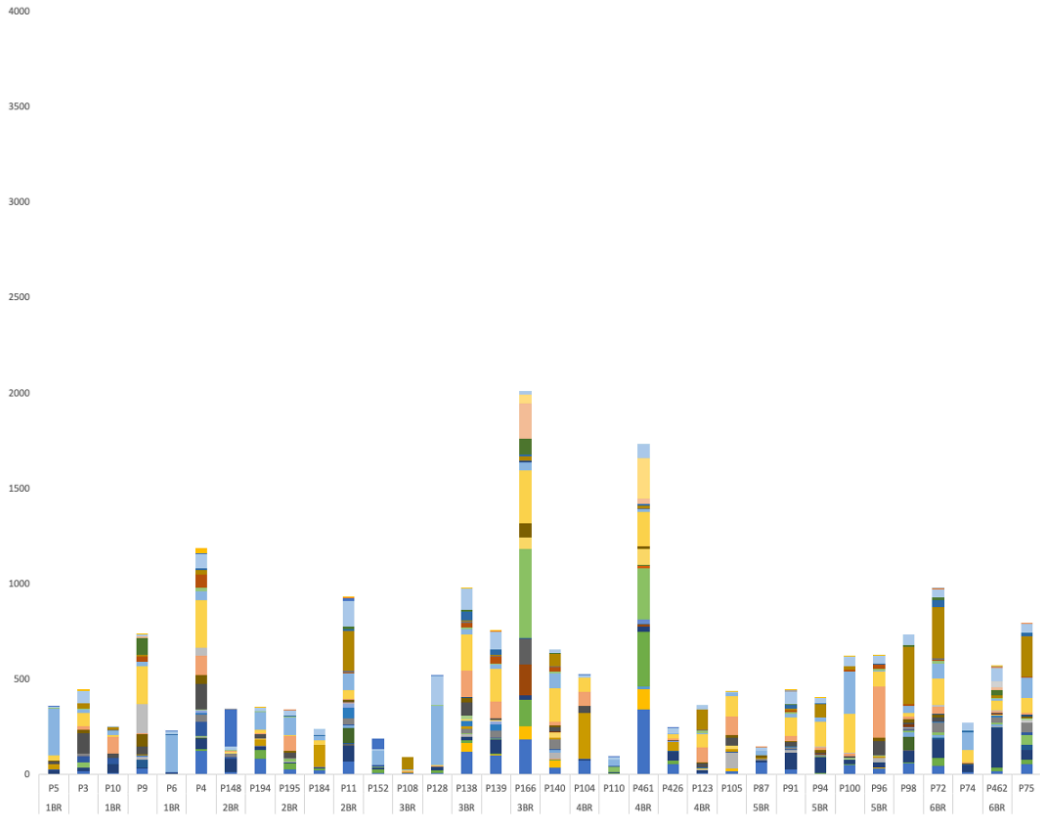


Figure D.7. Composition and distribution of bacterial phyla in Boiling River samples. These graphs include taxa with a 1% abundance or more.



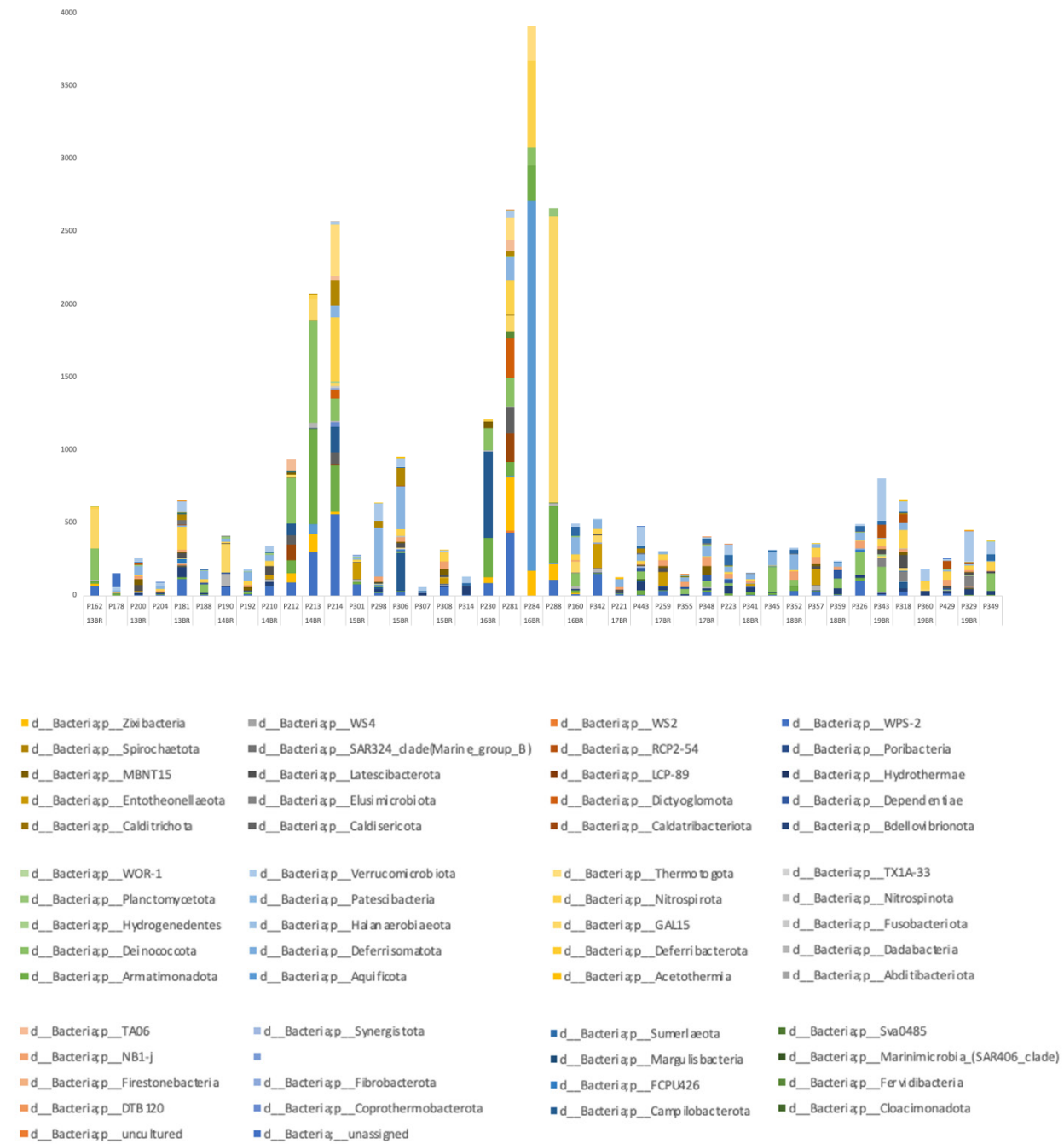


Figure D.8. Composition and distribution of less abundant bacterial phyla in Boiling River samples. These graphs include taxa with less than 1% abundance.

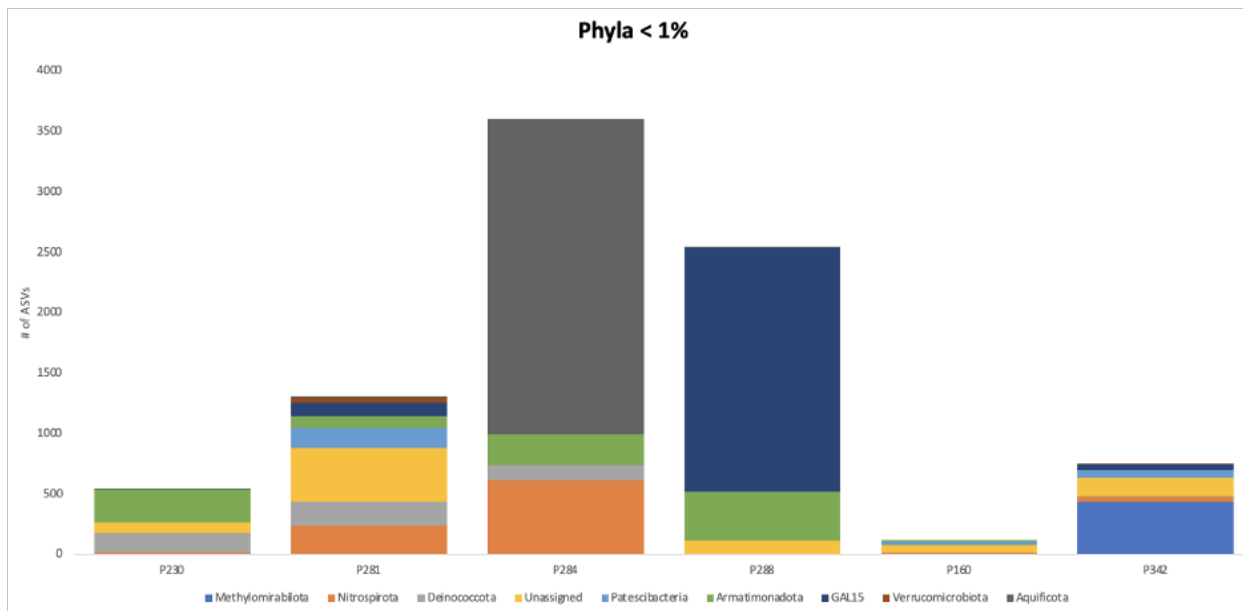
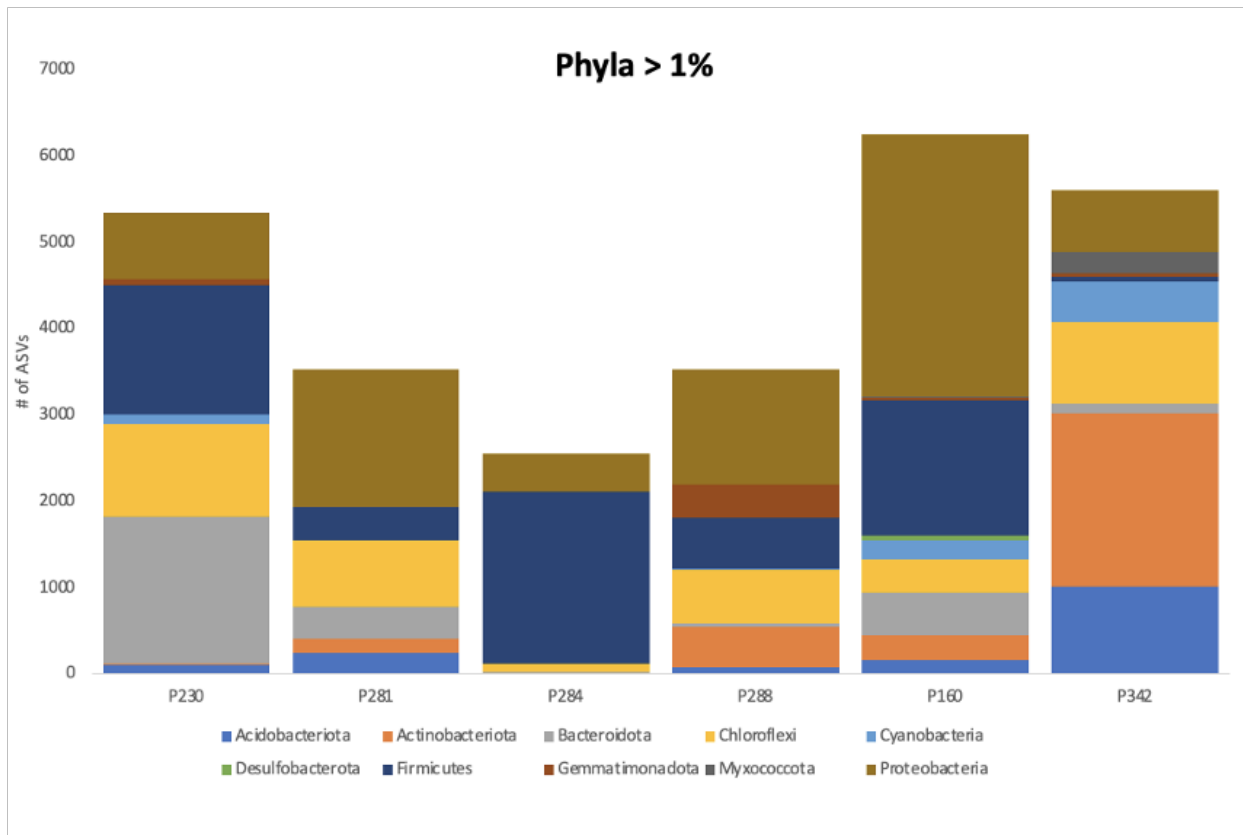


Figure D.9. Composition of bacterial phyla in samples from station 16BR at the phyla level.
 A) Phyla groups with abundance of 1% or more. B) Least numerous phyla (0.5-1% abundance).

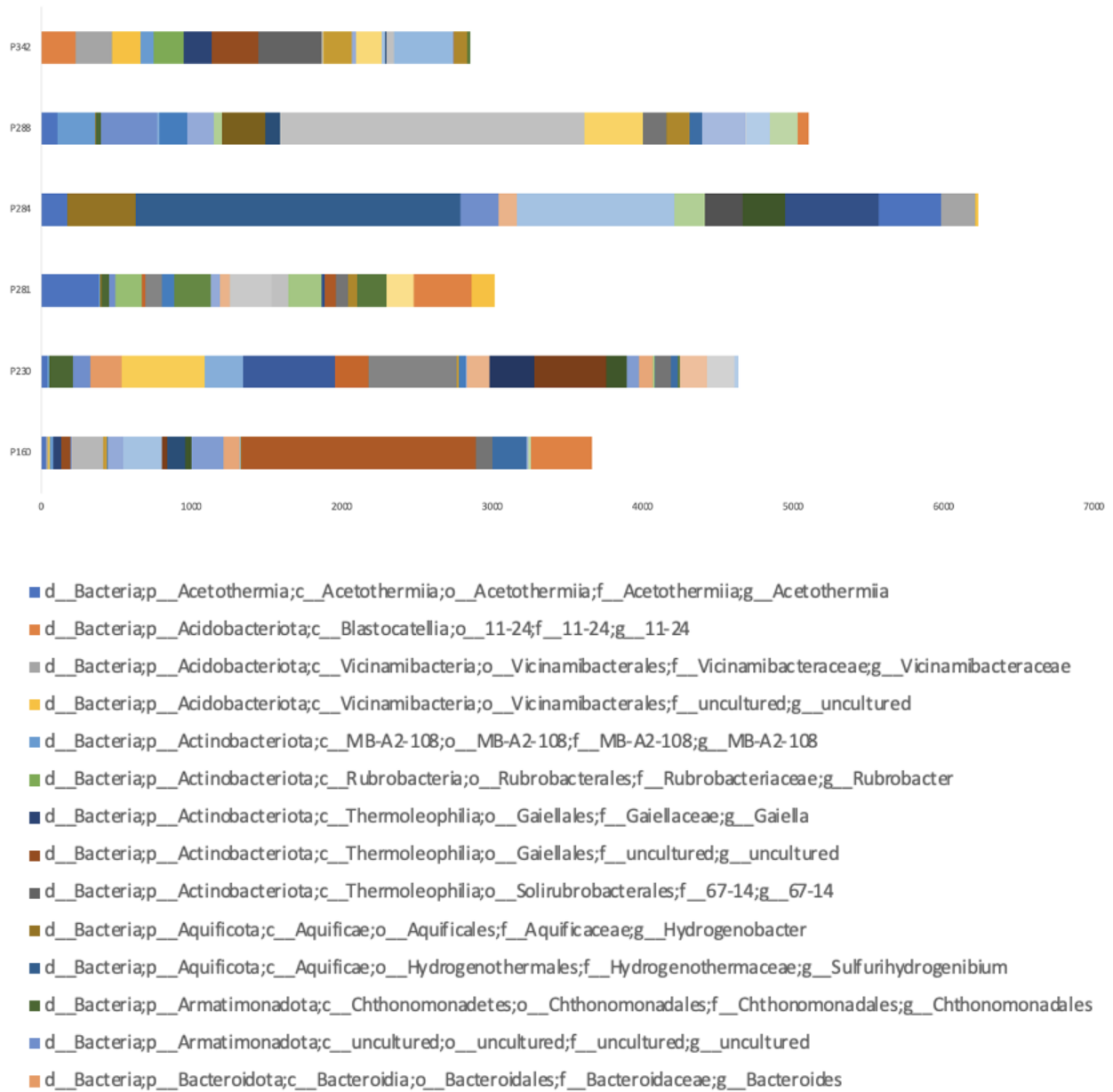


Figure D.10. Composition of bacterial phyla in samples from station 16BR at the genera level. Cutoff 0.5%

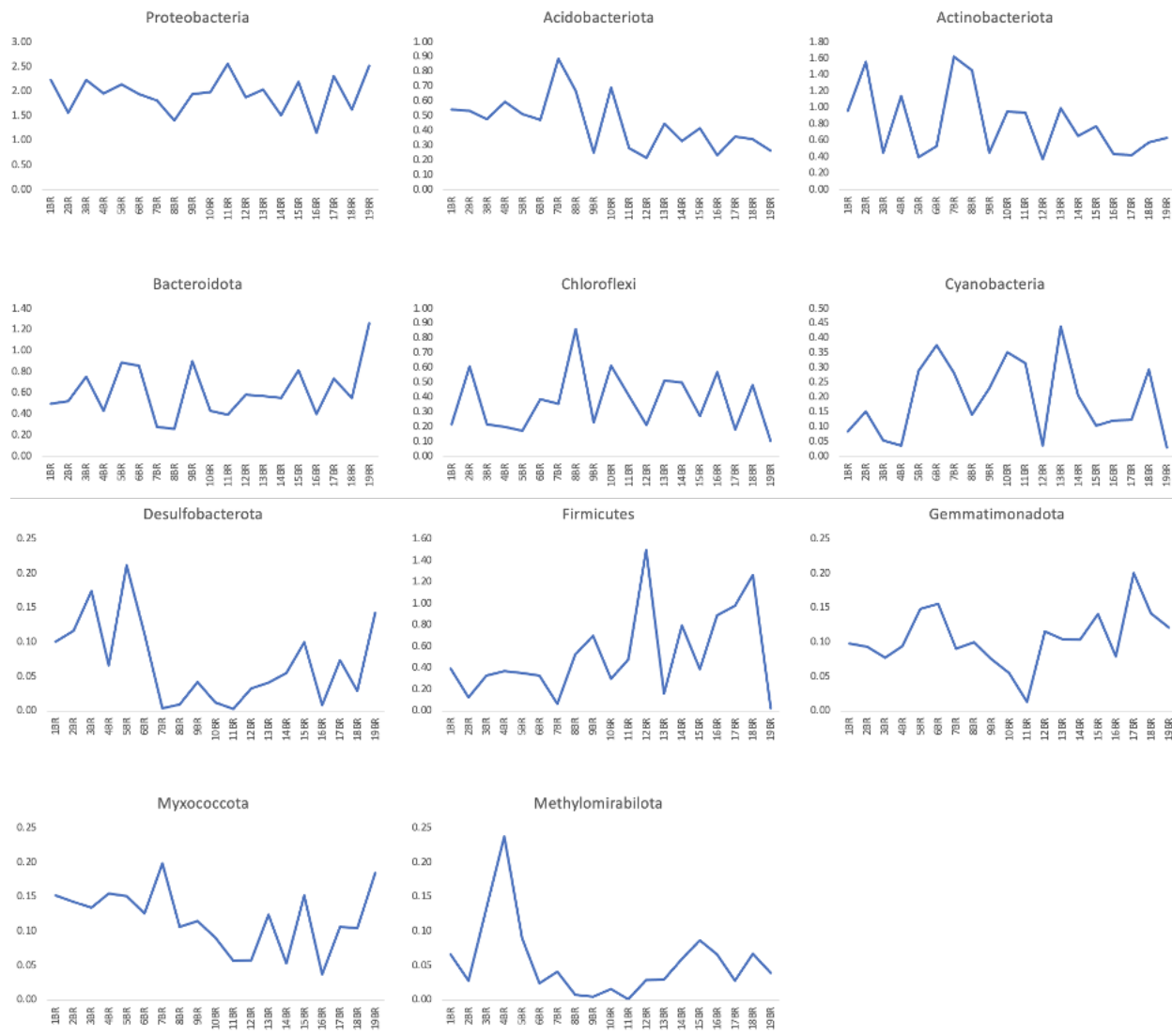
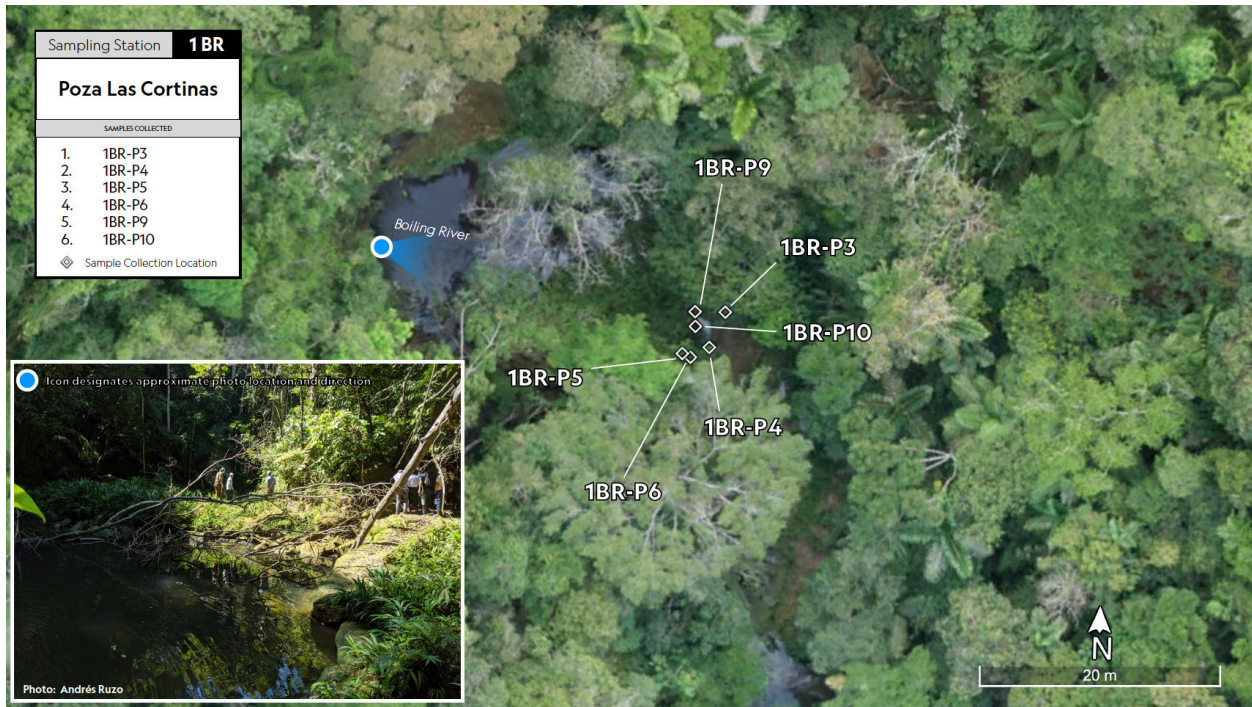
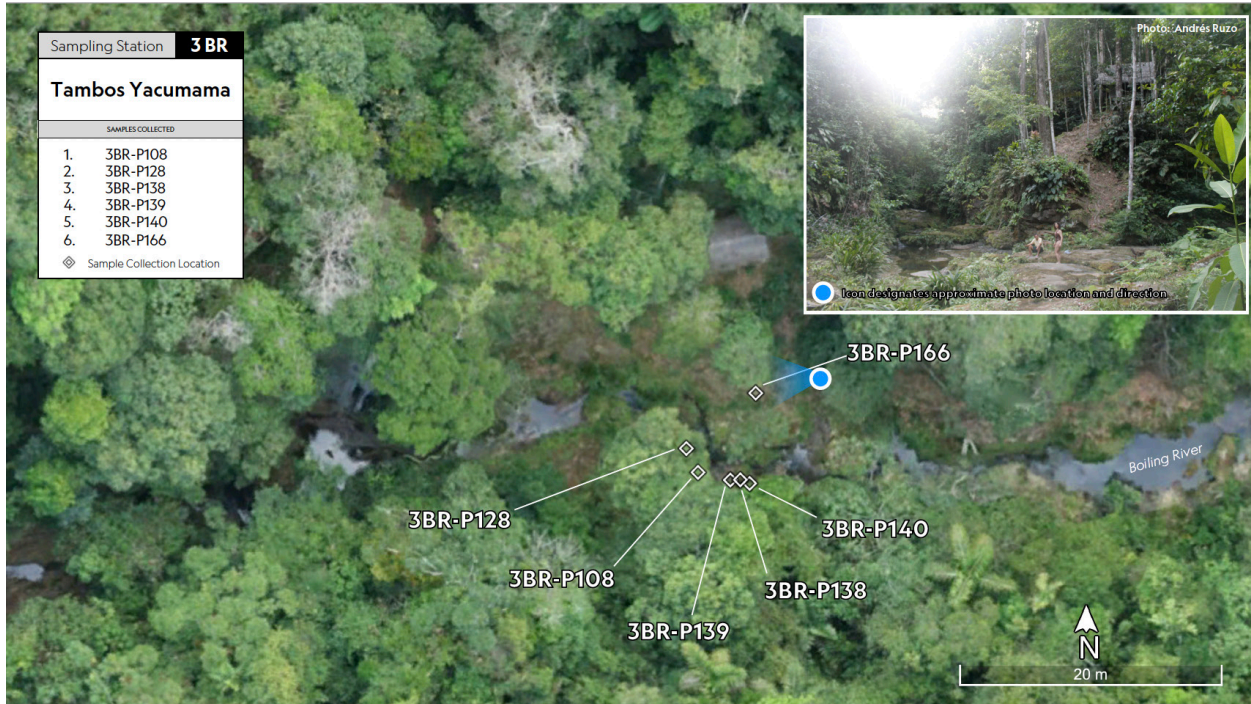
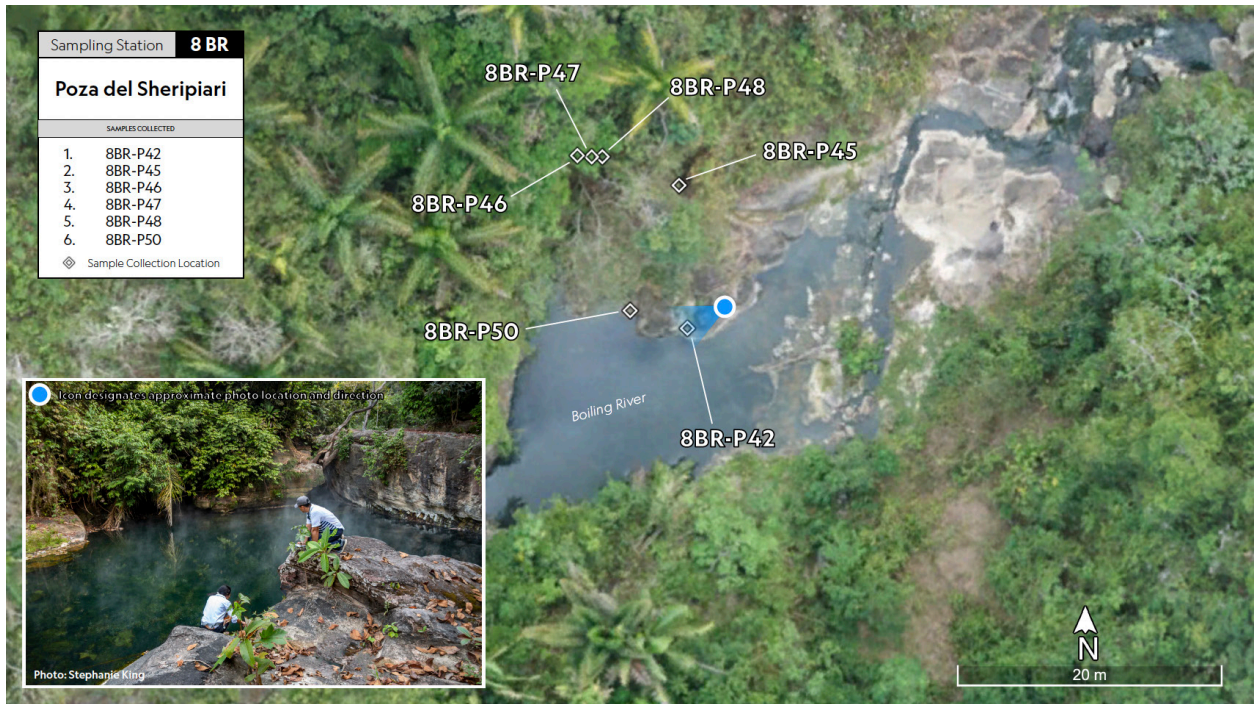


Figure D.11. Distribution of bacterial phyla in the Boiling River stations. Average relative abundance of the most numerous phyla groups (1% or more) including values from the six samples collected in each station.





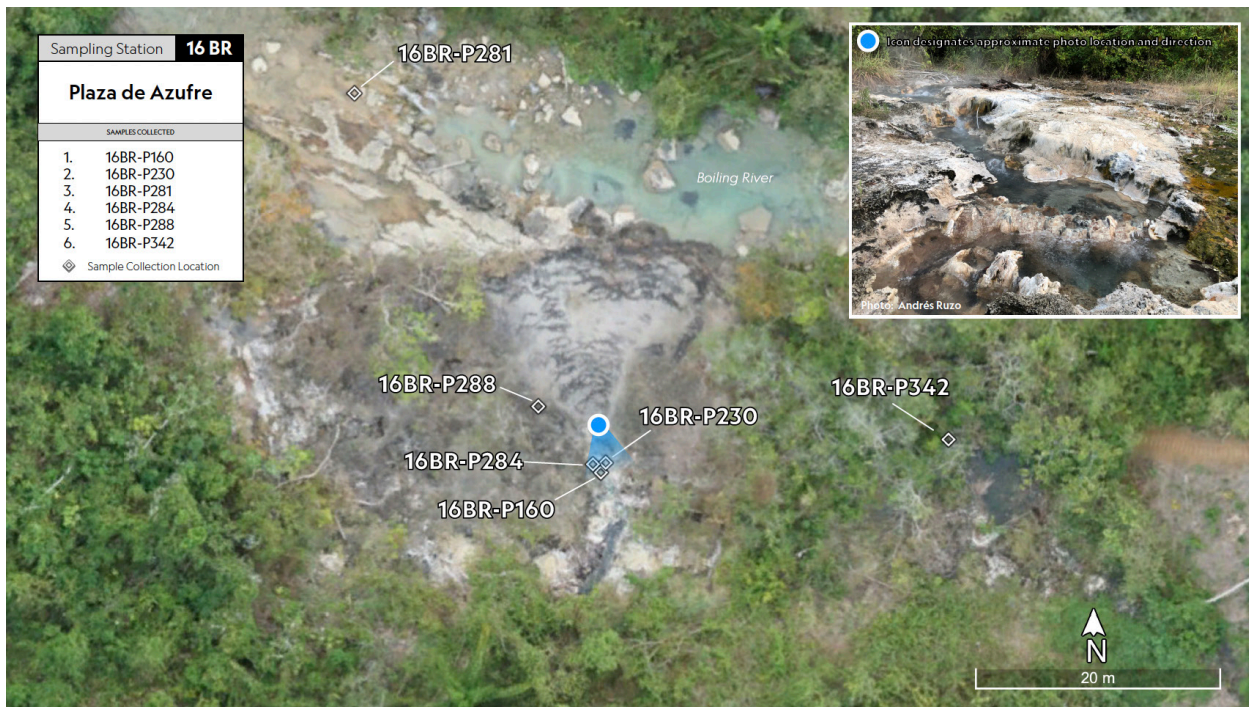
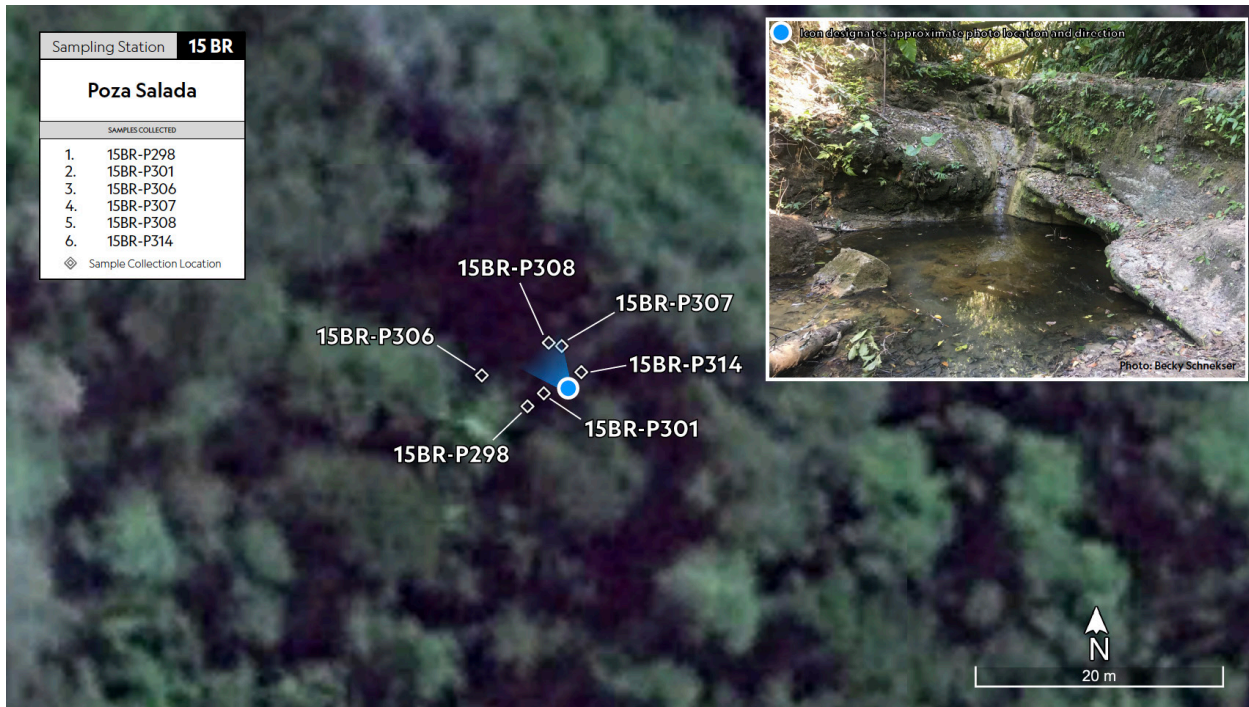


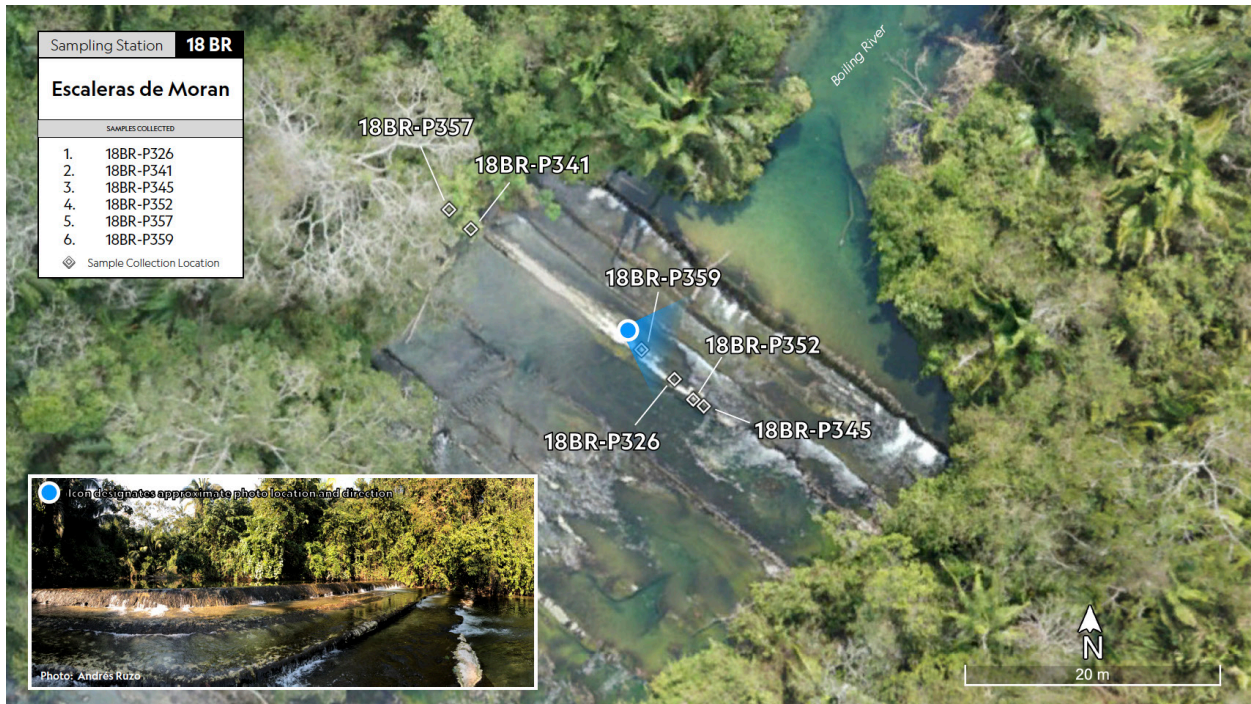
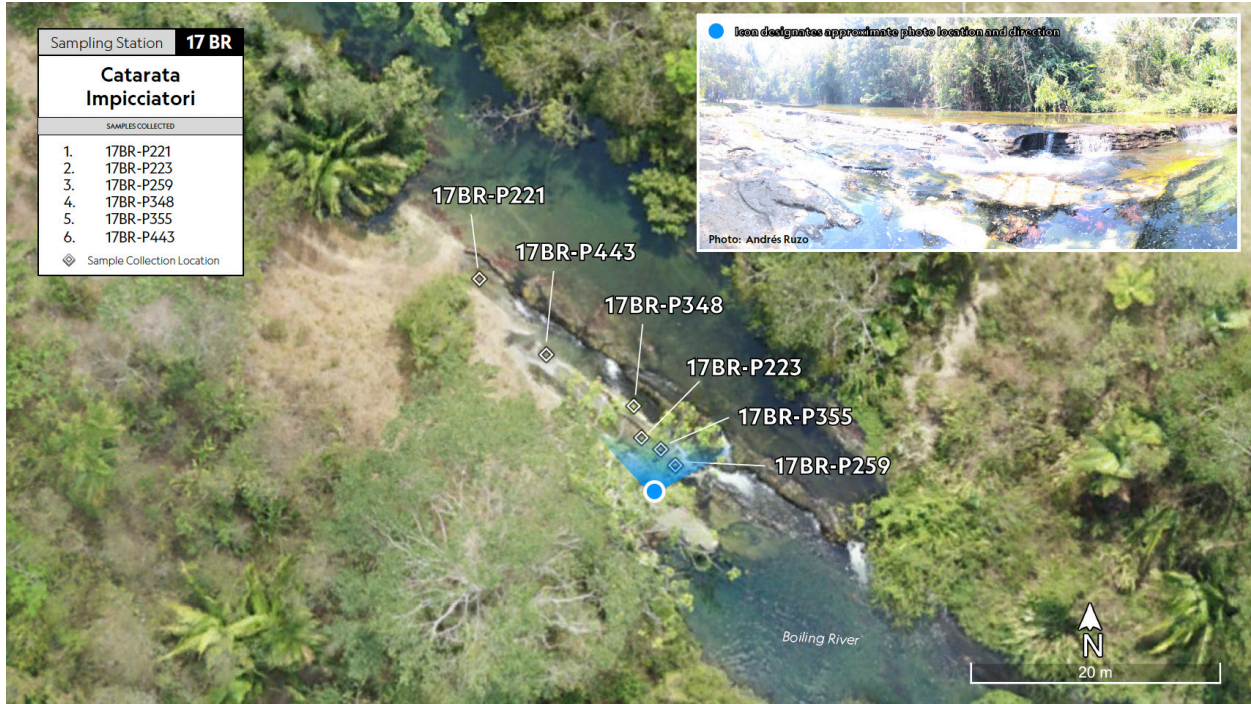












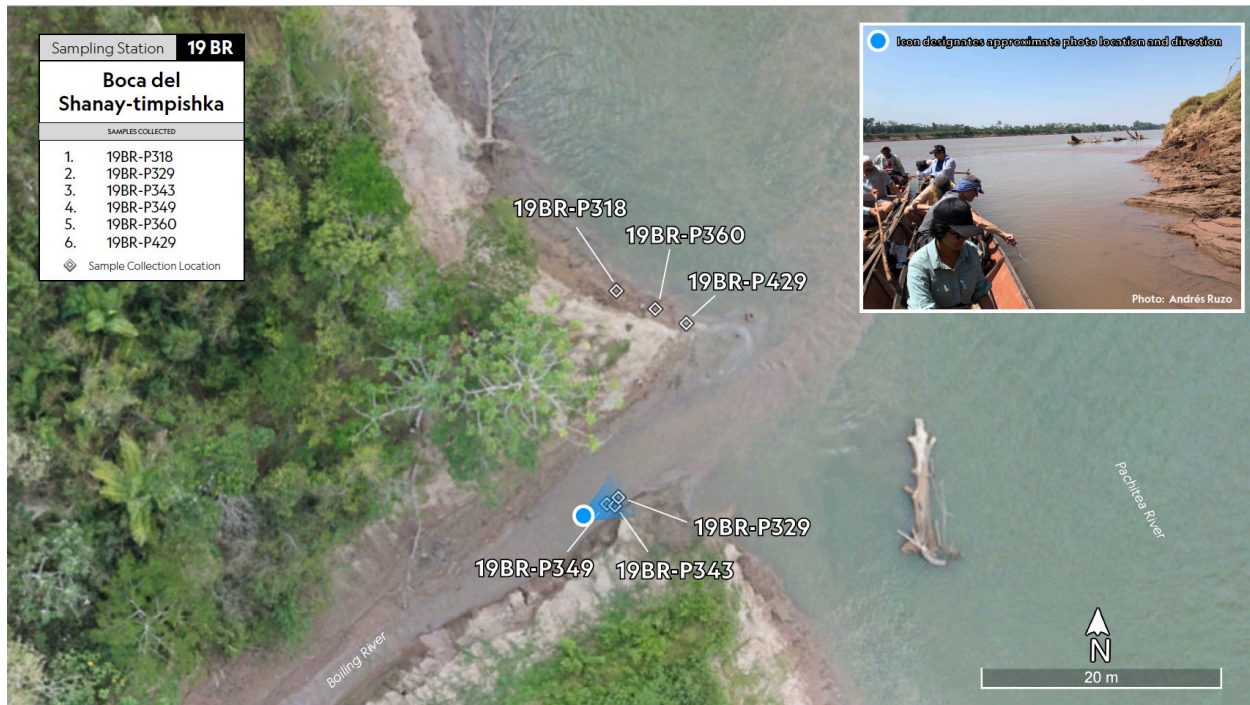


Figure D.12. Boiling River stations. The exact location of samples collected in each station are shown.

D.6 References

1. Smith, D.; da Silva, M.; Jackson, J.; Lyal, C. Explanation of the Nagoya Protocol on access and benefit sharing and its implication for microbiology. *Microbiology (Reading)* **2017**, 163, 289-296.
2. Ruzo, A. The Boiling river of the Amazon – Characterizing the Shanay-timpishka geothermal system. *In preparation*.
3. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F. O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **2013**, 41, D590-D596.