

**Understanding and Manipulating Alkaloid Biosynthesis**

by

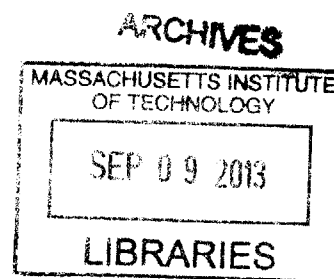
Weslee S. Glenn

B.S. Chemistry, Honors College  
Hampton University, 2008

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY  
AT THE  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2013



© 2013 Massachusetts Institute of Technology. All rights reserved.

Signature of Author \_\_\_\_\_  
Department of Chemistry  
June 28, 2013

Certified by \_\_\_\_\_  
Sarah E. O'Connor  
Professor  
Thesis Supervisor

Accepted by \_\_\_\_\_  
Robert W. Field  
Professor of Chemistry  
Chairman, Departmental Committee on Graduate Studies

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Professor Catherine L. Drennan

Chair \_\_\_\_\_

Professor Sarah E. O'Connor

Thesis Supervisor \_\_\_\_\_

Professor Barbara Imperiali

Committee Member \_\_\_\_\_

# Understanding and Manipulating Alkaloid Biosynthesis

Weslee S. Glenn

Submitted to the Department of Chemistry  
on June 28, 2013 in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy in Chemistry

## ABSTRACT

Humans have exploited plant alkaloids as medicines since at least the Neolithic Era. Today, alkaloids such as vinblastine (isolated from *Catharanthus roseus*) and morphine (isolated from *Papaver somniferum*) are prescribed to treat various cancers and relieve pain, respectively. Despite this storied use and palpable presence in the current pharmacopeia, relatively little is known about the biosynthesis, regulation and transport of these molecules.

For example, monoterpene indole alkaloid (MIA) biosynthesis, a set of metabolic pathways that produces hundreds of bioactive natural products, has not been fully elucidated in any organism. Here we examine the biosynthesis of secologanin, which contributes the monoterpene moiety to all MIAs. Specifically, we excavate *C. roseus* transcriptomic datasets to identify 10-hydroxygeraniol oxidoreductase, a missing step in secologanin biosynthesis. 10-hydroxygeraniol oxidoreductase catalyzes the oxidation of both hydroxyl moieties of 10-hydroxygeraniol to form 10-oxogeraniol, which is the substrate for iridoid synthase, the reductive cyclase that assembles the characteristic iridoid scaffold.

Despite having an incomplete understanding of MIA biosynthesis, several engineering strategies have been successfully deployed to incorporate halogenation into the MIA machinery and yield halogenated alkaloids. Although alkaloids and plant natural products have been used to treat various diseases, these compounds have not evolved specifically to do so. Therefore, these compounds frequently require editing to effectively tune their biological and pharmacological activities. We also describe efforts to reengineer tryptophan halogenase RebH to preferentially install chlorine onto tryptamine, the direct indole precursor for the MIAs. After reengineering RebH, we then over-expressed the tryptamine-specific mutant RebH Y455W and flavin reductase RebF in *C. roseus* and observed the *de novo* biosynthesis of a chlorinated unnatural natural product 12-chloro-19,20-dihydroakammicine.

Lastly, we describe the serendipitous discovery of a *P. somniferum* codeine-O-demethylase mutant that selectively demethylates codeine, a benzyloquinoline alkaloid involved in morphine biosynthesis, instead of both codeine and thebaine. This mutant may selectively disable a redundant route in the biosynthesis of morphine that has been associated with poor seed and licit opium quality.

Thesis Advisor: Sarah E. O'Connor

Title: Professor, The John Innes Centre and the University of East Anglia

## Acknowledgments

“Our crown has already been bought and paid for. All we have to do is wear it.” –James Baldwin

Receiving a Ph.D. is undoubtedly a crowning moment. This moment—this crowning moment—has been bought and paid for with intense love and sacrifice that, of course, defy a few short paragraphs. And, frankly, I’m unsure that five pure vowel sounds and the consonants that are used to modulate air pressure and form intelligible words can even hint at capturing the gratitude that this moment—this capstone—warrants. But, I will try.

First and foremost, thank you, Sarah, for selecting me as a graduate student. I honestly chuckle every time I think of how arrogant—and down right belligerent—I was when it came to being able to work in your lab. *I am only going to MIT if I can work for Sarah. As a matter of fact, don’t even admit me if I can’t work for Sarah.* I said it (well, I probably said ‘Professor O’Connor’ until you forced me to call you ‘Sarah’). And, I meant every word of it. I had cause for concern because I had heard that you were by far one of the most popular choices for biochemistry applicants. And, it is obvious why: Your crystal clear writing and absolute creative genius—which are what I first noticed—are enviable, but pale in comparison to your ability to assemble a truly dynamic team and exponentially amplify the talents of everyone you mentor. It truly has been an honor to work in your lab. I will wear ‘O’Connor Lab Alum’ as a badge of honor on a bosom swollen with pride. Thank you for strongly advocating for me. You are a divine gift.

Cathy, thank you. Thank you for teaching me about wise criticism. Thank you for really engaging me as a scholar and as a true citizen of the academy. I will never forget how you came to my first poster session in summer 2007 and really encouraged—and challenged—me. For some reason we, in the academy, love to leave praise on the shelf and put all the insecurities of others on the table. Thank you for making space to help me quash certain insecurities, work through them and put my talents on the table instead. Your work on stereotype threat has been transformative. Thank you for engaging me in that dialogue as well. Thank you for making space for the thousands of graduate students (myself included) who care deeply about science—and education. I am deeply indebted to you, and it is safe to say I likely wouldn’t have reached this crowning moment without you. Connection is the revolution, and you are leading it. Thank you. If I am ever privileged to assemble a team, you and Sarah have provided immaculate examples for how to do so.

Barbara and Prof. Kris Prather, thank you so much for serving on my committee. Your insights have been absolutely formative to my work and growth as a scholar. Barbara, thank you especially for developing the chemical biology community at MIT. The training grant and the courses really help(ed) defragment current knowledge structures and my own personal take on learning.

Being a student straddled across the Atlantic with feet on two continents can be difficult at times, but it also has its perks: I have two great cohorts. At MIT, I’d love to thank my awesome friends and classmates Mike Morrison, Alyssa Larson, Jen Yao, Stephanie Lam, Rachel Fleisher, Michelle Chang, JM, Jeremy Setser, Andrew Rajczewski and Mackenzie Parker. *I lived* for those Thanksgiving dinners together. I’ve learned so much from you all. I am so excited to watch all of your careers. At JIC, I give



a heartfelt thank you to Tony Maxwell and the Department of Biological Chemistry and all of Chatt for receiving me and really helping me to cultivate my knowledge of plants and science in general. Thank you Matt and Ellis and the rest of the Field Group for being so helpful once I arrived. I especially thank Dan ‘the Man’ Tromans, Tom Turner, John Steele and my bestie Farzana (Fuzz) Miah. John your scones are exquisite! Thank you so much for engaging my whims. Fuzz, you especially have been absolutely formative in Norwich’s transformation. Fuzz, your brilliance, sparkling smile and *Fuzz-y* personality effortlessly penetrate the steely grey British days and warm the soul. I love you, Fuzz!

I’ve also had a fantastic support network across two continents beyond the academy as well. Thank you so much, Mr. Gareth Williams. You’re a wonderful voice teacher; I have grown exponentially in both technique and confidence thanks to your meticulous ear. How are you so patient with me?! I give a special thanks to David Robertson, my wonderful counselor in Norwich. David, you are a saving grace, a bona fide life fountain. Thank you, Sue, Big Sue and Denice for always bringing a smile to my face at the canteen each morning. Seeing your faces always starts my day off right. And, thank you to the security staff, especially Ray, for protecting us and for being such wonderful people. It is a pleasure to see you each night and weekend. Father Maurice Charles (or should I say ‘Mr. Charles’? LOL!), you have been a wonderfully supportive friend. You have a special way of reminding me to look into the divinity of all life, to embrace solitude as a creative force, to love heartily and to forgive easily. So many important lessons of life and love I have learned from you. And, of course, I will never forget the day we sang together on a sleepy Sunday afternoon in autumn at a London park. You should’ve seen the looks everyone was giving you. You were amazing. And still are. Thank you for always reminding to sing—literally and figuratively—with everything I have within me. And, JBJ, thank you so much for keeping me accountable with this thesis. Thank you for helping me turn thoughts into words and words into chapters. Thank you so much for listening to me complain night after night about what didn’t go quite right. Thank you for helping me maintain perspective. Your advice is truly beyond your years. I am really happy to call you a friend, and I am so excited to see where life takes you! Dr. Sandra McGuire, my “Mom in Baton Rouge,” thank you for always being so encouraging and for providing warm inspiration, even from 6,732 miles away. You are one of my sheroes, and I love you. You bad. No verbs necessary. Thank you, Alana, for always being a treasured friend and for always providing loving advice and for telling me the things I need to hear, even if I don’t want to hear them. Can you believe we’ve known each other for ~15 years already? Bananas. Thank you so much for your help with my post-doc applications, Nick Ball! Thank you, Princess, for your help and encouragement as well. Your presence and success in the academy speaks to what is possible. Thank you so much. Both of you.

Dr. Ms. Joyce, words absolutely defy our relationship. Thank you so much for being a fantastic host mom. Thank you so much for encouraging me to be patient. Thank you so much for seeing so many things in me that I didn’t know were there, for helping me awaken the dormant states. Thank you so much for believing in me. Thank you so much for giving me a crash course on diversity training even when I didn’t know I needed it. Thank you so much for reminding me to breathe and remain present! Thank you, Dr. Ms Joyce and Mr. Leslie, for opening your beautiful home—and even more beautiful hearts—to a poor graduate student.

Thank you Deidre (my ‘gurrh’), Kenton, Hasani, Robert (Bobby), Joubert, Sam and Parson for making Boston such a wonderful place to live and study. Deidre, your smile is infectious; how can anyone see it and have a bad day? It’s impossible. I love you. Kenton, I will never forget the nights we studied together and watched *Girlfriends* during the breaks (I’m doing the dance right now as I write this). You’re a true friend, and I absolutely love you. It means so much to have someone at MIT who understands your struggle without you even having to give voice to it. Hasani, thank you for always inviting and encouraging me to go out; you really do build community very well. Bobby, thank you from the bottom of my heart. You’ve been a true friend ever since you ruined my shoes on the Boston Common. The queer people of color scene in Boston really does owe you a debt of gratitude for implementing real change and building lasting community. Your intelligence and fierce articulations are enviable, but are no match for the intensity and love you bring to your friendships. I love you. Thank you, Joubert, for being so awesome; you’ve become such a wonderful friend. Sam, you’ve become such a wonderful friend to me over the years, too. Thank you so much for introducing me to ‘Dear Sugar’ and all your favorite books, a number of which are now favorites of my own and which I have passed on to others as well. Parson, of course I knew you well before we both landed in Boston (in fact, I remember stalking you on the Hampton website while I was applying there as an undergrad). Your presence made Boston feel a little more like home. Whether we were sipping tea in Harvard Square, evaluating Boston Lyric Opera’s performances or getting down with the get down in your Zumba classes, we certainly created memorable moments. I always swell with pride when I get to show you off for appearing on TV as an analyst. *Hey, there’s my big sis!* Absolutely love and adore you. Thanks for being a great big sis all these years.

And of course I can’t fail to thank Dean Jones and all the staff of both MSRP and Converge for getting me to MIT in the first place. Thank you especially to Dean Jones, both Dean Statons (who both give the best hugs ever), Shawna Young, Monica Orta and Cyd McKenna! Thank you to chemistry Grad Ed office, too, especially Susan Brighton.

Dr. Urasa, Dr. Ndip, Dr. Paranawithana, Dr. Darby, Dr. Rankins, Mrs. Calloway, and Dr. Woods-Warrior, I really loved your classes at Hampton. And you are what makes Hampton a really special place. Your guidance, no doubt, has helped usher in this present moment. This is your moment, too. I am a proud son of Hampton University. I give a special thank you to the entire Department of Chemistry at Hampton and the Honors College, especially Isa and Kaiulani, who are both my sisters. Isa and Kai, you have been there from the very beginning. I love you both. How fantastic is it, Isa, that we both went to Hampton *and* MIT together?! I am no doubt in science today because of all your help and belief in my talents. And, Kai, I’ve no doubt stayed because I have a support sister in the sciences who understands my plight—and biochemistry is also much more enjoyable to study with you! Thank you, Dea, Chandra, Tashee, Walter, Aluko, Kashif, RaAnna and Tasha (I hope I didn’t miss anyone).

Marques Garrett, you sang, conducted and composed your way into my heart. You’ve been the best big brother a guy could ever ask for. Funny thing is I didn’t ask for you. You divinely entered my life instead; I am eternally grateful for the great fortune of sitting next to you in University Choir. Amon, you are one of my best friends. You really have been there from the beginning, too—through it ALL. I’m sure I’ve made you want to punch me at times, but I love you. Steven, thank you for always encouraging me to be

a dreamer. You constantly remind me through all of your hard work that faith must precede providence. Justin! I love you, despite that time you tried to beat me up in my own dorm room. You are positively crazy, and I love you for it. I couldn't imagine what science would have been like if I didn't have you to study with all those long hours at Hampton. Our time in differential equations especially has taught me that, even when we aren't at our best, chances are that things will turn out all right. Thank you, Sherea, Leon and Grover. Your friendship has been salvation. Sweet. And, Jessica A. Rivers! Excuse me, Dr. Jessica A. Rivers, M.D., you are my absolute best friend. Your companionship is invaluable. Thank you for listening to my super long stories (you dubbed me Father Goose, after all) and really helping to bring out the more fun and endearing qualities of my personality. I will never forget all the times we went out to dinners together at Hampton and took long walks...to the car to go to more dinners and on more dates. Your intellect is truly an inspiration to me. You are truly beautiful inside and out. I always love seeing your name pop up on Skype, but cannot wait until the day we land near each other again. An ear-to-ear smile always ensues when I think of you. Always. Outrageously loud laughter typically follows. I absolutely love you. Always. Again. And always.

Tavi (my BFF), Katlyn, Julia, Brittany, Jessica (Mrs. Boring), Kayl, Jill, Ms. Brown, thank you so much for being dear friends and for encouraging me over the years. Thank you for creating and experiencing sweet music with me. Thank you especially to Julia for help with these page breaks! Without you, this thesis would not be written. Thank you, Mrs. Holcomb, for encouraging my writing and for becoming a dear friend. Thank you, Mr. Redmon, Mr. McCarn, Mrs. Norman and all my science and math teachers who both encouraged and equipped me. I am deeply indebted to you. Thank you so much, Mrs. Hutchens (AKA 'Mom') for being our *de facto* guidance counselor and for always giving the best personal and professional advice and for always being so encouraging. You are a true inspiration. Thanks, Tara!!! California, here I come!

This 'coronation' would not be possible without having a fantastic team to work with and to learn science from over my career as an intern first, then later as a graduate student. Lesley, I am in this degree program because you mentored me the summer of 2007 and helped me construct a solid graduate school application. Thank you so much for deploying your fantastic teaching skills to tutor me in biochemistry before I took it my senior year. But, moreover, thank you for being such a fantastic friend and dedicated sister. Thank you for always allowing me a safe space to present my more vulnerable ideas, to let my guard down and to let love flow in. You are a gifted scientist educator. You enrich the lives of all you touch, and I am eternally indebted to you. Ricky, Ms. Girl, thank you so much for helping me get started in the lab as a graduate student. You are the standard of excellence. You are a true inspiration for me and a wonderful friend. I am so honored to have been able to author a number of papers with you and glean your wisdom. I am so excited to see how your career turns out. John, my straight lab husband, we really have become 'an old, married couple.' We've been co-captains on the volleyball team, eaten at multitudes of restaurants together across multiple cities, had spats across two continents and remained treasured friends through it all. You have become a wonderful friend and brother. I couldn't imagine a better person to share lab responsibilities with or to share an office with for 5 years. Thank you for helping me give birth to my visions—*and I'm having this baby!* Zeke, thank you so much for your help with the halogenase project. I really appreciate your life perspective and your constant reminders to stay

grounded. Thank you, Bettina (let's go on a F'real field trip), Nancy, Beth, Peter, Jenna, Dave, Cecilia, Aimee, Yeol, Dong and Anne. You all have been most helpful around the lab. Your presence made the O'Connor lab a dynamic and fun place to work and learn.

And, SOC@JIC, thank you. Thank you, Nat, Nando, Dorota (*HEEEEEEEY! I see you, girl!*), Richard, Stephanie, Steffie, Anna, Franziska, Leonie, Sarah H. and Greg Mann. Nat, you're a fantastic conversationalist; I love speaking to you, not the least of which because I always feel exponentially more learned after hearing your analyses. Thank you so much for helping me around the lab, especially with synthetic chemistry and presentation skills. You're a treasured friend. Nando, thank you for patiently teaching me about data mining (a more detailed shout out is provided in Chapter 2). Thank you for becoming a treasured friend and for always opening your home to us for movie nights. I love that we also share a passion and curiosity for great music! We must perform together. We *must*. Dorota, thank you for serving us life every day with your stunning brilliance and your infectious personality. I'm madly in love with you and your outfits. *Okay?* You're basically the social chair at JIC. You are a fantastic office mate and an even better friend. I look forward to seeing your meteoric rise in the academy. You are brilliance and confidence personified. Richard, I am so going to miss your daily taunts. You're Jerry to my Tom. Thank you so much for your friendship. Stephanie, Steffie and Anna, thank you all for being such great friends to me. Stephanie, thank you for remembering that I like salted microwave popcorn and gifting it to me on my birthday. That meant so much to me. Steffie, thank you for being such a great office mate and dancing with Dorota and me to any song that comes to mind. Franziska, thanks for being such a treasured confidant and for opening your home up to us on Christmas and every other holiday, too. Thank you also for helping with all the VIGS experiments. Your expertise is invaluable (Read: I could *not* have done any of this without you). Greg Mann you were such a delight to work with, even if you did leave behind HUGE messes! You 'Gregged' the place. We converted you into a verb. Thank you so much for your work on RebH! Leonie, you're such a joy to have around the lab as well. I think it goes without saying that you win the Award for Best Chocolate Cake! Delicious.

And, of course I wouldn't be here today without my family. Thank you, Grandma and Pops, for always believing in and enabling me. Your encouragement convinced me that I could achieve beyond any goal I could possibly imagine. Your love set a space for me to realize that not reaching a goal was okay and sometimes necessary to make way for better things that I could not yet imagine. I always love our talks, Grandma. I can't wait to be closer to you again physically to share more of those treasured moments. Uncle Harrison and Aunt Stephanie, I credit you more than anyone else with introducing the concept of higher education to me. Thank you both for always hiking to the mountaintops with me and shouting all of my accomplishments over the past 23+ years of schooling. These accomplishments are your accomplishments too because you helped me early on to imagine what *beyond* looked like. Aunt Babe, I will always wear 'Babe's Kid' as a badge of honor. I love you. We have had a lifetime of moments more precious than gems—usually served over absolutely delectable food you prepared. (Delicious!) Thank you Aunt Mamie and Aunt Lola (can I have a roll?) for taking care of my siblings and me when we were younger. Auntie Michelle, thank you for always being the cool aunt and for engaging my shenanigans, especially for joining in a duet to TLC's "Sleigh Ride." We're always the hit of the party (right, Grandma and B'anca?). Thank you, Cierra and

Jordan for being fantastic friends and cousins! Congratulations on all of your successes! I am so proud of both of you, and I hope I've had at least one-tenth of the impact on you that you've both had on me. I'll save all the embarrassment of stories for a later time...but the stories will come. Believe it!

B'anca, thank you for always being my favorite play mate as we were growing up—even if I went years without touching you! We had such a treasured childhood together playing Olympics around the yard (well, you were the cheerleader, at least—and you're still one of my biggest cheerleaders to date), choreographing dances to the songs on *TRL*'s countdown and sharing laughter at some of the ridiculous moments that went on around us. *EEEEEEEEEEEEEEEEEE! Seester. Mercy.* Proud of you. Alex, I am so proud of the young man you have become. I remember helping you with your homework and sports when you were much younger. Now you've grown into a very successful leader and I cannot even touch you in track, tennis or basketball. Well, maybe tennis and basketball. I still got a little skill. I'll have to play you before you head off to Hampton this fall. I'm so proud of you! And, Jonathan, you were always such a good big brother. You're a fantastic listener and very gentle—everything that I am not! Thank you for always showing me the ropes, so to speak. And thank you (and Christina) for bringing my beautiful nephew Jayden into the world. Jayden, I love you! Your matchless curiosity really does inspire me. I love how you always want to go search for something. (I guess Dora and Diego got you hooked, huh?) I will always treasure the moments you and I went on treasure hunts, flew around the house (well, I guess you did more of the flying and I did more of the carrying), or performed all the dances from *Yo Gabba Gabba!* Christina, you're such a wonderful mom to Jayden and a lifelong friend. Kelly, welcome to the family! I'm sure we are going to develop an epic repertoire of duets. Can't wait! Chris, you have always been more like a brother to me. We've spent a lot of time together growing up, no doubt because we were born a mere six weeks apart. It was always such a compliment to me for people to mistake us as twins when we were younger. And, of course it was a badge of honor at Hampton to have the cool and successful campus leader as a cousin. Having you at my side made that transition a lot less scary. I am eternally indebted to you, too! Michael, thank you—despite my oftentimes very vocal protests—for helping me to become a better craftsman with the use of my hands. I'm still not very good with them, but I'm practicing. Learning skills that require your hands really does help life run more smoothly, and every hiccup does not become cause for a major operation.

Thank you to Pastor Hunt and my New Birth Worship Center church family for providing a fantastic community to grow up in and to learn to seek divinity in every situation, including ones that seem impossible such as—I don't know—completing a Ph.D.. Thank you also to Dan Smith and First Church Cambridge for providing a safe and loving space in Cambridge, MA to openly seek God.

Marques Redd, we have become such fast friends! SHANDO! You are really like a brother to me now. I really enjoyed our brief time together in the UK and always look forward to speaking with you again! Thank you for helping me over the last few months to defragment my thinking and to engage new (superior) knowledge structures. Thank you for helping me to reclaim mental and spiritual spaces. You will build nations, and I am so excited to witness the fruition of all the beautiful work you have been doing. I

can't wait to visit you in Sedona, Arizona! Lyndon always said that I reminded him of you. What a compliment! Thank you, Marques!

And, Kamaal, more than anyone else in this acknowledgement, you have taught me how to forgive and be forgiven, to laugh and be laughed at and to love and be loved. You taught me the beauty of poetry, or perhaps you inspired it. I not so secretly love how you turn everything into an art project and how you take time to notice colors and words and people. Thank you for always encouraging me to take time for myself, to take time for tea, to take time to love, to take time to travel, to take time for the ones I love. I can't help but smile when remembering all the phone calls Monday through Thursday that started exactly at 9 PM, the Friday night movies you were consistently 6 minutes late for no matter what time they started, the Saturday night dinners at restaurants in town and the Sunday night meals we prepared together before we watched episodes of *The Cosby Show*. James Baldwin avers, '...love is a growing up.' And we certainly are growing. Thank you, Kamaal.

Thank you to all my brave ancestors, who built knowledge and fought for my freedoms. Thank you to Poppa and Ma who—with little formal education—knew more about plants than anyone with prestigious degrees could possibly ever know. Thank you, Ma and Poppa, for providing all of your grandchildren with such an enchanted childhood.

Oprah said something along the lines of 'Blessings come in the space of gratitude.' I reserve my richest gratitude for my richest blessing, Mom. You are my primal self. This is truly your crowning moment, too. Thank you so much for purchasing my crown with your love, sweat and tears over the past 27+ years. Your work has not gone unnoticed, and it is certainly not in vain. I only pray that I make you happy and that you are proud of the man I have become. Thank you so much for teaching me how to table my dreams. I am indebted to you for that. But, moreover, thank you for teaching me how to love. Thank you for teaching me how to be thankful for my gifts and be a good steward of them. Thank you so much for teaching me to sing loudly and to love my voice. But thank you even more for singing those songs to me when I had lost my way. They have absolutely been a compass, a North Star on life's raging seas. For all of these things and more, I thank you. This your shining moment, too. I am grateful for you; you are indeed a blessing.

My sincerest apologies to anyone I may have inadvertently forgotten to mention. Thank you so much to all of the friends who I could not mention in this thesis; listing everyone who deserves mention would require a lifetime. We have a saying at Hampton that dictates that we 'let our lives do the singing.' For everyone—those mentioned and not mentioned—please know that my life sings for you.

Any young black, gay boy or man reading this: You have no idea just how 'ok' things can be. To all those who came before me, I speak your names. too.

I dedicate this thesis to the memory of Dr. Freddie T. Davy, beloved honors college director, mentor and friend.

All my love,

Weslee

## Table of Contents

|   |           |
|---|-----------|
| <b>Chapter 1: Introduction: Recent Progress in the Metabolic Engineering of Alkaloids.....</b>                        | <b>19</b> |
| I. Introduction.....  | 20        |
| II. The Monoterpene Indole Alkaloids.....   | 23        |
| A. Introduction.....  | 23        |
| B. Obtaining the building blocks for metabolic engineering efforts..  | 24        |
| C. Engineering ‘unnatural’ natural products.....  | 27        |
| III. The Benzyloisoquinoline Alkaloids.....   | 28        |
| A. Introduction.....  | 28        |
| B. Enzyme discovery and engineering in the BIA pathway.....   | 29        |
| C. Engineering in native hosts.....   | 30        |
| D. Reconstituting BIA biosynthesis in microbial systems.....  | 33        |
| IV. The Glucosinolates.....   | 34        |
| A. Introduction.....  | 34        |
| B. Improving yields in non-native hosts.....  | 35        |
| V. Scope of Thesis.....   | 38        |
| VI. References.....   | 40        |
| VII. Acknowledgements.....  | 44        |
| <b>Chapter 2: Discovery of 10-hydroxygeraniol Oxidoreductase Activity in <i>C. roseus</i>.....</b>                    | <b>45</b> |
| I. Introduction.....  | 46        |
| II. Results and Discussion.....   | 51        |
| III. Future Work.....   | 65        |
| IV. Conclusions.....  | 69        |
| V. Methods.....   | 71        |
| A. Gene cloning.....  | 71        |
| B. Protein expression.....  | 74        |
| C. TLC assays.....  | 75        |
| D. GC-MS assays.....  | 75        |
| E. VIGS.....  | 77        |
| VI. Appendix A contents.....  | 78        |
| VII. References.....  | 79        |
| VIII. Acknowledgements.....   | 80        |
| <b>Chapter 3: Reengineering a Tryptophan Halogenase to Preferentially Chlorinate a Direct Alkaloid Precursor.....</b> | <b>82</b> |
| I. Introduction.....  | 83        |
| II. Results and Discussion.....   | 87        |
| III. Conclusions.....   | 103       |
| IV. Methods.....  | 105       |
| A. <i>In planta</i> over-expression of RebH wild type, RebF and TDC.....  | 105       |

|   |  |            |
|---|--|------------|
| B.  | Stable transformation protocol – RebH Y455W and RebF.....  | 107        |
| C.  | RebH mutant design and expression .....  | 108        |
| D.  | RebH activity assays.....  | 111        |
| E.  | RebH WT, RebH Y455W and RebH L113G competition assay .....   | 114        |
| F.  | RebH WT and RebH Y455W at different typtophan-to-trytamine ratios .....  | 114        |
| G.  | <i>In planta</i> over-expression of RebH Y455W and RebF .....  | 115        |
| H.  | Transient <i>in planta</i> expression of RebH Y455W and RebF .....   | 118        |
| V.  | References.....  | 119        |
| VI.   | Acknowledgements.....  | 120        |
| <b>Chapter 4: Redesign of a Dioxygenase Involved in Morphine Biosynthesis ...</b> |  | <b>122</b> |
| I.  | Introduction .....   | 123        |
| II.   | Results and Discussion .....   | 126        |
| III.  | Conclusions .....  | 143        |
| IV.   | Methods .....  | 144        |
| A.  | Construction of <i>PsCODM</i> mutant expression plasmids.....  | 144        |
| B.  | Heterologous expression of <i>PsCODM</i> mutants .....   | 145        |
| C.  | In vitro activity assay of <i>PsCODM</i> mutants.....  | 145        |
| D.  | Steady-State kinetic assay of wild type CODM and A1A2BC1*C2 mutant .....   | 146        |
| E.  | Computational modeling.....  | 147        |
| V.  | References.....  | 148        |
| VI.   | Acknowledgments .....  | 149        |
| <b>Chapter 5: Conclusions and Future Directions: The Alkaloids and Beyond ...</b> |  | <b>150</b> |
| I.  | Conclusions .....  | 151        |
| II.   | Future Directions—The Alkaloids and Beyond.....  | 155        |
| A.  | Effective mining strategies .....  | 156        |
| B.  | Metabolic engineering in native versus non-native hosts .....  | 156        |
| C.  | Controlling metabolic flux through new expression constructs, scaffolds and tunable regulatory elements .....    | 157        |
| D.  | Transport—Engineering in multiple dimensions .....   | 160        |
| E.  | Elucidating the physiological relevance of alkaloids .....   | 160        |
| F.  | Combinatorial biosynthesis in plants—mixing and matching pathways and engineering new enzyme specificities ..... | 161        |
| III.  | References.....  | 163        |
| IV.   | Acknowledgements.....  | 165        |
| <b>Appendix A: Chapter 2 Supplemental Data.....</b>                               |  | <b>166</b> |
| I.  | DNA and Protein Sequences.....   | 166        |
| II.   | TLC Controls .....   | 172        |
| III.  | GC-MS Spectra .....  | 174        |
| <b>Curriculum Vitae.....</b>  |  | <b>221</b> |



## Table of Figures

|  |     |
|--|-----|
| Figure 1.1: Monoterpene indole alkaloid biosynthesis and strategies to engineer<br>Halogenation .....    | 26  |
| Figure 1.2: Benzlisoquinoline alkaloid biosynthesis .....  | 31  |
| Figure 1.3: Glucosinolate biosynthesis .....   | 36  |
| Figure 2.1: Structures of representative plant iridoids.....   | 46  |
| Figure 2.2: Proposed biosynthesis of secologanin.....  | 48  |
| Figure 2.3: Representative hierarchical cluster.....   | 53  |
| Figure 2.4: SDS PAGE gels of candidate expression .....  | 58  |
| Figure 2.5: Candidate TLC screens .....  | 60  |
| Figure 2.6: Candidate and iridoid synthase coupled assay.....  | 61  |
| Figure 2.7: GC-MS of representative positive hit.....  | 63  |
| Figure 2.8: Two possible routes from 10-hydroxygeraniol to 10-oxogeraniol.....                           | 66  |
| Figure 2.9: Multiple peaks in GC-MS chromatograms of positive hits .....                                 | 67  |
| Figure 2.10: Substrate analogs to test with 10-hydroxygeraniol reductase .....                           | 69  |
| Figure 2.11: USER cloning scheme .....   | 73  |
| Figure 2.12: Candidate PCR inserts.....  | 71  |
| Figure 2.13: Colony PCR of candidates for VIGS.....  | 78  |
| Figure 3.1: RebH and RebF over-expression in <i>C. roseus</i> .....                                      | 84  |
| Figure 3.2: Strategies to alleviate TDC bottleneck.....  | 86  |
| Figure 3.3: <i>C. roseus</i> hairy roots over-expressing TDC or TDC, RebH and RebF .....                 | 87  |
| Figure 3.4: Tryptophan complexed with RebH (PDB code: 2E4G).....   | 89  |
| Figure 3.5: RebH wild type and RebH Y455W competition assay chromatogram.....                            | 92  |
| Figure 3.6: RebH L113G competition assay chromatogram .....  | 93  |
| Figure 3.7: RebH wild type and RebH Y455W competitive assay quantitative<br>comparison.....              | 94  |
| Figure 3.8: Molecular models of RebH wild type and RebH Y455W with L-<br>tryptophan and tryptamine ..... | 95  |
| Figure 3.9: Pictet-Spengler mechanism.....   | 97  |
| Figure 3.10: RebH WT and RebH Y455W chlorotryptamine LC-MS product<br>Characterization .....             | 98  |
| Figure 3.11: RebH Y455W and RebF hairy root lines .....  | 100 |
| Figure 3.12: Metabolite analysis of RebH Y455W and RebF hairy root line .....                            | 101 |
| Figure 3.13: Exogenous TDC genomic DNA amplicon .....  | 107 |
| Figure 3.14: RebH WT and RebH Y455W <i>E. coli</i> expression .....                                      | 110 |
| Figure 3.15: RebH WT chlorotryptamine product controls.....  | 112 |
| Figure 3.16: RebH Y455W chlorotryptamine product controls .....  | 113 |
| Figure 3.17: Hairy root genomic DNA analysis for RebH Y455W and RebF .....                               | 117 |
| Figure 4.1: Benzylisoquinoline alkaloid biosynthesis.....  | 124 |
| Figure 4.2: Sequence alignment of dioxygenases from <i>Papaver somniferum</i> .....                      | 127 |
| Figure 4.3: Homology models of T6ODM and CODM.....   | 128 |
| Figure 4.4: CODM mutant expression .....   | 133 |
| Figure 4.5: Activity of CODM mutants with codeine.....   | 134 |
| Figure 4.6: Activity of CODM mutants with thebaine.....  | 135 |
| Figure 4.7: CODM mutant competition assays .....   | 135 |

|  |     |
|--|-----|
| Figure 4.8: Michaelis-Menten curve of wild type CODM with codeine.....   | 137 |
| Figure 4.9: Michaelis-Menten curve of wild type CODM with thebaine.....  | 138 |
| Figure 4.10: Michaelis-Menten curve of CODM mutant A <sub>1</sub> A <sub>2</sub> BC <sub>1</sub> *C <sub>2</sub> with codeine... | 138 |
| Figure 4.11: Proposed dioxygenase demethylation mechanism.....   | 139 |
| Figure 4.12: Thebaine docked onto wild type CODM homology model.....   | 139 |
| Figure 4.13: Thebaine docked into mutant A <sub>1</sub> A <sub>2</sub> BC <sub>1</sub> *C <sub>2</sub> homology model.....       | 140 |
| Figure 4.14: Codeine docked onto wild type CODM homology model.....  | 140 |
| Figure 4.15: Codeine docked onto mutant A <sub>1</sub> A <sub>2</sub> BC <sub>1</sub> *C <sub>2</sub> homology model.....        | 141 |
| Figure 5.1: Chapter 2 summary figure – Secologanin biosynthesis.....   | 152 |
| Figure 5.2: Chapter 3 summary figure – RebH reengineering.....   | 154 |
| Figure 5.3: Chapter 4 summary figure – CODM reengineering.....   | 155 |
| Figure 5.4: Design of a nucleophilic aromatic fluorinase.....  | 162 |
| Figure A.1: TLC of candidate no NAD(P) <sup>+</sup> and no enzyme controls.....  | 172 |
| Figure A.2: TLC of candidate no zinc controls.....   | 173 |
| Figure A.3: TLC of candidate chelation controls.....   | 173 |
| Figure A.4: TLC of 10-hydroxygeraniol authentic standard.....  | 173 |
| Figure A.5: GC-MS chromatogram of 10-hydroxygeraniol authentic standard.....   | 174 |
| Figure A.6: Mass spectrum of 10-hydroxygeraniol authentic standard.....  | 175 |
| Figure A.7: GC-MS chromatogram of 10-oxogernial authentic standard.....  | 176 |
| Figure A.8: Mass spectrum of 10-oxogeraniol authentic standard.....  | 177 |
| Figure A.9: GC-MS chromatogram of Candidate 1786 with NAD <sup>+</sup> .....   | 178 |
| Figure A.10: Mass spectrum of Candidate 1786 with NAD <sup>+</sup> product peak.....   | 179 |
| Figure A.11: Chromatogram of Candidate 1786 with NAD <sup>+</sup> - No ZnCl <sub>2</sub> control.....                            | 180 |
| Figure A.12: Mass spectrum of Candidate 1786 with NAD <sup>+</sup> -No ZnCl <sub>2</sub> control product<br>Peak.....            | 181 |
| Figure A.13: Chromatogram of Candidate 1786 with NAD <sup>+</sup> - EDTA control.....  | 182 |
| Figure A.14: Mass spectrum of Candidate 1786 with NAD <sup>+</sup> - EDTA control product<br>Peak.....                           | 183 |
| Figure A.15: GC-MS chromatogram of Candidate 1786 with NADP <sup>+</sup> .....   | 184 |
| Figure A.16: Mass spectrum of Candidate 1786 with NADP <sup>+</sup> product peak.....  | 185 |
| Figure A.17: Chromatogram of Candidate 1786 with NADP <sup>+</sup> - No ZnCl <sub>2</sub> control.....                           | 186 |
| Figure A.18: Mass spectrum of Candidate 1786 with NADP <sup>+</sup> - No ZnCl <sub>2</sub> control<br>product peak.....          | 187 |
| Figure A.19: GC-MS chromatogram of Candidate 1786 with NADP <sup>+</sup> - Chelation<br>Control.....                             | 188 |
| Figure A.20: Mass spectrum of Candidate 1786 with NADP <sup>+</sup> - EDTA control.....  | 189 |
| Figure A.21: GC-MS chromatogram of Candidate 1786 no NAD(P) <sup>+</sup> control.....  | 190 |
| Figure A.22: GC-MS chromatogram of Candidate 26 with NAD <sup>+</sup> .....  | 191 |
| Figure A.23: GC-MS chromatogram of Candidate 26 with NADP <sup>+</sup> .....   | 192 |
| Figure A.24: Mass spectrum of Candidate 26 with NADP <sup>+</sup> product peak.....  | 193 |
| Figure A.25: Chromatogram of Candidate 26 with NADP <sup>+</sup> - No ZnCl <sub>2</sub> control.....                             | 194 |
| Figure A.26: Mass spectrum of Candidate 26 with NADP <sup>+</sup> - No ZnCl <sub>2</sub> control product<br>Peak.....            | 195 |
| Figure A.27: GC-MS of Candidate 26 – NADP <sup>+</sup> - EDTA control.....   | 196 |
| Figure A.28: Mass spectrum of Candidate 26 with NADP <sup>+</sup> - EDTA control product<br>peak.....                            | 197 |

|  |     |
|--|-----|
| Figure A.29: GC-MS chromatogram of Candidate 26 – No NAD(P) <sup>+</sup> control .....                                   | 198 |
| Figure A.30: GC-MS chromatogram of Candidate 4319 with NAD <sup>+</sup> .....  | 199 |
| Figure A.31: Mass spectrum of Candidate 4319 with NAD <sup>+</sup> product peak .....                                    | 200 |
| Figure A.32: GC-MS chromatogram of Candidate 5743 with NAD <sup>+</sup> .....  | 201 |
| Figure A.33: Mass spectrum of Candidate 5743 with NAD <sup>+</sup> product peak .....                                    | 202 |
| Figure A.34: Chromatogram of Candidate 5743 with NAD <sup>+</sup> - No ZnCl <sub>2</sub> control.....                    | 203 |
| Figure A.35: Mass spectrum of Candidate 5743 with NAD <sup>+</sup> - No ZnCl <sub>2</sub> control product<br>peak.....   | 204 |
| Figure A.36: Chromatogram of Candidate 5743 with NAD <sup>+</sup> - EDTA control.....                                    | 205 |
| Figure A.37: Mass spectrum of Candidate 5743 with NAD <sup>+</sup> - EDTA control product<br>peak.....                   | 206 |
| Figure A.38: GC-MS chromatogram of Candidate 5743 with NADP <sup>+</sup> .....   | 207 |
| Figure A.39: Mass spectrum of Candidate 5743 with NADP <sup>+</sup> product peak.....                                    | 208 |
| Figure A.40: Chromatogram of Candidate 5743 with NADP <sup>+</sup> - No ZnCl <sub>2</sub> control.....                   | 209 |
| Figure A.41: Mass spectrum of Candidate 5743 with NADP <sup>+</sup> - No ZnCl <sub>2</sub> control<br>product peak ..... | 210 |
| Figure A.42: Chromatogram of Candidate 5743 with NADP <sup>+</sup> - EDTA control .....                                  | 211 |
| Figure A.43: Mass spectrum of Candidate 5743 with NADP <sup>+</sup> - EDTA control product<br>Peak.....                  | 212 |
| Figure A.44: GC-MS chromatogram of Candidate 5743- No NAD(P) <sup>+</sup> control.....                                   | 213 |
| Figure A.45: GC-MS chromatogram of Candidate 7220 with NAD <sup>+</sup> .....  | 214 |
| Figure A.46: GC-MS chromatogram of Candidate 7220 with NADP <sup>+</sup> .....   | 215 |
| Figure A.47: Mass spectrum of candidate 7220 with NADP <sup>+</sup> .....  | 216 |
| Figure A.48: 10-hydroxygeraniol oxidoreductase activity assay – no enzyme control<br>– NAD <sup>+</sup> .....            | 217 |
| Figure A.49: Mass spectrum of no enzyme control – NAD <sup>+</sup> .....   | 218 |
| Figure A.50: 10-hydroxygeraniol oxidoreductase activity assay – no enzyme control<br>– NADP <sup>+</sup> .....           | 219 |
| Figure A.51: Mass spectrum of no enzyme control – NADP <sup>+</sup> .....  | 220 |

## List of Tables

|   |     |
|---|-----|
| Table 1.1: Engineering strategies for monoterpene indole alkaloids, benzyloquinoline alkaloids and glucosinolates ..... | 39  |
| Table 2.1: Top 25 contigs from G10H mutual rank.....  | 54  |
| Table 2.2: Candidate filter summary .....   | 55  |
| Table 2.3: TargetP 1.1 localization prediction.....   | 56  |
| Table 2.4: Summary of expression data .....   | 57  |
| Table 2.5: Summary of candidate screen.....   | 59  |
| Table 2.6: Candidate primers from expression in pET28a .....  | 72  |
| Table 2.7: 10-hydroxygeraniol oxidoreductase VIGS primers .....   | 77  |
| Table 3.1: Summary of mutant assays .....   | 90  |
| Table 3.2: RebH Y455W and RebF Line 31 metabolic analysis.....  | 100 |
| Table 3.3: RebH Y455W and RebF Line 74 metabolic analysis.....  | 100 |
| Table 3.4: TDC plant over-expression primers .....  | 106 |
| Table 3.5: TDC genomic DNA analysis primers .....   | 106 |
| Table 3.6: RebH mutant primers .....  | 108 |
| Table 3.7: Primers to amplify RebH Y455W from genomic DNA.....  | 117 |
| Table 4.1: Tested CODM mutants.....   | 129 |
| Table 4.2: Primers and templates for site-directed mutagenesis .....  | 130 |
| Table 4.3: Summary of CODM wild type and mutant kinetic parameters.....   | 137 |

## List of Abbreviations

This thesis uses standard abbreviations for nucleic acids (one-letter code) and amino acids (one and three-letter codes). Standard SI units are also employed unless otherwise noted.

| Abbreviation       | Meaning   |
|--------------------|---|
| <sup>1</sup> H NMR | proton nuclear magnetic resonance                     |
| 4'-OMT             | 3'-hydroxy-N-methyl coclaurine-4'-O-methyltransferase |
| 6-OMT              | norcoclaurin 6-O-methyltransferase                    |
| ADH                | alcohol dehydrogenase                                 |
| BBE                | berberine bridge enzyme                               |
| BIA                | benzylisoquinoline alkaloid                           |
| BLAST              | basic local alignment search tool                     |
| cDNA               | complementary deoxyribonucleic acid                   |
| CNMT               | coclaurine-N-methyltransferase                        |
| CODM               | codeine-O-demethylase                                 |
| COR                | codeinone reductase                                   |
| CYP                | cytochromes P450                                      |
| DMAPP              | dimethyl allyl pyrophosphate                          |
| DMSO               | dimethyl sulfoxide                                    |
| DRR                | 1,2-dehydroreticulene reductase                       |
| DRS                | 1,2-dehydroreticuline synthase                        |
| DTT                | dithiothreitol  |
| EDTA               | ethylene diamine tetraacetic acid disodium salt       |
| ER                 | endoplasmic reticulum                                 |
| ESPs               | epithiospecifier proteins                             |
| EtOAc              | ethyl acetate   |
| FAD                | flavin adenine dinucleotide                           |
| FDA                | Federal Drug Administration                           |
| FPKM               | fragments per kilobase of exon per million            |
| G10H               | geraniol 10-hydroxylase                               |
| GC-MS              | gas chromatography- mass spectrometry                 |
| gDNA               | genomic deoxyribonucleic acid                         |
| Glc                | glucose   |
| GLS                | glucosinolate   |
| GSH                | glutathione   |
| His6               | hexahistidine   |
| HPLC               | high performance liquid chromatography                |
| Immat.             | immature  |
| IPP                | isopentyl pyrophosphate                               |
| IPTG               | Isopropyl-β-D-galactopyranoside                       |
| kDa                | kiloDalton  |
| LB                 | Luria-Burtani media                                   |
| LC-MS              | liquid chromatography - mass spectrometry             |
| MAO                | bacterial monoamine oxidase                           |

|          |   |
|----------|---|
| MeJa     | methyl jasmonate  |
| MIA      | monoterpene indole alkaloid                               |
| MS       | mass spectrometry   |
| NAD(+)   | nicotinamide adenine dinucleotide                         |
| NADH     | nicotinamide adenine dinucleotide (reduced)               |
| NADP(+)  | nicotinamide adenine dinucleotide phosphate               |
| NADPH    | nicotinamide adenine dinucleotide phosphate (reduced)     |
| NCS      | norcochlorine synthase                                    |
| NFPs     | nitrile-forming proteins                                  |
| P450s    | cytochromes P450  |
| PAPS     | 3'-phosphoadenosine-5'-phosphosulfate                     |
| PCC      | Pearson Correlation Coefficient                           |
| PDB      | Protein Data Bank   |
| RNAi     | ribonucleic acid interference                             |
| rpm      | revolutions per minute                                    |
| SDS PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SMT      | scoulerine 9-O-methyltransferase                          |
| STS      | strictosidine synthase                                    |
| T6ODM    | thebaine-6-O-demethylase                                  |
| TDC      | tryptophan decarboxylase                                  |
| TDCi     | tryptophan decarboxylase interference line                |
| TFPs     | thiocyanate-forming proteins                              |
| TLC      | thin layer chromatography                                 |
| TRV      | tobacco rattle virus                                      |
| VIGS     | virus-induced gene silencing                              |
| VIGS     | virus-induced gene silencing                              |
| WT       | wild type   |

# Chapter 1

Introduction: Recent Progress in the Metabolic Engineering of Alkaloids

Part of this chapter published as a perspective in

Glenn WS\*, Runguphan W\*, O'Connor SE. *Curr Opin Biotechnol.* 2013 Apr;  
24(2): 354-65 (\*indicates equal contribution)

## I. Introduction

Medicinal plants and the natural products derived from them have been exploited for thousands of years. For example, opium poppy has been employed both as an anaesthetic and a conduit to the spiritual world since at least the Neolithic Era.<sup>1,2</sup> Independently, the blue petals of some periwinkle varieties are said to invoke a sense of calm in Hoodoo practices (traditional African-American folk magic), while the leaves are believed to strengthen conjugal vows if sewn into a couple's mattress.<sup>2</sup>

The alkaloids themselves—isolated from poppy, periwinkle and other medicinal plants—have a particularly long and storied narrative as well. This history is highlighted in the life of Cleopatra, who used alkaloid-containing extracts from belladonna (Italian for 'beautiful woman') to dilate her pupils so as to increase her beauty and thereby disarm her enemies.<sup>3</sup> Far from being confined to ancient chronicles, the alkaloids retain a palpable presence in today's clinics. For example, optometrists still apply eye drops containing the alkaloid atropine, an active component of belladonna, to dilate the pupil during routine eye exams.<sup>3</sup>

It is unsurprising that most alkaloids are bioactive given that evolutionary processes select for the biosynthesis of products that confer an advantage to the producing organism. Despite the rich ethnopharmacological tradition and high usage of alkaloids in the modern era, relatively little is known about the biosynthesis, regulation and transport of these molecules. Access to these potent pharmaceuticals frequently pivots upon isolation from their native producers;



isolation typically requires laborious separation techniques that often result in low yields. Lacking a more sophisticated understanding of alkaloid biosynthesis significantly impedes our ability to co-opt nature's machinery in order to overproduce—that is, to metabolically engineer—these valuable compounds.

Notably, many drug screening efforts exclude plant natural products because of their high production costs and instead screen larger numbers of simpler synthetic molecules, which can be produced inexpensively and in fewer chemical steps.<sup>4</sup> Typically these high throughput screens are enriched with aromatic,  $sp^2$ -rich compounds that are obtained from commercial vendor libraries.<sup>5</sup> While these compounds have been useful for certain targets such as kinase inhibition, this area of chemical space is not always ideal for drug-like molecules.<sup>4,5</sup> In contrast, natural products and compounds inspired by natural products occupy a 'privileged' drug-like space, and comprise nearly half of all FDA approved drugs.<sup>5-7</sup> Importantly, natural products nearly always have more chiral carbons—a metric of complexity—than compounds typically found in commercial vendor libraries.<sup>5</sup> Given that natural products have specifically evolved to bind to cellular targets and exhibit some bioactivity that is beneficial to the host organism, it is hardly surprising that natural products are enriched for bioactivity, while relatively flat compounds uninspired by natural product structures typically fail pharmaceutical screens.<sup>8</sup>

Given the successes of getting natural products through the drug pipeline, we contend that more natural products—including plant alkaloids—should be included in drug screens. Plant natural products have a high success rate as

candidates and leads.<sup>6,7</sup> While the chemical syntheses of plant natural products, particularly the alkaloids, are dramatically improving,<sup>9</sup> many syntheses are still too lengthy for commercial production or require industrially impractical separation steps. Therefore, alternative production platforms must be developed, evaluated and instituted. An increasing body of work enlists microbes as well as cell and tissue cultures to produce these valuable plant-derived products.<sup>7</sup> Biological systems have the potential to be scalable and selective, while simultaneously being more environmentally friendly and—importantly—less expensive than synthetic reactions.<sup>7</sup>

In this chapter, we highlight recent metabolic engineering efforts designed to improve production of selected plant-derived alkaloids. We focus on the monoterpene indole alkaloids (MIAs), the benzyloisoquinoline alkaloids (BIAs) and the glucosinolates. Though not classically classified as alkaloids, the glucosinolates are nitrogen-containing compounds that have been the subject of a compelling body of research that will inform the forward engineering of all plant natural products. In total, these three classes of plant-derived nitrogen-containing natural products have been the subject of recent research efforts aimed at discovering and manipulating cellular activities, which include enzymatic function, metabolite transport and regulatory control. Ultimately this work may lead to biotechnologically useful enzymes and new drug candidates. Throughout this chapter, we also highlight the challenges that arise in attempting to chart the underexplored landscape of plant biosynthesis. Lastly, the chapter concludes by contextualizing the scope of this thesis.

## II. The Monoterpene Indole Alkaloids

### A. Introduction

The monoterpene indole alkaloids (MIAs) have garnered interest over the past few decades largely because of vinblastine **8** and vincristine **9**, two potent and widely prescribed anti-cancer agents that are currently produced solely through harvest from the leaves of mature periwinkle plants (*Catharanthus roseus*).<sup>10</sup> The concentrations of vinblastine and vincristine per gram of dry leaf material are approximately 0.01% and 0.003%, respectively, and are greatly dependent upon plant growth conditions.<sup>11</sup> Their low yields and lengthy production timeline have elicited intense efforts to engineer higher titers of these medicinally important MIAs.

The MIAs are encountered most commonly in the Apocynaceae, Loganiaceae and Rubiaceae families.<sup>12</sup> Most MIAs are built from the secoiridoid secologanin **3** and the indole-containing molecule tryptamine **2** (Figure 1a) [9]. Strictosidine synthase (STS) condenses these two molecules via a Pictet-Spengler condensation that forms strictosidine **4**. The  $\beta$ -carboline backbone that is formed via STS exhibits over 25 unique activities, highlighting the 'privileged' status of this class of compounds.<sup>7,13</sup>

Strictosidine **4**—the central precursor in MIA metabolism—is believed to ultimately succumb to either of two chemical fates.<sup>14,15</sup> If the plant is not under herbivore attack, strictosidine **4** is deglycosylated and rearranged into the over 3000 MIAs found in nature.<sup>15</sup> Madagascar periwinkle contains a subset of

approximately 130 MIAs. Alternatively, if the plant is under herbivore attack, the strictosidine 4 pool (estimated to be approximately 10 mM in periwinkle leaf epidermal cells hormonally treated to mimic herbivore attack) can be directed to the nucleus for mass deglycosylation, leading to a reactive dialdehyde species capable of cross-linking proteins.<sup>15</sup> This mechanism has been dubbed the strictosidine nuclear ‘bomb’ in reference to the ‘mustard oil bomb’ mechanism of glucosinolate biosynthesis (see below).<sup>15</sup> Importantly, many of the MIA metabolites themselves have also been implicated in plant defense strategies.<sup>16</sup>

#### *B. Obtaining the Building Blocks for Metabolic Engineering Efforts—A Case Study on the Discovery of P450s Involved in MIA biosynthesis*

The enzymatic pathways leading to the MIAs have not yet been fully elucidated in any organism. These uncharacterized biochemical steps may utilize novel chemistries or possess informative and interesting specificities that enable the enzymes to be employed in various synthetic metabolic pathway designs.<sup>17,18</sup> Notably, many plants—including MIA producers—are predicted to contain a high percentage of cytochromes P450 (P450s). Some estimates place P450s at approximately 1% of representative plant genomes,<sup>19</sup> over 5-fold higher than the proportion of P450s found in the human genome.<sup>20</sup> By using molecular oxygen to tailor hydrocarbon skeletons, P450s facilitate a panel of difficult chemical transformations and are consequently utilized in many alkaloid biosynthetic pathways.<sup>18</sup>

P450s have also been successfully engineered for biotechnological purposes.<sup>20,21</sup> Various technologies, such as nanodiscs,<sup>22</sup> and N-terminus reengineering efforts<sup>21</sup> have improved the expression of membrane-bound P450s, making this class of enzymes accessible to a full suite of biochemical and biophysical characterization techniques. Given the high sequence similarity of P450s, identifying a P450 that facilitates a specific biochemical reaction within a biosynthetic pathway remains a challenge. This has greatly slowed the discovery and characterization of new P450s within the plant kingdom. However, Giddings *et al.* recently used co-expression analysis to identify P450s with expression profiles similar to known MIA biosynthetic genes.<sup>17</sup> By functionally assaying these candidates in *Saccharomyces cerevisiae*, Giddings *et al.* discovered one P450 (CYP71BJ1) that hydroxylated the **19** position of either lochnericine or tabersonine **7**, an intermediate that is positioned at a metabolic branch point.<sup>17</sup> Hydroxylation of tabersonine **7** at the 16 position commits the intermediate to vindoline and vinblastine **8** biosynthesis, whereas hydroxylation at the 19 position commits the molecule to 19-O-acetylhörhammericine formation.<sup>17</sup> Controlling this switch may be important in engineering efforts designed to improve the titres of vindoline and vinblastine **8**. Moreover, similar strategies must be employed to find other missing pathway steps in MIA biosynthesis (Chapter 2) and other natural product pathways.

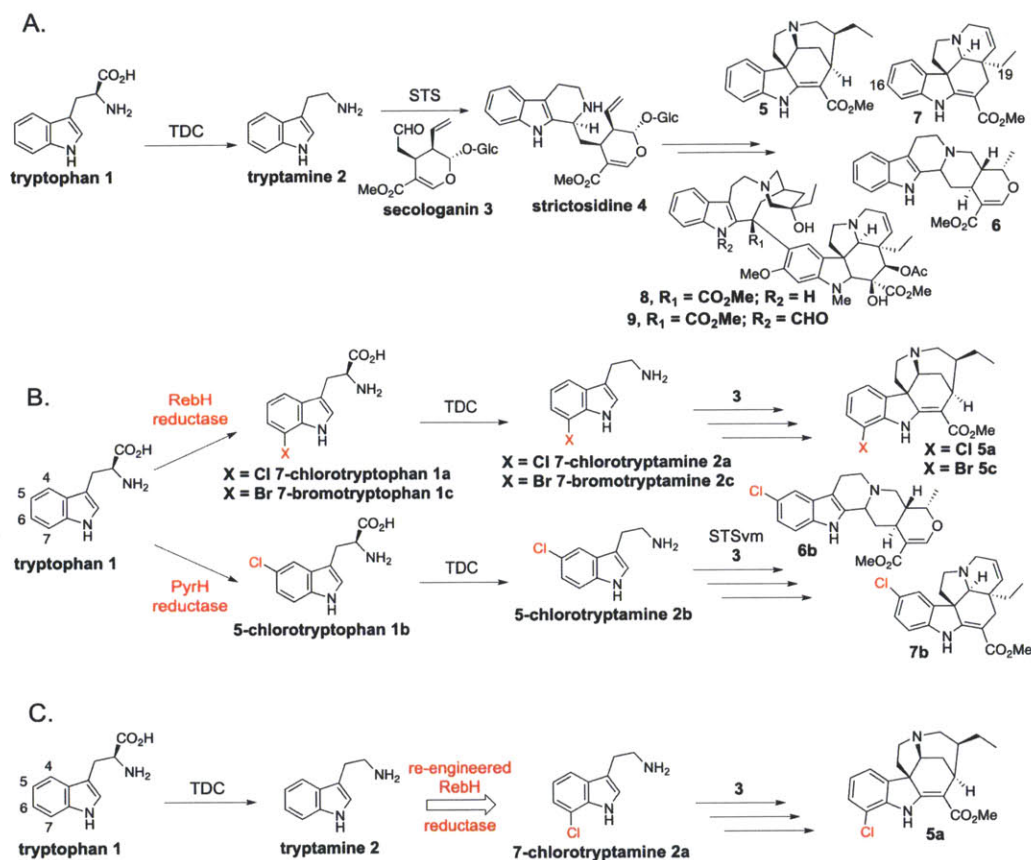


Figure 1.1: (A) The monoterpene indole alkaloid (MIA) pathway. TDC, tryptophan decarboxylase; STS, strictosidine synthase; Glc, glucose. (B) Introduction of halogenation into the MIA pathway. RebH and PyrH are both flavin-dependent halogenases from actinomycetes species; STS<sub>vm</sub>, strictosidine synthase Val214Met mutant. (C) Reengineering of halogenase to preferentially chlorinate tryptamine **2** over the natural substrate tryptophan **1** (Chapter 3).

### C. Engineering 'Unnatural' Natural Products

Plant alkaloids often require modification to improve their pharmacological properties prior to human consumption. Halogenation, particularly fluorination and chlorination, is a pervasive modification in successful pharmaceutical candidates.<sup>23,24</sup> Halogens often confer the potency of a drug, alter its pharmacokinetics or function as site-specific handles for subsequent modification.<sup>23-25</sup> Halogens can be introduced into MIA pathways by a number of methods. One particular example employed mutasynthesis, a process whereby natural biosynthesis is first blocked by genetic silencing of the natural precursor, and then rescued by feeding with structural analogs of the precursor.<sup>26</sup> In this case, in conjunction with the RNAi-mediated knockdown of tryptophan decarboxylase (TDC), unnatural tryptamine 2 analogs were added to a chemically 'silent'—non-alkaloid producing—background and fluorinated MIA analogs were observed.<sup>26</sup> In a separate engineering strategy, strictosidine synthase (STS)—the enzyme situated at the first committed step of MIA biosynthesis—was engineered to accept an expanded range of halogenated tryptamine 2 precursors;<sup>27</sup> the utility of this enzyme was demonstrated in planta by feeding previously unaccepted unnatural precursors to *C. roseus* hairy roots.<sup>28</sup>

Finally, Runguphan *et al.* interfaced RebH and PyrH—two tryptophan halogenases isolated from soil-dwelling actinomycetes species—with the MIA metabolism of periwinkle to produce halogenated natural products de novo [24] (Figure 1b). However, the lines overexpressing both RebH and RebF also displayed a brown and slow growth morphology.<sup>29</sup> Runguphan *et al.*

hypothesized that this morphology was the result of the accumulation of 7-chlorotryptan **1a**, an analog of primary metabolite L-tryptophan **1** that is somewhat structurally similar to 4-chloroindole-3-acetic acid, an auxin known to be involved in regulating plant growth.<sup>29</sup>

To circumnavigate this problem, Glenn *et al.* employed structure-guided protein design to engineer a halogenase that preferentially chlorinated tryptamine **2**, a more direct MIA precursor (Figure 1C) (Chapter 3).<sup>30</sup> Microgram per gram fresh weight quantities of 12-chloro-19,20-dihydroakuamicine **5a** were observed with this strategy, but neither 7-chlorotryptophan **1a** nor 7-chlorotryptamine **2a** accumulated in planta, indicating the chlorinated precursor was being effectively shuttled into MIA metabolism [25]. Engineering halogenation into MIA metabolism highlights an important need to interface specialized metabolism with primary carbon and nitrogen metabolism.

### III. The benzyloquinoline alkaloids

#### A. Introduction

Benzyloquinoline alkaloids (BIAs) are found mainly in the Papaveraceae, Ranunculaceae, Berberidaceae and Menispermaceae plant families. Approximately 2500 BIAs have been isolated to date.<sup>31</sup> This class of compounds has been used throughout human history and contains pharmaceuticals that are still widely used today, including the narcotic and analgesic morphine **28**, the cough suppressant codeine **26**, the muscle relaxant papaverine, and the anti-microbial agents sanguinarine **19** and berberine **29**. All known BIAs, like the



MIAs, are derived from a single intermediate, which, for this class of compounds, is norcoclaurine **13**. Norcoclaurine synthase (NCS) catalyzes the Pictet-Spengler condensation between dopamine **11** and 4-hydroxyphenylacetaldehyde **12** to yield the central intermediate, norcoclaurine **13**.

Notably, the biosynthetic pathways of several benzyloquinoline alkaloids—morphine **28**, sanguinarine **19** and berberine **29**—have been fully elucidated at the genetic level, which has enabled sophisticated metabolic engineering approaches. The application of metabolic engineering strategies for BIAs has focused predominantly on improving the yields of specific alkaloid compounds that exhibit medicinal value (Table 1.1).

### *B. Enzyme Discovery and Engineering in BIA Pathways*

Several outstanding efforts in enzyme discovery have been reported for BIA biosynthetic pathways. In a recent effort, Hagel *et al.* characterized two O-demethylases that are involved in morphine biosynthesis, completing the characterization of the morphinan pathway (Figure 2B).<sup>32</sup> This work also clearly highlighted how co-expression analysis can be used to discover enzymes with unprecedented catalytic function. These enzymes offer the first examples of non-heme iron(II) oxoglutarate dioxygenases capable of catalyzing O-demethylation. Codeine-O-demethylase (CODM) regioselectively demethylates codeine **26** and thebaine **22** at the 3-position, while thebaine-6-O-demethylase (T6ODM) demethylates thebaine **22** and oripavine **24** at the 6-position. Swapping amino acid regions between the two demethylases resulted in a CODM mutant that

selectively demethylates codeine (Figure 2B) (Chapter 4).<sup>33</sup> This mutant—which effectively sidesteps oripavine **24** production by committing thebaine **22** to just one of two possible routes—could potentially impact titers of codeine **26** and morphine **28** in subsequent metabolic engineering efforts. Collectively, these studies highlight how characterizing individual pathways steps and understanding their specificity and selectivity can both inform and enable metabolic engineering efforts.

While transcript analysis has proven to be spectacularly successful in elucidating the demethylases of morphine **28** biosynthesis, Winzer *et al.* provide a rare example of gene clustering in a BIA pathway.<sup>34</sup> The authors describe a 10-gene cluster in the poppy genome that putatively encodes the entire biosynthetic pathway of the BIA noscapine **15**. This is the first gene cluster discovered for an alkaloid pathway, and it is the largest plant gene cluster discovered to date. The authors further successfully silenced six of the ten proposed genes using VIGS to validate their role in noscapine **15** biosynthesis.<sup>34</sup> This study indicates that genomic data, in addition to expression data, can be used to decipher alkaloid pathways in plants.

### *C. Engineering in Native Hosts*

In one of the earliest attempts to engineer BIA-producing plants, RNA interference (RNAi) was used to silence the expression of codeinone reductase

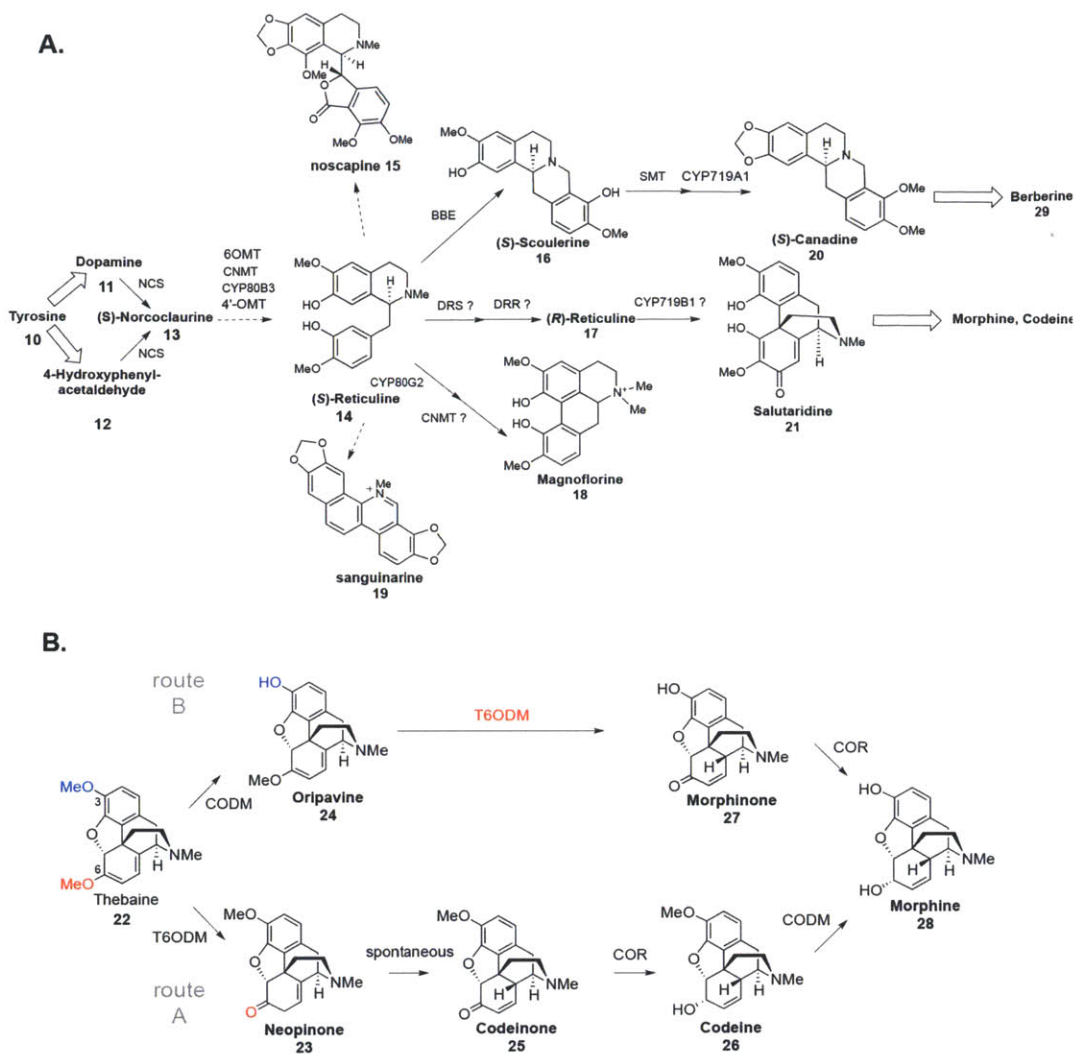


Figure 1.2: (A) The benzylisoquinoline alkaloid (BIA) pathway. NCS, norcoclaurine synthase; 6-OMT, norcoclaurine 6-O-methyltransferase; CNMT, coclaurine-N- methyltransferase; 4'-OMT, 3'-hydroxy-N-methylcoclaurine-4'-O-methyltransferase; DRS, 1,2-dehydroreticuline synthase; DRR, 1,2-dehydroreticulene reductase; BBE, berberine bridge enzyme; SMT, scoulerine 9-O-methyltransferase; MAO, bacterial monoamine oxidase; CYP2D6, human cytochrome P450 enzyme; CYP80G2, CYP719B1 and CYP719A1, plant cytochrome P450 enzymes. (B) The morphinan alkaloid pathway. CODM, codeine O-demethylase; T6ODM, thebaine 6-O-demethylase; COR, codeinone reductase.

in the opium poppy.<sup>35</sup> COR, the penultimate enzyme of morphine biosynthesis, converts codeinone **25** to codeine **26** (Figure 2b). While one might anticipate that the silencing of COR would lead to elevated levels of codeinone **25**, the study found instead that COR-silenced plants accumulated reticuline **14**—an intermediate seven steps upstream of codeinone **25**—at the expense of morphine **28**, codeine **26**, oripavine **24** and thebaine **22**. A feedback mechanism was proposed as an explanation for the elevated levels of reticuline **14**, though testing this hypothesis has yielded conflicting results.<sup>35</sup>

Other early attempts to improve the yields of BIA alkaloids include the overexpression of berberine bridge enzyme (BBE) in *Eschscholzia californica* root cultures. This effort resulted in elevated levels of downstream alkaloids and decreased levels of amino acids, though notably levels of tyrosine **10**—the amino acid employed in BIA synthesis—were unaltered.<sup>36</sup> Conversely, the antisense suppression of BBE expression led to the effective silencing of BIA production and increased cellular amino acid levels, though, again, tyrosine **10** levels went largely unchanged (less than two-fold higher than in control lines).<sup>37</sup> Nonetheless, these two studies highlight how perturbations in alkaloid metabolism can impact primary metabolism.<sup>36,37</sup> More recent studies suggest, however, that the RNAi suppression of BBE in *E. californica* leads to increased accumulations of (S)-reticuline **14** instead of various canonical amino acids.<sup>38</sup> These contradictory results are surprisingly common in the metabolic engineering of alkaloids in plants and cell cultures and provide us with the impetus to understand these

pathways in greater detail, paying specific attention to their biochemical and molecular regulatory elements.

#### *D. Reconstituting BIA Biosynthesis into Microbial Systems*

Many pathways in BIA biosynthesis are fully characterized, which opens the possibility of transplanting entire alkaloid pathways into microbial hosts. Though relatively difficult, the reconstitution of entire metabolic pathways into microbial hosts confers a number of advantages, including rapid biomass accumulation, facile purification and access to the host of tools available for workhorse organisms like *E. coli* and *S. cerevisiae*. A number of recent reports have successfully reconstituted portions of BIA pathways into *S. cerevisiae*, *E. coli* and combinations thereof in co-culture systems.<sup>39-41</sup> For example, Hawkins *et al.* were able to produce reticuline **14** as well as sanginarine/berberine-type and morphinan-type BIAs in yeast by over-expressing genes from mixed plant sources and human.<sup>40</sup> Notably, they were also able to tune enzyme expression levels through use of a glucocorticoid-inducible promoter and in situ promoter titration.<sup>35</sup> This level of tuning enables maximal pathway flux and minimal enzyme expression. The expression system is nominally taxed under these conditions, since valuable cellular resources are not used on the biosynthesis of supernumerary proteins and nucleic acids.

## IV. The Glucosinolates

### *A. Introduction*

The glucosinolates are not classified as alkaloids, although, along with the alkaloids, these compounds are amino acid-derived, nitrogen-containing small molecules of plant origin. The glucosinolates are included in this chapter because the recent and creative metabolic engineering studies performed on this class of compounds will undoubtedly inform the forward engineering of all plant natural products, particularly the alkaloids, which also contain nitrogen. Glucosinolates are sulfur- and nitrogen-containing compounds that are derived from glucose and various amino acids (Figure 3A).<sup>42</sup> They are found in cruciferous vegetables (the Brassicaceae plant family) and have been shown to possess a range of bioactivities.<sup>42</sup> The glucosinolates occupy an essential space in the chemical ecology of their host organisms by attracting specialist crucifer pollinators and insects and deterring predatory herbivores.<sup>43</sup> Specifically, crucifers employ myrosinases (hydrolases) to cleave the glucose moiety of glucosinolates in response to predation and herbivory (Figure 3B).<sup>44</sup> The myrosinases and glucosinolates are physically segregated within the plant, coming into contact only upon disruption of the plant tissue (Figure 3B).<sup>44</sup> Upon hydrolysis, the resultant unstable aglycone intermediate spontaneously rearranges into the corresponding isothiocyanate via a Lossen-type rearrangement.<sup>44</sup> The three known types of specifier proteins, Thiocyanate-Forming Proteins (TFPs), Nitrile-Forming Proteins (NFPs) and Epithiospecifier Proteins (ESPs)—which can be

found in planta or in various specialist insects—can redirect glucosinolate hydrolysis from isothiocyanate products toward thiocyanate, simple nitrile and epithionitrile products, respectively (Figure 3B).<sup>44</sup> Notably, many specifiers can direct glucosinolate hydrolysis to more than one product.<sup>44</sup> Early workers on this plant defense and pollination system dubbed it ‘The Mustard Oil Bomb’.<sup>43</sup> To date, over 120 glucosinolates have been identified.<sup>45</sup>

### *B. Improving Yields in Non-Native Hosts*

The reconstitution of entire metabolic pathways into heterologous plant hosts requires the use of efficient and facile ‘gene stacking’ methodologies. A spectacularly successful example is the engineering of benzylglucosinolate biosynthesis into *Nicotiana benthamiana*. Benzylglucosinolate was reconstituted in *N. benthamiana* using a transient expression system. In this study, Geu-Flores *et al.* identified a  $\gamma$ -glutamyl peptidase bottleneck, suggesting that reduced sulfur is incorporated into glucosinolates via glutathione conjugation (Figure 3A).<sup>42</sup> The co-expression of this peptidase augmented the yield of benzylglucosinolate 5.7-fold, indicating how consideration of primary metabolite resources can impact natural product yield.<sup>42</sup>

In a separate metabolite analysis, Møldrup *et al.* monitored the accumulation of desulfobenzylglucosinolate, the penultimate product in the benzylglucosinolate pathway.<sup>46</sup> Directing sulfur from primary to secondary metabolism through the co-expression of adenosine 5-phosphosulfate kinase—which provides the 3'-phosphoadenosine-5'-phosphosulfate (PAPS) co-

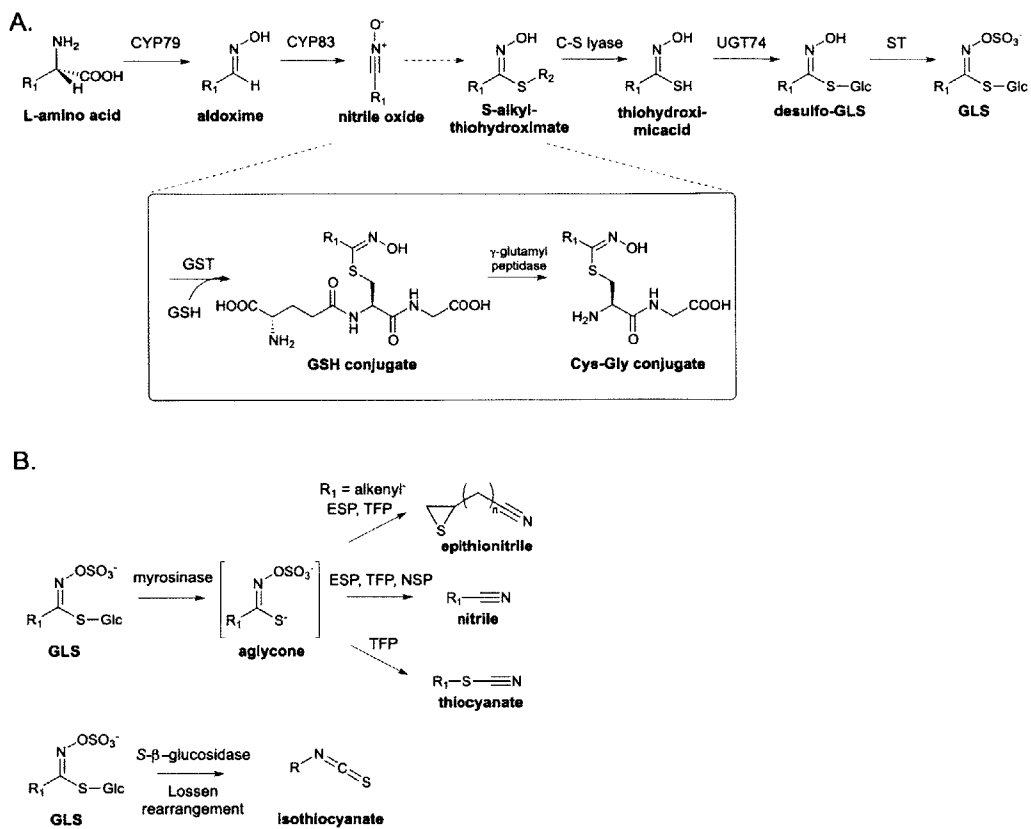


Figure 1.3: (A) The glucosinolate (GLS) pathway. GSH, glutathione. (B) Glucosinolate hydrolysis to form epithionitrile, nitrile, thiocyanate and isothiocyanate.



-substrate necessary for the final step of benzylglucosinolate biosynthesis (Figure 3A)—in the *N. benthamiana* expression system alleviated the subsequent bottleneck and increased the benzylglucosinolate yield by 16-fold.<sup>46</sup>

In yeast, Mikkelsen *et al.* were able to reconstitute the biosynthesis of indolyglucosinolate.<sup>45</sup> This example was a proof-of-concept study for a technology that enables the stacking of large numbers of genes, a requirement for total pathway reconstitution. Notably, the benzylglucosinolate biosynthetic pathway has also been stably transformed into *Nicotiana tabacum*, another non-cruciferous plant, which does not normally produce glucosinolates.<sup>47</sup> This reengineered plant has been shown to attract the diamondback moth (*Plutella xylostella*) and encourage oviposition (the deposition of eggs), highlighting its potential utility as a dead-end trap crop to deter predatory insects and prevent billion-dollar damages to cruciferous crops worldwide.<sup>47</sup>

Forward engineering in non-native hosts is particularly attractive if the product distribution converges to one or a few products. This obviates the need for taxing and costly separation procedures and can allow for rapid biomass accumulation. Moreover, forward engineering can increase product gains, as the engineering takes place in a nearly or completely chemically silent background. In contrast, over-expressing or silencing single genes in the context of normal plant primary and secondary metabolism typically does not significantly alter the product profile.

## V. Scope of Thesis

This thesis—*Understanding and Manipulating Alkaloid Biosynthesis*—commences with an effort to identify 10-hydroxygeraniol oxidoreductase activity in Madagascar periwinkle (Chapter 2). 10-hydroxygeraniol oxidoreductase is an enzyme involved in the biosynthesis of secologanin, the terpenoid precursor for all MIAs. A high proportion of genes in the *C. roseus* transcriptomic datasets are predicted to facilitate oxidation or reduction steps, making the discovery of the physiologically relevant enzymes(s) involved in this transformation difficult.

Despite having an incomplete understanding of MIA biosynthesis, the pathway has still been amenable to various engineering strategies, most notably incorporation of halogens. Chapter 3 discusses our efforts to engineer halogenation into periwinkle by redesigning RebH, a tryptophan halogenase, to preferentially chlorinate tryptamine, a direct MIA precursor. We subsequently incorporated this reengineered halogenase into alkaloid biosynthesis and observed the de novo biosynthesis of a halogenated ‘unnatural’ natural product.

The final research chapter (Chapter 4) explores mixing-and-matching closely related protein sequences in benzyloquinoline alkaloid biosynthesis to generate an enzyme with novel activity. Specifically, we systematically swapped residues from *PsT6ODM*—a dioxygenase that demethylates the 6 position of oripavine and thebaine—into *PsCODM*, a dioxygenase that demethylates the 3

| Class          | Engineering Strategy  | System   | Observations  | Ref. |
|----------------|---|--|---|------|
| MIAs           | Overexpression of TDC   | <i>C. roseus</i> crown gall  | Did not significantly alter alkaloid yield  | 49   |
| MIAs           | Overexpression of transcription factor CrWRKY1  | <i>C. roseus</i> hairy root cultures   | 3-fold increase in serpentine<br>10-fold increase in ajmalicine<br>2-fold decrease in catharanthine   | 50   |
| MIAs           | Overexpression of transcription factor CrWRKY1  | <i>C. roseus</i> hairy root cultures   | 300-fold increase in tryptophan<br>10-fold increase in tryptamine<br>1.8-fold in lochnericine   | 51   |
| MIAs           | Overexpression of alpha or alpha and beta subunits of anthranilate synthase and feeding of 10-deoxy-D-xylose, loganin and secologanin |  | 2.3-fold increase in horhammericine<br>1.5-fold increase in catharicine<br>1.3-fold increase in catharanthine<br>1.8-fold increase in ajmalicine<br>2.1-fold increase in lochnericine<br>4.5-fold increase in tabersonine   | 52   |
| MIAs           | Methyl jasmonate elicitation<br>Fed loganin and/or tryptamine   | <i>C. roseus</i> hairy root tissue cultures  | 3.2-fold increase in strictosidine<br>8.8-fold increase in ajmalicine<br>8.4-fold increase in tabersonine<br>Substrate feeding did not increase yield in elicited cultures  | 53   |
| MIAs           | Feeding with unnatural tryptamine precursors  | <i>C. roseus</i> tissue cultures   | observed up/ gram fresh weight quantities of many modified MIAs   | 54   |
| MIAs           | Mutasynthesis   | Fed tryptamine isomers to TDC-silenced petiwinkle hairy roots                                      | observed up/ gram fresh weight quantities of many modified MIAs<br>Achieved "alkaloid free" background  | 26   |
| MIAs           | Substrate feeding to roots harboring STS with expanded substrate specificity  | <i>C. roseus</i> hairy root cultures   | observed up/gram fresh weight quantities of previously metabolically inaccessible modified products   | 28   |
| MIAs           | Systematic knockdown of vindoline pathway enzymes through VIGS  | <i>C. roseus</i> seedlings   | Observed accumulation of direct precursors to silenced genes  | 11   |
| MIAs           | Overexpression of RebH/RebF and PyrH/RebF in planta   | <i>C. roseus</i> hairy root cultures   | observed the de novo production of 12-chloro-19,20-dihydroakuammicine (25 ug/ gram fresh weight)<br>Observed accumulation of 7-chlorotryptophan (50 ug/g fresh weight)  | 29   |
| MIAs           | Overexpression of RebF and reengineered RebH in planta  | <i>C. roseus</i> hairy root cultures   | Alleviated TDC bottleneck (no accumulation of 7-chlorotryptophan)<br>observed the de novo production of 12-chloro-19,20-dihydroakuammicine (2.7 ug/ gram fresh weight)<br>over-expression of TDC led to slow-growing hairy roots  | 30   |
| BIAs           | heterologous expression system for production of 5-reticuline and downstream alkaloids  | Artificial pathway in yeast  | Produced 5-reticuline in yields up to 150 mg/L<br>Demonstrated production of (S)-tetrahydroberberine (<5 mg/L)<br>Demonstrated production of (S)-scoulerine<br>Demonstrated production of (S) tetrahydrocolumbarine (60 mg/L)   | 40   |
| BIAs           | heterologous expression system for production of 5-reticuline   | Artificial pathway in <i>E. coli</i>   | Produced (S)-reticuline in yields up to 40 mg/L   | 41   |
| BIAs           | Suppression of codeinone reductase expression   | RNAi in Opium poppy  | Accumulation of (S)-reticuline (7 metabolic steps upstream)   | 35   |
| BIAs           | Overexpression of BBE in <i>E. californica</i>  | <i>E. californica</i> root cultures  | 5.8-fold increase of total downstream alkaloids<br>3.2-fold decrease in leucine concentration<br>2.4-fold decrease in threonine concentration<br>2.1-fold decrease in valine concentration<br>Tyrosine levels unaltered   | 36   |
| BIAs           | Suppression of BBE expression   | <i>E. californica</i> root cultures  | Elevated levels of (S)-reticuline to 310 ug/g cell fresh weight<br>(S)-reticuline levels at 6 mg/20 mL in the media   | 38   |
| BIAs           | Systematic knockdown of morphinan pathway enzymes through VIGS  | Opium poppy seedlings  | Suppression of SalSyn, SalR, T6ODM, CDDM resulted in concomitant increase of direct precursors<br>Suppression of SalAT resulted in the accumulation of salutaridine (not direct precursor)<br>Suppression of COR resulted in increases of (S)-reticuline (not direct precursor) | 55   |
| Glucosinolates | Benzylglucosinolate pathway reconstitution  | Transient expression in <i>N. benthamiana</i><br>Coexpression of gamma-glutamyl peptidase          | Produced benzylglucosinolate in yields up to 0.57 nmol/ mg freshweight leaf tissue<br>coexpression of peptidase raised yield 5.7-fold   | 42   |
| Glucosinolates | Benzylglucosinolate pathway reconstitution  | Transient expression in <i>N. benthamiana</i><br>coexpression of adenosine 5-phosphosulfate kinase | Production of benzylglucosinolate in yields up to 1.8 nmol/ mg freshweight<br>expression of kinase elevated yield by 16-fold  | 46   |
| Glucosinolates | Indolylglucosinolate pathway reconstitution   | yeast  | Reach indolylglucosinolate titers as high as 1.07 mg/ ml<br>product excreted into the media   | 47   |
| Glucosinolates | Benzylglucosinolate pathway reconstitution  | Stable expression in <i>N. tabacum</i>   | benzylglucosinolate titers as high as 0.5 nmol/mg fresh weight<br>Increase oviposition on non-cruciferous plants  | 48   |

Table 1.1 Engineering strategies across the three classes of compounds and their outcomes.

position of thebaines and codeine. The resulting enzyme was a *PsCODM* mutant that was specific for codeine. This switch in selectivity does not readily correlate with the substrate specificity of the parent enzymes. This chapter highlights the difficulty of rationally redesigning enzymes. Nonetheless, because the mutant is specific for codeine, it could presumably be used in reconstitution efforts to disable a redundant route in morphine biosynthesis.

The thesis closes with major conclusions of the work discussed within this text followed by the grand challenges and future work of the field (Chapter 5). From folk magic to clinics, plant-derived natural products have an exciting and storied past and hopefully a rich and expansive future.

## VI. References

1. Bernhardt, P. and O'Connor, S.E., Opportunities for Engineering in Natural Product Biosynthesis. *Curr Opin Chem Biol.* 2009, 13(1), 35-42.
2. Yronwode, C., *Hoodoo Herb and Root Magic: A Materia Magica of African-American Conjure and Traditional Formulary Giving the Spiritual Uses of Natural Herbs, Roots, Minerals and Zoological Curios*, 1<sup>st</sup> Edition. The Lucky Mojo Curio Co. Forestville, California 2002.
3. Shibamoto T. and Bjeldanes L.F., *Introduction to Food Toxicology*. Academic Press (Elsevier) 2009.
4. Lipinski C. and Hopkins A., Navigating Chemical Space for Biology and Medicine. *Nature* 2004, 432: 855-861.
5. Dandapani, S. and Marcaurelle, L. A., Accessing New Chemical Space for 'Undruggable' Targets. *Nat. Chem. Biol.* 2010, 6, 861-3.
6. Rosén, J.; Gottfries, J.; Muresan, S.; Backlund, A. and Oprea, L.A., Novel chemical space exploration via natural products. *J Med Chem.* 2009, 52, 1953-1962.
7. Leonard, E.; Runguphan, W.; O'Connor, S. E. and Prather K.J., Opportunities in metabolic engineering to facilitate scalable alkaloid production. *Nat Chem Biol.* 2009, 5: 292-300.
8. Lovering, F.; Bikker, J. and Humblet, C., Escape from Flatland: Increasing Saturation as an Approach to Increasing Clinical Success. *J. Med. Chem.*

- 2009, 52, 4, 6752-6.
9. Martin, D.B.C; Nguyen, L.Q. and Vanderwal, C.D., Syntheses of strychnine, norfluorocurarine, dehydrodesacetylretuline, and valparicine enabled by intramolecular cycloadditions of Zincke aldehydes. *J Org Chem.* 2012, 77, 17-46.
  10. Murata, J; Roepke, J; Gordon, H and De Luca, V: The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* 2008, 20, 524-542.
  11. Liscombe, D.K. and O'Connor, S.E., A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. *Phytochemistry* 2011, 72, 1969-1977.
  12. Seigler, D.S: *Plant Secondary Metabolism*. Kluwer Academic Publishers; 1998.
  13. Beghyn, T.; Deprez-Poulain, R.; Willard, N.; Folleas, B. and Deprez, -B. Natural Compounds: Leads or Ideas? Bioinspired Molecules for Drug Discovery. *Chem Biol Drug Des.* 2008, 72, 3-15.
  14. O'Connor S.E. and Maresh J. J., Chemistry and Biology of Monoterpene Indole Alkaloid Biosynthesis. *Nat Prod Rep.* 2006, 23, 532-547.
  15. Guirimand, G.; Courdavault, V.; Lanoue, A.; Mahroug, S.; Guihur, A.; Blanc, N.; Giglioli-Guivarc'h, N.; St-Pierre, B. and Burlaut, V., Strictosidine Activation in Apocynaceae: Towards a "Nuclear Time Bomb"? *BMC Plant Biology* 2010,10.
  16. Dewick P.M., *Medicinal Natural Products A Biosynthetic Approach 2<sup>nd</sup> Edition*. John Wiley & Sons, West Sussex, England 2001.
  17. Giddings, L.-A.; Liscombe, D.K.; Hamilton, J.P.; Childs, K.L.; DellaPenna, D.; Buell, C.R. and O'Connor, S.E., A Stereoselective Hydroxylation Step of Alkaloid Biosynthesis by a Unique Cytochrome P450 in *Catharanthus roseus*. *J Biol Chem.* 2011, 286, 16751-16757.
  18. Mizutani, M. and Sato, F., Unusual P450 Reactions in Plant Secondary metabolism. *Arch Biochem Biophys* 2011, 507, 194-203.
  19. Mizutani, M. and Ohta, D., Diversification of P450 Genes during Land Plant Evolution. *Annu Rev Plant Biol.* 2010, 61, 291-315.
  20. Sawayama, A.M.; Chen, M. M.Y; Kulanthaivel, P.; Kuo, M.-S.; Hemmerle, H. and Arnold, F.H., A panel of Cytochrome P450 BM3 Variants to Produce Drug Metabolites and Diversify Lead Compounds. *Chem Eur J.* 2009, 15, 11723-11729.
  21. Leonard, E and Koffas M.A.G., Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Appl Environ Microbiol.* 2007, 73, 7246-7251.
  22. Denisov, I. and Sligar, S.G., Cytochromes P450 in Nanodiscs. *Biochemica et*

- Biophysica* 2011, 1814, 223-229.
23. Neumann, C.S., Fujimori, D.G. and Walsh, C.T., Halogenation Strategies in Natural Product Biosynthesis. *Chem Biol.* 2008, 15: 99-109.
  24. Herrera-Rodriguez, L.N.; Khan, F.; Robins, K.T. and Meyer, H.-P. Perspective on biotechnological halogenation Part I: Halogenated Products and Enzymatic Halogenation. *Chemistry Today* 2011, 29, 31-33.
  25. Blasiak, L.C. and Drennan, C.L., Structural Perspective on Enzymatic Halogenation. *Acc Chem Res.* 2009, 42, 147-155.
  26. Runguphan, W.; Maresh, J.J. and O'Connor, S.E., Silencing of Tryptamine Biosynthesis for Production of Nonnatural Alkaloids in Plant Cultures. *Proc Natl Acad Sci. U S A* 2009, 106: 13673-13678.
  27. Bernhardt, P.; McCoy, E. and O'Connor, S.E., Rapid identification of Enzyme Variants for Reengineered Alkaloid Biosynthesis in Periwinkle. *Chem Biol.* 2007, 14: 888-897.
  28. Runguphan, W and O'Connor, S.E., Metabolic Reprogramming of Periwinkle Plant Culture. *Nat Chem Biol.* 2009, 5:151-153.
  29. Runguphan, W.; Qu, X. and O'Connor, S.E., Integrating Carbon-Halogen Bond Formation into Medicinal Plant Metabolism. *Nature* 2010, 18: 461-464.
  30. Glenn, W.S., Nims, E. and O'Connor, S.E., Reengineering a Tryptophan Halogenase to Chlorinate a Direct Alkaloid Precursor. *J Am Chem Soc.* 2011, 133, 19346-19349.
  31. Facchini, P and De Luca, V., Opium Poppy and Madagascar Periwinkle: Model Non-Model Systems to Investigate Alkaloid Biosynthesis in Plants. *Plant J.* 2008, 54, 763-784.
  32. Hagel, J.M. and Facchini, P.J. Dioxygenases catalyze the O-demethylation Steps of Morphine Biosynthesis in Opium Poppy. *Nat. Chem. Biol.* 2010, 6, 273-275.
  33. Runguphan, W.; Glenn, W.S. and O'Connor, S.E., Redesign of a Dioxygenase in Morphine Biosynthesis. *Chem Biol.* 2012, 19: 674-678.
  34. Winzer, T.; Gazda, V.; He, Z.; Kaminski, F.; Kern, M.; Larson, T.R.; Li, Y.; Meade, F.; Teodor, R.; Vaistij FE *et al.*: A *Papaver Somniferum* 10-gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. *Science* 2012, 336, 1704-1708.
  35. Allen, R.S.; Milligate, A.G.; Chilty, J.A.; Thisleton, J.; Miller, J.A.; Fist, A.J.; Gerlach, W.L. and Lancin, P.J. RNAi-Mediated Replacement of Morphine with the Nonnarcotic Alkaloid Reticuline in Opium Poppy. *Nat. Biotechnol.* 2004, 22, 1559-66.
  36. Park, S.-U.; Yu, M. and Facchini, P.J., Modulation of Berberine Bridge Enzyme Levels in Transgenic Root Cultures of California Poppy Alters the Accumulation of Benzophenanthridine Alkaloids. *Plant mol biol.* 2003, 51, 153-

164.

37. Park, S.-U. and Facchini, P.J., Antisense RNA-Mediated Suppression of Benzophenanthridine Alkaloid Biosynthesis in Transgenic Cell Cultures of California Poppy. *Plant Physiol.* 2002, 128, 696-706.
38. Fujii, N.; Inui, T.; Iwasa, K.; Morishige, T. and Sato, F., Knockdown of Berberine Bridge Enzyme by RNAi Accumulates (S)-reticuline and Activates a Silent Pathway in Cultured California Poppy Cells. *Transgenic Res.* 2007, 16, 363-375.
39. Minami, H.; Kim, J.S.; Ikezawa, N.; Takemura, T.; Katayama, T.; Kumagai, H. and Sato, F.: Microbial Production of Plant Benzylisoquinoline Alkaloids. *Proc Natl Acad Sci U S A* 2008, 105, 7393-7398.
40. Hawkins, K.M. and Smolke, C.D., Production of Benzylisoquinoline Alkaloids in *Saccharomyces cerevisiae*. *Nat Chem Biol.* 2008, 4: 564-573.
41. Nakagawa, A.; Minami, H.; Kim, J.-S.; Koyangi, T.; Katayama, T.; Sato, F. and Kumagai, H.: A Bacterial Platform for Fermentative Production of Plant Alkaloids. *Nat Commun.* 2011, 2.
42. Geu-Flores, F.; Nielsen, M.T.; Nafisi, M.; Møldrup, M.E.; Olsen, C.E.; Motawia, M.S. and Halkier, B.A.: Glucosinolate Engineering Identifies a  $\gamma$ -glutamyl Peptidase. *Nat Chem Biol.* 2009, 5, 575-577.
43. Ratzka, A.; Vogel, H.; Kliebensten, D.J.; Mitchell-Olds, T. and Kroymann, J., Disarming the Mustard Oil Bomb. *Proc Natl Acad Sci U S A.* 2002, 99: 11223-11228.
44. Wittstock, U. and Burow, M.: Tipping the Scales--Specifier Proteins in Glucosinolate Hydrolysis. *IUBMB Life* 2007, 59: 744-751.
45. Mikkelsen, M.D.; Buron, L.D.; Salomonse, B.; Olsen, C.E.; Hansen, B.G.; Mortensen, U.H.; and Halkier, B.A., Microbial Production of Indolylglucosinolate through Engineering of a Multi-Gene Pathway in a Versatile Yeast Expression Platform. *Metab Eng.* 2012, 14, 104-111.
46. Møldrup, M.E.; Geu-Flores, F.; Olsen, C.E. and Halkier, B.A., Modulation of Sulfur Metabolism Enables Efficient Glucosinolate Engineering. *BMC Biotechnol.* 2011,11.
47. MØldrup, M.E.; Geu-Flores, F.; de Vos, M.; Olsen, C.E.; Sun, J.; Jander, G.; Halkier, B.A., Engineering of Benzylglucosinolate in Tobacco Provides Proof-of-Concept for Dead-End Trap Crops Genetically Modified to Attract *Plutella xylostella* (diamondback moth). *Plant Biotechnol J.* 2012, 10: 435-442.
48. Goddijn, O.J.; Pennings, E.J.; van der Helm, P.; Schilperoot, R.A.; Verpoorte, R. and Hoge, J.H., Overexpression of a Tryptophan Decarboxylase cDNA in *Catharanthus roseus* Crown Gall Calluses Results in Increased Tryptamine Levels but Not in Increased Terpenoid Indole Alkaloid Production. *Transgenic Res.* 1995, 4: 315-323.

49. Suttipanta, N.; Pattanaik, S.; Kulshrestha, M.; Patra, B.; Singh, S. K.; Yuan, L., The transcription Factor CrWRKY1 Positively Regulates the Terpenoid Indole Alkaloid Biosynthesis in *Catharanthus roseus*. *Plant Physiol.* 2011, 157, 2081-2093.
50. Hughes, E.H.; Hong, S.B.; Gibson, S.I.; Shanks, J.V. and San, K.Y., Expression of a Feedback-Resistant Anthranilate Synthase in *Catharanthus roseus* Hairy Roots Provides Evidence for Tight Regulation of Terpenoid Indole Alkaloid Levels. *Biotechnol Bioeng.* 2004, 86: 718-727.
51. Peebles, C.A.; Hong, S.B.; Gibson, S.I.; Shanks, J.V. and San K.Y., Effects of terpenoid precursor feeding on *Catharanthus roseus* hairy roots over-expressing the alpha or the alpha and beta subunits of anthranilate synthase. *Biotechnol Bioeng.* 2006, 93, 534-540.
52. Goklany, S.; Loring, R.H.; Glick, J.; Lee-Parsons, C.W.; Assessing the Limitations to Terpenoid Indole Alkaloid Biosynthesis in *Catharanthus roseus* Hairy Root Cultures through Gene Expression Profiling and Precursor Feeding. *Biotechnol Prog.* 2009, 25: 1289-1296.
53. McCoy, E. and O'Connor, S.E., Directed Biosynthesis of Alkaloid Analogs in the Medicinal Plant *Catharanthus roseus*. *J Am Chem Soc.* 2006, 128: 14276-14277.
54. Wijekoon, C.P.; Facchini, P.J.; Systematic Knockdown of Morphine Pathway Enzymes in Opium Poppy using Virus-Induced Gene Silencing. *Plant J.* 2012, 69: 1052-1063.

## VII. Acknowledgements

This chapter is based on an opinion article published in *Curr. Opin. Biotech.* Dr. Weerawat Rungphan and I collaboratively wrote this article.



## Chapter 2

Discovery of 10-hydroxygeraniol Oxidoreductase Activity in *C. roseus*

## I. Introduction

The iridoids constitute a sizable class of natural products and boast an equally impressive repertoire of biological activities (Figure 2.1).<sup>1,2</sup> Despite their utility as both pharmaceuticals and pest repellents, many steps of iridoid biosynthesis are unknown.<sup>1,2</sup> The enzymatic reactions within this pathway have both captivated and challenged scientists for decades. Ultimately, understanding this pathway could potentially decrypt novel enzymatic function. Moreover, understanding iridoid biosynthesis will aid in the production of these valuable fine chemicals in tractable heterologous hosts.

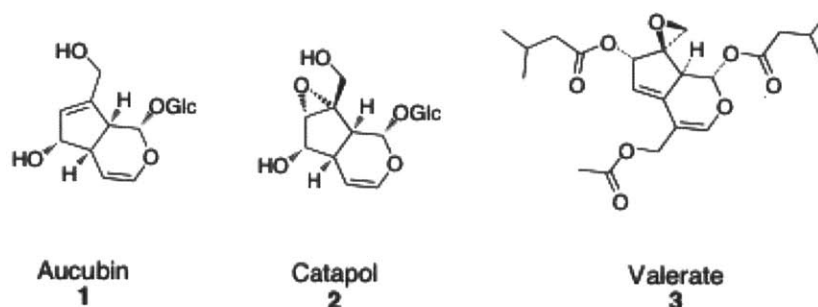


Figure 2.1: Structures of three representative plant iridoids. Glc, glucose.

Iridoids are composed of a bicyclic 10-carbon skeleton that is derived from the condensation of dimethylallyl pyrophosphate (DMAPP) **4** and isopentyl pyrophosphate (IPP) **5**.<sup>2</sup> Subsequently, geraniol synthase dephosphorylates this 10-carbon unit—geranyl pyrophosphate **6**—to form geraniol **7**.<sup>3</sup> In the first committed step of iridoid terpene biosynthesis, geraniol 10-hydroxylase (G10H)—a cytochrome P450—hydroxylates geraniol **7** at position 10, forming 10-hydroxygeraniol **8**.<sup>4</sup>

The pathway from 10-hydroxygeraniol **8** to secologanin **15**, the direct precursor to monoterpene indole alkaloid formation, contains approximately eight steps, at least three of which are unknown (Figure 2.2).<sup>2</sup> Namely, the enzyme(s) that facilitates oxidation of the di-alcohol 10-hydroxygeraniol **8** to the di-aldehyde 10-oxogeraniol **9** is unknown.<sup>2</sup> However, alcohol dehydrogenases are known to catalyze the oxidation of alcohols and are therefore strongly implicated in this transformation. Likewise, the enzyme(s) that facilitates conversion of nepetalactol **10** to 7-deoxyloganic acid aglycone **12** is unknown.<sup>2</sup> Notably, this conversion may require as many as three unique enzymes. The first predicted step in this conversion—a hydroxylation reaction—is likely cytochrome P450-dependent. The oxidation of the resultant alcohol **11** to the carboxylic acid **12** may employ the same cytochrome P450 or require separate dehydrogenases. Namely, a dehydrogenase may convert the alcohol **11** to an aldehyde, then the aldehyde to the carboxylic acid. Alternatively, a dedicated alcohol dehydrogenase may convert the alcohol **11** to an aldehyde, and a disparate aldehyde dehydrogenase may convert the resulting aldehyde to the carboxylic acid **12**. Finally, the enzyme responsible for the glucosylation of the 7-deoxyloganin aglycone **13** is unknown, though a glucosyl transferase is strongly implicated in this transformation.<sup>2</sup> Importantly, it is perfectly feasible to imagine that the glucosyl transferase could also act on nepetalactol **10** or any intermediate prior to acid formation. Intriguingly, the glucosylation of nepetalactol **10** directly—prior to P450 oxidation—would prove strongly redolent of flavonoid biosynthesis.<sup>5</sup> While some

substrate specificity studies have been performed with crude plant lysates, the order of these reactions has not yet been established definitively.<sup>6</sup>

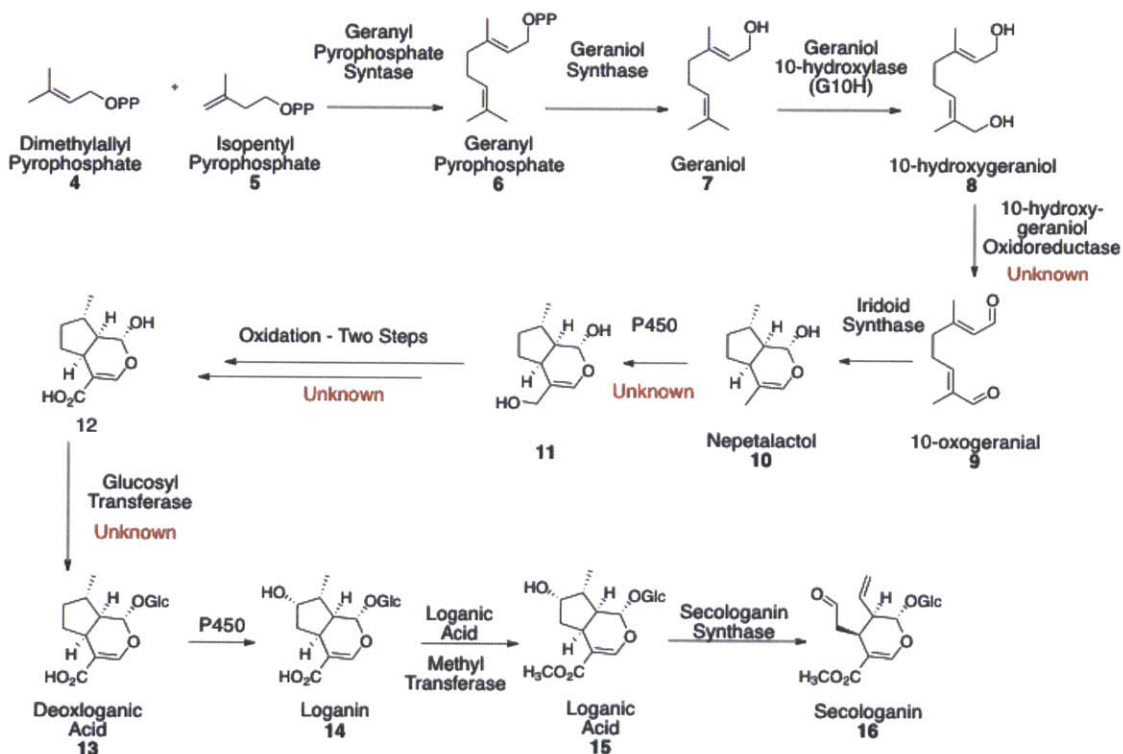


Figure 2.2: The proposed biosynthesis of iridoid secologanin **16**. From 10 hydroxygeraniol **8** to secologanin **16**, a direct monoterpene indole alkaloid precursor, at least three enzymatic steps are unknown. Glc, glucose; OPP, pyrophosphate.

As outlined above, chemical logic allows us to sensibly hypothesize the enzyme class responsible for each unknown pathway step. Further, advances in sequencing technologies and statistical methods equip us to identify and assess the biological function of candidate genes of interest. This chapter focuses on the

discovery and characterization of 10-hydroxygeraniol oxidoreductase, the enzyme that catalyzes the conversion of 10-hydroxygeraniol **8** to 10-oxogeraniol **9**. This enzyme acts on the substrate positioned at the seat of iridoid biosynthesis and produces the substrate for iridoid synthase, a reductive cyclase that assembles the bicyclic iridoid scaffold.

Previous attempts to identify the 10-hydroxygeraniol oxidoreductase were confined to protein isolation from iridoid-containing plants.<sup>7,8</sup> Specifically, Hallahan *et al.* demonstrated that oxidoreductase activity found in *Nepeta racemosa* (catmint) leaf extracts converted 10-hydroxygeraniol **8** to 10-oxogeraniol **9** in an NADP<sup>+</sup>-dependent fashion (NAD<sup>+</sup> is not accepted). The enzyme of interest—a heterodimer comprised of one 40 kDa subunit and one 42 kDa subunit—was purified 1150-fold, though not to apparent homogeneity. Consequently, no protein sequence is reported.<sup>7</sup>

In a separate study, Ikeda *et al.* purified to apparent homogeneity a 44-kDa oxidoreductase from *Rauwolfia serpentina* cell cultures that converted 10-hydroxygeraniol **8** to 10-oxogeraniol **9**.<sup>8</sup> This enzyme was shown to be NADP<sup>+</sup>-dependent (while NAD<sup>+</sup> is not accepted) and, based on atomic absorption spectroscopy, to bind zinc ions. Moreover, in contrast to previously characterized alcohol dehydrogenases from higher plants, this enzyme appears to function as a monomer based on gel filtration chromatography and SDS PAGE analysis. Strikingly, the first 21 residues of alcohol dehydrogenases isolated from maize, pea and *Arabidopsis* exhibit high sequence homology, with 13 of the N-terminal residues conserved.<sup>8</sup> However, the N-terminus of the purified protein from *R.*

*serpentina*—NH<sub>2</sub>-NQQXTKVTKMVYKLVLVNTY—did not display significant homology to previously characterized alcohol dehydrogenases or to proteins registered in the Protein Data Bank. To the best of our knowledge, the 10-hydroxygeraniol oxidoreductase gene from *R. serpentina* has not been cloned, and the full sequence has not been reported.

While previous studies relied upon protein isolation—an arduous, but state-of-the-art technique when sequence data is unavailable—we trawled recently acquired *C. roseus* transcriptomic datasets in this study to ‘fish out’ genes of interest.<sup>9</sup> Specifically, we selected genes of interest by employing the ‘guilt-by-association’ principle, whereby uncharacterized, but functionally annotated genes are baited with genes of known function—in this case, genes known to be involved in *C. roseus* iridoid biosynthesis.<sup>1,10</sup> Genes functionally predicted to catalyze alcohol dehydrogenation that also cluster with known iridoid biosynthetic genes in hierarchical clustering algorithms are assigned as prime 10-hydroxygeraniol oxidoreductase candidates. These candidates are subsequently assessed for length completeness, cloned from the complementary DNA (cDNA) of *C. roseus*, expressed heterologously, then assayed for the desired activity. Importantly, the N-terminus of the oxidoreductase isolated from *R. serpentina* displays no significant sequence homology to any transcripts in the *C. roseus* assemblies or to genes registered within the Plant Genomic Database.

Efficiently mining large datasets for genes of interest is currently a major challenge in the field of plant enzyme discovery.<sup>11</sup> In the most recent *C. roseus* transcriptomic assembly, approximately 1.2% (about 400) of the gene transcripts

(0.6% heme-dependent and 0.6% non-heme-dependent) are functionally predicted to facilitate redox reactions!—a figure that undergirds terrestrial plants' expansive oxidoreductive landscape. Notably, however, the 'guilt-by-association' principle has been successfully employed previously with *C. roseus* transcriptomic data sets to bait the genes of missing pathway steps, including the reductive cyclase responsible for iridoid scaffold assembly.<sup>1,10</sup> This chapter describes our efforts to excavate *C. roseus* transcriptomic datasets to unearth the alcohol dehydrogenase(s) responsible for the oxidation of 10-hydroxygeraniol **8** to 10-oxogeraniol **9**, a missing step in iridoid biosynthesis.

## II. Results and Discussion

The oxidation of the di-alcohol 10-hydroxygeraniol **8** to the di-aldehyde 10-oxogeraniol **9** strongly invokes catalysis via an alcohol dehydrogenase. Therefore, in a principle known as 'guilt-by-association,' we mined *C. roseus* transcriptome assemblies for alcohol dehydrogenases with similar expression patterns to known genes in iridoid biosynthesis.

The most recent *C. roseus* transcriptome assembly contains approximately 33,000 transcripts.<sup>9</sup> To facilitate the mining process, we employed various transcript filtering conditions. For example, Dr. Fernando Geu-Flores filtered the data set based on gene expression in leaves and methyl jasmonate elicitation. Genes either not expressed or only poorly expressed [fragments per kilobase of exon per million (FPKM) values < 2] in leaves can be discarded, as iridoid biosynthesis is known to occur in leaves. We postulate that 10-

hydroxgeraniol oxidoreductase expression is high in leaves. Similarly, methyl jasmonate is known to upregulate iridoid biosynthesis.<sup>12</sup> Therefore, genes non-elicited by methyl jasmonate may also be discarded. Approximately 4,500 transcripts remain after these constraints are enforced.

Additionally, in three separate filters, I retained only the 5000 most highly expressed genes in *C. roseus* immature leaves, mature leaves and hairy roots, respectively. Because iridoid-containing molecules are isolated from each of these three tissues, we conjecture that iridoid biosynthetic genes should be highly expressed in these tissues as well. Transcripts with FPKM values less than zero in each of these tissue samples were discarded.

We then employed hierarchical clustering analyses on each of the four filtered assemblies to correlate gene expression levels across the 17 different tissue samples. Transcripts were imported into Multiple Experiment Viewer 4\_7, then clustered using the hierarchical clustering algorithm based on Pearson Correlation as the distance metric and average linkage clustering as the linkage selection method. A representative cluster (cluster based on both high leaf expression and methyl jasmonate elicitation) is shown below (Figure 2.3).

In addition to hierarchical clustering, we also treated the data with another type of analysis called mutual ranking analysis. Using a preliminary transcriptome dataset (*C. roseus* transcriptome assembly 1), Geu-Flores *et al.* generated a mutual ranking list based on gene expression likeness to geraniol 10-hydroxylase, the enzyme directly upstream of 10-hydroxygeraniol oxidoreductase.<sup>1</sup> Specifically, Geu-Flores *et al.* calculated the Pearson



Correlation Coefficients (PCCs) for each contig that passed the methyl jasmonate elicitation and high leaf expression filters. Genes were ranked in descending order according to their PCCs for geraniol 10-hydroxylase. This list is the forward ranking list. To obtain the reverse ranking list, Geu-Flores *et al.* computed the PCCs for the 200 best-correlated contigs in the forward list against each other. The mutual rank (Table 2.1) is the square root of the forward and reverse product:

$$\text{Equation 2.1: Mutual Rank} = [(\text{forward rank})(\text{reverse rank})]^{1/2}$$

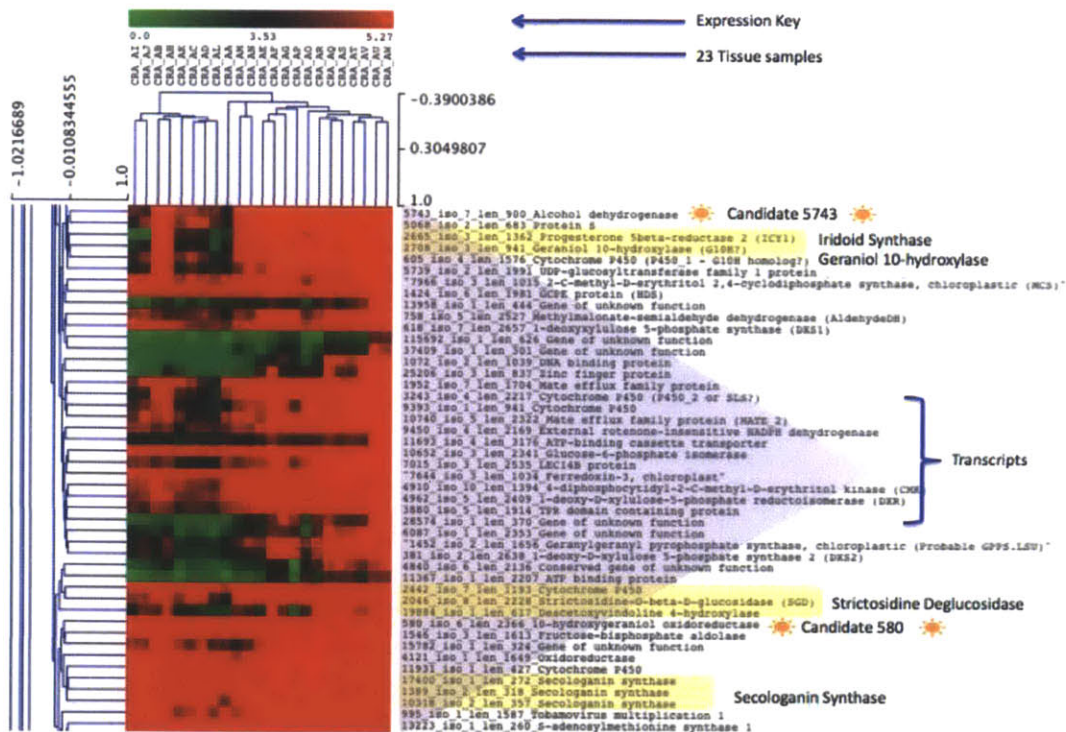


Figure 2.3: Representative cluster from hierarchical clustering analysis (high leaf expression and methyl jasmonate elicitation filter). Red represents high expression; green signifies low expression. The cluster is highlighted in purple. Yellow highlighted transcripts are known iridoid biosynthetic genes. Candidates 5743 and 580 are found in this cluster.

| ID          | Mut Rank | PCC         | Rev Rank | Fwd Rank | MRV   | Annotation    | Comments   |
|-------------|----------|-------------|----------|----------|-------|---------------|--|
| 2665_iso=1  | 1        | 0.989099123 | 2        | 1        | 1.414 | p450          | Cytochrome P450  |
| 9933_iso=1  | 2        | 0.974834433 | 3        | 2        | 2.449 | ---           | ---  |
| 34690_iso=1 | 3        | 0.972471793 | 3        | 3        | 3     | ---           | ---  |
| 9710_iso=1  | 4        | 0.965799646 | 4        | 4        | 4     | p450          | Cytochrome P450  |
| 13334_iso=1 | 5        | 0.964354354 | 5        | 5        | 5     | GcpE          | 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, bacterial-type |
| 729_iso=1   | 6        | 0.957657585 | 4        | 10       | 6.325 | ---           | Iridoid Synthase   |
| 42065_iso=1 | 7        | 0.962529328 | 7        | 6        | 6.481 | ---           | ---  |
| 17224_iso=1 | 8        | 0.9620476   | 6        | 7        | 6.481 | DXP_redisom   | 1-deoxy-D-xylulose 5-phosphate reductoisomerase, C-terminal          |
| 5089_iso=1  | 9        | 0.955826582 | 5        | 11       | 7.416 | ADH_zinc_N    | Candidate 5743   |
| 15157_iso=2 | 10       | 0.959631421 | 7        | 8        | 7.483 | MTHFR         | Methylenetetrahydrofolate reductase                                  |
| 37738_iso=3 | 11       | 0.953962234 | 5        | 12       | 7.746 | ---           | ---  |
| 12882_iso=1 | 12       | 0.9340208   | 3        | 20       | 7.746 | UDPGT         | UDP-glucuronosyl/UDP-glucosyltransferase                             |
| 17468_iso=1 | 13       | 0.958511244 | 8        | 9        | 8.485 | zf-Dof        | Zinc finger, Dof-type  |
| 18942_iso=1 | 14       | 0.947575109 | 8        | 13       | 10.2  | polyprenyl_sy | Polyprenyl synthetase  |
| 36044_iso=1 | 15       | 0.942874239 | 9        | 15       | 11.62 | DUF581        | Protein of unknown function DUF581                                   |
| 39905_iso=1 | 16       | 0.945957994 | 13       | 14       | 13.49 | ---           | ---  |
| 2437_iso=1  | 17       | 0.933503619 | 9        | 21       | 13.75 | YgbB          | 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, core          |
| 3272_iso=1  | 18       | 0.929045736 | 8        | 26       | 14.42 | GcpE          | 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, bacterial-type |
| 7526_iso=1  | 19       | 0.916690108 | 6        | 37       | 14.9  | ---           | ---  |
| 3938_iso=1  | 20       | 0.94046914  | 15       | 17       | 15.97 | GHMP_kinase   | GHMP kinase, C-terminal  |
| 1136_iso=1  | 21       | 0.921921584 | 9        | 33       | 17.23 | Aldedh        | Aldehyde dehydrogenase   |
| 43316_iso=1 | 22       | 0.936586676 | 19       | 18       | 18.49 | ---           | ---  |
| 9680_iso=2  | 23       | 0.887353468 | 4        | 98       | 19.8  | HLH           | Helix-loop-helix DNA-binding domain                                  |
| 8952_iso=1  | 24       | 0.907762934 | 7        | 57       | 19.97 | ---           | ---  |
| 31515_iso=1 | 25       | 0.941437245 | 32       | 16       | 22.63 | LRRNT_2       | Leucine-rich repeat-containing N-terminal domain, type 2             |

Table 2.1 The top 25 contigs from a mutual rank with Geraniol 10-hydroxylase. Candidate 5743 ranks 9<sup>th</sup> and is shown in red. ID, transcript number in *C. roseus* transcriptome dataset 1; Mut Rank, Mutual Rank; PCC, Pearson Correlation Coefficient; Rev, Reverse; Fwd, Forward; MRV, Mutual Ranking Value.

After selecting alcohol dehydrogenases from each of these analyses, we then assessed each candidate for length by comparing the longest open reading frame across all *C. roseus* assembly data with the top-ranking BLAST (Basic Local Alignment Search Tool) hit. Partial length genes were discarded. Next, we assessed each gene for its predicted sub-cellular localization by submitting sequences to the TargetP 1.1 server.<sup>13</sup> Genes strongly (i.e. confidence level 1) predicted to localize to the chloroplast, mitochondria or endoplasmic reticulum (ER) were discarded, as all early iridoid biosynthetic steps are cytosolic, including the characterized steps directly before (i.e. G10H) and after (i.e. iridoid synthase) 10-hydroxygeraniol oxidoreductase. Notably, the P450-dependent enzyme G10H is anchored to the ER membrane, but catalysis occurs within the cytoplasm.

Eight candidates were retained after all constraints were enforced; the filter(s) and hierarchical cluster(s) in which the candidates appear are shown in Table 2.2. Table 2.2 also displays the mutual rank order, if the candidate appeared on the mutual ranking list. Candidate 5743 had a mutual rank of 9, making it the highest-ranking candidate on the mutual ranking list.

| Candidate | MeJa+Leaf | Immat. Leaf | Hairy Root | Root | Mutual Rank | Cytosol |
|-----------|-----------|-------------|------------|------|-------------|---------|
| 1786      | X         |             | X          | X    | Not Found   | X       |
| 26        | X         |             | X          | X    | 103         | X       |
| 4319      | X         | X           |            |      | Not Found   |         |
| 5743      | X         | X           | X          | X    | 9           |         |
| 7220      |           |             | X          |      | Not Found   | X       |
| 8694      |           |             |            |      | 165         |         |
| 2041      | X         |             |            |      | 91          |         |
| 580       | X         | X           | X          | X    | Not Found   | X       |

Table 2.2: Candidate filter summary. 'X' signifies that the candidate passed the filter restraints and was identified in the corresponding cluster. Geraniol 10-hydroxylase mutual rank is provided if the candidate is found on the list. MeJa, methyl jasmonate; Immat., immature.



Notably, four of the final eight candidates are predicted with varying confidence levels to localize to the ER, according to TargetP 1.1 (Table 2.3). The candidates were retained, however, because the signal peptide predictor SignalP 4.1 server failed to identify any regions likely to function as signal peptides for any of these candidates.<sup>14</sup> Moreover, none of the candidates are strongly predicted to localize to the ER. Significantly, Candidate 5743 is one of the candidates predicted, albeit with low confidence (i.e. confidence level 5), to enter the secretory pathway and localize to the ER. We postulate that the high cysteine content (15 cysteine residues in the 378-residue primary structure) likely relegates this candidate to the ER in the prediction algorithm. Notably, however, Ikeda *et al.* indirectly implicated sulfhydryl groups in 10-hydroxygeraniol oxidoreductase catalysis by incubating the oxidase from *R. serpentina* with a thiol-reactive Michael acceptor (i.e. N-ethyl maleimide) and observing significant activity attenuation.<sup>8</sup> Therefore, Candidate 5743—which co-expresses well with known iridoid biosynthetic genes, according to both hierarchical clustering analyses and mutual ranking—was retained in the screen even though the prediction algorithms suggested it was unlikely to localize to the cytosol.

| Candidate | Chloroplast | Mitochondria | ER    | Other | Confidence |
|-----------|-------------|--------------|-------|-------|------------|
| 1786      | 0.193       | 0.130        | 0.122 | 0.519 | 4          |
| 26        | 0.180       | 0.165        | 0.035 | 0.406 | 4          |
| 4319      | 0.102       | 0.023        | 0.850 | 0.042 | 2          |
| 5743      | 0.589       | 0.028        | 0.607 | 0.063 | 5          |
| 7220      | 0.197       | 0.261        | 0.024 | 0.557 | 4          |
| 8694      | 0.140       | 0.038        | 0.321 | 0.645 | 4          |
| 2041      | 0.268       | 0.078        | 0.278 | 0.092 | 5          |
| 580       | 0.009       | 0.326        | 0.159 | 0.606 | 4          |

Table 2.3: TargetP1.1 localization prediction. Strongest predictions are highlighted in yellow. High confidence = 1; low confidence = 5. Confidence values reflect the difference between the two highest predictions.

Faced with cloning multiple genes, we employed USER cloning, a ligation-free, cassette-based expression method.<sup>15</sup> Each gene, except for Candidate 580, which could not be cloned from *C. roseus* cDNA, was successfully cloned into the USER cassette (gene and protein sequences are provided in Appendix A). All candidate expression trials were run in Rosetta™ 2 *E. coli*, a host strain optimized for rapid and robust eukaryotic protein expression. Proteins were expressed and purified with an N-terminal histidine tag (His<sub>6</sub>). Five of the seven cloned candidates—candidates 1786, 26, 4319, 5743 and 7220—expressed well

| Candidate | Theoretical Mass (kDa) | Predicted pI | Expression? |
|-----------|------------------------|--------------|-------------|
| 1786      | 40.7                   | 5.27         | yes         |
| 26        | 32.5                   | 5.22         | yes         |
| 4319      | 40.0                   | 7.63         | yes         |
| 5743      | 40.4                   | 6.27         | yes         |
| 7220      | 33.4                   | 5.71         | yes         |
| 8694      | 42.9                   | 5.76         | no/poor     |
| 2041      | 42.9                   | 6.57         | no/poor     |

Table 2.4: Summary of data from candidate expression in Rosetta 2 cells. Five of the 7 cloned candidates expressed robustly. Theoretical masses listed are the predicted average masses using the ExPASy server.

(Table 2.4; Figure 2.4) and were subsequently assayed for 10-oxogeraniol **9** formation from 10-hydroxygeraniol **8**.

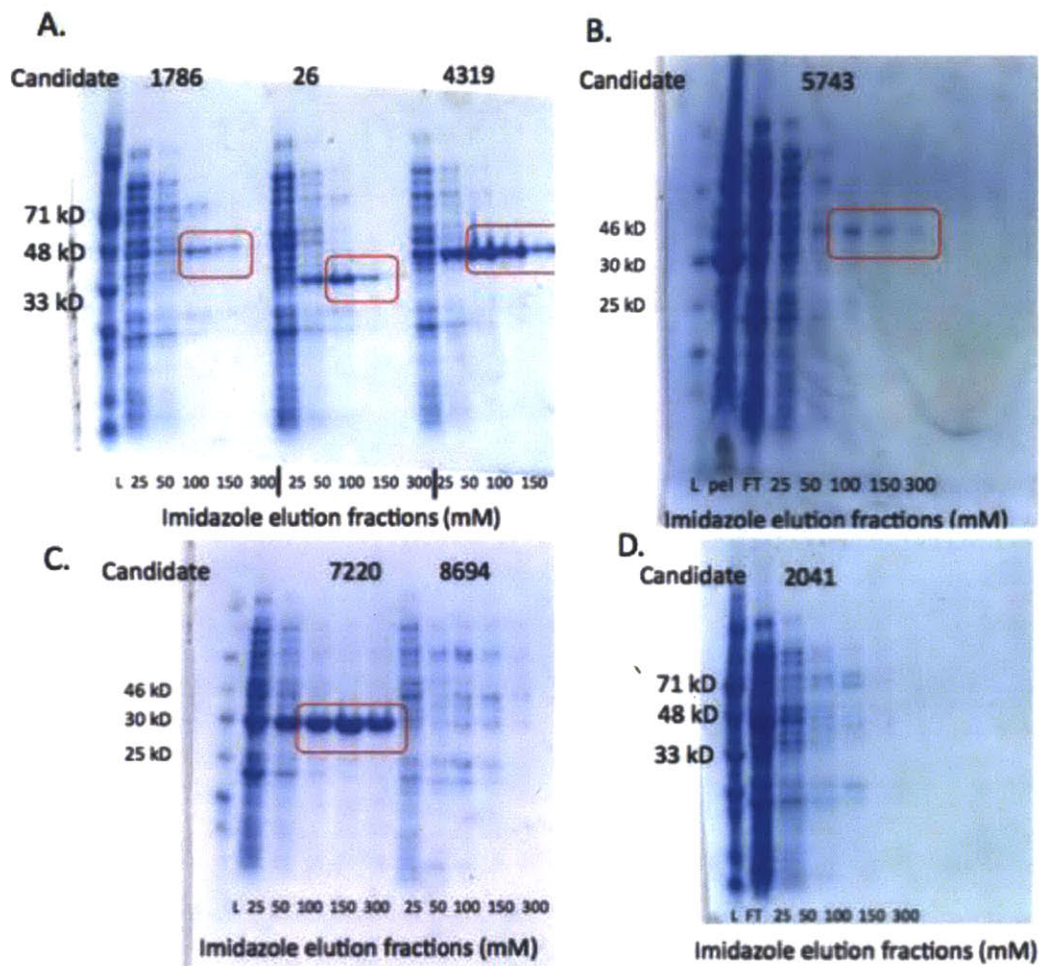


Figure 2.4: SDS PAGE gels of candidate expression and purification. A. Expression of Candidates 1786, 26 and 4319. Each of these candidates expressed well. B. Expression of Candidate 5743. C. The expression trials of Candidates 7220 and 8694. Candidate 7220 expressed well under the conditions highlighted in the method section, whereas Candidate 8604 did not. D. Expression trial of Candidate 2041. Candidate 2041 did not express well under the conditions of the screen (methods section). Red boxes highlight the fractions that were pooled, concentrated and buffer exchanged into 20 mM MOPS (pH 7.0) for subsequent assay.

Each of the five well-expressed candidates was screened for 10-oxogeraniol **9** formation by both thin layer chromatography (TLC) and gas

chromatography – mass spectrometry (GC-MS). Positive hits—candidates that produced 10-oxogeranial **9**—were further assessed with a battery of controls to test for zinc and cofactor (NAD<sup>+</sup> or NADP<sup>+</sup>) dependence (Appendix A). Namely, we conducted the following control experiments on positive hits: (1) no enzyme controls, where enzyme was omitted from the assay; (2) no NAD<sup>+</sup> or NADP<sup>+</sup> controls, where these cofactors were omitted from the assay; (3) no zinc controls, where no ZnCl<sub>2</sub> was added to the assay mixture; and (4) chelator controls, where 1 mM EDTA (ethylene diamine tetraacetic acid) was added to the assay and ZnCl<sub>2</sub> was omitted. The summary of those experiments is shown below in Table 2.5; the TLC controls and all GC-MS data are supplied in Appendix A.

| Candidate Screen           | TLC  |       | GC-MS |       |
|----------------------------|------|-------|-------|-------|
|                            | NAD+ | NADP+ | NAD+  | NADP+ |
| <b>1786</b>                | Yes  | Yes   | Yes   | Yes   |
| <b>1786 No Zinc</b>        | Yes  | Yes   | Yes   | Yes   |
| <b>1786 EDTA + No Zinc</b> | Yes  | No    | Yes   | No    |
| <b>26</b>                  | No   | No    | No    | Yes   |
| <b>26 No Zinc</b>          | N/A  | N/A   | N/A   | Yes   |
| <b>26 EDTA + No Zinc</b>   | N/A  | N/A   | N/A   | Yes   |
| <b>4319</b>                | No   | No    | No    | No    |
| <b>5743</b>                | Yes  | Yes   | Yes   | Yes   |
| <b>5743 No Zinc</b>        | Yes  | Yes   | Yes   | Yes   |
| <b>5743 EDTA + No Zinc</b> | Yes  | Yes   | Yes   | Yes   |
| <b>7220</b>                | No   | No    | No    | No    |
| <b>No NAD(P)+ Controls</b> | No   | No    | No    | No    |
| <b>No Enzyme Controls</b>  | No   | No    | No    | No    |

Table 2.5: Summary of the candidate screen with TLC and GC-MS. ‘Yes’ indicates that 10-oxogeranial **9** formation was observed. ‘No’ indicates that 10-oxogeranial **9** formation was not observed. ‘N/A’ indicates that the candidate was not screened under said conditions (No controls with Candidate 26 were conducted on TLC because product formation was not observed via TLC.) Experiments for each candidate are grouped.

The TLC screen identified two candidates—Candidates 1786 and 5743—that can utilize either NAD<sup>+</sup> or NADP<sup>+</sup> to form 10-oxogeranial **9**, based on co-migration with an authentic standard (Figure 2.5). To confirm 10-oxogeranial **9**



formation via TLC, we also performed a coupled assay with a representative positive hit (Candidate 5743 with  $\text{NAD}^+$ ) and iridoid synthase, the enzyme directly downstream of 10-hydroxygeranial oxidoreductase. In aqueous media, nepetalactol **10**—the iridoid synthase product—forms a range of hydrates, which most standard TLC stains fail to detect readily. Therefore, we monitored 10-oxogeranial disappearance rather than nepetalactol **10** formation. The spot that co-elutes with 10-oxogeranial **9** was metabolized in the coupled assay, whereas the same spot was retained in the control experiment with only Candidate 5743 and  $\text{NAD}^+$ , confirming 10-oxogeranial **9** formation (Figure 2.6).

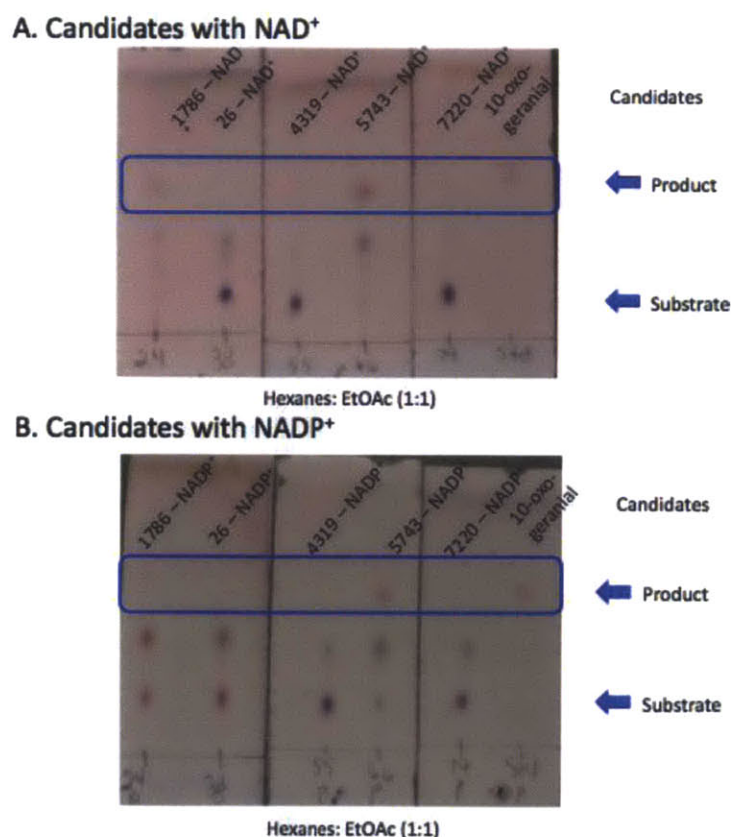


Figure 2.5: TLC screens of candidates with (A)  $\text{NAD}^+$  and (B)  $\text{NADP}^+$ . Candidates 1786 and 5743 accept both  $\text{NAD}^+$  and  $\text{NADP}^+$  to form 10-oxogeranial, the authentic standard of which is on the right of both sets of TLCs. TLC plates are stained with anisaldehyde.



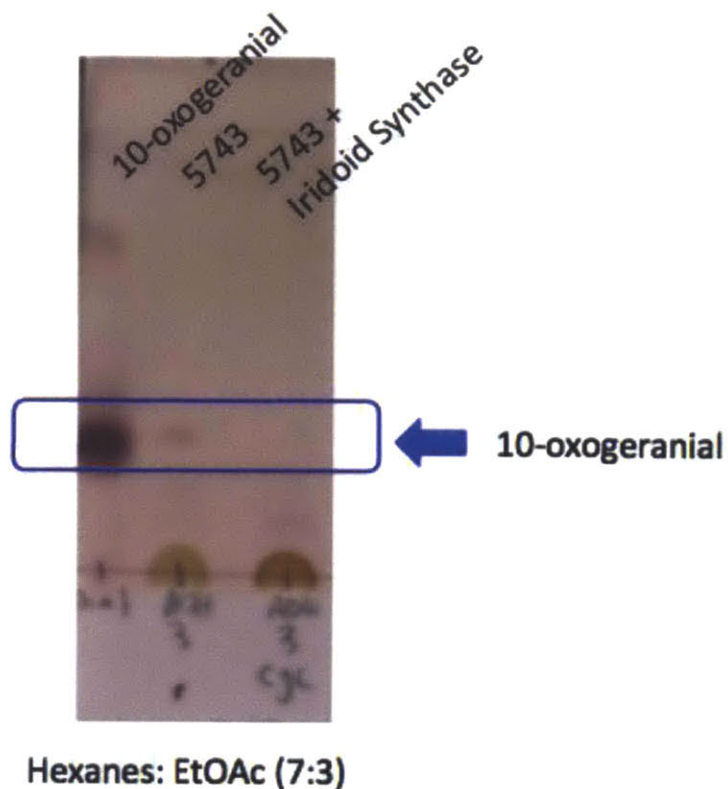


Figure 2.6: Coupled assay with Candidate 5743 (NAD<sup>+</sup>) and iridoid synthase. In the presence of iridoid synthase and NADPH (right), the spot that co-migrates with 10-oxogeranial **9** (middle) disappears. The 10-oxogeranial standard is shown on the left. The iridoid synthase product, nepetalactol **10**, forms a range of hydrates in aqueous media, which are not readily detected by anisaldehyde.

For both Candidate 1786 and Candidate 5743, product formation occurs only in the presence of the enzyme and NAD<sup>+</sup> or NADP<sup>+</sup>. Product formation was still observed when zinc was omitted from the assay, suggesting that either Zn<sup>2+</sup> is unnecessary for catalysis, or more likely that the enzyme co-purifies with metal bound during expression in LB media supplemented with 100  $\mu$ M ZnCl<sub>2</sub>. PredZinc, a zinc binding prediction server with 75% accuracy for known zinc-binding proteins, predicts zinc-binding sites at positions C43, H64, C94, C97 and C100 for Candidate 1786 and positions C51, H72, C102, C105 and C108 for Candidate 5743.<sup>17</sup>

Intriguingly, the addition of EDTA to the assay mixture in conjunction with the omission of ZnCl<sub>2</sub> only quelled product formation with Candidate 1786 when NADP<sup>+</sup> was employed, suggesting Zn<sup>2+</sup> binds more weakly with this co-factor. The addition of chelator EDTA did not disrupt product formation for Candidate 1786 with NAD<sup>+</sup> or for Candidate 5743 with either NAD<sup>+</sup> or NADP<sup>+</sup>.

We also confirmed 10-oxogeraniol **9** formation with GC-MS, specifically by co-elution with an authentic standard and spectral similarity to the standard's fragmentation pattern. Representative traces of positive hits are shown in Figure 2.7; all other traces are provided in Appendix A.

The GC-MS results were in good agreement with the TLC results, except GC-MS—a more sensitive technique than TLC—identified an additional positive hit. Specifically, GC-MS identified that Candidate 26 exclusively employs NADP<sup>+</sup> to form 10-oxogeraniol (NAD<sup>+</sup> is not accepted). Candidate 26 also forms product when zinc is omitted from the assay and when the EDTA is added to the assay and zinc is omitted. Again, these results suggest that either the candidate does not require Zn<sup>2+</sup> for catalysis or more likely that the enzyme co-purifies with metal bound during protein expression LB media supplemented with 100 μM ZnCl<sub>2</sub>. PredZinc predicted no zinc-binding residues for Candidate 26.<sup>16</sup> Intriguingly, however, Candidate 26 is the only candidate that accepts NADP<sup>+</sup> exclusively, which is in agreement with previously reported 10-hydroxygeraniol oxidoreductase activities purified from other plants.<sup>7,8</sup>

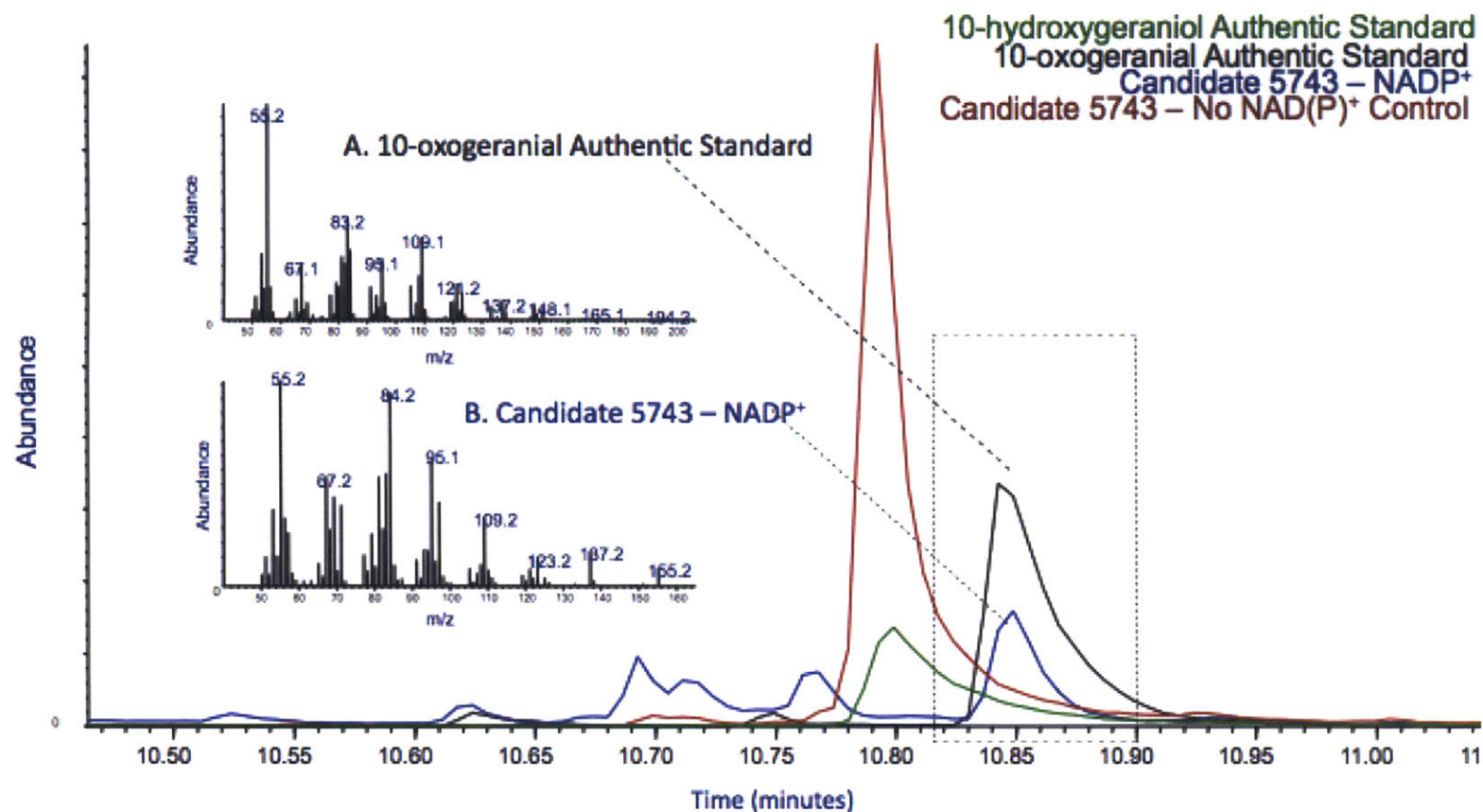


Figure 2.7: Gas Chromatography – Mass Spectroscopy (GC-MS) chromatograms and spectra of a representative positive hit (Candidate 5743 with NADP<sup>+</sup> - Blue). Candidate 5743 with NADP<sup>+</sup>, shown in blue, co-elutes with an authentic standard of 10-oxogeraniol (Black). The GC spectra of both the authentic standard of 10-oxogeraniol (black) and Candidate 5743 with NADP<sup>+</sup> (blue) are shown in insets A and B. Magnified spectra are provided in Appendix A. The fragmented ions are identical in the assay and the authentic standard, which confirms 10-oxogeraniol **9** formation in the assay. Notably, 10-oxogeraniol **9** is not formed in the control lacking NAD(P)<sup>+</sup> (shown in red). The enzymatic substrate, 10-hydroxygeraniol **8** is shown in green.

Having confirmed product formation via GC-MS, we have successfully identified three candidates (Candidates 1786, 26 and 5743) that may prove useful in reconstituting iridoid biosynthesis, most notably secologanin biosynthesis, for biotechnological purposes. Importantly, preliminary experiments suggest that candidate 5743 functions well in efforts to reconstitute nepetalactol (Sherden and O'Connor, unpublished). However, we also aim to understand the physiological relevance of these candidates. A silencing effect—whereby the candidate gene is knocked down and a distinct metabolic phenotype emerges—would physiologically validate or revoke candidates in this screen, proving or disproving their relevancy to iridoid biosynthesis in *C roseus*.

Assuming no functional redundancy, silencing the 10-hydroxygeraniol oxidoreductase could result in the accumulation of substrate, 10-hydroxygeraniol **8**, and a decrease in downstream iridoids and iridoid-derived alkaloids. It is important to note, however, that the plant could potentially derivatize 10-hydroxygeraniol **8** to prevent its accumulation. In this case, we would not observe an accumulation of 10-hydroxygeraniol **8**, but would still expect a decrease in downstream iridoid or monoterpene indole alkaloid production. We can use Virus-Induced Gene Silencing (VIGS) to test this hypothesis. VIGS utilizes the tobacco rattle virus platform, a bipartite vector system that hijacks the plant's own defense system to degrade the cognate mRNA of interest and thereby potentially induce a transient gene silencing effect.<sup>17</sup> Importantly, VIGS has been shown to be effective in *C. roseus* for a number of genes involved in monoterpene indole alkaloid biosynthesis.<sup>17</sup>

Here, 478-base pair regions in both candidates 1786 and 5743 were targeted for knockdown (candidate 26 was not prepared in time for the first round of VIGS experiments and will be assessed at a later date). The region to be silenced is cloned and inserted into vector pTRV2 (tobacco rattle virus). To facilitate the cloning, inserts were flanked with adapters compatible with the pTRV2 USER cassette described in Geu-Flores *et al.*<sup>1</sup> The metabolic profiles of these silenced lines remain to be measured. Additionally, quantitative PCR must also confirm that the gene candidates are in fact silenced. These experiments are ongoing.

In addition to understanding the physiological relevance of these candidates, we also wish to understand why the enzyme—which produces a presumably toxic dialdehyde intermediate capable of cross-linking proteins—evolved in the first place. The answer is as yet unknown. Speculation leads us to ask if the 10-hydroxygeraniol oxidoreductase and iridoid synthase rapidly co-evolved and if a protein-protein interaction is required to shuttle the dialdehyde product into nepetalactol **10** formation in order to prevent cell damage by the highly reactive aldehyde.

### **III. Future Work**

We have successfully mined *C. roseus* transcriptomic datasets for an enzyme that catalyzes a missing step in iridoid biosynthesis. Specifically, using hierarchical clustering algorithms and the mutual ranking list of G10H—the enzyme directly upstream of 10-hydroxygeraniol oxidoreductase—we identified

three enzymes with 10-hydroxygeraniol oxidoreductase activity. While we believe these candidates could prove to be promising for a number of biotechnology applications, most notably the reconstitution of secologanin biosynthesis, a number of experiments remain to be completed in the full characterization of these candidates. First, a number of peaks in addition to the substrate and

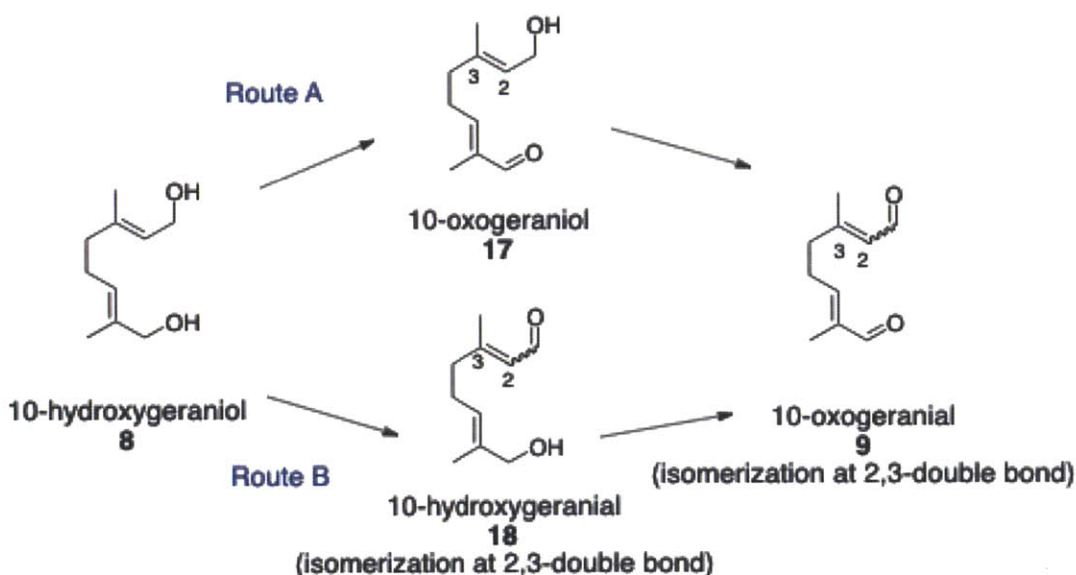


Figure 2.8: The two possible routes from 10-hydroxygeraniol **8** to 10-oxogeraniol **9**. The possible intermediate 10-hydroxygeraniol **18** is known to isomerize at the 2,3-double in aqueous solution. Similarly, the enzymatic product 10-oxogeraniol **9** is known to isomerize at the 2,3-double bond in aqueous solution. The isomers of the potential intermediate at the product could account for four peaks on a representative GC-MS chromatogram (Candidate 1786 with NAD<sup>+</sup>).

product are found on GC-MS chromatograms; these must be properly characterized. Co-injecting authentic standards of the intermediates will aid in peak identification. Dr. Nathaniel Sherden is synthesizing and characterizing both possible enzymatic intermediates, 10-oxogeraniol **17** and 10-hydroxygeraniol **18** (Figure 2.8). The intermediate 10-hydroxygeraniol **18** is



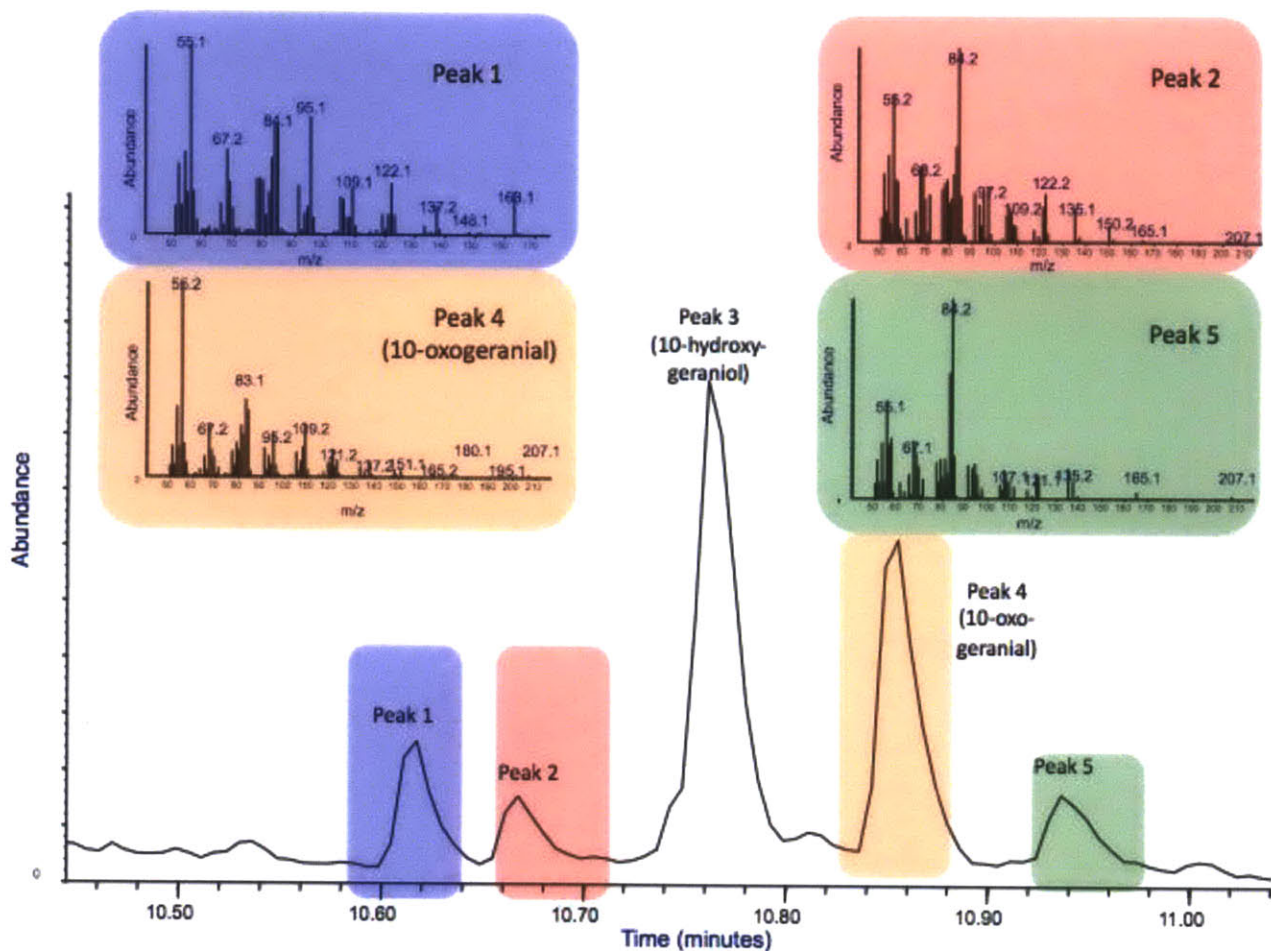


Figure 2.9: Representative trace (Candidate 1786 with NAD<sup>+</sup>) illustrating multiple peaks in addition to the substrate not highlighted, 10-hydroxygeraniol **8**) and product (orange, 10-oxogeraniol **9**) peak. Peak 1 (blue) and Peak 4 (orange, 10-oxogeraniol) have similar GC spectra, suggesting they could be isomers of each other. Likewise, Peak 2 (red) and Peak 5 (green) have similar GC spectra, which suggests they could be isomers of each other. Both 10-oxogeraniol **9** and potential intermediate 10-hydroxygeraniol **17** are known to isomerize in aqueous solution.

known to isomerize at the 2,3-double bond in aqueous solution and could account for two peaks observed in the Candidate 1786 with NAD chromatogram, for example (Figure 2.9).<sup>18</sup> Peaks 2 and 5 have similar fragmentation patterns, suggesting that they are related (red and green insets, respectively) and could be isomers of 10-hydroxygeraniol **18**. An authentic standard of 10-hydroxygeraniol **18** will confirm this assignment. Similarly, the enzymatic product 10-oxogeraniol **9** is also isomerically unstable at the 2,3-double bond, which could potentially explain the presence of Peak 1.<sup>19</sup> Peak 1 (blue inset) has a similar fragmentation pattern to 10-oxogeraniol **9** (orange inset, Peak 4, E isomer at the 2,3-double bond), suggesting the peaks are related. These assignments would account for all 5 major peaks in the assay with Candidate 1786 and NAD<sup>+</sup> (Figure 2.9). Similar assignments must be made for all other positive hits in the screen as well.

After identifying the enzymatic intermediate(s), steady state kinetic parameters will be measured via spectrophotometry of the NADP<sup>+</sup> absorbance (increase in absorbance at 340 nm) for 10-hydroxygeraniol **8** and a number of commercially available substrate analogs, including geraniol **7**, nerol **19**, citronellol **20** and linalool **21**. These compounds were selected for their structural similarity to the native substrate, 10-hydroxygeraniol **8**. Notably, linalool **21** cannot undergo dehydrogenation because the alcohol is tertiary. However, it is possible that linalool **21** may function as a competitive alcohol dehydrogenase inhibitor by binding, but not being converted to product.

Understanding the substrate scope of these enzymes will inform future biotechnological applications. To understand the physiological relevance of these



candidates, however, the metabolic and gene expression profiles of the VIGS experiments must be thoroughly analyzed. These experiments are underway.

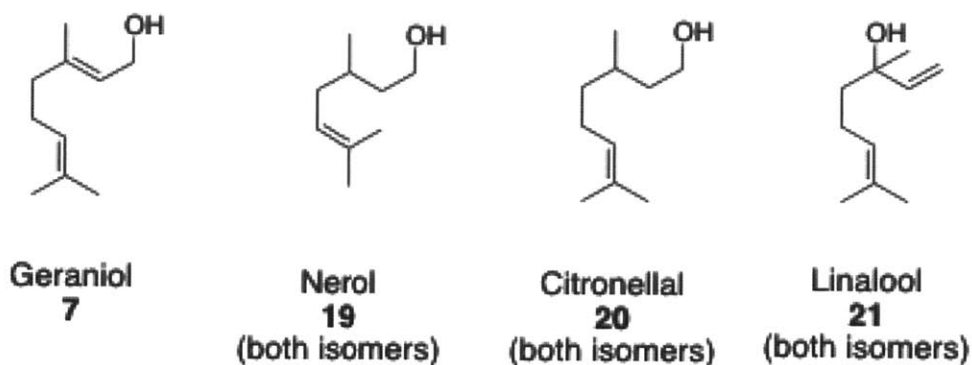


Figure 2.10: Substrate analogs to test with 10-hydroxygeraniol oxidoreductase. Analogs were chosen based on structural similarity to 10-hydroxygeraniol **8**. Linalool **21** possesses only a tertiary alcohol that cannot be oxidized to the corresponding aldehyde. Importantly, **21** may function as an inhibitor.

#### IV. Conclusions

In this chapter, I have described how we employed ‘guilt-by-association’ through the use of hierarchical clustering to discover enzymes with 10-hydroxygeraniol oxidoreductase activity. 10-hydroxygeraniol reductase is a missing enzyme in *C. roseus* iridoid biosynthesis and as such has never been cloned and characterized. This discovery will aid in biotechnological efforts to reconstitute iridoid biosynthesis in tractable heterologous hosts. Three candidates from our screen—Candidates 1786, 5743 and 26—were able to oxidize both hydroxyl moieties of 10-hydroxygeraniol. Interestingly, Candidates 1786 and 5743 can utilize both  $\text{NADP}^+$  and  $\text{NAD}^+$ , whereas candidate 26 uses  $\text{NADP}^+$  exclusively. Interestingly, in both *N. racemosa* and *R. serpentina*—the two plants in the literature with partially characterized 10-hydroxygeraniol oxidoreductases—the enzyme exclusively accepts  $\text{NADP}^+$ .<sup>7,8</sup>

For all three candidates, in assays to test  $Zn^{2+}$  dependency,  $ZnCl_2$  was omitted from the reaction. However, product formation was still observed, suggesting that either  $Zn^{2+}$  is unnecessary for catalysis or more likely that the enzyme co-purifies with metal bound during expression in LB media supplemented with 100  $\mu M$   $ZnCl_2$ . Intriguingly, the addition of EDTA to the assay mixture in conjunction with the omission of  $ZnCl_2$  only quelled product formation in the assay with Candidate 1786 and  $NADP^+$ , suggesting  $Zn^{2+}$  binds more weakly in these conditions. The zinc-binding prediction server ZincPred predicted zinc-binding sites for both Candidate 1786 and Candidate 5743 (bolded, underlined and in red font in Appendix A; listed in Results and Discussion), but did not identify any zinc-binding residues for Candidate 26.<sup>16</sup> More complete characterizations of these enzymes are underway.

Though the discovery of these enzymes is poised to aid in various biotechnological endeavors, most importantly, the heterologous reconstitution of secologanin, the physiological relevance of the three active enzymes is yet to be determined. We have elected to study the candidates' physiological relevance using VIGS; these experiments are currently underway. Theoretically, three outcomes are plausible: (1) These candidates could be physiologically irrelevant, meaning no silencing effect is observed at the metabolite or transcript levels; (2) one (or more) candidate could be physiologically relevant, but functional redundancy mutes any silencing effect at the metabolite level (candidates could potentially be knocked down combinatorially to address this issue); or (3) one (or

more) could be physiologically relevant, and that relevancy is displayed in both the gene expression profile and metabolic profile.

## V. Methods

### A. Gene Cloning

The cassette—consisting of GFP flanked on either side by *PacI* sites, variable nucleotide regions (which enable directional cloning) and *Nt.BbvCI* sites—was generated according to Nour-Eldin *et al.* then inserted into the *BamHI* and *HindIII* sites of pET28a (Figure 2.11; Page 29).<sup>15</sup> Candidates were cloned from *C. roseus* cDNA with X7 phusion polymerase ® and primer annealing temperatures of 55 °C and flanked with the corresponding USER adapters. Primers to amplify each candidate are listed in Table 6.2. The linearized vector (digested with *PacI* and *Nt.BbvCI*) was incubated with gel purified and USER-digested 10-hydroxygeraniol oxidoreductase candidate genes prior to

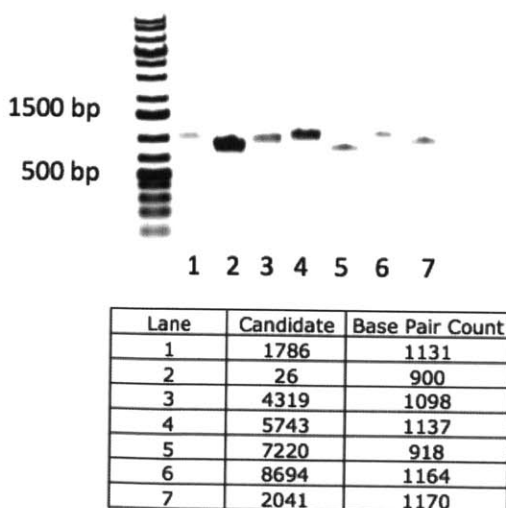


Figure 2.12: USER cloning PCR-purified inserts and expected base pair counts.

transformation into chemically competent Top 10 ® cells (Figure 2.12). Plasmids were then isolated and sequenced to check for the correct insert sequence prior to transformation into chemically competent Rosetta™ 2 cells for expression. All candidates were cloned successfully except candidate 580 (Figure 2.12).

| Primer Name    | Sequence (5'-3')                                    |
|----------------|---|
| 1786_sense     | <u>GGCCTTAAUG</u> CAGATCATAACTTGCAAGGCTGTGG         |
| 1786_antisense | GGTTTAAU TCACAATGTGATGAGAACCTTCACGC                 |
| 26_sense       | <u>GGCCTTAAUG</u> GCCGCCATGGGTACC                   |
| 26_antisense   | GGTTTAAU TCATTCAAACGATGACTCCTCGCTG                  |
| 4319_sense     | <u>GGCCTTAAUG</u> GCCAGAAAATCACCAGAAGATGAACAT       |
| 4319_antisense | GGTTTAAU TCACACCTCTGATGGAAGAGTGAG                   |
| 5743_sense     | <u>GGCCTTAAUG</u> ACCAAGACCAATCCCCTGC               |
| 5743_antisense | GGTTTAAU TTAGAACTTGATAACAACCTTGACACAATCAG           |
| 7220_sense     | <u>GGCCTTAAUG</u> GAGATTAATGTTGAAGTTGCTCCAGTAAG     |
| 7220_antisense | GGTTTAAU TCAAACTCGGATAGTTTTGTCTGATCAAAGT            |
| 8694_sense     | <u>GGCCTTAAUG</u> ACGTCGTCATCCTCGCCGTC              |
| 8694_antisense | GGTTTAAU TTA CTCTTGAGAAGCCCCATATCTGC                |
| 2041_sense     | <u>GGCCTTAAUG</u> GGATACTACCATTATTATATTAGACAACCACTC |
| 2041_antisense | GGTTTAAU TCAACATCTGCAACTATGTTGTGCTTCG               |
| 580_sense      | <u>GGCCTTAAUG</u> CATCTGCAGCACCCCATCCG              |
| 580_antisense  | GGTTTAAUTCATTCTCAAATTTCAATGTATTTCCAATGTCAAT         |

Table 2.6: Primers used to amplify candidates in this study for expression with pET28a in Rosetta 2 cells. Sense and antisense adaptors for the pET28a USER cassette are shown in red and green, respectively. Black letters represent gene-specific sequences. The underlined portion of the sense primers shows the start codon.

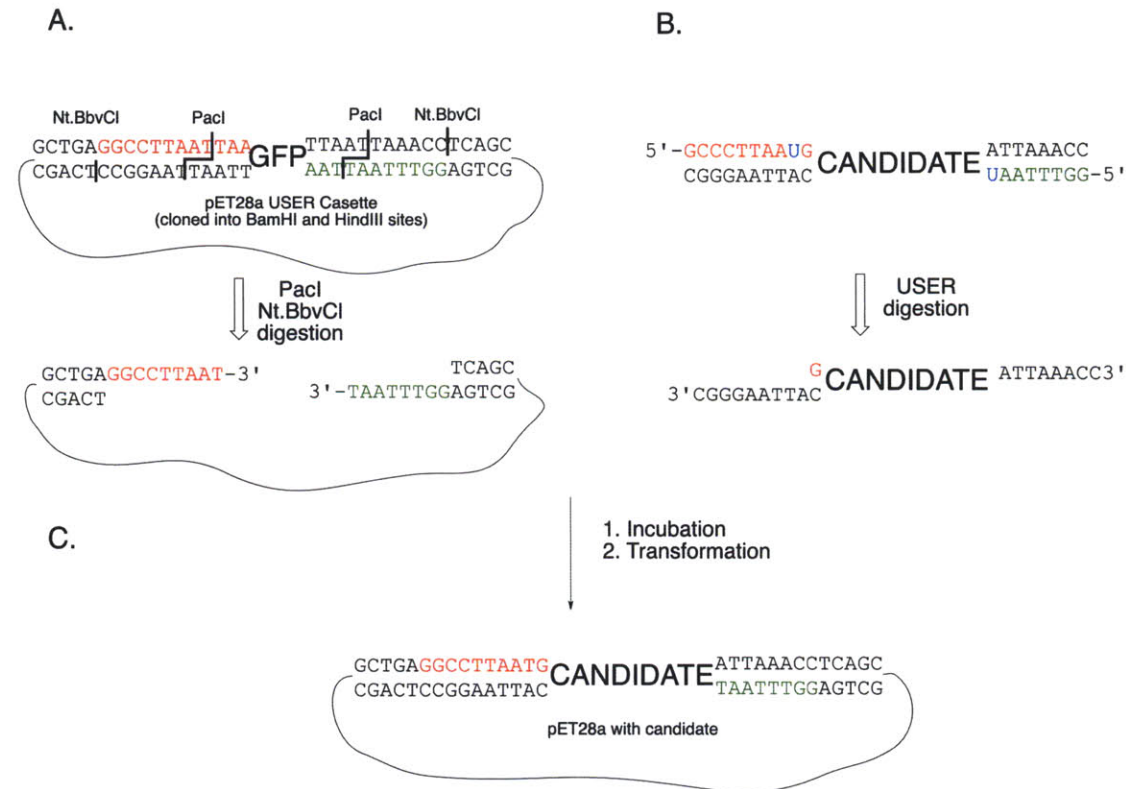


Figure 2.11. Scheme for USER cloning. Adapted from Nour-Eldin. (A) pET28a USER cassette, cloned into the BamHI and HindIII sites. Digestion with PacI and Nt.BbvCI results in directional overhang adaptors. B. 'CANDIDATE' amplified with uracil-containing adaptors that are complementary to the overhangs of the USER cassette (shown in A). Digestion with USER enzyme removes the uracil residues and allows short fragments to dissociate. C. After incubating the linearized cassette and the USER enzyme-digested candidate at 37 °C for 15 minutes and at room temperature for 15 minutes, the plasmid can be directly transformed into *E. coli* without prior ligation.

## *B. Protein Expression*

All candidates were cloned into pET28a and in frame with an N-terminal His<sub>6</sub> tag. Cultures were seeded from a single colony harboring the gene of interest; colonies were grown overnight at 37 °C and shaking at 200 rpm in LB media supplemented with kanamycin and chloriamphenicol. Ten milliliters of the seed culture were added to 500 mL of LB media supplemented with kanamycin and chloramphenicol and shaken at 200 rpm at 37 °C until the optical density at 600 nm (OD<sub>600</sub>) reached 0.5-0.7. After reaching the appropriate OD<sub>600</sub>, the cultures were incubated at room temperature for 1 hour. Subsequently, the LB media was supplemented with IPTG to a final concentration of 0.1 mg/mL to induce expression and 100 μM ZnCl<sub>2</sub> to potentially aid in protein folding. Candidates were expressed at 18 °C for 16 hours.

Cells were then spun down at 8000 rpm for 20 minutes. The cells were then resuspended in chilled PBS buffer [50 mM KH<sub>2</sub>PO<sub>4</sub> (pH 8), 300 mM NaCl, 10 mM imidazole, 10 mM dithiothreitol and 10% glycerol] and subjected to sonication (1 second on and 4 seconds off for 1.5 minutes; 60% intensity). The cellular refuse was spun down at 18000 rpm for 1 hour. The supernatant was incubated with 600 μL of Ni-NTA resin for 2.5 hours at 4 °C. The enzyme was eluted with stepwise fractions of increasing imidazole concentrations (25 mM, 50 mM, 100 mM, 150 mM and 300 mM).

Imidazole fractions were then subjected to SDS PAGE. Fractions shown to contain protein (red boxes, Figure 2.4) at the correct molecular weight were pooled and concentrated to 1 mL using a 10 kDa molecular weight cutoff Amicon

® centrifugal filter (15 mL). Concentrated samples were then buffered exchanged three times with 20 mM MOPS and immediately stored at -20 °C in single-thaw aliquots.

### *C. Thin Layer Chromatography (TLC) Assays*

Assays for oxidase activity that were monitored by TLC contained the following components: 700  $\mu$ M 10-hydroxygeraniol, 9 mM NAD(P)<sup>+</sup>, 100  $\mu$ M ZnCl<sub>2</sub>, 100 mM phosphate (pH 7.2), 3  $\mu$ M enzyme (concentration based on Bradford Assay and the protein's predicted average molecular weight on the ExpASY server).<sup>19</sup> Distilled and deionized water was added to a final volume of 100  $\mu$ L in each reaction. The reactions were run at 30 °C for approximately 16 hours. To quench the reactions and extract the product mixtures, 90  $\mu$ L of the assay mixture were added to 200  $\mu$ L of dichloromethane in a glass vial and mixed thoroughly with a syringe. After the phases settled, approximately 100  $\mu$ L of the dichloromethane layer were recovered. Samples were then concentrated to approximately 15  $\mu$ L under vacuum. To screen for product formation, approximately 5  $\mu$ L of the sample or control were spotted onto thin layer chromatography (TLC) plates and run in hexanes: ethyl acetate (1:1). All plates were stained with anisaldehyde and analyzed for product formation (Appendix A).

### *D. Gas Chromatography – Mass Spectrometry (GC-MS) Assays*

Assays were set up with the following components: 700  $\mu$ M 10-hydroxygeraniol, 9 mM NAD(P), 100  $\mu$ M ZnCl<sub>2</sub>, 100 mM Phosphate (pH 7.2), 3

$\mu\text{M}$  Enzyme (based on Bradford Assay and the protein's predicted average molecular weight on the ExPASy server).<sup>19</sup> Deionized and distilled water was added to a final volume of 100  $\mu\text{L}$  in each reaction. The reactions were run at 30  $^{\circ}\text{C}$  for approximately 16 hours. To quench the reactions and extract the product mixtures, 90  $\mu\text{L}$  of the assay mixture were added to 400  $\mu\text{L}$  of dichloromethane (in a glass vial) and mixed thoroughly with a glass syringe. After the phases settled, approximately 300  $\mu\text{L}$  of the dichloromethane layer were recovered. Samples were concentrated to dryness under vacuum then re-suspended with dichloromethane to exactly 50  $\mu\text{L}$  in glass vial inserts.

GC-MS analyses were carried out using an Agilent 6890N GC system that was connected to an Agilent 5973 MS detector. The separations were performed on a Zebron ZB-5 HT column (30 m X 0.25 mm X 0.10 mm) using helium as the carrier gas at a rate of 1  $\text{mL min}^{-1}$  (linear velocity of 37  $\text{cm s}^{-1}$ ) and an injector temperature of 220  $^{\circ}\text{C}$ . The program initiated with an isothermal phase at 60  $^{\circ}\text{C}$  for 5 minutes, followed by a 20  $^{\circ}\text{C min}^{-1}$  gradient up to 150  $^{\circ}\text{C}$ , a 45  $^{\circ}\text{C min}^{-1}$  gradient up to 280  $^{\circ}\text{C}$ , then a 4-min isothermal phase at 280  $^{\circ}\text{C}$ . The total run time was 16.39 min. GC-MS chromatograms and product spectra are shown in Appendix A.



### E. Virus-Induced Gene Silencing (VIGS)

478-base pair regions from Candidates 1786 and 5743 were cloned using the following primer pairs:

| Primer              | Sequence (5'-3')                       |
|---------------------|--|
| 1786_VIGS_Sense     | GGCGCGAUTCCTCAAGAATGTCTTCCGCCAAAG      |
| 1786_VIGS_Antisense | GGTTGCGAUATTC AAGAGCTTCATTAACCAAGTCAGG |
| 5743_VIGS_Sense     | GGCGCGAUTCCTCTTCTGTCATCACTTGCAAAG      |
| 5743_VIGS_Antisense | GGTTGCGAUTGGATCTACCTTCACTGCATAAGCTG    |

Table 2.7: Primers used to amplify 478-base pair regions of open reading frames from Candidates 1786 and 5743. Red and green show the sense and antisense adapters for the pTRV2 cassette. Black represents gene specific sequences.

pTRV2 vectors harboring the inserts of interest were subjected to sequencing prior to being transformed into electrocompetent *Agrobacterium tumefaciens* (strain Gv3101). The agrobacterium was prepared and handled according to Liscombe *et al.* Colony PCR (using primers shown in Table 2.6) was performed on overnight cultures to ensure that the agrobacterium retained the plasmid (Figure 2.13).<sup>18</sup>

Subsequently, cultures harboring pTRV1 and pTRV2 with the region to be silenced were pelleted and resuspended in inoculation solution (10 mM MES, 20  $\mu$ M acetosyringone, 10 mM MgCl<sub>2</sub>) to a final OD<sub>600</sub> of 0.7. pTRV1 and pTRV2 harboring the insert of interest were mixed in a 1:1 ratio (according to OD<sub>600</sub>) and incubated at 30 °C with shaking for 4 hours. Eight-week-old *C. roseus* plants (Little Bright Eyes)—grown in a walk-in growth chamber at 26 °C and a photoperiod of 16 hours light and 8 hours dark—were inoculated via pinching just

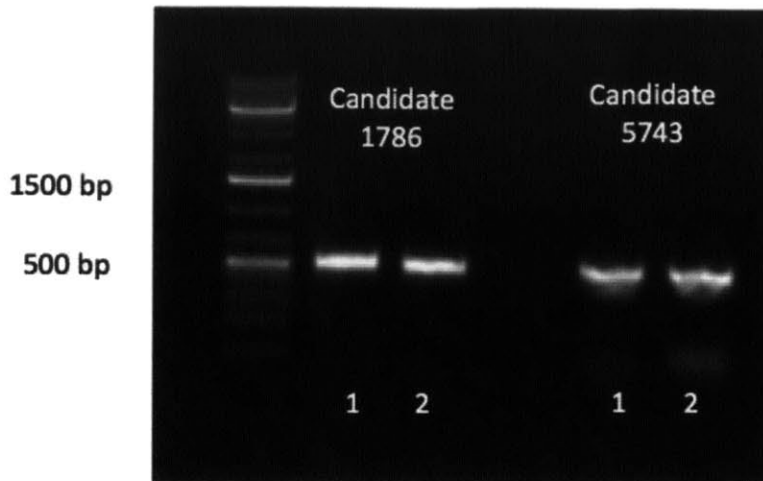


Figure 2.13: Colony PCR of Candidates 1786 and 5743 (in duplicate) after *Agrobacterium tumefaciens* (Gv3101) harboring the construct containing the 478-base pair insert had been cultured overnight. This experiment was conducted to ensure the *A. tumefaciens* retained the construct prior to plant inoculation.

below the apical meristem. Eight plants were inoculated with each construct harboring regions from either Candidate 1786 or Candidate 5743. Silenced magnesium chelatase—which displays a leaf yellowing phenotype resulting from the disruption of photosynthesis—was used as a silencing marker proxy. Additionally, an empty vector was also infiltrated to enable comparison of silencing due to the insert alone. Eight plants were inoculated with each of the magnesium chelatase and empty vector control constructs as well.

## VI. Appendix A Contents

Appendix A, affixed to the end of this thesis, contains gene and protein sequences for each of the 7 cloned candidates, TLC controls and all GC-MS spectra.

## VII. References

1. Geu-Flores, F.; Sherden, N.H.; Courdavault, V.; Burlat, V.; Glenn, W.S.; Wu, C.; Nims, E.; Cui, Yuehua and O'Connor, S. E., An Alternative Route to Cyclic Terpenes by Reductive Cyclization in Iridoid Biosynthesis. *Nature* 2012, 492 (7427), 138-42.
2. Dewick, P. M.; *Medicinal Products: A Biosynthetic Approach, 2<sup>nd</sup> Edition* John Wiley and Sons, Ltd: New York, 2003.
3. Simkin A.J; Miertiren, K.; Claude, P.; Burlat, V.; Guirimand, G.; Courdavault, V.; Papon, V.; Meyer, S.; Godet, S.; St-Pierre, B.; Giglioli-Guivarc'h, N.; Fisch, M.J.; Memelink, J. and Clastre, M., Characterization of the Plastidal Geraniol Synthase from Madagascar Periwinkle Which Initiates the Monoterpenoid Branch of the Alkaloid Pathway in the Internal Phloem Associated Parenchyma. *Phytochemistry* 2013, 85, 36-43.
4. [A] Collu, G.; Unver N.; Peltenburg-Looman, A.M.; van der Heijden, R.; Verpoorte, and Memelink, J., Geraniol 10-hydroxylase, a Cytochrome P450 Enzyme Involved in Terpenoid Indole Alkaloid Biosynthesis. *FEBS Lett.* 2001, 508 (2), 215-20. [B] Guirimand, G.; Burlat, V.; Oudin, A.; Lanoue, A; St-Pierre, B. and Courdavault, V., Optimization of the Transient Transformation of *Catharanthus roseus* cells by Particle Bombardment and its Application to the Subcellular Localization of Hydroxymethylbutenyl-4-diphosphate Synthase and Geraniol 10-hydroxylase. *Plant Cell Rep* 2009, 28(8), 1215-34.
5. Winkel-Shirley, B., Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiology* 2001, 126, 485-493.
6. O'Connor, S.E. and Maresh, J. J, Chemistry and Biology of Monoterpene Indole Alkaloid Biosynthesis. *Nat. Prod. Rep.* 2006, 23, 532-547.
7. Hallahan, D.L.; West, J. M; Wallsgrove, R. M.; Smiley, D. W.; Dawson, G. W.; Pickett, J.A. and Hamilton, J. G., Purification and Characterization of an Acyclic Monoterpene Primary Alcohol: NADP<sup>+</sup> Oxidoreductase from Catmint (*Nepeta racemosa*). *Arch Bioch Biophys* 1995, 318 (1), 105-112.
8. Iketa, H.; Esaki, N.; Nakai, S.; Hashimoto, K.; Uesato, S.; Soda, K. and Fuita, T., Acyclic Monoterpene Primary Alcohol: NADP<sup>+</sup> Oxidoreductase of *Rauwolfia serpentina* Cells: The Key Enzyme in Biosynthesis of Monoterpene Indole Alcohols. *J Biochem* 1991, 109(2), 341-7.
9. Góngura-Castillo, G.; Childs, K. L.; Fedua, G.; Hamilton, J.P.; Liscombe, D. K.; Magallanes-Lindback, M.; Mandad, K. K.; Nims, E.; Runguphan, W.; Vallaincourt, B.; Varbanova-Herde, M.; Dellapenna, D.; McKnight, T. D.; O'Connor, S and Buell, C.K., Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. *PLoS One* 2012, 7 (12), e52506.
10. Giddings, L.A.; Liscombe, D. K.; Hamilton, J. P.; Childs, K.L.; Dellapenna, D.; Buell, C. K. and O'Connor, S. E., A Stereoselective Hydroxylation Step of Alkaloid Biosynthesis by a Unique Cytochrome P450 in *Catharanthus roseus*. *J Biol Chem* 2011, 286 (9), 16751-7.
11. Glenn, W. S.; Runguphan, W. and O'Connor, S. E., Recent Progress in the

- Metabolic Engineering of Alkaloids in Plant Systems. *Curr Opin Biotechnol.* 2013. 24(2), 354-65.
12. [A] Goklany, S.; Loring, R. H.; Glick, J. and Lee-Parsons, G. W., Assessing the Limitations to Terpenoid Precursor Feeding on *Catharanthus roseus* Hairy Root Cultures through Gene Expression Profiling and Precursor Feeding. *Biotechnol Prog.* 2009. 25, 1289-1296. [B] Cheong, J-J and Choi, D.Y., Methyl Jasmonate as a Vital Substance in Plants. *Trends Genet.* 2003. 19(7), 409-13.
  13. Emanuelsson, O.; Nielsen, H.; Brunat, S. and von Heijne, G. Predicting Subcellular Localization of Proteins Based on Their N-terminal Amino Acid Sequence. *J. Mol. Biol.* 2000, 300, 1005-1016.
  14. Petersen, T. N., Brunk, S.; von Heijne G. and Nielsen, H., SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nature Methods.* 2011. 8, 785-789.
  15. Nour-Eldin, H. H.; Geu-Flores, F. and Halkier, B.A., USER Cloning and USER Fusion: The Ideal Cloning Technique for Small and Big Laboratories. *Methods Mol Biol.* 643, 2010, 185-200.
  16. Shu N.; Zhou T. and Hovmöller S, Prediction of Zinc-binding Sites in Proteins from Sequence. *Bioinformatics* 2008, 24(6): 775-82.
  17. Liscombe, D. K. and O'Connor, S. E., A Virus-Induced Gene Silencing Approach to Understanding Alkaloid Metabolism. *Phytochemistry* 2011, 72, 1969-1977.
  18. Wolken W.A.M.; ten Have K. and van der Werf, M.J., Amino Acid-Catalyzed Conversion of Citral: Cis-Trans Isomerization and its Conversion into 6-methyl-5-hepten-2-one and acetaldehyde. *J. Agric. Food Chem* 2000, 48, 5401-5405.
  19. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaus, S.; Wilkins, M. R.; Appel, R. D. and Bairoch, A., *Protein Identification and Analysis Tools on the ExPASy Server.* Humana Press. 2005.

### **VIII. Acknowledgements**

Dr. Fernando Geu-Flores contributed significantly to the design and implementation of this work and will be listed as co-first author on any publication from this study. Dr. Nathaniel H. Sherden provided all substrates and authentic standards of products. Franziska Kellner aided tremendously in the design and implementation of VIGS experiments. Professor Sarah E. O'Connor directed all research activities and contributed significantly to the design of this project. I am

deeply indebted to each person listed in this section. Any publication resulting from this work will reflect these contributions with authorship.

## Chapter 3

Reengineering a Tryptophan Halogenase to Preferentially Chlorinate a Direct Alkaloid Precursor

Part of this chapter is published as a communication in  
Glenn WS, Nims E and O'Connor SE. J Am Chem Soc. 2011 Dec  
7;133(48):19346-9

## I. Introduction

Though natural products have been used to treat human disease for thousands of years, they have not evolved specifically to do so. Therefore, before reaching the clinic, these co-opted natural products frequently require structural modification—an exigent enterprise given their often awe-inspiring complexity.<sup>1-4</sup> Installing halogens regioselectively onto natural products can generate compounds with novel or improved properties. What is more, a significant fraction (roughly 25%) of pharmaceuticals in the clinic contains halogens.<sup>2,4</sup> Notably, several classes of halogenases have been discovered and characterized, which enables environmentally friendly halogenation; however, applications are limited because of the oftentimes narrow substrate specificity of these enzymes.<sup>5</sup> Importantly, it has been demonstrated across different classes of natural products that incorporating simple precursor analogs into biosynthetic pathways is an effective strategy to produce complex natural products that are modified in a site-specific manner.<sup>6-9</sup> Therefore, we reasoned that enzymatically generating a modified precursor *in situ* would enable de novo production of complex natural products that are site-specifically modified.

This chapter describes strategies to interface RebH—a halogenase that regioselectively installs chlorine atoms onto the 7 position of tryptophan **1**—with monoterpene indole alkaloid (MIA) metabolism to yield chlorinated alkaloids de novo. MIA biosynthesis commences with the decarboxylation of L-Tryptophan **1** by tryptophan decarboxylase to yield tryptamine **2** (Figure 1A),<sup>10</sup> which condenses with secologanin **3** via a Pictet-Spengler-type reaction mechanism

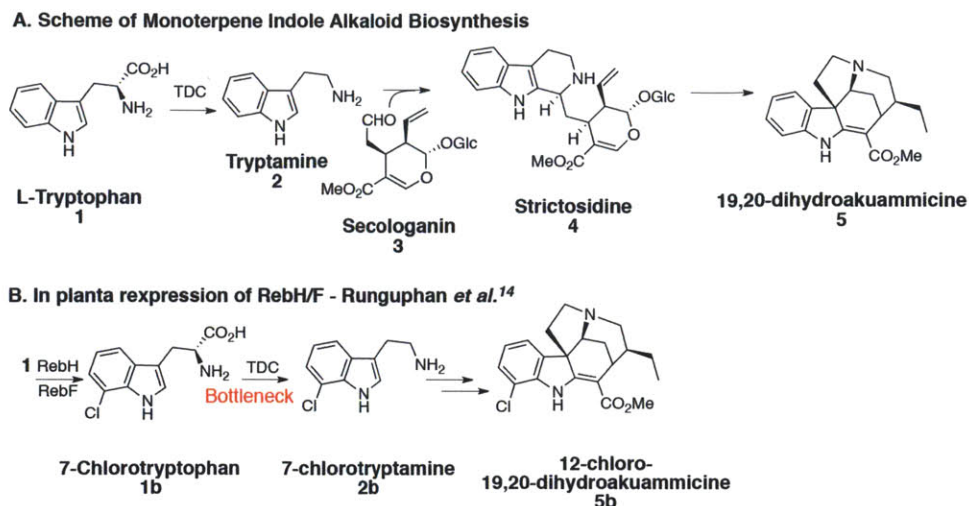


Figure 3.1: A. Scheme of MIA biosynthesis. Tryptophan decarboxylase (TDC) decarboxylates L-Tryptophan **1** to form tryptamine **2**, which condenses with secologanin via a Pictet-Spenglerase-type mechanism to form strictosidine **4**, the central MIA intermediate. 19, 20-dihydroakuammicine is a MIA found in Madagascar periwinkle that harbors opiate activity. B. Runguphan *et al.* expressed RebH/F in planta and observed the de novo production of chlorinated alkaloid 12-chloro-19,20-dihydroakuammicine at 25  $\mu\text{g/g}$  fresh weight. However, 7-Chlorotryptophan **1b** accumulated at 50  $\mu\text{g/g}$  fresh weight because of the TDC bottleneck.

to form strictosidine **4**, which is the central intermediate in the biosynthesis of over 3000 MIAs, including 19,20-dihydroakuammicine **5**, a strychnos-type MIA with opiate activity (Figure 1A). Previously, Runguphan *et al.* integrated RebH and its partner flavin reductase RebF into the metabolism of *C. roseus* (Madagascar periwinkle), a plant that produces a subset of approximately 130 MIAs (Figure 1B).<sup>13,14</sup> RebH and RebF generate 7-chlorotryptophan, which endogenous tryptophan decarboxylase decarboxylates to form 7-chlorotryptamine **2b**. 7-chlorotryptamine **2b**—a modified direct MIA precursor—is then incorporated into MIA metabolism, generating a chlorinated MIA, namely 12-chloro-19,20-dihydroakuammicine **5b**. However, in addition to accumulating 12-chloro-19,20-dihydroakuammicine **5b**, a halogenated “unnatural” natural product,



the resulting tissue cultures also accumulated substantial levels of 7-chlorotryptophan **1b** (Figure 1B).<sup>14</sup> Tryptophan decarboxylase, the enzyme that converts L-Tryptophan **1** to tryptamine **2**, accepts 7-chlorotryptophan **1b** at only 3% of the efficiency of the native substrate L-Tryptophan **1**, thereby creating a metabolic bottleneck .<sup>14</sup>

This bottleneck is undesirable because 7-chlorotryptophan **1b** could be shuttled into the production of the more valuable halogenated alkaloid final product 12-chloro-19,20-dihydroakuaammicine **5b**. Moreover, L-Tryptophan **1** is an essential metabolite that is involved in many central metabolic processes, including protein biosynthesis and, in the case of plants, auxin (growth hormone) biosynthesis. The accumulation of a halogenated primary metabolite seemingly has adverse effects on the growth rate of the tissues, perhaps because it is incorporated into protein or auxin biosynthesis. In contrast, chlorinated tryptamine analogs can be fed to seedlings and hairy roots at concentrations up to 1 mM without adverse effects.<sup>8</sup>

We hypothesized that this bottleneck could be alleviated in two ways: (1) over-express the endogenous periwinkle tryptophan decarboxylase to increase the conversion of 7-chlorotryptophan **1b** to 7-chlorotryptamine **2b** (Figure 3.2A); and (2) reengineer the halogenation machinery, namely RebH, to preferentially chlorinate a substrate downstream of the decarboxylase bottleneck (Figure 3.2B). Here we demonstrate that while over-expressing tryptophan decarboxylase failed to fracture the metabolic bottleneck (Figure 3.2A), reengineering RebH to install chlorine preferentially onto tryptamine **2**

successfully circumvented the bottleneck altogether (Figure 3.2B). To validate the function of this engineered enzyme *in vivo*, we transformed the tryptamine-specific RebH mutant (Y455W) into the alkaloid-producing plant Madagascar

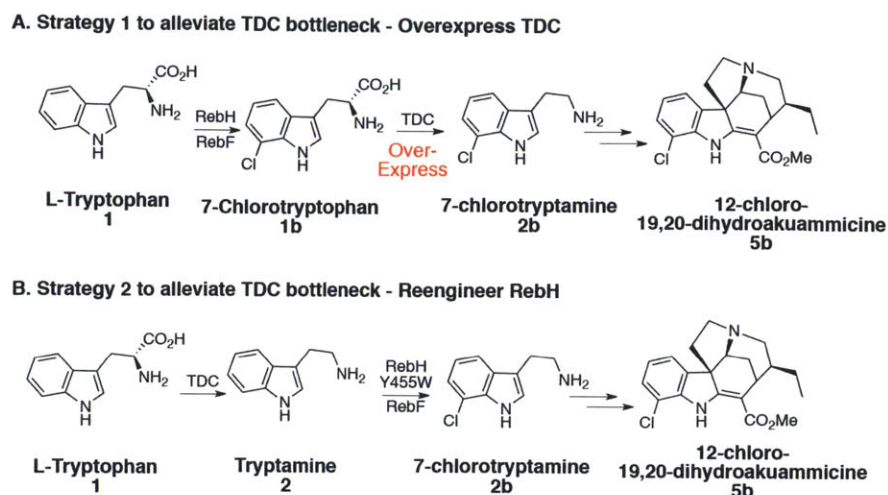


Figure 3.2: A. Strategy 1 to alleviate the TDC bottleneck by over-expressing TDC in planta. B. Strategy 2 to alleviate the TDC entails reengineering RebH to preferentially chlorinate tryptamine **2**, which is downstream of the bottleneck.

periwinkle. Despite having an incomplete understanding of MIA biosynthesis at the genetic level, we observed the *de novo* production of the halogenated alkaloid 12-chloro-19,20-dihydroakuammicine **5b**. In comparison with cultures harboring wild-type RebH and RebF, tissue cultures containing mutant RebH Y455W and RebF also accumulate microgram per gram fresh-weight quantities of 12-chloro-19,20-dihydroakuammicine **5b** but, in contrast, do not accumulate 7-chlorotryptophan **1b**, demonstrating the selectivity and potential utility of this mutant in metabolic engineering applications.

## II. Results and Discussion

We initially hypothesized that the over-expression of tryptophan decarboxylase in periwinkle would increase the conversion of 7-chlorotryptophan **1b** to 7-chlorotryptamine **2b** and thereby alleviate the bottleneck (Figure 3.2A). However, all efforts at the constitutive over-expression of tryptophan decarboxylase, RebH and RebF in periwinkle resulted in plant tissue that failed to survive selection (Figure 3.3A). Additionally, tissues transformed with only tryptophan decarboxylase also failed to survive selection and could not be

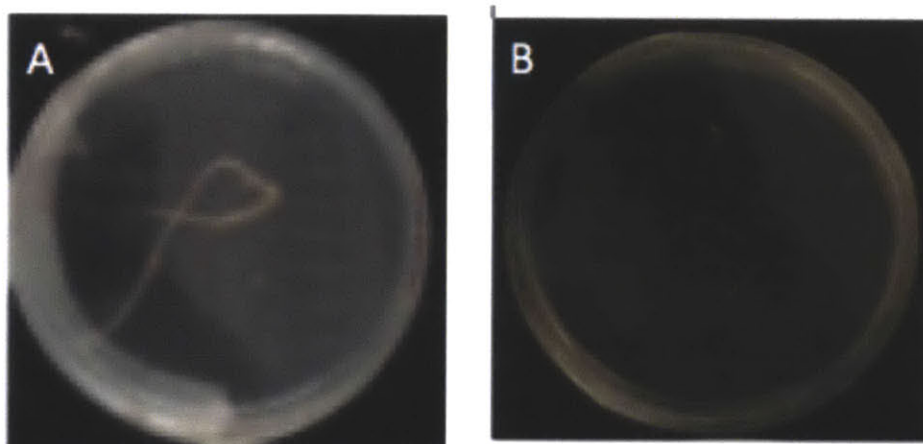


Figure 3.3: A. Madagascar periwinkle hairy roots over-expressing RebH, RebF and TDC on Gamborg's B5 media after 21 days. B. Periwinkle hairy roots on Gamborg's B5 media after 21 days over-expressing TDC only. Both sets of roots failed to survive selection. Roots only over-expressing TDC could not be rescued by being placed on media supplemented with tryptophan.

rescued through transfer to growth medium supplemented with 500  $\mu$ M L-Tryptophan **1** (Figure 3.3B). The apparent lethality of tryptophan decarboxylase over-expression suggests, perhaps unsurprisingly, that disrupting the flux of L-Tryptophan **1** is detrimental to plant survival.

The alternative strategy—bypassing the tryptophan decarboxylase bottleneck—hinges upon our ability to reengineer substrate specificity. Two strategies are plausible, namely reengineering tryptophan decarboxylase to selectively decarboxylate 7-chlorotryptophan **1b** or reengineering RebH to selectively chlorinate a downstream intermediate. Because the structure of tryptophan decarboxylase is unknown and the structure of RebH is known<sup>15</sup> we elected to focus our efforts solely on the halogenase. Enzyme engineering efforts are substantially enhanced when the protein structure is known and the mechanism is well understood. Then, the enzyme can be subjected to structure-guided techniques, such as domain swapping and site-directed mutagenesis, which create smaller protein libraries that are enriched for functional mutants. RebH, a structurally characterized enzyme whose mechanism is well understood, is a prime candidate with which to undertake various enzyme engineering efforts for these reasons.

With various enzyme engineering strategies in hand, we can begin to envision ways to engineer halogenases to accept non-native substrates for various metabolic engineering and biocatalysis efforts. Specifically, we envisioned reengineering the halogenase to preferentially chlorinate tryptamine **2**, a downstream biosynthetic intermediate that is the direct precursor to MIA biosynthesis and removed from primary metabolism (Figure 3.2B). We targeted tryptamine, specifically, because chlorinated tryptamine analogs have been shown to be non-toxic in feeding experiments with seedlings and hairy roots, even when concentrations were as high as 1 mM.

To reengineer RebH for tryptamine **2** selectivity, we examined the crystal structure of RebH complexed with L-Tryptophan **1** (PDB entry 2E4G) and proposed 17 mutations to the active site, specifically targeting residues proximal to the carboxylate moiety of the native substrate L-Tryptophan **1** (Figure 3.3).<sup>15</sup> We employed LC-MS to monitor mutant activity for both Tryptophan **1** and tryptamine **2**. Gratifyingly, one RebH mutant, RebH Y455W, preferentially

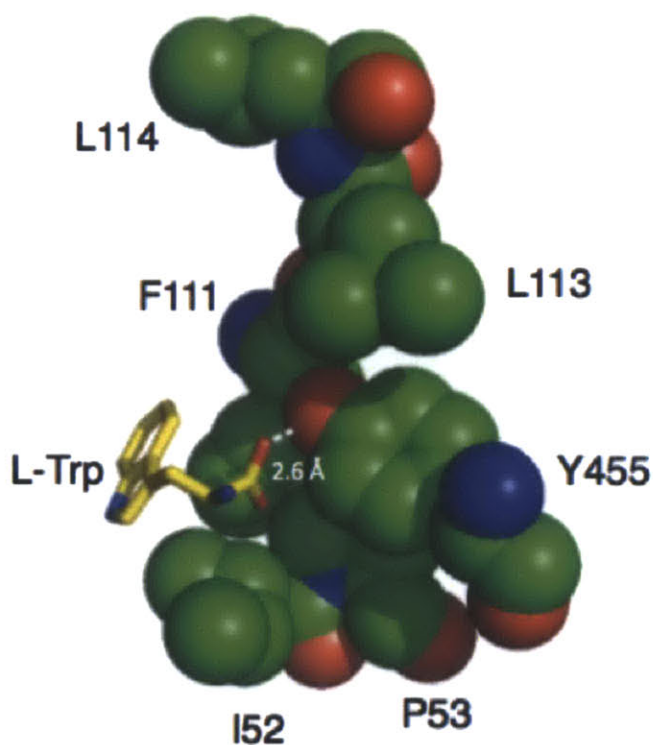


Figure 3.4: Tryptophan **1** complexed with RebH (2E4G). Tryptophan **1** is highlighted in yellow sticks. The residues proximal to the carboxylate moiety of **1** that were targeted for reengineering the substrate selectivity of RebH are shown in green space filling models.

| Mutant | Substrate at 20 $\mu$ M |            |
|--------|-------------------------|------------|
|        | L-Trp                   | Tryptamine |
| WT     | ✓ (80%)                 | ✓ (<10%)   |
| I52F   | X                       | X          |
| I52Y   | X                       | X          |
| P53F   | X                       | X          |
| P53G   | X                       | X          |
| P53W   | X                       | X          |
| F111L  | X                       | X          |
| F111W  | X                       | X          |
| F111Y  | X                       | X          |
| L113F  | X                       | X          |
| L113G  | ✓ (20%)                 | ✓ (20%)    |
| L113W  | X                       | X          |
| L114F  | X                       | X          |
| L114G  | X                       | X          |
| L114W  | X                       | X          |
| Y455F  | X                       | X          |
| Y455L  | X                       | X          |
| Y455W  | ✓ (<10%)                | ✓ (40%)    |

Table 3.1: Summary of RebH WT and mutant assays. Only RebH L113G and RebH Y455W retained activity for either substrate. Only RebH Y445W is selective for tryptamine **2**. Percent substrate conversions are shown in parentheses.



accepts tryptamine **2** as opposed to the natural substrate L-tryptophan **1** (Figure 3.5). Only one other mutant, RebH L113G, retained activity for either tryptophan **1** or tryptamine **2** (Table 3.1). Since RebH L113G converted both L-Tryptophan **1** and tryptamine **2** to the respective chlorinated products, we subjected this mutant to competition assays as described in Methods. RebH L113G does not have the desired selectivity, converting approximately 20% of both L-Tryptophan **1** and tryptamine **2** to product after an incubation period of 16 hours in the competition assay (Figure 3.6; Table 3.1). In contrast, RebH Y455W converts approximately 40% of tryptamine **2** to product, but less than 5% of L-Tryptophan **1** to product (Figure 3.5B; Table 3.1).

Slow conversion to product *in vitro* prevented accurate measurement of the steady-state enzyme kinetics parameters for WT RebH, RebH Y455W and RebH L113G. Thus, to rigorously assess the substrate selectivity of RebH Y455W—the mutant that preferentially chlorinates tryptamine **2** instead of tryptophan **1**—we utilized competition assays where either WT RebH or the Y455W mutant was incubated with different ratios of L-Tryptophan **1** and tryptamine **2**. In total, we tested three different L-Tryptophan **1**: tryptamine **2** ratios (500  $\mu$ M:500  $\mu$ M, 1000  $\mu$ M:500  $\mu$ M, and 500  $\mu$ M:1000  $\mu$ M). WT RebH chlorinated both L-Tryptophan **1** and tryptamine **2** under these assay conditions (Figure 3.7). However, we observed an approximately 30-fold higher accumulation of 7-chlorotryptamine **2b** than 7-chlorotryptophan **1b** in RebH Y455W assays across all three substrate ratios (Figure 3.7). Relative to WT RebH in these competition assays, the production of 7-chlorotryptophan **1b** was

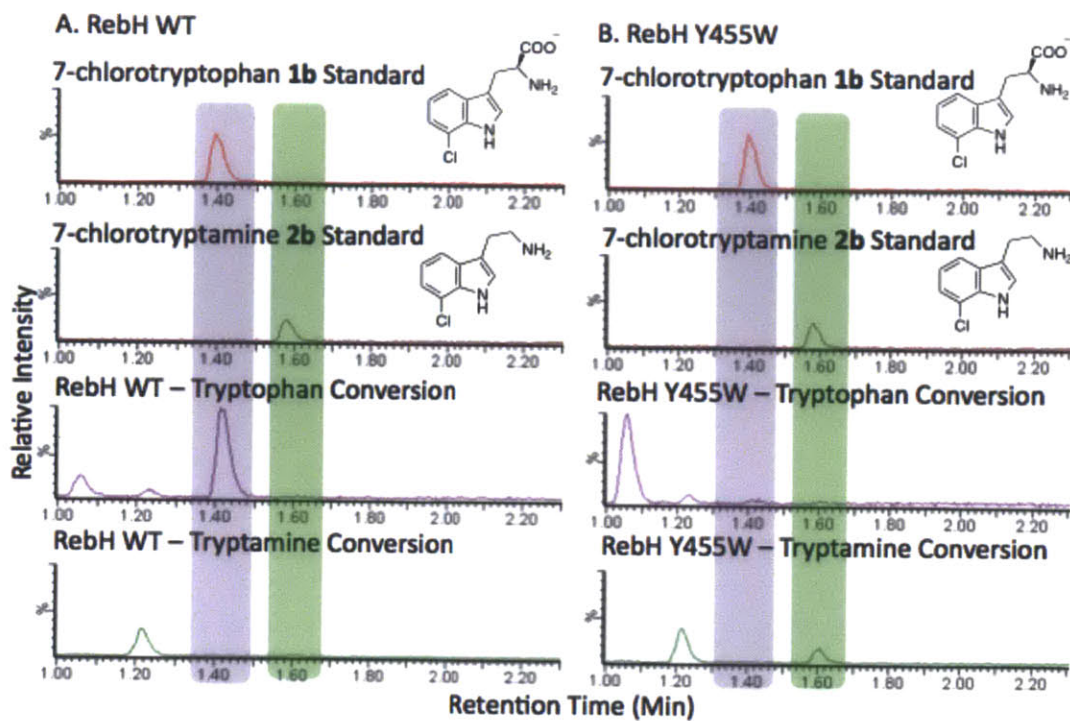


Figure 3.5: A. Extracted Ion Chromatograms for RebH WT competition assay with tryptophan and tryptamine, each at 20  $\mu\text{M}$  concentration. The top two chromatograms display 7-chlorotryptophan and 7-chlorotryptamine authentic standards. The third chromatogram displays the conversion of tryptophan, where 80% of the substrate L-tryptophan is converted to product after 16 hours. The bottom pane displays RebH WT tryptamine conversion in the competition assay, with less than 5% of tryptamine converted to product after 16 hours. B. Extracted Ion Chromatograms for RebH Y455W competition assay with tryptophan and tryptamine, each at 20  $\mu\text{M}$  concentration. The top two chromatograms display 7-chlorotryptophan and 7-chlorotryptamine authentic standards. The third chromatogram displays the RebH Y455W tryptophan conversion, where less than 5% of the substrate L-tryptophan is converted to product after 16 hours. The bottom pane displays RebH Y455W tryptamine conversion, with approximately 40% of tryptamine converted to product after 16 hours. These results illustrate that the substrate specificity of RebH has been altered successfully.



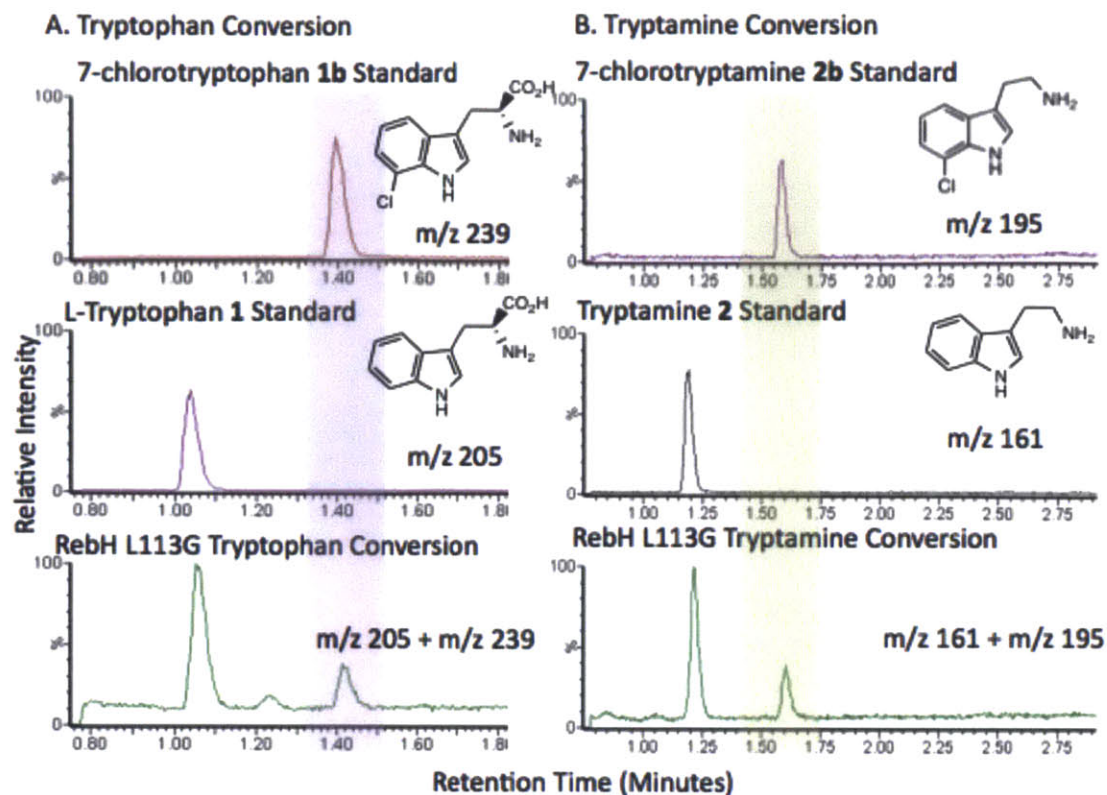


Figure 3.6: A. RebH L113G competition assays with tryptophan and tryptamine (tryptophan conversion). At concentrations of 20  $\mu\text{M}$  for both substrates, RebH L113 converts approximately 20% of tryptophan to product after an incubation period of 16 hours (bottom pane). Top and middle panes, respectively, display 7-chlorotryptophan and L-tryptophan authentic standards. B. RebH L113G competition assays with tryptophan and tryptamine (tryptamine conversion). At concentrations of 20  $\mu\text{M}$  for both substrates, RebH L113 converts approximately 20% of tryptamine to product after an incubation period of 16 hours (bottom pane). The top and middle panes, respectively, display 7-chlorotryptamine and tryptamine authentic standards.

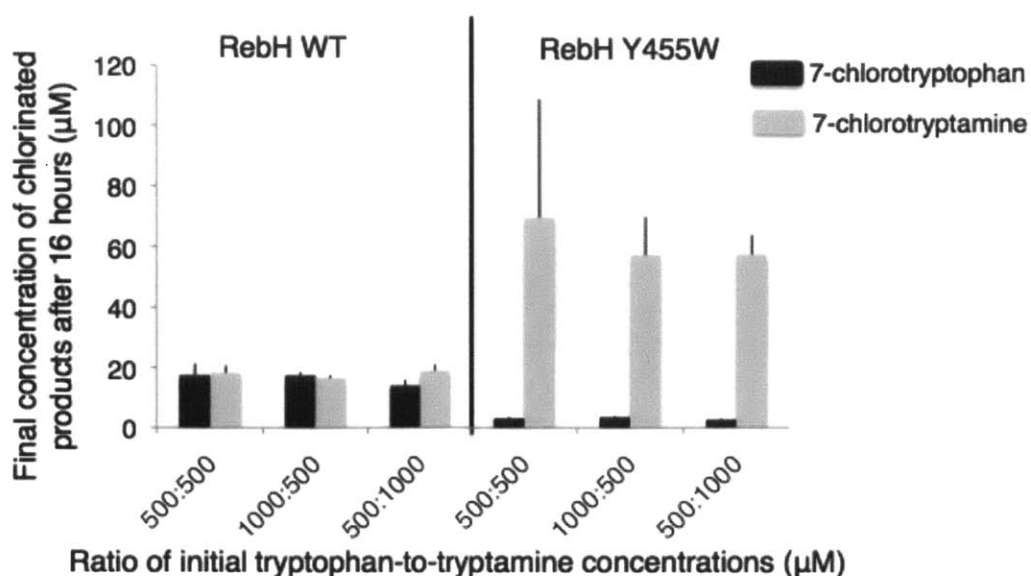


Figure 3.7: RebH WT (wild type) and RebH Y455W competition assays with tryptophan and **1** tryptamine **2**. Both RebH WT (left) and RebH Y455W (right) were incubated with three different ratios of tryptophan-to-tryptamine. RebH Y455W has a 30-fold higher accumulation of 7-chlorotryptamine (60  $\mu\text{M}$ ) over 7-chlorotryptophan (2  $\mu\text{M}$ ), indicating that this mutant is highly selective for **2**.

diminished 10-fold while the production of 7-chlorotryptamine **2b** was augmented approximately 3-fold with RebH Y455W (Figure 3.7).

Mutational analyses in PyMol suggest that RebH Y455W partially occludes L-tryptophan **1** from the redesigned active site while not impeding access for tryptamine **2** (Figure 3.8). This mutational analysis is congruent with the observation that the RebH Y455W mutant shows a clear preference for tryptamine **2** as a substrate, even when L-Tryptophan **1** is present at initial concentrations twice as high as tryptamine **2** (Figure 3.7). These results demonstrate that we successfully altered the substrate specificity of RebH in vitro to make it highly specific for tryptamine **2**, a direct MIA precursor.

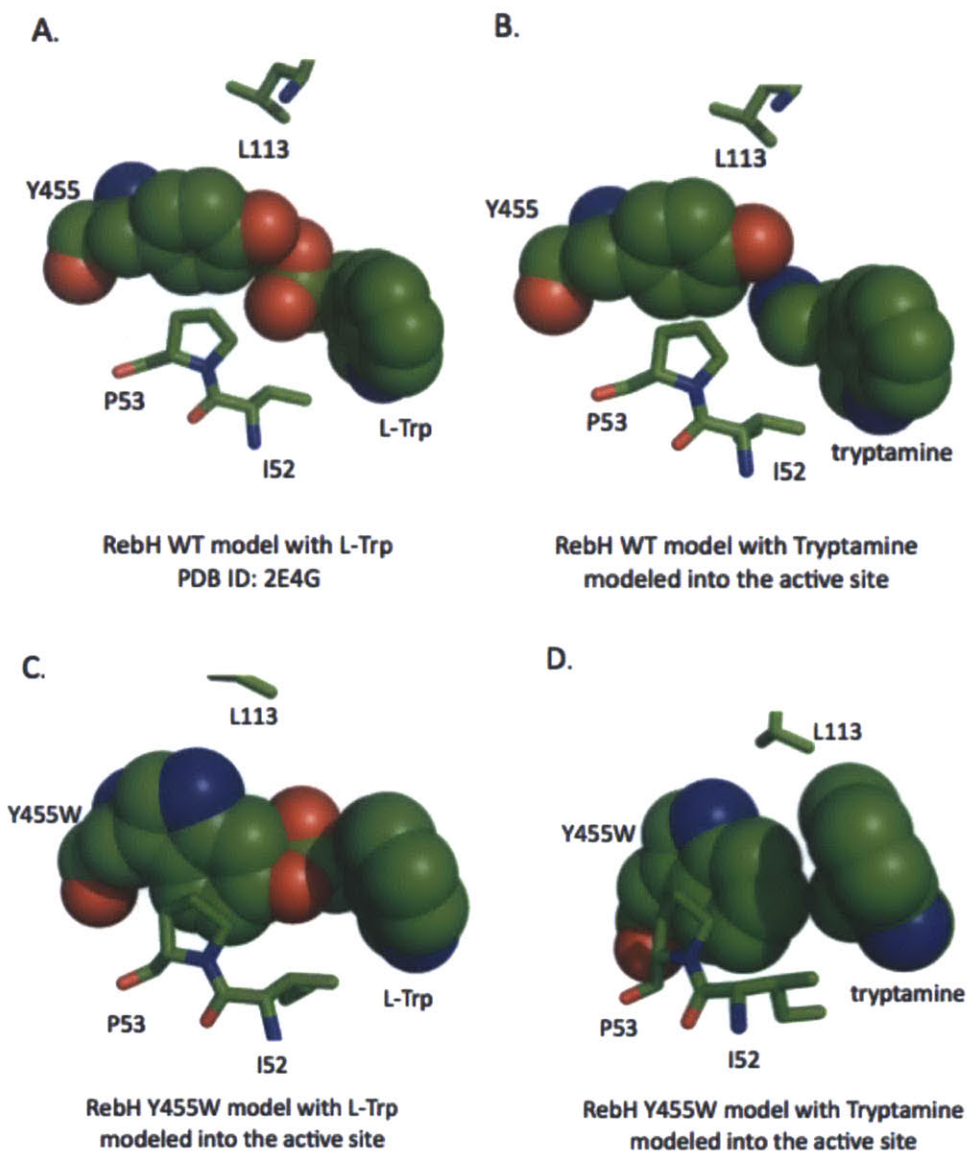


Figure 3.8. A. RebH WT model with L-tryptophan from the PDB code: 2E4G. B. Model of RebH WT with tryptamine. Tryptamine was modeled into the RebH active site by removing the carboxylate moiety of the tryptophan substrate of PDB code: 2E4G. C. RebH Y455 model with L-tryptophan. L-Trp buttresses RebH Y455W. The mutant RebH Y455W model was constructed via *in silico* mutational analysis from the PDB code: 2E4G in MacPyMOL. The right panel shows a model of RebH Y455W with tryptamine. Tryptamine was modeled into the RebH Y455W active site by removing the carboxylate moiety of the tryptophan substrate of the RebH Y455W model with L-Tryptophan (left). Models were constructed in MacPyMOL.

Notably, Hölzer and co-workers demonstrated that PrnA, a tryptophan 7-halogenase involved in pyrrolnitrin biosynthesis (55% sequence identity to RebH) does accept tryptamine analogues.<sup>16</sup> However, PrnA was shown to install chlorine atoms at the more nucleophilic 2-position of various analogues of tryptamine **2**, not at the 7-position, suggesting that the non-native substrates bind differently to PrnA (Figure 3.9).<sup>16</sup> To ensure that the regioselectivity of RebH Y455W was unaltered, and also to determine whether WT RebH possesses the same regioselectivity for the non-native substrate tryptamine **2**, we compared the tryptamine **2** enzymatic products of WT RebH and RebH Y455W with authentic standards of all possible monochlorinated tryptamine isomers, namely, tryptamine **2** chlorinated at the 2-, 4-, 5-, 6-, or 7-position of the indole ring. Using LC-MS to monitor the retention times of the various chlorotryptamine isomers ( $m/z$  195), we noted that the WT RebH and RebH Y455W tryptamine enzymatic products co-eluted exclusively with the 7-chlorotryptamine authentic standard (Figure 3.10). Thus, WT RebH and RebH Y455W retained regioselectivity for the 7-position of the indole ring with tryptamine **2**.

Though the mutant enzyme was sluggish *in vitro*, we rationalized that a steady supply of fresh enzyme in the plant cell may allow the mutant enzyme to function adequately over extended periods to yield isolable quantities of chlorinated alkaloids. To test this reengineered enzyme in the context of a biosynthetic pathway *in vivo*, we introduced RebH Y455W and RebF into

periwinkle (*C. roseus*) via *Agrobacterium rhizogenes* to yield stably transformed root cultures. To streamline the engineering process, neither gene was codon

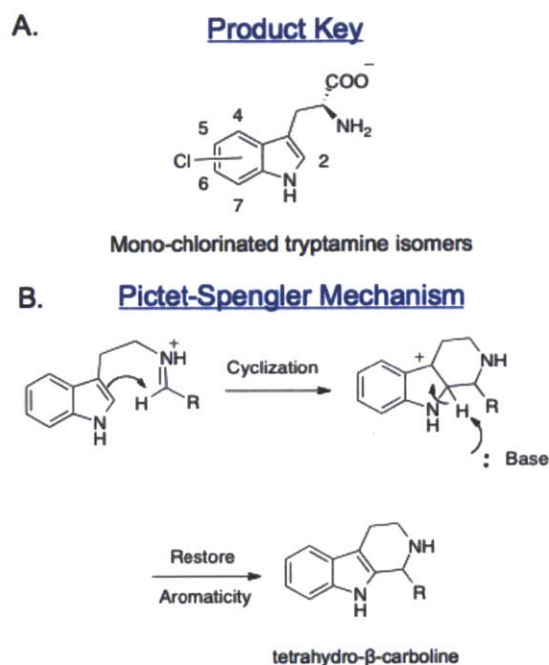
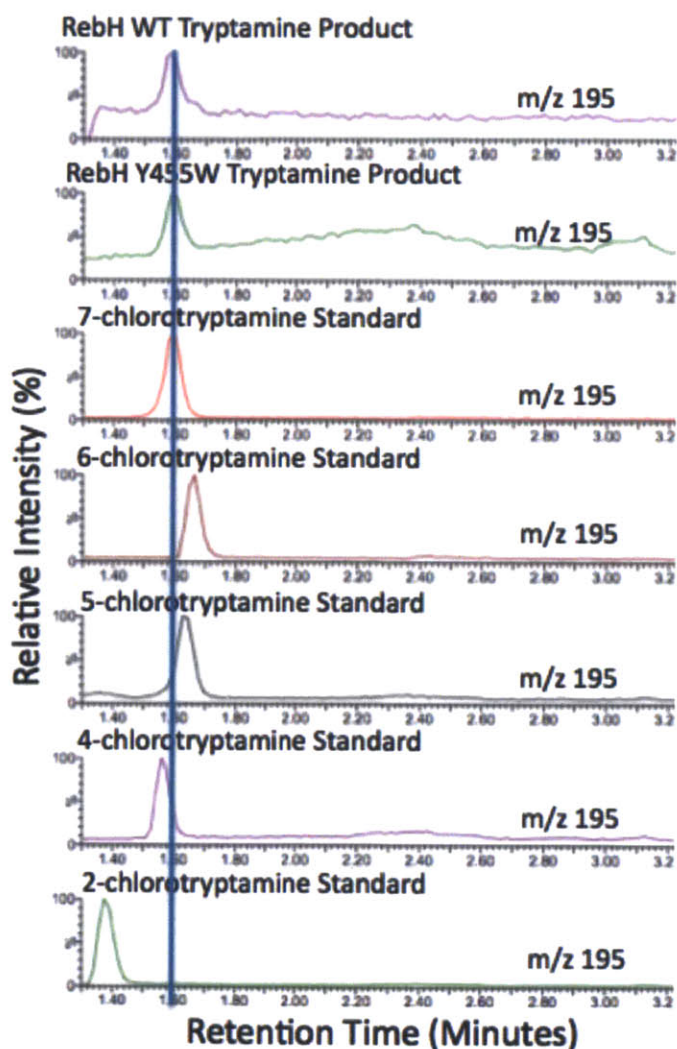


Figure 3.9: A. Chlorotryptamine product key. B. Proposed Pictet-Spengler mechanism of strictosidine synthase. Hölzer and co-workers demonstrated that PrnA, a tryptophan 7 halogenase involved in pyrrolnitrin biosynthesis (55% sequence identity to RebH) does accept tryptamine analogues.<sup>16</sup> However, PrnA was shown to install chlorine atoms at the more nucleophilic 2-position of various analogues of tryptamine **2**, not at the 7-position. Chlorination at the 2 position of tryptamine **2** would preclude tetrahydro- $\beta$ -carboline formation via a Pictet-Spengler-type mechanism, a necessary step in MIA biosynthesis.

optimized for expression in periwinkle. Each gene was placed into a commercially available plant vector (pCAMBIA1305.1) and under the control of the constitutive promoter CaMV 35S.



Precursor-directed biosynthesis studies in periwinkle with 7-chlorotryptamine **2b**, as well as the prior studies in which periwinkle was transformed with WT RebH and RebF, indicated that analogues of 19,20-



dihydroakuammicine **5** are the major alkaloid analogue products when the 7-position of the indole ring is modified. Notably, when the tryptamine precursor is chlorinated at other positions of the indole ring and integrated into periwinkle metabolism, the resulting chlorinated alkaloid profiles are drastically different,

and chlorinated 19,20-dihydroakuammicine **5** is not a major product.<sup>8</sup> Methanolic extracts of the transformed roots were analyzed with selected ion monitoring for 12-chloro-19,20-dihydroakuammicine **5b** (m/z 359) and 7-chlorotryptophan **1b** (m/z 239). We observed several root lines harboring RebH Y455W and RebF that produced 12-chloro-19,20-dihydroakuammicine **5b** (Figure 3.11). RebH Y455W/RebF (line 13), for example, accumulated  $2.65 \pm 1.08$   $\mu\text{g}$  per gram fresh weight of the product 12-chloro-19,20-dihydroakuammicine **5b** (averaged across three biological replicates) with no measured accumulation of 7-chlorotryptophan **1b**, indicating that RebH Y455W displays the desired substrate selectivity in planta as well as in vitro (Figure 3.12; Tables 3.2 and 3.3). The accumulation of 12-chloro-19,20-dihydroakuammicine **5b** (characterized by co-elution with an authentic standard), the major metabolite expected from 7-chlorotryptamine **2b**, provides further confirmation that RebH Y455W installs chlorine regioselectively at the 7-position of the indole ring of tryptamine **2**. Chlorinated alkaloids aside from 12-chloro-19,20-dihydroakuammicine **5b** were not observed in this study, as evidenced by selected ion monitoring. Moreover, chlorination at the 2-position of the indole ring of tryptamine, as was observed in the study with Hölzer and co-workers,<sup>16</sup> would preclude the formation of the tetrahydro- $\beta$ -carboline via a Pictet-Spengler mechanism, a necessary step in the biosynthesis of the monoterpene indole alkaloids (Figure 3.9).<sup>1</sup>

| Line 13            | 12-chloro-19,20-dihydroakuammicine                    | 7-chlorotryptophan                                     |
|--------------------|---|--|
| m/z                | 359   | 239  |
| Subculture         | Accumulation ( $\mu\text{g} / \text{g}$ fresh weight) | Accumulation ( $\mu\text{g} / \text{g}$ fresh weight ) |
| Subculture 1       | 1.41  | not detected   |
| Subculture 2       | 3.23  | not detected   |
| Subculture 3       | 3.32  | not detected   |
| Average            | 2.65  | not detected   |
| Standard Deviation | 1.08  | not detected   |

Table 3.2: Metabolic analysis of RebH Y455W and RebF line 13 across 3 subcultures.

| Line 74            | 12-chloro-19,20-dihydroakuammicine                    | 7-chlorotryptophan                                    |
|--------------------|---|---|
| m/z                | 359   | 239   |
| Subculture         | Accumulation ( $\mu\text{g} / \text{g}$ fresh weight) | Accumulation ( $\mu\text{g} / \text{g}$ fresh weight) |
| Subculture 1       | 2.15  | 1.25  |
| Subculture 2       | 2.24  | not detected  |
| Subculture 3       | 1.83  | not detected  |
| Average            | 2.07  | 0.415   |
| Standard Deviation | 0.215   | 0.722   |

Table 3.3: Metabolic analysis of RebH Y455W and RebF line 74 across 3 subcultures.

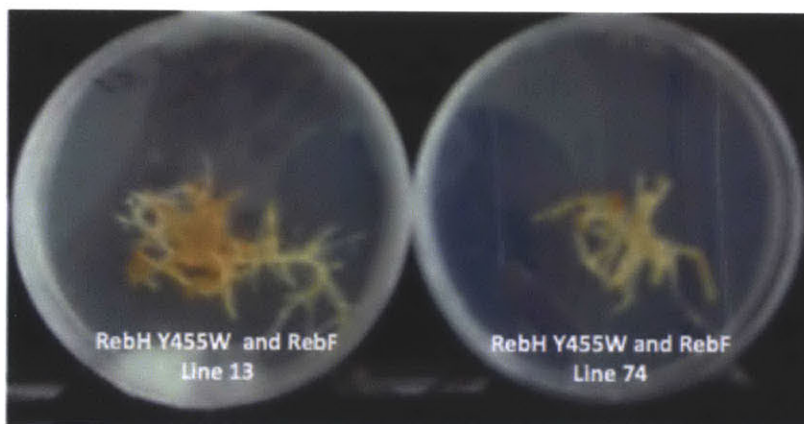


Figure 3.11: RebH Y455W and RebF lines 13 and 74 at 21 days.



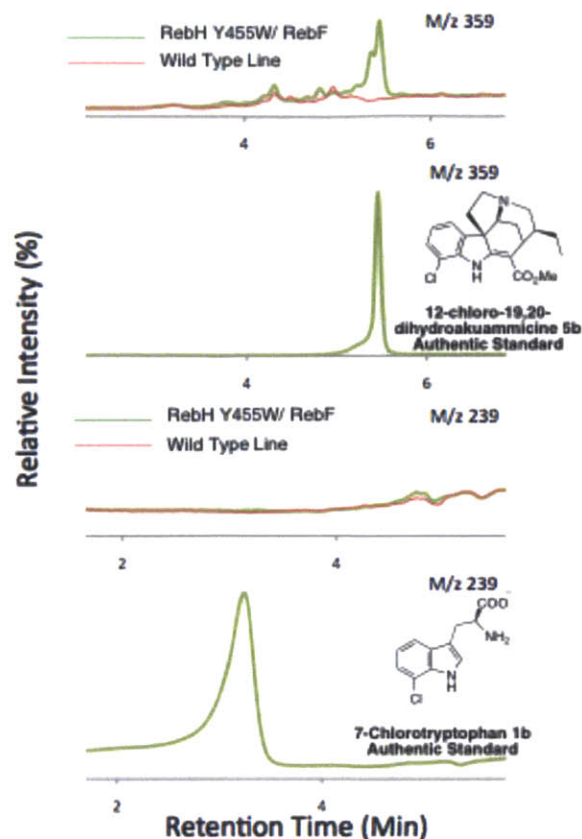


Figure 3.12: Metabolic analysis of lines harboring RebH Y455W and RebF. A. Selected ion monitoring for 12-chloro-19,20-dihydroakuammicine **5b** ( $m/z$  359) in transgenic root lines. This line accumulates  $2.65 \pm 1.08 \mu\text{g}$  per gram fresh weight of **5b**. B. The authentic standard of **5b**.<sup>14</sup> C. Selected ion monitoring of 7-chlorotryptophan **1b** ( $m/z$  239), D. Authentic standard of **1b**.<sup>11</sup> Lines expressing RebH Y455W and RebF do not display a peak corresponding to 7-chlorotryptophan **1b**, but do produce alkaloid **5b**.

Notably, no 7-chlorotryptamine **2b** accumulated in these transformed hairy root lines, suggesting that 7-chlorotryptamine **2b** is readily shuttled into the alkaloid metabolism of periwinkle. Moreover, the lines harboring RebH Y455W and RebF survived selection and grew more rapidly than lines over-expressing tryptophan decarboxylase, RebH WT, and RebF (Figures 3.3A and 3.11), demonstrating that reengineering the halogenase was the superior method of alleviating the tryptophan decarboxylase bottleneck. However, the yield of unnatural alkaloid 12-chloro-19,20-dihydroakuammicine **5b** in this study remained low (approximately 1% of the total alkaloid content),<sup>14</sup> indicating that this system is not yet at a stage where large-scale production of 12-chloro-19,20-dihydroakuammicine **5b** is practical.

We also transiently expressed RebH Y455W and RebF in leaf. Gene constructs were separately transformed into *Agrobacterium tumefaciens*, then the two *Agrobacterium* strains were mixed in a 1:1 ratio just prior to transfection. Periwinkle leaves were transfected via either vacuum infiltration or syringe injection. We did not observe the accumulation of either 7-chlorotryptamine **2b** or halogenated alkaloids under these conditions, suggesting perhaps that MIA biosynthesis—which is removed from primary metabolism—was not highly active at that stage of growth. Alternatively, build up of halogenated alkaloids may require an extended period of time, which in turn would require a constitutive rather than transient expression system. Free tryptamine **2** levels in periwinkle leaf were not measured in this study, but it is plausible that low levels of free tryptamine **2** prevent the detection of the direct RebH Y455W and RebF product,

7-chlorotryptamine **2b**. Because we did not see formation of 7-chlorotryptamine **2b**, we also screened for the accumulation of downstream alkaloids, specifically 12-chloro-19,20-dihydroakuammicine **5b**, which we did not observe either. Notably, 19,20-dihydroakuammicine **5** does not accumulate at detectable levels in periwinkle plant leaf, presumably because the genes involved in 19,20-dihydroakuammicine biosynthesis—which have not been identified as yet—are at best only poorly expressed in leaf.

We hypothesize that as tractable heterologous hosts are developed to produce plant-derived alkaloid pathways in high yields, incorporation of this redesigned biosynthetic enzyme (along with selected downstream biosynthetic enzymes that have also been engineered to favor chlorinated substrates) may play a crucial role in improving the production of chlorinated alkaloids.

### III. Conclusions

Installing halogens onto natural products can generate compounds with novel or improved properties.<sup>1-4</sup> Notably, enzymatic halogenation is now possible as a result of the discovery of several classes of halogenases;<sup>5</sup> however, applications are limited because of the narrow substrate specificity of these enzymes. Here we demonstrate that the flavin-dependent halogenase RebH can be engineered to install chlorine preferentially onto tryptamine **2** rather than the native substrate L-Tryptophan **1**. Tryptamine **2** is a direct precursor to many alkaloid natural products, including approximately 3000 monoterpene indole alkaloids. To validate the function of this engineered enzyme *in vivo*, we

transformed the tryptamine-specific RebH mutant (Y455W) into the alkaloid-producing plant Madagascar periwinkle (*Catharanthus roseus*) and observed the de novo production of the halogenated alkaloid 12-chloro-19,20-dihydroakuammicine **5b**. While wild type RebH has been integrated into periwinkle metabolism previously, the resulting tissue cultures accumulated substantial levels of 7-chlorotryptophan **1b**—a metabolite that potentially disrupts primary metabolism, including auxin (growth hormone) biosynthesis. Tryptophan decarboxylase, the enzyme that converts L-Tryptophan **1** to tryptamine **2**, accepts 7-chlorotryptophan **1b** at only 3% of the efficiency of the native substrate, thereby creating a bottleneck.<sup>14</sup> The RebH Y455W mutant circumvents this bottleneck by installing chlorine directly onto tryptamine **2**, a downstream substrate. In comparison with cultures harboring RebH and RebF, tissue cultures containing mutant RebH Y455W and RebF also accumulate microgram per gram fresh-weight quantities of 12-chloro-19,20-dihydroakuammicine **5b** but, in contrast, do not accumulate 7-chlorotryptophan **1b**, demonstrating the selectivity and potential utility of this mutant in metabolic engineering applications.

Halogen moieties in natural products have been shown to confer potency and modulate molecular bioactivity and pharmacokinetics.<sup>1,2,4,6,13,17</sup> Additionally, halogens offer unique, site-specific handles that can be utilized in cross-coupling methodology for further derivatization.<sup>18</sup> Notably, halogens appear in 25% of pharmaceutical compounds.<sup>2,4</sup> We have demonstrated the de novo biosynthesis of a halogenated “unnatural” plant natural product by redesigning a halogenase to preferentially install a chlorine atom onto a direct alkaloid precursor, tryptamine

2, and subsequently integrating this redesigned enzyme into the alkaloid biosynthesis of periwinkle. Interestingly, RebH can brominate the 7-position of L-Tryptophan **1** in the presence of bromide ions.<sup>14,15</sup> However, bromination does not occur selectively in the presence of chloride sources (such as sodium chloride), which to date has prevented selective formation of brominated products using chlorinase enzymes in whole-cell systems.<sup>14</sup>

This mutant allows us to circumvent the metabolic bottleneck positioned at tryptophan decarboxylase. Moreover, this work, along with other recently reported studies<sup>19</sup>—for example, Payne *et al.* demonstrated the use of wild-type RebH to halogenate an array of arenes regioselectively<sup>20</sup>—highlights the potential use of halogenases for more widespread applications. Notably, the work presented in this chapter was conducted without a complete understanding of MIA biosynthesis at the genetic level. One of the efforts to further elucidate MIA biosynthesis is described in Chapter 2. Understanding the pathway completely would potentially enable the efficient shuttling of chlorinated alkaloids into different pathway branches through the reengineering of enzymes that do not accept halogenated substrate analogs.

#### **IV. Methods**

##### *A. The in planta Over-Expression of RebH WT, RebF and TDC*

To over-express RebH WT, RebF and TDC *in planta*, we used the pCAMBIA vector system. Specifically, the plant transformation vector harboring codon optimized RebH WT and RebF (pCAMBIA1300-RebHRebF) and the plant

|                       |   |
|-----------------------|---|
| TDC_NcoI_pCAM1305.1   | 5'-AAAAA <u><b>CCATGG</b></u> ATGGGCAGCATTGA-3' |
| TDC_BstEII_pCAM1305.1 | 5'-AAAAA <u><b>GGTGACCT</b></u> CAAGCTTTTGG-3'  |

Table 3.4: Primers to clone Tryptophan Decarboxylase into pCAMBIA1305.1. Restriction sites are underlined and bolded.

transformation vector harboring tryptophan decarboxylase (pCAMBIA1305.1-TDC) were transformed separately into *A. rhizogenes*. pCAMBIA1300-RebHRebF was a gift from Runguphan and Qu.<sup>14</sup> *C. roseus* Tryptophan decarboxylase (TDC) was amplified from TDC in pGEM (pGEM-TDC was a gift from Runguphan) and flanked with NcoI and BstEII then placed into the vector pCAMBIA1305.1 using the primers listed in Table 3.4. These vectors were used to transform *C. roseus* according to Methods Section B. *C. roseus* roots overexpressing Tryptophan Decarboxylase (TDC) along with codon optimized RebH and RebF were grown, selected and propagated according to Methods Section B. To ensure that transgenic TDC was integrated into the genome of *C. roseus*, primers were designed to amplify from the CaMV 35S promoter of pCAMBIA1305.1 to the middle of TDC to give an amplicon of 500 basepairs. These primers are listed in Table 3.5. The amplicon from genomic DNA is shown in Figure 3.13. Roots expressing RebH, RebF and transgenic TDC are shown in Figure 3.3A. Control lines overexpressing TDC only are shown in Figure 3.3B.

|                        |                               |
|------------------------|-------------------------------|
| TDC_transgenic_forward | 5'-CTCTTGACCATGGATGGGCAGC-3'  |
| TDC_transgenic_reverse | 5'-GTGGTGTTTTGGATGACGCCGCC-3' |

Table 3.5: Primers for 500-bp TDC amplicon from CaMV 35S promoter to center of TDC gene.

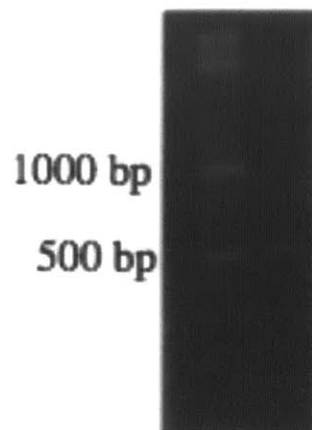


Figure 3.13: Agarose electrophoresis gel (1%) of TDC amplicon from RebH, RebF and TDC-overexpressing line. Lane 1: DNA ladder. Lane 2: TDC amplicon from constitutive eukaryotic CaMV 35S promoter to middle of TDC gene (amplicon 500 base pairs).

#### *B. Stable Transformation Protocol – RebH Y455W and RebF*

*C. roseus* seedlings were germinated aseptically on solid Gamborg's B5 media (full strength basal salts, full strength vitamins, 30 g/L sucrose, 6 g/L agar, pH 5.7) and grown for 4-6 weeks in a 16-hour light 8-hour dark cycle at 26 °C. Liquid cultures of *A. rhizogenes* containing the plasmid of interest were grown in Yeast Extract and Mannitol (YEM) media supplemented with kanamycin at 30 °C for 48 hours just prior to infection. Seedlings were wounded at the stem tip using forceps freshly dipped in the inoculant. Hairy roots formed on the seedlings at approximately 3 weeks on approximately 80% of the punctured seedlings. Hairy roots were grown on the seedlings for an additional 3 weeks then excised and placed on solid Gamborg's B5 media (half strength basal salts, full strength vitamins, 30 g/L sucrose, 6 g/L agar, pH 5.7) containing hygromycin (0.03 mg/mL) for selection and cefotaxime to a final concentration of 0.25 mg/mL to kill

remaining *A. rhizogenes*. Cultures were grown in the dark for 1 month at 26 °C during the selection process. Following the selection process, roots were propagated by transferring actively growing portions of the root after 21 days onto Gamborg's B5 media lacking hygromycin and cefotaxime.

### C. *RebH* Mutant Design and Expression

The construct containing the *RebH* gene (pET28a-*RebH*) was expressed in BL21 (DE3) pLysS and purified with Ni-NTA resin as previously described.<sup>14</sup> Single-thaw aliquots were stored at -80 °C for no more than 2 weeks. Primers to introduce 17 mutations (I52F, I52Y, P53F, P53G, P53W, F111L, F111W, F111Y, L113F, L113G, L113W, L114F, L114G, L114W, Y455F, Y455L, Y455W) into the pET28a-*RebH* construct were designed using the Stratagene (Agilent) online site-directed mutagenesis tool (<http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Tool&SubPageType=ToolQCPD&PageID=15>). Mutations to the pET28a-*RebH* construct were made using the QuikChange Site-Directed Mutagenesis Kit according to manual protocol specifications. The sequences of all vector constructs were verified by DNA sequencing. Mutant genes were expressed and purified following the same procedure for *RebH* wildtype.<sup>1</sup> Primers to introduce mutations are listed in Table 3.6.

| Primer Name    | Primer Sequence (5'-3')            |
|----------------|------------------------------------|
| I52F_SENSE     | GGCGAGGCCACGTCCCCCAATCTGC          |
| I52F_ANTISENSE | GCAGATTGGGGAACGTGGCCTCGCC          |
| I52Y_SENSE     | GTCGGCGAGGCCACGTATCCCCAATCTGCAGACG |



|                 |   |
|-----------------|---|
| I52Y_ANTISENSE  | CGTCTGCAGATTGGGATACGTGGCCTCGCCGAC         |
| P53F_SENSE      | GGCGAGGCCACGATCTTCAATCTGCAGACGGC          |
| P53F_ANTISENSE  | GCCGTCTGCAGATTGAAGATCGTGGCCTCGCC          |
| P53G_SENSE      | GCGAGGCCACGATCGGCAATCTGCAGACGG            |
| P53G_ANTISENSE  | CCGTCTGCAGATTGCCGATCGTGGCCTCGC            |
| P53W_SENSE      | GGCGAGGCCACGATCTGGAATCTGCAGACGGCG         |
| P53W_ANTISENSE  | CGCCGTCTGCAGATTCCAGATCGTGGCCTCGCC         |
| F111L_SENSE     | CCACTTCTACCACTCCTTAGGTCTGCTCAAGTACC       |
| F111L_ANTISENSE | GGTACTTGAGCAGACCTAAGGAGTGGTAGAAGTGG       |
| F111W_SENSE     | ACCACTTCTACCACTCCTGGGGTCTGCTCAAGTACC      |
| F111W_ANTISENSE | GGTACTTGAGCAGACCCCAGGAGTGGTAGAAGTGGT      |
| F111Y_SENSE     | GACCACTTCTACCACTCCTATGGTCTGCTCAAGTACCA    |
| F111Y_ANTISENSE | TGGTACTTGAGCAGACCATAGGAGTGGTAGAAGTGGTC    |
| L113F_SENSE     | CTACCACTCCTTCGGTTTCCTCAAGTACCACGAGC       |
| L113F_ANTISENSE | GCTCGTGGTACTTGAGGAAACCGAAGGAGTGGTAG       |
| L113G_SENSE     | CTTCTACCACTCCTTCGGTGGGCTCAAGTACCACGAG     |
| L113G_ANTISENSE | CTCGTGGTACTTGAGCCCACCGAAGGAGTGGTAGAAG     |
| L113W_SENSE     | CTTCTACCACTCCTTCGGTTGGCTCAAGTACCACGAG     |
| L113W_ANTISENSE | CTCGTGGTACTTGAGCCAACCGAAGGAGTGGTAGAAG     |
| L114F_SENSE     | CACTCCTTCGGTCTGTTCAAGTACCACGAGC           |
| L114F_ANTISENSE | GCTCGTGGTACTTGAACAGACCGAAGGAGTG           |
| L114G_SENSE     | CCACTCCTTCGGTCTGGGCAAGTACCACGAGCAG        |
| L114G_ANTISENSE | CTGCTCGTGGTACTTGCCCAGACCGAAGGAGTGG        |
| L114W_SENSE     | CTACCACTCCTTCGGTCTGTGGAAGTACCACGAGCAGATT  |
| L114W_ANTISENSE | GAATCTGCTCGTGGTACTTCCACAGACCGAAGGAGTGGTAG |
| Y455F_SENSE     | ACGACGCCCAGCTCTACTTCGGCAACTTC             |
| Y455F_ANTISENSE | GAAGTTGCCGAAGTAGAGCTGGGCGTCGT             |
| Y455L_SENSE     | GACGACGCCCAGCTCTACTTAGGCAACTTCGAGG        |

|                 |                                    |
|-----------------|------------------------------------|
| Y455L_ANTISENSE | CCTCGAAGTTGCCTAAGTAGAGCTGGGCGTCGTC |
| Y455W_SENSE     | ACGACGCCAGCTCTACTGGGGCAACTTCGAG    |
| Y455W_ANTISENSE | CTCGAAGTTGCCCCAGTAGAGCTGGGCGTCGT   |

Table 3.6: Primers to introduce 17 separate RebH mutations.

RebH WT and RebH Y455W expression are shown in Figure 3.13.

**A. RebH WT Expression**

**B. RebH Y455W Expression**

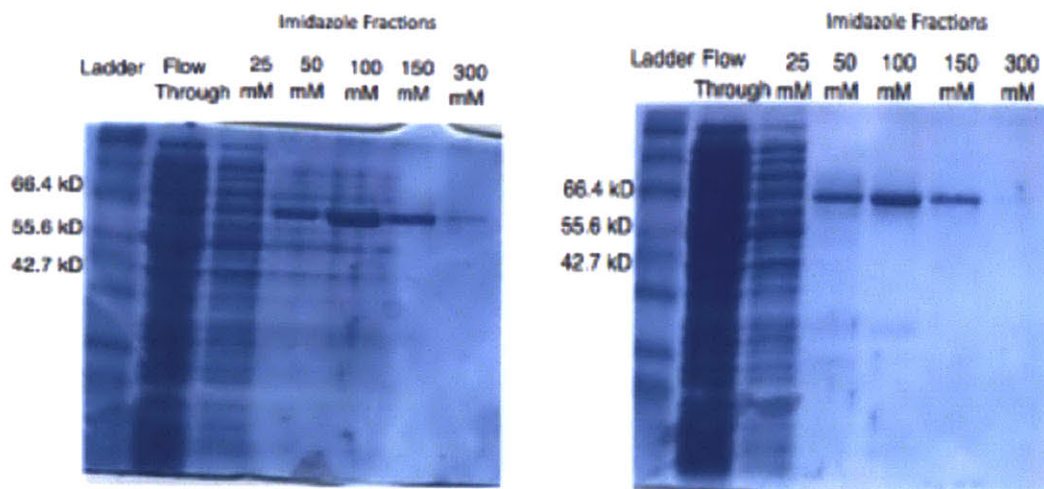


Figure 3.14: A. RebH WT expression (expected molecular weight 62.3 kDa). B. RebH Y455W expression (expected molecular weight 62.3 kDa).

#### D. RebH Activity Assays

RebH or RebH mutant enzyme at a final concentration of 1  $\mu\text{M}$  (estimated by Bradford assay) was incubated with 50  $\mu\text{M}$  Flavin Adenine Dinucleotide (FAD), 50 mM NaCl, and substrate (either 20  $\mu\text{M}$  L-tryptophan, 20  $\mu\text{M}$  tryptamine or 1250  $\mu\text{M}$  tryptamine), in 100 mM  $\text{K}_2\text{HPO}_4$  buffer (pH 7.2) in a final volume of 100  $\mu\text{L}$ . For ease of screening, dithiothreitol (20 mM) was added to all *in vitro* assays as a reductant for FAD. The assay mixtures were incubated at 30  $^\circ\text{C}$  for 12 hours, after which aliquots (25  $\mu\text{L}$ ) were quenched with methanol (975  $\mu\text{L}$ ) and centrifuged at 13,000 rpm for 5 minutes to remove any particulates.

Liquid chromatography was performed on an Acquity Ultra Performance BEH C18, 1.7  $\mu\text{m}$ , 2.1 x 100 mm column. The gradient was 10-90% acetonitrile over 4.1 minutes with water and 0.1% formic acid as the second solvent. The flow rate was 0.6 mL/min. Ionization was performed by ESI with a Micromass LCT Premier TOF Mass Spectrometer in positive ionization V- mode. The formation of 7-chlorotryptophan was monitored by selected ion monitoring at  $m/z$  239. The identity of the product was characterized by co-elution with an authentic standard.<sup>1</sup> The formation of 7-chlorotryptamine was monitored by selected ion monitoring at  $m/z$  195. The halogenated tryptamine product was characterized by co-elution with an authentic 7-chlorotryptamine standard.<sup>14</sup> The following control experiments were performed for both RebH WT and RebH Y455W: (1) boiled enzyme control, where the enzyme was boiled to deactivate it, (2) no DTT control, where 20 mM DTT was omitted from the assay and (3) no substrate control, where tryamine was omitted from the assay mixture. Product formation

was not observed for RebH WT (Figure 3.15) or RebH Y455W (Figure 3.16) when active enzyme or reductant (20 mM DTT) was removed from the assay.

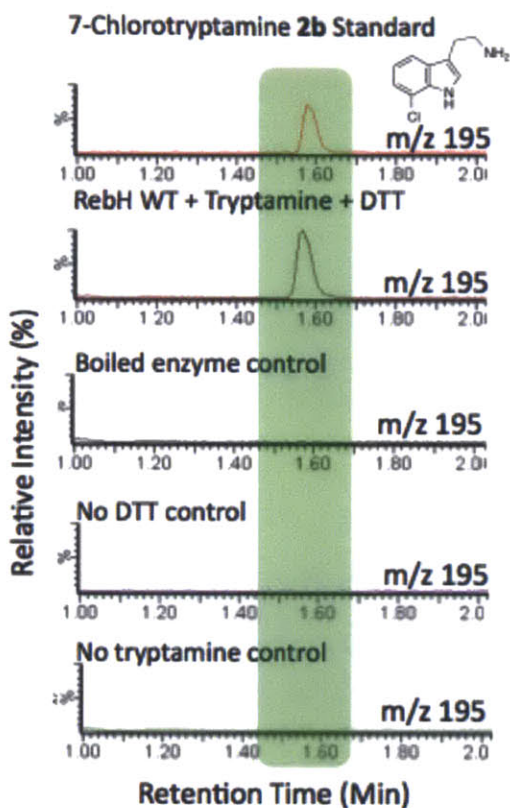


Figure 3.15: RebH WT tryptamine product co-elutes with an authentic 7-chlorotryptamine standard. The RebH tryptamine product is not observed in any of the negative controls (boiled enzyme, no DTT, no tryptamine).

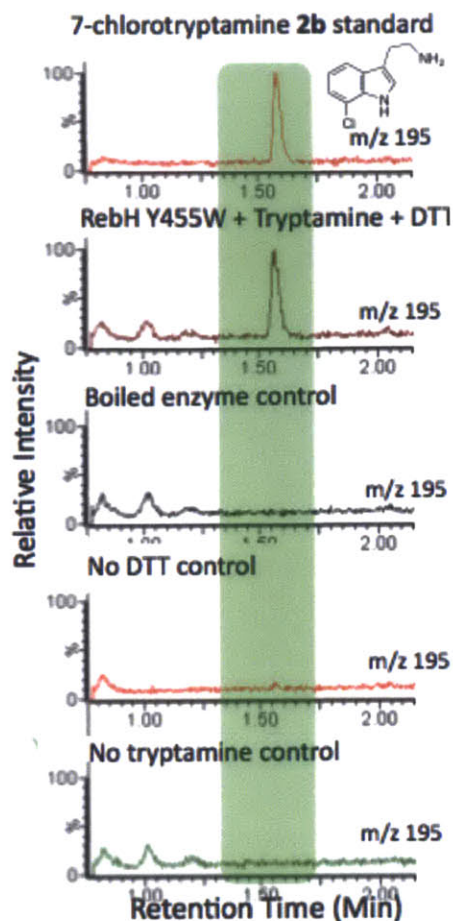


Figure 3.16: RebH Y455W chlorotryptamine product ( $m/z$  195) co-elutes with an authentic 7-chlorotryptamine standard. The RebH Y455W tryptamine product is not observed in any of the negative controls (boiled enzyme, no DTT, no tryptamine).

For more rigorous characterization, we also used authentic standards of each mono-chlorinated tryptamine isomer (2-chlorotryptamine, 4-chlorotryptamine, 5-chlorotryptamine, 6-chlorotryptamine and 7-chlorotryptamine). Chlorinated tryptamine isomers 4-chlorotryptamine and 6-chlorotryptamine were a gift from Elizabeth McCoy.<sup>8</sup> Weerawat Rungphan

provided authentic standards of chlorinated tryptamine isomers 5-chlorotryptamine and 7-chlorotryptamine that were previously described.<sup>14</sup> The 2-chlorotryptamine isomer was synthesized.<sup>21</sup> (<sup>1</sup>HNMR, 500 MHz, DMSO: 7.51 (1H, d, *J* = 7.5 Hz), 7.24 (1H, d, *J* = 7.5 Hz), 7.07 (1H, t, *J* = 7.5 Hz), 7.02 (1H, t, *J* = 7.5 Hz) 2.77 (multiplet s, 4H); <sup>13</sup>CNMR, 125 MHz, DMSO: 134.6, 126.7, 121.7, 121.1, 119.4, 117.8, 110.9, 39.2, 22.9). The RebH WT and RebH Y455W halogenated tryptamine product co-eluted exclusively with 7-chlorotryptamine.

#### *E. RebH Wild Type, RebH Y455W and RebH L113G Competition Assay Screen*

Slow conversion to product in vitro prevented accurate measurement of the steady-state enzyme kinetics parameters for wild-type (WT) RebH, RebH Y455W and RebH L113G. RebH, RebH Y455W and RebH L113G were assessed by competition assay with a mixture of L-tryptophan and tryptamine. These assays contained 1  $\mu$ M enzyme, 50  $\mu$ M FAD, 50 mM NaCl, 20 mM DTT, 20  $\mu$ M L-tryptophan and 20  $\mu$ M tryptamine, in 100 mM K<sub>2</sub>HPO<sub>4</sub> buffer (pH 7.2) in a final volume of 100  $\mu$ L. Assays were quenched and analyzed by LC-MS as described above in Methods Section D.

#### *F. RebH WT and RebH Y455W Selectivity at Different Tryptophan-to-Tryptamine Ratios*

Further competition assays with different substrate concentrations, as described in the main text, were also performed with RebH WT and RebH Y455W. Enzyme at a final concentration of 2  $\mu$ M (estimated by Bradford Assay)

was incubated with 20 mM DTT, 1 mM FAD, 50 mM NaCl, L-tryptophan, tryptamine and 100 mM K<sub>2</sub>HPO<sub>4</sub> buffer (pH 7.2) at 30 °C for 16 hours. Ratios of L-tryptophan:tryptamine were 500:500 μM, 500:1000 μM and 1000:500 μM. The final assay volume was 100 μL. Aliquots (25 μL) from enzyme assays were quenched with 975 μL methanol and then subjected to LC-MS analysis as outlined in Methods Section D. Conversion of substrate to product was monitored by simultaneous selected ion monitoring of both substrate and product masses (L-Tryptophan *m/z* 205), 7-chlorotryptophan (*m/z* 239), tryptamine (*m/z* 161) and 7-chlorotryptamine (*m/z* 195).

#### *G. The in planta Over-Expression of RebH Y455W and RebF*

Approximately 350 *C. roseus* seedlings were germinated aseptically on solid Gamborg's B5 media as described according to Methods Section B. The plant transformation vectors pCAMBIA1305.1-RebHY455W and pCAMBIA1305.1-RebF were separately transformed into *Agrobacterium rhizogenes* ATCC 15834 via electroporation according to manual specifications (BioRad electroporator). Liquid cultures of *A. rhizogenes* harboring pCAMBIA1305.1RebHY455W and *A. rhizogenes* harboring pCAMBIA1305.1RebF were mixed just prior to seedling infection.

All hairy root lines surviving hygromycin selection were macerated in methanol (between 10 and 40 mL/g of fresh weight hairy roots) using a mortar and pestle with 106 μm acid washed glass beads. Crude product mixtures were filtered through a 0.2 μm cellulose acetate membrane (VWR). Prior to analysis,

samples were centrifuged at 13,000 rpm for 5 minutes to remove any particulates. Alkaloid methanolic extracts (60  $\mu$ L) from hairy root tissues were added to 700  $\mu$ L HPLC grade methanol and subjected to LC-MS analysis as follows. Liquid chromatography was performed on a single quadrupole Agilent 1100 HPLC-MS. A Phenomenex C18, 2.0  $\mu$ m, 2 x 50 mm column was employed. The gradient was 10 to 90% acetonitrile over 8 minutes with water and 0.1% formic acid as the second solvent. The solvent composition was returned to 10% acetonitrile by 8.1 minutes, and the column was self-equilibrated at initial run conditions until 12 minutes. The flow rate was 0.4 mL/min. Ionization was performed by an 1100 MSD Mass Spectrometer in positive ionization mode. The following specifications were employed: 4000 V capillary voltage, 350 °C drying gas temperature, 30 psig nebulizer pressure. Single ion monitoring, 50% *m/z* 239 for 7-chlorotryptophan and 50% *m/z* 359 for 12-chloro-19,20-dihydroakuammicine, was used. Each run injected 20  $\mu$ L.

No halogenated alkaloids (either 12-chloro-19,20-dihydroakuammicine or 7-chlorotryptophan) were detected in negative control lines [wild type hairy roots, hairy roots with knocked down tryptophan decarboxylase (TDCi), pCAMBIA1305.1 empty vector]. TDCi lines were a gift from Weerawat Runguphan.<sup>14</sup> An authentic standard of 12-chloro-19,20-dihydroakuammicine was previously reported.<sup>14</sup> Quantities of 12-chloro-19,20-dihydroakuammicine and 7-chlorotryptophan (averaged across 3 subcultures) from two representative lines harboring both RebHY455W and RebF are shown in Table 3.2 and Table 3.3.



Two lines producing 12-chloro-19,20-dihydroakuammicine **5b** (Lines 13 and 74) were subjected to isolation of genomic DNA following manual specifications (Qiagen DNeasy kit). Both lines contained both RebH Y455W and RebF. A 660 basepair amplicon was amplified in RebH Y455W. The primers for the RebH Y455W amplicon are shown in Table 3.7. Full length RebF (503 basepairs) was amplified. The primers for the full-length RebF amplicon (513 basepairs) are provided in Weerawat *et al.*<sup>14</sup> These lines co-migrated with the corresponding amplicons from pET28aRebHY455W and pET28aRebF. The negative control lines (TDCi) did not contain the amplicon (Figure 3.17). Exposure was set at 040 ms. Contrast was set at 1.00.

|                        |                              |
|------------------------|------------------------------|
| RebHY455W_gDNA_forward | 5'-GTCTTCGATGCCGACCTCTTC-3'  |
| RebHY455W_gDNA_reverse | 5'-GTACATGTCGATCTTCTCCTGC-3' |

Table 3.7: Primers for RebH Y455W amplicon from genomic DNA (gDNA) extraction.

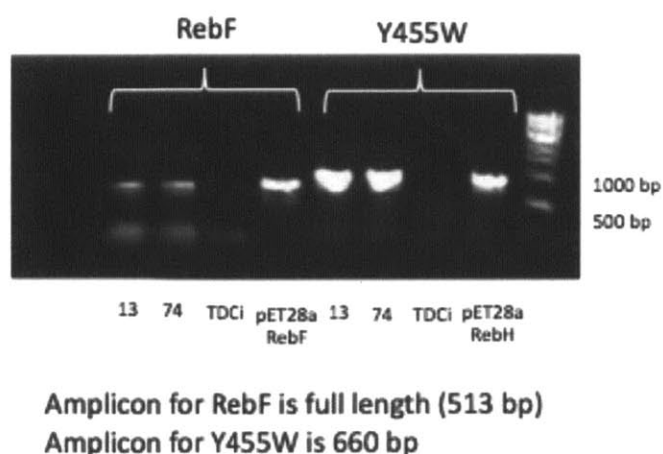


Figure 3.17: Agarose gel (1%) of genomic DNA analysis of RebH Y455W and RebF lines. RebF and RebH Y455W amplicons are shown in root lines and in plasmid (positive control), but not in the TDCi line (negative control).

#### *H. Transient Expression of RebH Y455W and RebF*

The plasmids pCAMBIA1305.1-RebHWT, pCAMBIA1305.1-RebHY455W and pCAMBIA1305.1-RebF were each transformed separately into both *Agrobacterium tumefaciens* strain GV 3101 and *A. tumefaciens* LB 4404 for a total of 6 bacterial transformations. Liquid cultures (5 mL) of *A. tumefaciens* containing the plasmid of interest were grown in Yeast Extract and Mannitol (YEM) media supplemented with kanamycin at 30 °C for 48 hours just prior to infection. Since plasmids were transformed into separate *A. tumefaciens* strains, the cultures harboring them were mixed (either pCAMBIA1305.1-RebHWT and pCAMBIA1305.1-RebF or pCAMBIA1305.1-RebHY455W and pCAMBIA1305.1-RebF) just prior to plant transformation. (1) pCAMBIA1305.1-RebHWT and pCAMBIA1305.1-RebF and (2) pCAMBIA1305.1-RebHY455W and pCAMBIA1305.1-RebF were transiently expressed *in planta* via two different methods (1) syringe injection (without a needle) and (2) vacuum infiltration for each gene combination. Transformations were conducted also by testing both *A. tumefaciens* strains (GV3101 and LB 4404) using each infiltration method (vacuum infiltration or syringe injection). All transiently transformed mature leaves, each day for 3 days after transformation, were macerated in methanol (between 10 and 40 mL/g of fresh weight hairy roots) using a mortar and pestle with 106 µm acid washed glass beads. Crude product mixtures were filtered through a 0.2-µm cellulose acetate membrane (VWR). Alkaloid methanolic extracts (30 µL) from mature leaves were diluted with 700 µL HPLC grade

methanol and subjected to LC-MS analysis as outlined in Methods Section D. Each run used 2  $\mu$ L of extract. No chlorinated alkaloids were detected in any transient expressions under these conditions.

## V. References

1. Kutchan, T.M., Alkaloid Biosynthesis – The Basis for Metabolic Engineering of Medicinal Plants. *The Plant Cell* 1995, 7, 1059-1070.
2. Harvey, A.L., Natural Products in Drug Discovery. *Drug Discovery Today* 2008, 19-20, 894-901.
3. Patterson, L. and Anderson, E.A., The Renaissance of Natural Products as Drug Candidates. *Science* 2005, 310, 451-453.
4. Thomas, C.J.; Fluorinated Natural Products with Clinical Significance. *Current Topics in Medicinal Chemistry* 2006, 6, 1529-1543.
5. Blasiak, L.C. and Drennan, C.L., Structural Perspective on Enzymatic Halogenation. *Acc. Chem. Research* 2009, 42(1), 147-155.
6. Eustáquio, A.S.; O'Hagan, D. and Moore, B.S., Engineering Fluorometabolite Production: Fluorinase Expression in *Salinospora tropica* Yields Fluorosalinosporamide. *Nat. Prod.* 2010, 73, 378-382.
7. Josephson, K.; Hartman, M.C.T. and Szostak, J.W., Ribosomal Synthesis of Unnatural Peptides. *J. Am. Chem. Soc.* 2005, 127, 11727-35.
8. McCoy, E. and O'Connor, S.E., Directed Biosynthesis of Alkaloid Analogs in the Medicinal Plant *Catharanthus roseus*. *J. Am. Chem. Soc.* 2006, 128, 14276-77.
9. Jacobsen, J.R.; Hutchinson, C.R.; Cane, D.E. and Khosla, C., Precursor-directed Biosynthesis of Erythromycin Analogs by an Engineered Polyketide Synthase. *Science* 1997, 277, 367-369.
10. Guillet, G.; Poupard, J.; Basucro, J. and de Luca, V., Expression of Tryptophan Decarboxylase and Tyrosine Decarboxylase Genes in Tobacco Results in Altered Biochemical and Physiological Phenotypes. *Plant Physiol.* 2000, 122(3), 933-944.
11. Maresh, J.J.; Giddings, L.A.; Friedrich, A.; Loris, E.A.; Panjikar, S.; Trout, B.L.; Stöckigt, J.; Peters, B.; O'Connor, S.E., Strictosidine Synthase: Mechanism of a Pictet-Spengler Catalyzing Enzyme. *J. Am. Chem. Soc.* 2008, 130(2), 710-23.
12. Dewick, P.M., *Medicinal Products, A Biosynthetic Approach, 2<sup>nd</sup> Edition* John Wiley and Sons, Ltd. New York, 2003.
13. Yeh, E.; Gameau, S.; Walsh, C.T., Robust In Vitro Activity of RebF and RebH, a two component reductase halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl Acad. Sci USA* 2005, 102(11), 3960-5.
14. Runguphan, W.; Qu, X. and O'Connor, S.E., Integrating Carbon-Halogen

- Bond Formation into Medicinal Plant Metabolism. *Nature* 2010, 468 (7322), 461-4.
15. [A] Yeh, E.; Blasiak, L.C.; Koglin, A.; Drennan, C.L. and Walsh, C.T., Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-dependent Halogenases. *Biochemistry* 2007, 46 (5), 1284-92. [B] Bitto, E.; Huang, Y.; Bingman, C.A.; Singh, S.; Thorson, J.S. and Phillips, G.N., The Structure of Flavin-dependent Tryptophan 7-Halogenase RebH. *Proteins* 2008, 70, 289-93.
  16. Hölzer, M.; Burd, W.; Reißig, U.W. and van Pée, K-H., Substrate Specificity and Regioselectivity of Tryptophan 7-halogenase from *Pseudomonas fluorescens* BL915. *Adv. Synth. Cat.* 2001, 343, 591-595.
  17. Onaka, H., Taniguchi, S.; Igarashi, Y.; Furumai, T., Characterization of the Biosynthetic Gene Cluster of Rebeccamycin from *Lechevaliera aerocolonigenes* ATCC 39243. *Biosci. Biotechnol. Biochem.* 2003, 67(1), 127-138.
  18. [A] *Organometallics in Synthesis: A manual 2<sup>nd</sup> Edition*. Smith, K.; Schlosser, M.; Eds.; Wiley: Chichester, England 2002; 512. [B] Deb Roy, A.; Grünschow, S.; Cairns, N.; Goss, R.J.M., Gene Expression Enabling Synthetic Diversification of Natural Products: Chemogenetic Generation of Pacidamycin Analogs. *J. Am. Chem. Soc.* 2010, 132, 12243-5. [C] Runguphan, W. and O'Connor, S.E., Diversification of Monoterpene Indole Alkaloid Analogs through Cross-Coupling. *Org. Lett.* Publication Date: 28 May 2013 (web). DOI: 10.1021/ol401179k
  19. Lang, A.; Polnic, S.; Nicke, T.; William, P.; Patallo, E.P.; Naismith, J.H. and van Pée, K-H., Changing the Regioselectivity of the Tryptophan 7-halogenase PrnA by Site-directed Mutagenesis. *Angew. Chem. Int. Ed.* 2011, 50, 2951-2953.
  20. Payne J.T.; Andorfer, M.C., and Lewis J.C., Regioselectivity of Arene Halogenation using the FAD-dependent Halogenase RebH. *Angew Chem. Int. Ed. Engl.* 2013, 52(20), 5271-4.
  21. Miyake, F.Y.; Yakushijin, K. and Hone, D. A., Preparation and Synthetic Applications of 2-Halotryptamines: Synthesis of Elacomine and Isoetacomine. *Org. Lett.* 2004, 6, 711-713.
  22. Runguphan, W.; Maresh J.J.; and O'Connor, S.E., Silencing of Tryptamine Biosynthesis for Production of Nonnatural Alkaloids in Plant Culture. *Proc. Natl. Acad. Sci. USA.* 2009, 106(33), 13673-8.

## VI. Acknowledgments

Dr. Ezekiel Nims performed all transient expression experiments and half of the stable transformation experiments. He is listed as a co-author on the manuscript.

We give special thanks Bettina M. Ruff for assistance in preparing and

characterizing an authentic standard of 2-chlorotryptamine. Additionally, we thank Dr. Weerawat Runguphan and Dr. Elizabeth McCoy for providing authentic standards of 4-chlorotryptamine, 5-chlorotryptamine, 6-chlorotryptamine and 7-chlorotryptamine. We gratefully acknowledge Dr. Xudong Qu for providing the pET28a-RebH expression construct. We also thank Dr. Weerawat Runguphan for TDCi root lines. Thank you to Dr. Lionel Hill for assistance with the single-quad LC-MS. Lastly, we thank Dr. Lesley-Ann Giddings for her critical reading of the manuscript.

## Chapter 4

Redesign of a Dioxygenase Involved in Morphine Biosynthesis

Part of this chapter is published as a communication in

Runguphan W\*, Glenn WS\* and O'Connor SE. Chem Biol. 2012  
Jun 22;19(6):674-8. (\*equal contribution)

## I. Introduction

Opium poppy (*Papaver somniferum*) produces an array of medicinally important benzyloisoquinoline alkaloids, including the analgesics codeine **10** and morphine **11** (Figure 4.1).<sup>1,2</sup> The biosynthesis of these alkaloids commences with the Pictet-Spengler condensation of dopamine **1** and 4-hydroxyphenylacetaldehyde **2** to form (S)-norcoclaurine **3**, which is further modified to form (S)-reticuline **4**, the pivotal biosynthetic intermediate of all benzyloisoquinoline alkaloids. (S)-reticuline **4** is subsequently converted to thebaine **5**, the intermediate at the entry point of the morphinan alkaloid pathway.

Two biosynthetic routes have been proposed for the conversion of thebaine **5** to morphine **11**.<sup>3</sup> In the first route (route A; see Figure 4.1), thebaine 6-O-demethylase (*PsT6ODM*) demethylates thebaine **5** at the 6 position to form neopinone **7**, which spontaneously isomerizes to form codeinone **8**. Codeinone **8** is then enzymatically reduced to yield codeine **10**. Codeine O-demethylase (*PsCODM*) demethylates codeine **10** at the 3 position to yield morphine **11**. Alternatively, in the second route (route B; see Figure 4.1), *PsCODM* demethylates thebaine **5** at the 3 position to form oripavine **6**. *PsT6ODM* catalyzes the second demethylation of oripavine **6**—this time at the 6 position—to form morphinone **9**, which is then reduced to form morphine **11**. Morphinan alkaloids thebaine **5**, oripavine **6**, codeine **10**, and morphine **11** all accumulate in opium poppy, suggesting that both routes are operative *in vivo*.<sup>1,2</sup>

The discovery of O-demethylases *PsCODM* and *PsT6ODM* completed the genetic characterization of the morphinan pathway.<sup>3</sup> These genes were

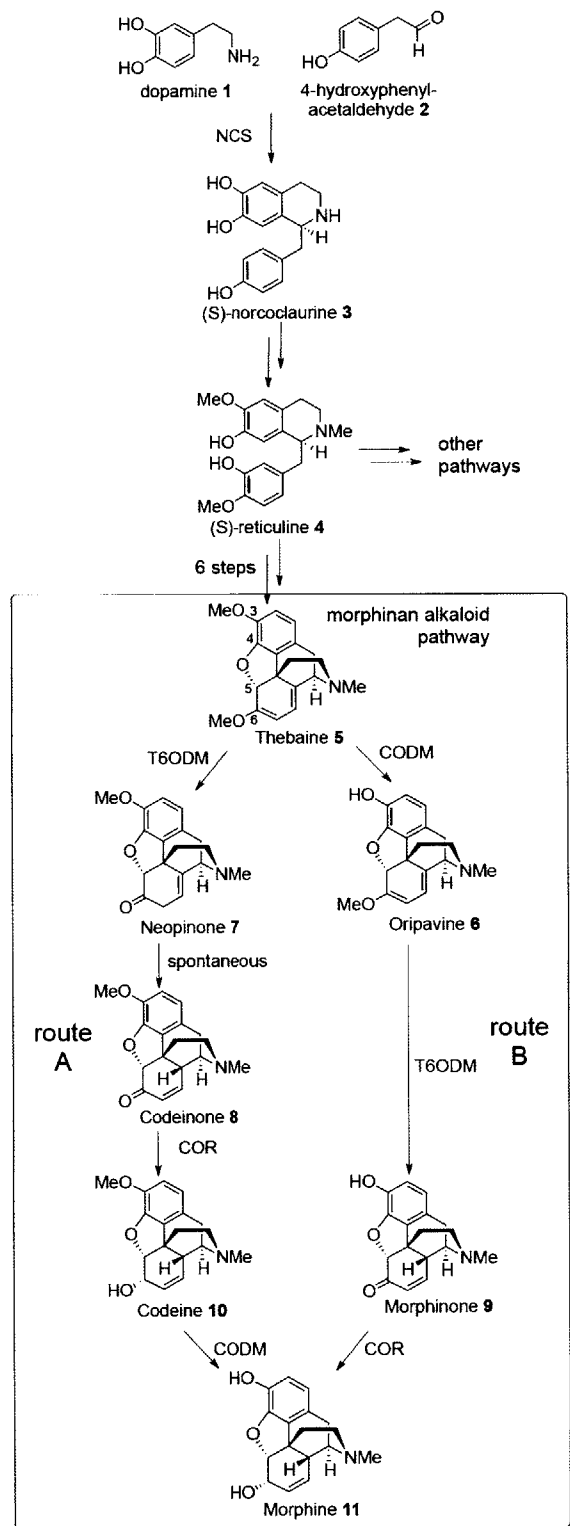


Figure 4.1: Biosynthesis of benzyisoquinoline alkaloids in *P. somniferum*. The biosynthetic pathways leading to morphine 11 via morphine 9 and codeine 10.



discovered through a functional genomics approach by differentially comparing transcripts found in native opium poppy (plants that accumulate morphine) versus poppy mutant varieties that are devoid of morphinan alkaloids.<sup>4</sup> The differentially expressed genes (i.e. genes expressed in morphinan alkaloid-containing plants that were not expressed in varieties lacking morphinan alkaloids) could then be assayed for the desired demethylase activity.<sup>4,5</sup> Surprisingly, non-heme iron [Fe(II)]- and  $\alpha$ -ketoglutarate-dependent dioxygenases (*PsCODM* and *PsT60DM*), were identified from this study. Notably, *PsCODM* and *PsT60DM* are the only members of the non-heme iron- and  $\alpha$ -ketoglutarate-dependent dioxygenase family capable of catalyzing O-demethylation; all others are P450s.<sup>4</sup>

The morphinan pathway provides a highly attractive reengineering target at the enzyme and pathway levels because all late pathway genes are known. Here we demonstrate how exploiting natural enzyme variation through systematically mixing and matching the non-conserved amino acid regions of these two recently discovered demethylases (*PsCODM* and *PsT60DM*) has led to an enzyme with new specificity: a *PsCODM* mutant that is highly selective for codeine. The unique selectivity of the reengineered demethylase enzyme may allow us to explore how closing the metabolic valve to route B and redirecting substrate exclusively through route A (Figure 4.1) will impact downstream product yields in this branched natural product pathway.

## II. Results and Discussion

A clustal alignment of *PsCODM*, *PsT6ODM*, and *PsDIOX2* (a *P. somniferum* 2-oxoglutarate/Fe(II)-dependent dioxygenase for which the native substrate has not yet been identified, but which is known to be able to O-demethylate a number of protoberberine alkaloids)<sup>4</sup> revealed five specific regions where these dioxygenases differ significantly at the amino acid level: A<sub>1</sub> (residues 145–149), A<sub>2</sub> (residues 150–152), B (residues 334–336), C<sub>1</sub> (residues 338–342), and C<sub>2</sub> (residues 343–352) (Figure 4.2). Lacking a crystal structure of a *P. somniferum* dioxygenase, we built a homology model of *PsCODM* based on the crystal structure of *Arabidopsis thaliana* anthocyanidin synthase (AtANS), a structurally characterized 2-oxoglutarate/Fe(II)-dependent dioxygenase with moderate amino acid sequence similarity to *PsCODM* (32% identity at the amino acid level).<sup>6</sup> We employed SWISS-MODEL, an automated protein homology-modeling server, to generate the model (Figure 4.3).<sup>7</sup> Key residues likely to be involved in Fe(II)/2-oxoglutarate binding are highlighted in the (Figure 4.2). The five regions—grouped as follows, A<sub>1</sub> and A<sub>2</sub> (residues 145–152); B (residues 334–336); and C<sub>1</sub> and C<sub>2</sub> (residues 338–346)—were then mapped on to this homology model, where they appeared to be located proximal to the anthocyanin binding site (Figure 4.3).

```

DIOX1      MEKAKLMKLGNGMEIPSVQELAKLTLAEIPSRVVCANENLLLPMGASVINDHETIPVIDI 60
DIOX2      METAKLMKLGNGMSIPSVQELAKLTLAEIPSRVICTVENLQLPVGASVIDDHETVPVIDI 60
DIOX3      METPILIKLGNGLSIPSVQELAKLTLAEIPSRYTCTGESPLNNIGASVTDD-ETVPVIDL 59
          *.. :*****.***** : * . :**** :* *****:

DIOX1      ENLLSPEPIIGKLELDRLHFACKEWGFFQVNVHGVDASLVDSVKSEIQGFFNLSMDEKTK 120
DIOX2      ENLISSEPVTEKLELDRLHSACKEWGFFQVNVHGVDTSLVDNVKSIDIQGFFNLSMNEKIK 120
DIOX3      QNLLSPEPVVGKLELDKLSACKEWGFFQLVNHGVDALLMDNIKSEIKGFFNLPMEKTK 119
          :****.*: *****.***** :*.:*****:**** *

          Region A
DIOX1      YEQEDGDVEGFGQGFIESEDQTLDWADIFMIFTLPLHLRKPFLFSKLPVPLRETIESYSS 180
DIOX2      YGQKDGDEVEGFGQAFVASEDQTLDWADIFMILTLPPLHLRKPFLFSKLPVPLRETIESYSS 180
DIOX3      YGQQDGDVEGFGQPYIESEDQRLDWTEVFSMLSLPLHLRKPFLFPPELPLPFRETLESYLS 179
          * :****.* : : **** * : :***** :***** :*.:*****:**** *
          ♦
DIOX1      EMKKLSMVLFNKMEKALQVQAAEIKGMSEVVIDGTQAMRMNYYPPCPQPNLAIGLTSBSD 240
DIOX2      EMKKLSMVLFEKMEKALQVQAVEIKEISEVFKDMTQVMRMNYYPPCPQPELAIGLTPHSD 240
DIOX3      KMKKLSLVVVFEMLEKSLQL--VEIKGMTDLFEDGLQIMRMNYYPPCPRPELVGLTSBSD 237
          :***** :*.: ***** .*** :**** * * .***** :*.:*****:**** *
          ♣
DIOX1      FGGLTILLQINEVEGLQIKREGTWISVKPLPNAFVNVVGDILEIMTNGIYHSVDHRAVVN 300
DIOX2      FGGLTILLQLNEVEGLQIKNEGRWISVKPLPNAFVNVVGDVLEIMTNGMYRSVDHRAVVN 300
DIOX3      FSGLTILLQLNEVEGLQIRKEERWISIKPLPDAFIVNVGDILEIMTNGIYRSVEHRAVVN 297
          *.:***** :* .***** :***** :***** :*.:*****:**** *
          ♦ ♦
          Region B Region C
DIOX1      STNERLSIATFFHDPSPLESVIGPISSLITPETPALFKSGSTYGDLVEECKTRKLDGKSFLD 360
DIOX2      STKERLSIATFFHDPNLESEIGPISSLITPNTPALFRSGSTYGELVEEFHSRKLKGKSFLD 360
DIOX3      STKERLSIATFFHDSKLESEIGPISSLVTPETPALFKRG-RYEDILKENLSRKLDGKSFLD 356
          *.:***** :* .***** :***** :* * :**** :*****

DIOX1      SMRI- 364
DIOX2      SMRM- 364
DIOX3      YMRM- 360
          **:
```

Figure 4.2: Sequence alignment of *Ps*DIOX1 (*Ps*T6ODM), *Ps*DIOX2 and *Ps*DIOX3 (*Ps*CODM). ♦ indicates residues that have been proposed to be important in 2-oxoglutarate binding. ♣ indicates residues that have been proposed to be important in coordinating Fe(II).

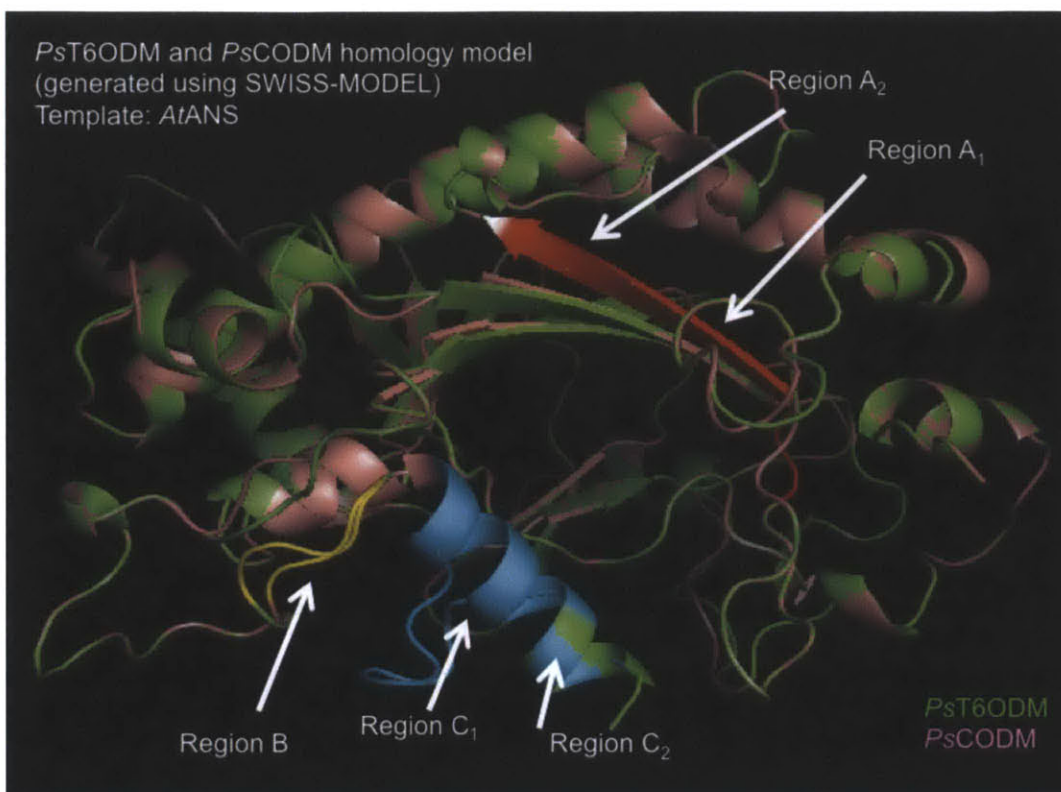


Figure 4.3: Homology Models of *PsT6ODM* and *PsCODM* Based on the Crystal Structure of *AtANS* *PsT6ODM* is indicated in green, and *PsCODM* is indicated in pink. The models were created using SWISS-MODEL. Regions A1 and A2 (residues 145–152) are shown in red, B (residues 334–336) is in yellow, and C1 and C2 (residues 338–346) are in blue.

We hypothesized that the residues' proximity to the primary substrate binding site implicated their involvement in dictating regioselectivity, which could be readily reengineered through mutagenesis. Specifically, we envisioned using site-directed mutagenesis to reverse the regioselectivity of *PsCODM*, which demethylates the 3 position of morphinan alkaloids, and *PsT6ODM*, which demethylates the 6 position of morphinan alkaloids. Initial protein expression screening revealed that all *PsT6ODM* mutants are expressed at low levels. Therefore, in this study, we focused our efforts on developing engineered

*PsCODM* enzymes. We systematically replaced the native *PsCODM* sequence with the corresponding sequence from *PsT6ODM* at the five nonconserved regions using standard site-directed mutagenesis. In total, we constructed 16 *PsCODM* mutants (Table 4.1).

|     | A <sub>1</sub>   | A <sub>2</sub>   | B  | C <sub>1</sub>   | C <sub>2</sub>   | (wild type <i>PsCODM</i> ) |
|-----|------------------|------------------|----|------------------|------------------|----------------------------|
| 1)  | A <sub>1</sub> * | A <sub>2</sub> * | B  | C <sub>1</sub> * | C <sub>2</sub> * |                            |
| 2)  | A <sub>1</sub> * | A <sub>2</sub>   | B  | C <sub>1</sub> * | C <sub>2</sub> * |                            |
| 3)  | A <sub>1</sub> * | A <sub>2</sub> * | B  | C <sub>1</sub>   | C <sub>2</sub> * |                            |
| 4)  | A <sub>1</sub> * | A <sub>2</sub> * | B  | C <sub>1</sub> * | C <sub>2</sub>   |                            |
| 5)  | A <sub>1</sub> * | A <sub>2</sub> * | B  | C <sub>1</sub>   | C <sub>2</sub>   |                            |
| 6)  | A <sub>1</sub> * | A <sub>2</sub>   | B  | C <sub>1</sub>   | C <sub>2</sub> * |                            |
| 7)  | A <sub>1</sub> * | A <sub>2</sub>   | B  | C <sub>1</sub> * | C <sub>2</sub>   |                            |
| 8)  | A <sub>1</sub> * | A <sub>2</sub>   | B  | C <sub>1</sub>   | C <sub>2</sub>   |                            |
| 9)  | A <sub>1</sub>   | A <sub>2</sub> * | B  | C <sub>1</sub> * | C <sub>2</sub> * |                            |
| 10) | A <sub>1</sub>   | A <sub>2</sub>   | B  | C <sub>1</sub> * | C <sub>2</sub> * |                            |
| 11) | A <sub>1</sub>   | A <sub>2</sub> * | B  | C <sub>1</sub>   | C <sub>2</sub> * |                            |
| 12) | A <sub>1</sub>   | A <sub>2</sub> * | B  | C <sub>1</sub> * | C <sub>2</sub>   |                            |
| 13) | A <sub>1</sub>   | A <sub>2</sub> * | B  | C <sub>1</sub>   | C <sub>2</sub>   |                            |
| 14) | A <sub>1</sub>   | A <sub>2</sub>   | B  | C <sub>1</sub>   | C <sub>2</sub> * |                            |
| 15) | A <sub>1</sub>   | A <sub>2</sub>   | B  | C <sub>1</sub> * | C <sub>2</sub>   |                            |
| 16) | A <sub>1</sub>   | A <sub>2</sub>   | B* | C <sub>1</sub>   | C <sub>2</sub>   |                            |

**Table 4.1:** *PsCODM* mutants that were constructed in this study. \* indicates that the residues in that region were mutated to those of *PsT6ODM*

Table 4.2: Primers and templates for site-directed mutagenesis.

| Plasmid  | Template used in SDM  | Primers used in SDM  |
|--|---|--|
| pQECODM A <sub>1</sub> * A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>    | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>     | <b>Forward:</b> 5'-ggaccaaagacttgattgggctgatgtgttagcatgtaagtc-3'<br><b>Reverse:</b> 5'-gacttaacatgctaaacacatcagcccaatcaagctttggcc-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-ccaaagacttgattgggctgatataatgcatgtaagcttctc-3'<br><b>Reverse:</b> 5'-gaggaagacttaacatgctaaatatacagcccaatcaagctttgg-3'                 |
| pQECODM A <sub>1</sub> * A <sub>2</sub> * B C <sub>1</sub> C <sub>2</sub>  | pQECODM A <sub>1</sub> * A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>   | <b>Forward:</b> 5'-aaagacttgattgggctgatataatgcatgtaagcttctcctcatt-3'<br><b>Reverse:</b> 5'-aatggagaggaagacttaacatcataatatacagcccaatcaagcttt-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-cttgattgggctgatataatgcatgctcacttctcctcattagg-3'<br><b>Reverse:</b> 5'-tccttaaatggagaggaagtgaaacatcataatatacagcccaatcaag-3' |
| pQECODM A <sub>1</sub> A <sub>2</sub> * B C <sub>1</sub> C <sub>2</sub>    | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>     | <b>Forward:</b> 5'-caaagacttgattggactgaagtgttatgatgttcacttctcctcattaaaggaagcc-3'<br><b>Reverse:</b> 5'-ggcttccttaaatggagaggaagtgaaacatcataaacacttcagccaatcaagctttg-3'  |
| pQECODM A <sub>1</sub> A <sub>2</sub> B* C <sub>1</sub> C <sub>2</sub>     | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>     | <b>Forward:</b> 5'-gacacctgctttgtcaaaagtgatctacatagaggatatttgaagg-3'<br><b>Reverse:</b> 5'-ccttcaaaatcctcatatgtagatccactttgaacaaagcaggtg-3'  |
| pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub>    | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>     | <b>Forward:</b> 5'-ttgttcaaaagaggtaggtatgggatctttgaaggaaaatcttcaagg-3'<br><b>Reverse:</b> 5'-ccttgaagatttcttcaaaagatccccatacctcctttgaacaaa-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-caaaagaggtaggtatgggatcttggaggaaaatcttcaagg-3'<br><b>Reverse:</b> 5'-tcctgaaagatttctccacaagatccccatacctcctttg-3'             |
| pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub> *  | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub>   | <b>Forward:</b> 5'-gtatgggatcttggaggaaatgtcttcaaggaaagcttga-3'<br><b>Reverse:</b> 5'-tcaagctccttgaaagacattcctccacaagatccccatac-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-agagtaggtatgggatcttggaggaaatgaagacgaggaagcttgatggaaa-3'<br><b>Reverse:</b> 5'-ttccatcaagcttctctgtctacattcctccacaagatccccatacctcct-3'    |
| pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub> *    | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>     | <b>Forward:</b> 5'-gacacctgctttgtcaaaagaggtaggtatgaggatatttgaaggaaatgaagacgaggaagcttgatgaaaaatcttct-3'<br><b>Reverse:</b> 5'-agaatgatttccatcaagcttctctgtctacattcctcaaaatcctcatacctcctttgaacaaagcaggtg-3'   |
| pQECODM A <sub>1</sub> * A <sub>2</sub> BC <sub>1</sub> * C <sub>2</sub> * | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub> * | <b>Forward:</b> 5'-ggaccaaagacttgattgggctgatgtgttagcatgtaagtc-3'<br><b>Reverse:</b> 5'-gacttaacatgctaaacacatcagcccaatcaagctttggcc-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-ccaaagacttgattgggctgatataatgcatgtaagcttctc-3'<br><b>Reverse:</b> 5'-gaggaagacttaacatgctaaatatacagcccaatcaagctttgg-3'                 |

Table 4.2 (continued): Primers and templates for site-directed mutagenesis

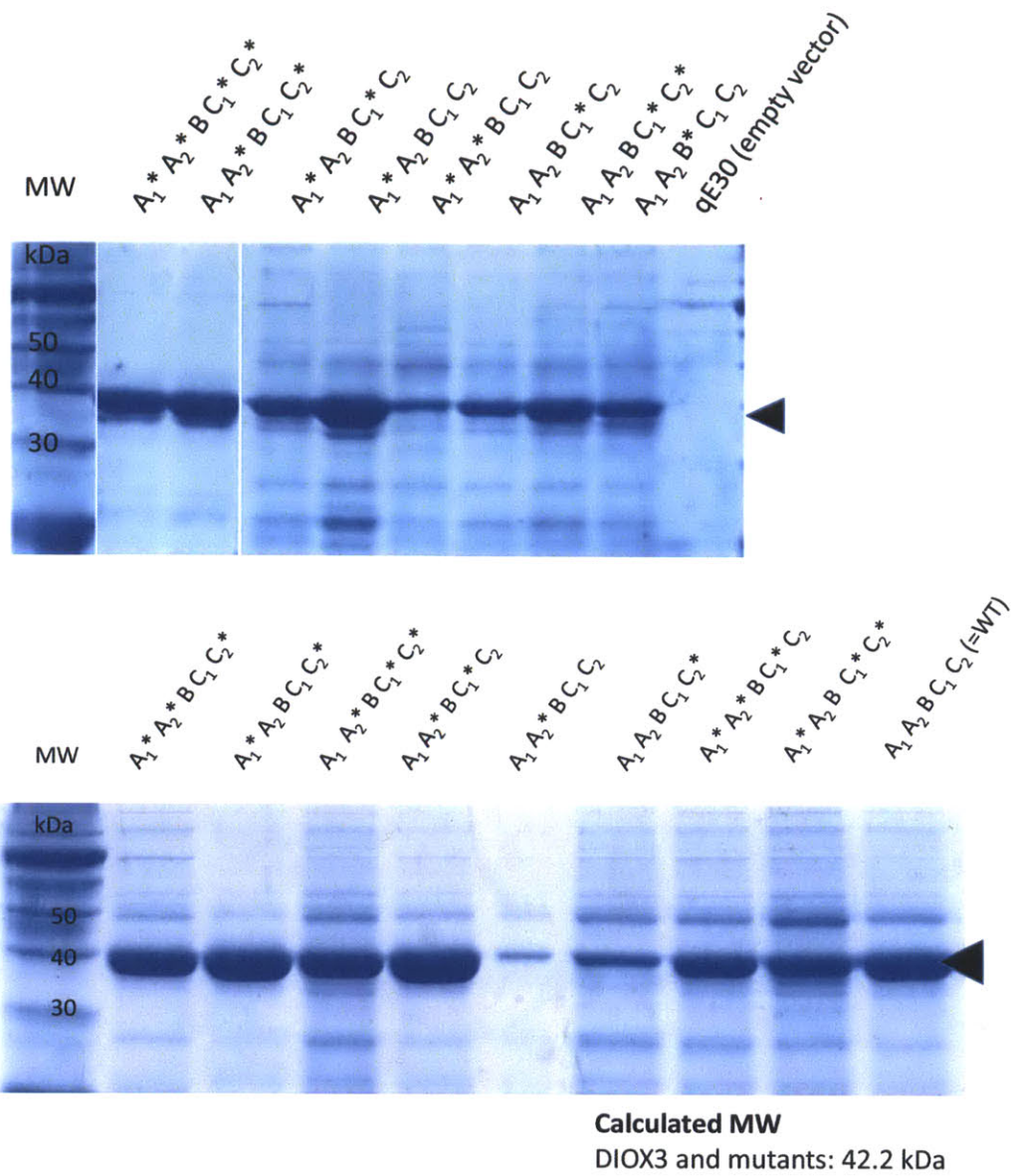
| Plasmid   | Template used in SDM   | Primers used in SDM   |
|---|--|---|
| pQECODM A <sub>1</sub> *A <sub>2</sub> * BC <sub>1</sub> *C <sub>2</sub> *  | pQECODM A <sub>1</sub> * A <sub>2</sub> BC <sub>1</sub> * C <sub>2</sub> * | <b>Forward:</b> 5'-aaagacttgattgggctgatatattatgatgtaagtcttctctccatt-3'<br><b>Reverse:</b> 5'-aatggagaggaagacttaacatcataaataatcagcccaatcaagctttt-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-cttgattgggctgatatattatgatgttcaactcttctctccatttaagga-3'<br><b>Reverse:</b> 5'-tccttaaatggagaggaagagtgaacatcataaataatcagcccaatcaag-3' |
| pQECODM A <sub>1</sub> A <sub>2</sub> * B C <sub>1</sub> * C <sub>2</sub>   | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub>    | <b>Forward:</b> 5'-caaagacttgattggactgaagtgtttatgatgttcaactcttctctccatttaaggaagcc-3'<br><b>Reverse:</b> 5'-ggcttccttaaatggagaggaagagtgaacatcataaacacttcagtccaatcaagcttttg-3'  |
| pQECODM A <sub>1</sub> * A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub> *   | pQECODM A <sub>1</sub> * A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>    | <b>Forward:</b> 5'-gacacctgctttgttcaaaagaggttaggtatgaggatattttgaa<br>ggaatgtaagacgaggaagcttgatggaaaatcatttct-3'<br><b>Reverse:</b> 5'-agaaatgattttccatcaagcttctctcttcttaccctctcaa<br>aatactctacatacctacctcttttgaacaaagcaggtgtc-3'   |
| pQECODM A <sub>1</sub> A <sub>2</sub> * B C <sub>1</sub> C <sub>2</sub> *   | pQECODM A <sub>1</sub> A <sub>2</sub> * B C <sub>1</sub> C <sub>2</sub>    | <b>Forward:</b> 5'-gacacctgctttgttcaaaagaggttaggtatgaggatattttgaa<br>ggaatgtaagacgaggaagcttgatggaaaatcatttct-3'<br><b>Reverse:</b> 5'-agaaatgattttccatcaagcttctctcttcttaccctctcaa<br>aatactctacatacctacctcttttgaacaaagcaggtgtc-3'   |
| pQECODM A <sub>1</sub> * A <sub>2</sub> * BC <sub>1</sub> C <sub>2</sub> *  | pQECODM A <sub>1</sub> * A <sub>2</sub> * B C <sub>1</sub> C <sub>2</sub>  | <b>Forward:</b> 5'-gacacctgctttgttcaaaagaggttaggtatgaggatattttgaa<br>ggaatgtaagacgaggaagcttgatggaaaatcatttct-3'<br><b>Reverse:</b> 5'-agaaatgattttccatcaagcttctctcttcttaccctctcaa<br>aatactctacatacctacctcttttgaacaaagcaggtgtc-3'   |
| pQECODM A <sub>1</sub> A <sub>2</sub> * B C <sub>1</sub> * C <sub>2</sub> * | pQECODM A <sub>1</sub> A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub> *  | <b>Forward:</b> 5'-caaagacttgattggactgaagtgtttatgatgttcaactcttctctccatttaaggaagcc-3'<br><b>Reverse:</b> 5'-ggcttccttaaatggagaggaagagtgaacatcataaacacttcagtccaatcaagcttttg-3'  |
| pQECODM A <sub>1</sub> * A <sub>2</sub> * BC <sub>1</sub> * C <sub>2</sub>  | pQECODM A <sub>1</sub> * A <sub>2</sub> * B C <sub>1</sub> C <sub>2</sub>  | <b>Forward:</b> 5'-tttgtcaaaagaggttaggtatggggatctttgaaaggaaaatcttcaagg-3'<br><b>Reverse:</b> 5'-ccttgaaagattttcctcaaaagatccccatacctacctcttttgaacaaa-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-caaaagaggttaggtatggggatctttgaaaggaaaatcttcaagga-3'<br><b>Reverse:</b> 5'-tccttgaaagattttcctccacaagatccccatacctacctcttttg-3'     |
| pQECODM A <sub>1</sub> * A <sub>2</sub> B C <sub>1</sub> * C <sub>2</sub>   | pQECODM A <sub>1</sub> * A <sub>2</sub> B C <sub>1</sub> C <sub>2</sub>    | <b>Forward:</b> 5'-tttgtcaaaagaggttaggtatggggatctttgaaaggaaaatcttcaagg-3'<br><b>Reverse:</b> 5'-ccttgaaagattttcctcaaaagatccccatacctacctcttttgaacaaa-3'<br>Then perform the second round of SDM with the following primers<br><b>Forward:</b> 5'-caaaagaggttaggtatggggatctttgaaaggaaaatcttcaagga-3'<br><b>Reverse:</b> 5'-tccttgaaagattttcctccacaagatccccatacctacctcttttg-3'     |

Dr. Jillian Hagel and Professor Peter Facchini (University of Calgary, Calgary, Alberta, Canada) provided *Escherichia coli* expression plasmids pQEDIOX1 and pQEDIOX3, which contain the open reading frames of *P. somniferum* T6ODM and CODM, respectively. Primers to design the mutant constructs are listed in Table 4.2. We adapted heterologous expression conditions for *E. coli* from a previously reported protocol.<sup>3</sup> Protein expression of the majority of PsCODM mutants was robust (Figure 4.4). Only the A<sub>1</sub>A<sub>2</sub>\*BC<sub>1</sub>C<sub>2</sub> mutant (S149M L151F S152T mutant; the asterisk designates the mutated region) was expressed at low levels, perhaps due to improper folding. We screened each of the mutant enzymes with substrates thebaine **5** and codeine **10** at a concentration of 0.25 mM. Assay conditions are provided in Methods.

Product formation was monitored using liquid chromatography-mass spectrometry (LC-MS). LC-MS chromatograms of representative in vitro enzymatic assays of wild type and mutant PsCODM enzymes are shown in Figures 4.5 and 4.6, respectively. These endpoint assays indicated that the majority of the mutants lost O-demethylase activity toward both thebaine **5** and codeine **10**. Out of the total of 16 mutants, only two, A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> (E338G I340L L341V K342E) and A<sub>1</sub>A<sub>2</sub>B\*C<sub>1</sub>C<sub>2</sub> (R334S R336S+T), retained O-demethylase activity toward either thebaine **5** or codeine **10** (Figure 4.5). The A<sub>1</sub>A<sub>2</sub>B\*C<sub>1</sub>C<sub>2</sub> mutant was similar to the wild type enzyme in that it turned over both thebaine **5** and codeine **10** to yield oripavine **6** and morphine **11**, respectively.

The co-elution of the wild type PsCODM and PsCODM A<sub>1</sub>A<sub>2</sub>B\*C<sub>1</sub>C<sub>2</sub> mutant thebaine **5** demethylation products suggests that the A<sub>1</sub>A<sub>2</sub>B\*C<sub>1</sub>C<sub>2</sub> mutant





**Figure 4.4:** SDS-PAGE of purified wild type and mutant *PsCODM* enzymes from Talon cobalt affinity column (Clontech). Arrow indicates band corresponding to the correct molecular weight.

retained regioselectivity for the 3 position of thebaine **5** (Figure 4.5). In contrast, the  $A_1A_2BC_1^*C_2$  mutant displayed only negligible O-demethylation activity for thebaine **5** but selectively turned over codeine **10** (Figure 4.5). Since codeine **10** only contains a methoxy group at the 3 position, the C3 regioselectivity of the  $A_1A_2BC_1^*C_2$  mutant is also clearly unchanged from the wild type PsCODM. Competitive assay conditions with both thebaine **5** (0.25 mM) and codeine **10** (0.25 mM) were also employed to assess the activity of PsCODM mutants (Figure 4.5). Morphine **11** and oripavine **6** products formed at a 4:1 ratio when wild type PsCODM was subjected to these assay conditions. Similarly, the  $A_1A_2B^*C_1C_2$  mutant also yielded morphine **11** and oripavine **6** at approximately a 4:1 ratio, though notably at lower concentrations than the wild type enzyme,

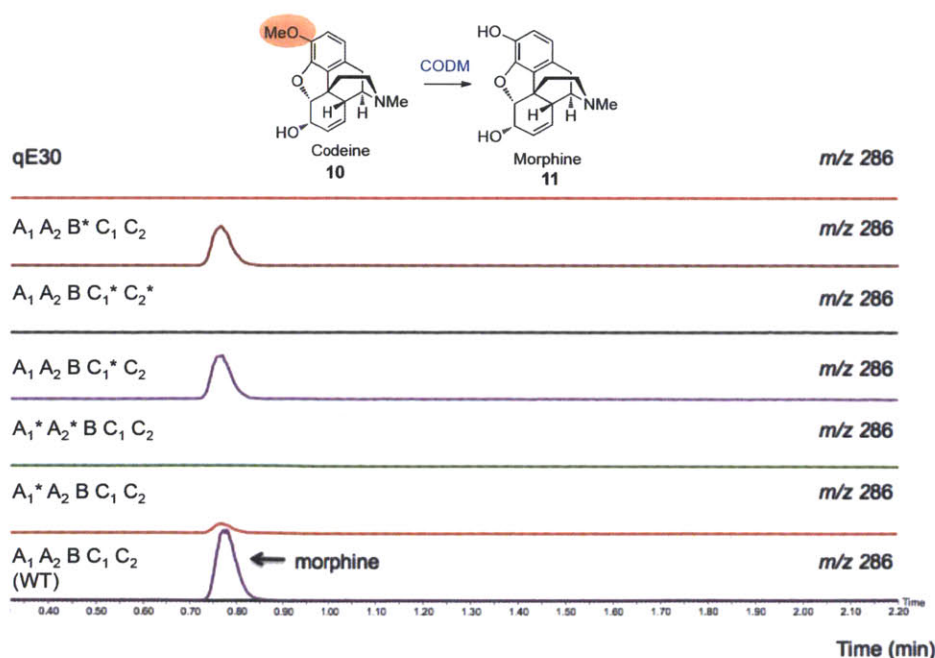


Figure 4.5: PsCODM and representative PsCODM mutants with codeine **10**. Only  $A_1A_2B^*C_1C_2$  and  $A_1A_2BC_1^*C_2$  retain activity for codeine **10**. qE30 is empty vector.

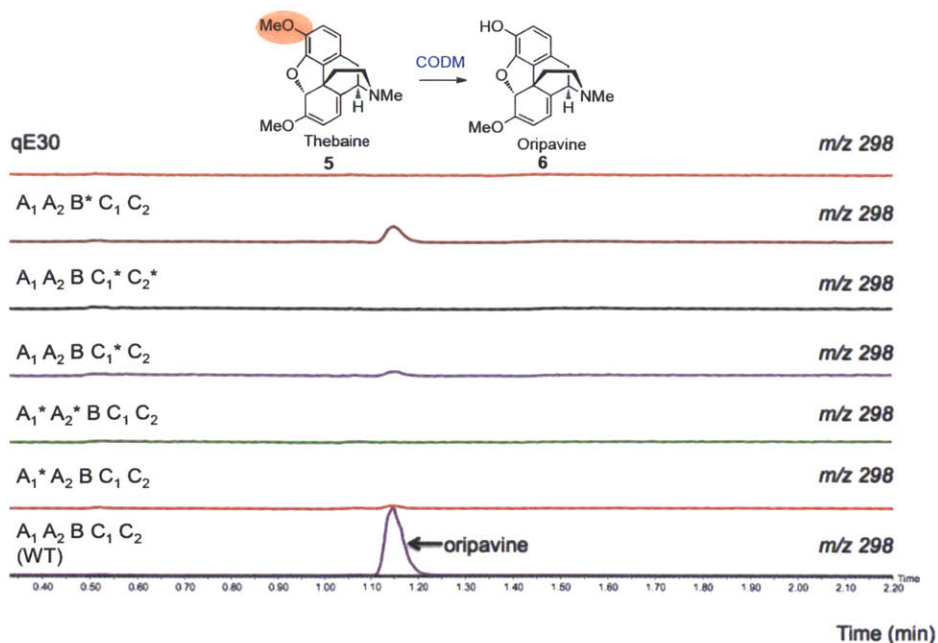


Figure 4.6: *PsCODM* wild type and representative *PsCODM* mutants with thebeine 5. Only  $A_1 B_2 B^* C_1 C_2$  and  $A_1 A_2 B C_1^* C_2$  (negligible) retain activity for thebeine 5. qE30 is the empty vector control.

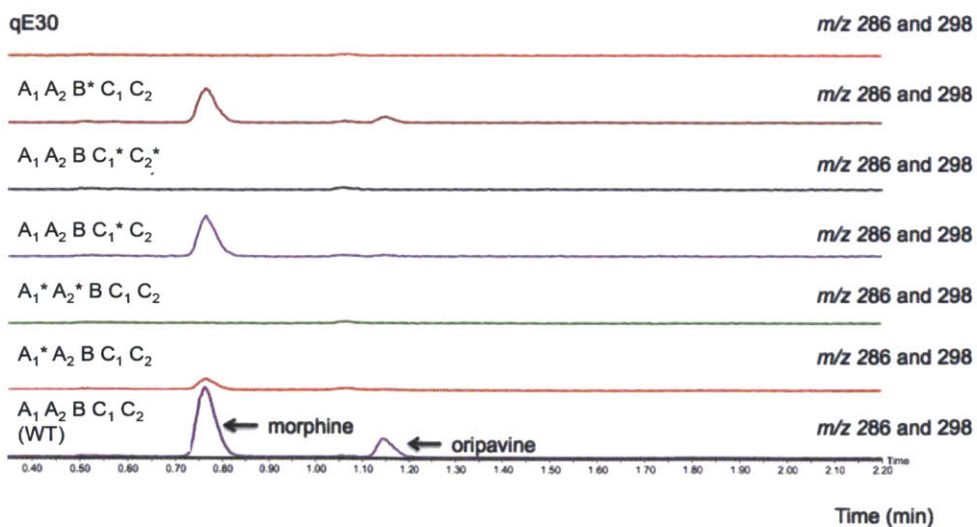


Figure 4.7: *PsCODM* wild type and representative *PsCODM* mutants in a competition assay with both codeine 10 and thebeine 5. Only  $A_1 B_2 B^* C_1 C_2$  and  $A_1 A_2 B C_1^* C_2$  retain activity for either substrate;  $A_1 A_2 B C_1^* C_2$  is specific for codeine 10. qE30 is the empty vector control.

indicating lower catalytic efficiency for the mutated enzyme. However, while the A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> mutant produced only negligible amounts of oripavine **6**, morphine **11** was produced at levels similar to those observed with wild type enzyme, confirming the stringent selectivity of the A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> mutant.

We measured the steady-state kinetic parameters of wild type *PsCODM* and mutant *PsCODM* A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> by monitoring the rate of product formation. In total, we measured kinetic parameters for the following combinations: (1) wild type *PsCODM* with codeine ( $k_{\text{cat}}/K_M = 9.54 \times 10^{-7} \text{ s}^{-1} \mu\text{M}^{-1}$  and  $K_M = 72.3 \pm 33.4 \mu\text{M}$ ); (2) wild type *PsCODM* with thebaine ( $k_{\text{cat}}/K_M = 1.52 \times 10^{-8} \text{ s}^{-1} \mu\text{M}^{-1}$  and  $K_M = 216 \pm 76.2 \mu\text{M}$ ); and (3) *PsCODM* mutant A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> with codeine ( $k_{\text{cat}}/K_M = 3.62 \times 10^{-7} \text{ s}^{-1} \mu\text{M}^{-1}$  and  $K_M = 99.0 \pm 22.4 \mu\text{M}$ ). Notably, the mutant enzyme formed only negligible amounts of the thebaine **5** demethylation product after 4 hr even when substrate concentrations were as high as 2,000  $\mu\text{M}$ . Steady-state kinetic data are shown in Figures 4.8-4.10; Table 4.3).

Although structural information is not yet available for these demethylases, we could build a homology model for both wild type and A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> mutant *PsCODM* based on the anthocyanidin synthase enzyme docked with thebaine **5**. While these computational results must be interpreted with caution, docking studies suggest that the binding orientation of thebaine differs substantially between the wild type enzyme and the A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> mutant (Figures 4.10 and 4.11). Specifically, in the model of the mutant with thebaine **5**, the histidine/aspartate facial triad seemingly anchors the iron-oxo complex away from the O-methyl moiety. It is interesting that the homology model (Figure 4.2) also

predicts that the C<sub>1</sub> region switches from an alpha helix in the wild type enzyme to a random coil in the codeine-specific mutant, suggesting that the mutation in the C<sub>1</sub> region introduces changes to the *Ps*CODM secondary structure.

| Parameters   | WT with Codeine         | A <sub>1</sub> A <sub>2</sub> BC <sub>1</sub> *C <sub>2</sub> with Codeine | WT with Thebaine        | A <sub>1</sub> A <sub>2</sub> BC <sub>1</sub> *C <sub>2</sub> with Thebaine |
|--|-------------------------|--|-------------------------|---|
| V <sub>max</sub> (μM/min)  | 0.299 ± 0.044           | 0.213 ± 0.015  | 0.043 ± 0.001           | ND  |
| K <sub>M</sub> (μM)  | 72.2 ± 33.4             | 99.0 ± 22.4  | 216.1 ± 76.2            | ND  |
| k <sub>cat</sub> (s <sup>-1</sup> )                                  | 6.90 × 10 <sup>-5</sup> | 3.58 × 10 <sup>-5</sup>  | 3.28 × 10 <sup>-6</sup> | ND  |
| k <sub>cat</sub> /K <sub>M</sub> (s <sup>-1</sup> μM <sup>-1</sup> ) | 9.54 × 10 <sup>-7</sup> | 3.62 × 10 <sup>-7</sup>  | 1.52 × 10 <sup>-8</sup> | ND  |
| r <sup>2</sup>   | 0.92                    | 0.97   | 0.96                    | ND  |
| replicates   | 2                       | 3  | 2                       | 3   |

Table 4.3: Summary of kinetic parameters for *Ps*CODM and Mutant A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> with codeine **10** and thebaine **5**.

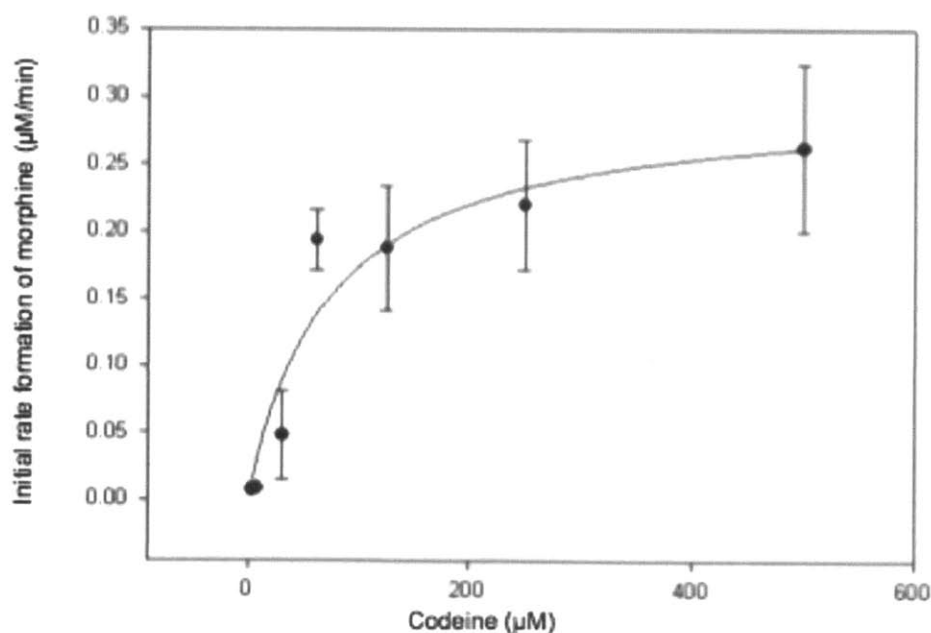


Figure 4.8: Michaelis-Menten curve of WT *Ps*CODM with codeine **10**.

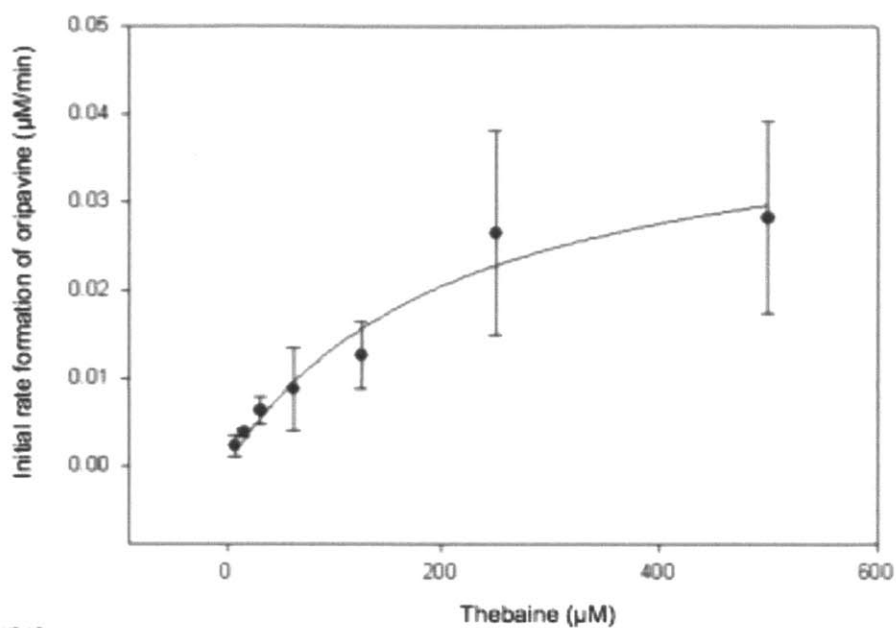


Figure 4.9: Michaelis-Menten curve of WT *PsCODM* with thebaine **5**.

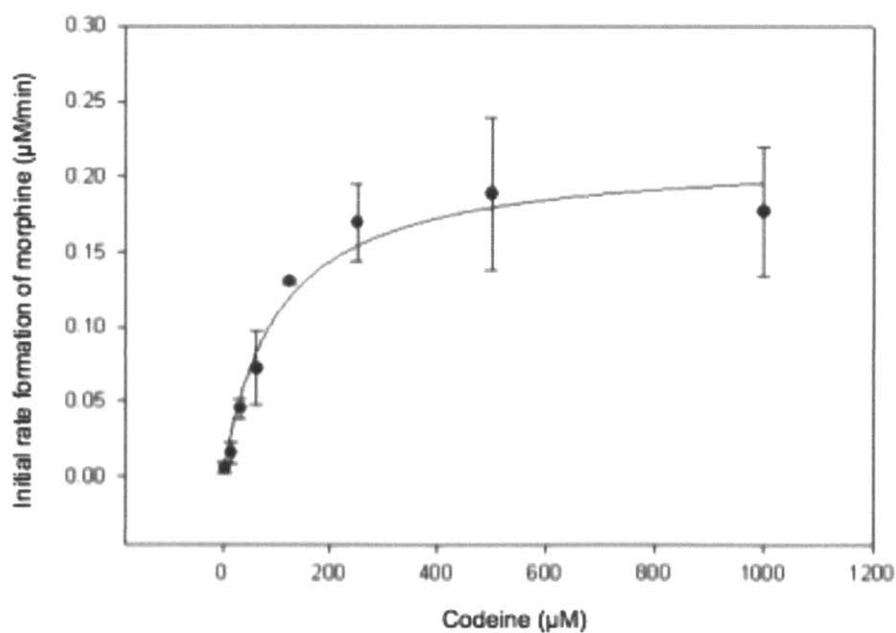


Figure 4.10: Michaelis-Menten curve of Mutant  $A_1A_2BC_1^*C_2$  with codeine **10**.

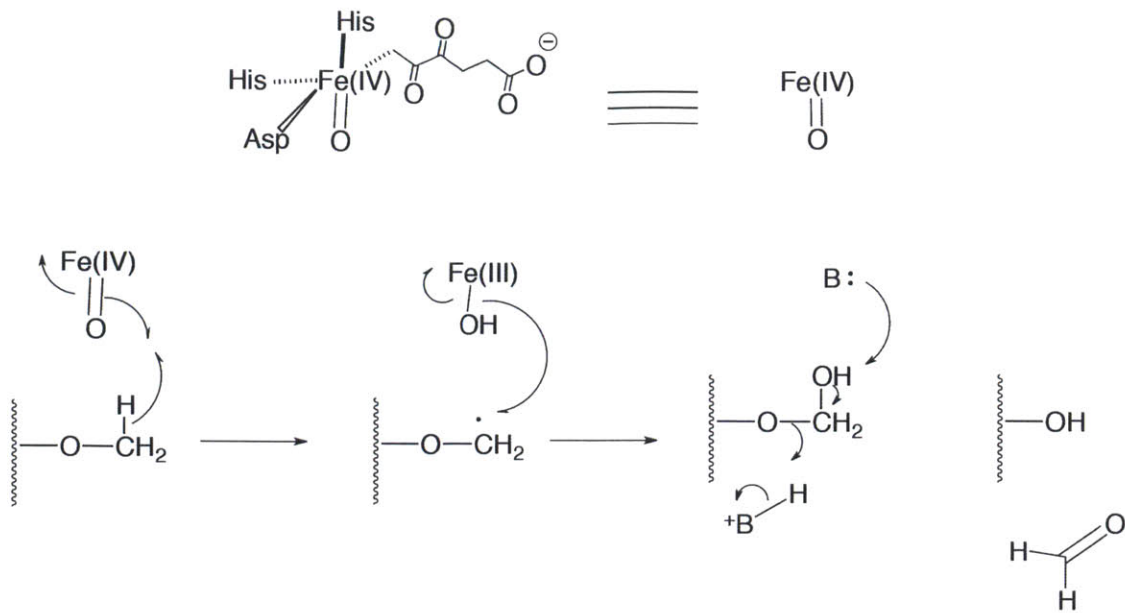


Figure 4.11: Proposed dioxygenase demethylation mechanism.

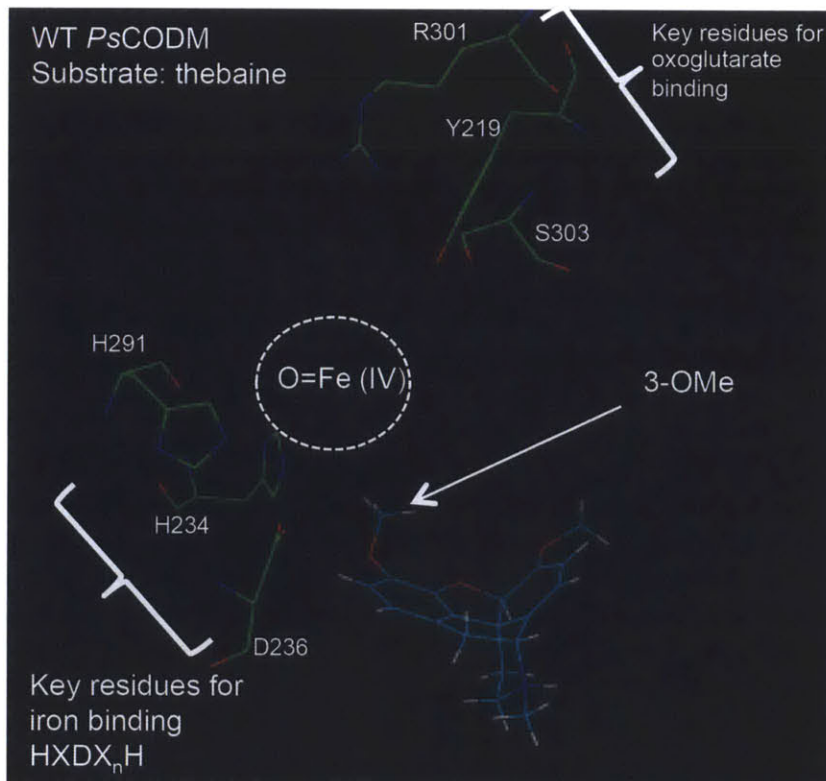


Figure 4.12: Thebaine docked into wild type *PsCODM* (DIOX3).



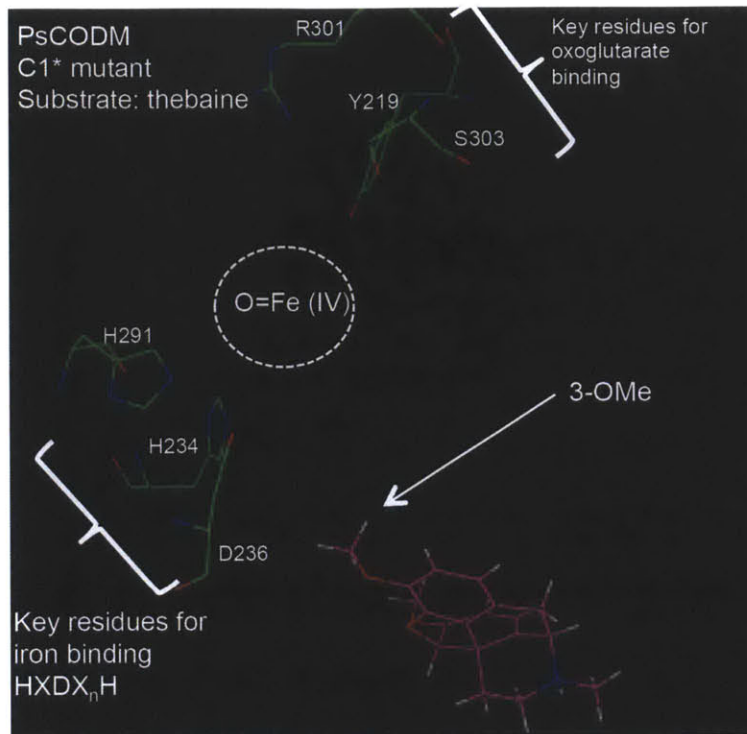


Figure 4.13: Thebaine docked into *PsCODM* (DIOX3) mutant  $A_1A_2BC_1^*C_2$

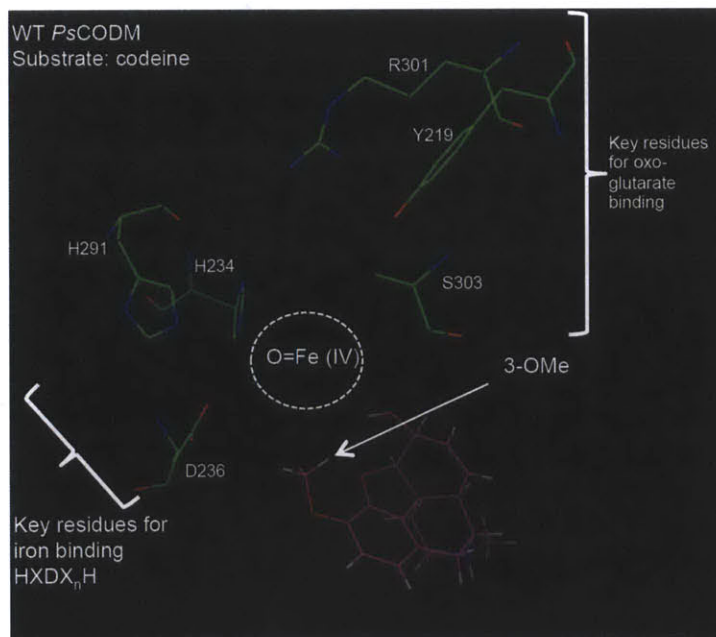


Figure 4.14 Codeine docked into wild type *PsCODM* (DIOX3).



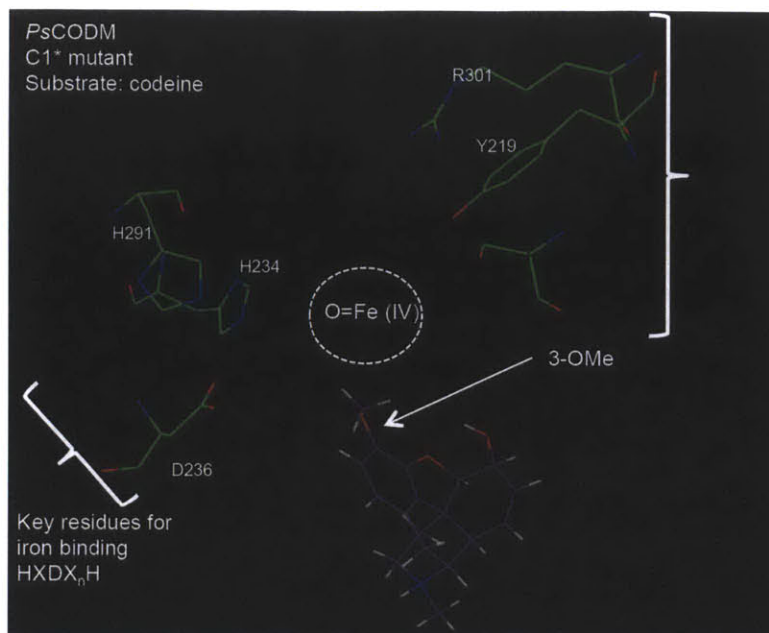


Figure 4.15: Codeine docked into *PsCODM* (DIOX3) Mutant A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub>.

The mutation in the C<sub>1</sub>\* region appears to prevent productive binding of the thebaine **5** substrate without greatly altering the catalytic efficiency for codeine **10** ( $k_{\text{cat}}/K_{\text{M}} = 9.54 \times 10^{-7} \text{s}^{-1} \mu\text{M}^{-1}$  for wild-type,  $k_{\text{cat}}/K_{\text{M}} = 3.62 \times 10^{-7} \text{s}^{-1} \mu\text{M}^{-1}$  for mutant). In the models, the binding orientation of codeine **10** does not appear to differ substantially between the wild type enzyme and the A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> mutant.

At the outset of this study, we hypothesized that the amino acid differences between *PsCODM*, a C3 O-demethylase, and *Pst6ODM*, a C6 O-demethylase, would control the distinct regioselectivity of the demethylation reactions. Therefore, we expected that swapping the non-conserved amino acid regions of these two enzymes would alter the regioselectivity of the mutated enzymes. Instead, the mutations altered the substrate selectivity in an unpredictable

manner. Moreover, the  $A_1A_2BC_1^*C_2$  *PsCODM* mutant retained C3 regioselectivity.

Notably, *PsT6ODM* turns over thebaine **5** and oripavine **6**, and *PsCODM* turns over thebaine **5** and codeine **10**. Despite both wild type enzymes accepting thebaine **5**, the introduction of residues from *PsT6ODM* into *PsCODM* at the C<sub>1</sub> position (Figure 4.3) yields a mutant *PsCODM* enzyme that is selective for codeine **10**. This change in substrate specificity does not readily correlate with the substrate specificity of the parent wild type enzymes. Particularly in the absence of experimental structural data, it is difficult to rationalize what structural changes these mutations confer to *PsCODM* and how these changes impact substrate specificity. Nevertheless, while this work highlights the difficulty of rational, structure-based protein design, we successfully demonstrate how sequence alignment of enzymes with subtle differences in specificity can be used to readily generate swapped sequences that are functionally distinct from naturally occurring wild-type variations.

One goal of metabolic engineering is the removal of shunt or redundant pathways that adversely impact production yields of a desired compound.<sup>8</sup> This  $A_1A_2BC_1^*C_2$  *PsCODM* mutant may fulfill an engineering function by providing a means to shut off a redundant route (route B in Figure 1) in morphinan biosynthesis. Because  $A_1A_2BC_1^*C_2$  *PsCODM* fails to demethylate thebaine **5** to form oripavine **6**, the first committed step of route B, replacement of wild-type *PsCODM* with the  $A_1A_2BC_1^*C_2$  mutant would presumably force the morphine pathway to proceed via route A.

Notably, growers in India, a major cultivator of licit opium, inadvertently counter-selected for *P. somniferum* cultivars low in oripavine **6** when selecting for the highest seed quality and opium yields, suggesting that low oripavine **6** production levels may correlate with economically desirable traits.<sup>9</sup> Therefore, in addition to assessing how knock-outs of wild type *PsCODM* and *PsT6ODM* affect flux into morphine production, it will be of interest to observe how using the *PsCODM* A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> mutant—which effectively sidesteps oripavine **6** production by committing thebaine **5** to route A—instead of the wild type *PsCODM* enzyme would impact titers of codeine **10** and morphine **11**.<sup>9</sup> Moreover, substantial interest lies in reconstituting morphinan alkaloid biosynthesis in yeast and *E. coli*.<sup>10-12</sup> Mutants with altered specificity such as *PsCODM* A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> could provide important building blocks for these synthetic biology efforts.

### III. Conclusions

We have altered the substrate specificity of a morphinan pathway enzyme codeine O-demethylase (*PsCODM*). One *PsCODM* mutant, mutant A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> (E338G I340L L341V K342E), exhibits demethylase activity exclusively toward codeine **10**, whereas the wild type *PsCODM* exhibits demethylase activity with both codeine **10** and thebaine **5**. These results provide a starting point for rationalizing how *PsCODM* and *PsT6ODM*, two enzymes involved in morphinan biosynthesis with 73% identity at the amino acid level, facilitate O-demethylation regioselectivity on separate sets of substrates. In addition to providing insight into

O-demethylase substrate selectivity, this mutant could also be useful in biotechnological efforts to provide *P. somniferum* strains with augmented yields of codeine **10** and morphine **11** and diminished titers of oripavine **6**, an intermediate in a redundant pathway branch that has been associated with poor seed quality and low opium yields.<sup>9</sup> In short, mutants with enhanced enzyme selectivity will allow us to explore how the targeted disruption of a redundant pathway branch affects downstream product yields and may enable more efficient production of these high value compounds. In addition to the potential biotechnological applications of this enzyme, these protein engineering efforts also provide a starting point for understanding how the subtle sequence differences of highly similar enzymes can impact substrate and regioselectivity.

#### **IV. Methods**

##### *A. Construction of PsCODM Mutant Expression Plasmids*

pQEDIOX1 (pQET6ODM) and pQEDIOX3 (pQECODM), which contain the open reading frames of *PsT6ODM* and *PsCODM*, respectively, were provided by Professor Peter Facchini and Dr. Jillian Hagel (University of Calgary). To obtain the 16 CODM mutants (Table 4.1), site-directed mutagenesis (SDM) was performed using the Stratagene QuikChange Site-Directed Mutagenesis kit. SDM primers are listed in the Supplemental Information (Table 4.1). The *PsCODM* mutant constructs were sequenced to verify the DNA sequence and were subsequently transformed for expression into *E. coli* strain SG13009 (QIAGEN) via electroporation using standard protocols.

### *B. Heterologous Expression of PsCODM Mutants*

A single transformed *E. coli* colony (strain SG13009 [QIAGEN]) was inoculated in 10 ml Luria-Bertani (LB) media supplemented with kanamycin (0.05 g/l) and ampicillin (0.1 g/l) and incubated overnight at 37°C with shaking at 225 rpm. Subsequently, 800 ml LB media supplemented with kanamycin (0.05 g/l) and ampicillin (0.1 g/l) was inoculated with an overnight culture (8 ml) and incubated at 37°C with shaking at 225 rpm until reaching an optical density 600 (OD<sub>600</sub>) of 0.6. Protein expression was induced by the addition of isopropyl-β-D-galactopyranoside (IPTG; final concentration 0.3 mM). Following induction, cells were incubated at 6°C with shaking at 225 rpm for 24 hr. Cells were harvested by centrifugation and lysed by sonication. The hexa-histidine-tagged *PsCODM* mutants were purified using Talon cobalt affinity column (Clontech) using the manufacturer's protocols. Eluted enzyme was subsequently buffer-exchanged into Tris buffer (100 mM Tris-HCl, pH 7.4, 10% [v/v] glycerol, and 14 mM 2-mercaptoethanol) and immediately assayed for activity. These enzymes were not stable upon extended storage.

### *C. In Vitro Activity Assay of PsCODM Mutants*

The in vitro activity assay protocol was adapted from a previous report (Hagel and Facchini, 2010). Briefly, the assay for 2-oxoglutarate/Fe(II)-dependent dioxygenase activity was performed using a 100 ml reaction mixture of 100 mM Tris-HCl (pH 7.4), 10% (v/v) glycerol, 14 mM 2-mercaptoethanol, 0.25

mM alkaloid(s), 10 mM 2-oxoglutarate, 10 mM sodium ascorbate, 0.5 mM FeSO<sub>4</sub>, and 1 mM purified enzyme. Assays were carried out for 4 hr at 30\_ C. Aliquots (25 ml) were quenched in 1 ml methanol containing yohimbine (500 nM) as an internal standard. The samples were centrifuged in a microcen- trifuge (13,000 rpm, 5 min) to remove particulates and then analyzed by LC-MS. Samples were ionized by ESI with a Micromass LCT Premier TOF Mass Spectrometer. The LC was performed on an Acquity Ultra Performance BEH C18, 1.7 μM, 2.1 x 100 mm column on a gradient of 10%–90% acetoni- trile/water (0.1% formic acid) over 5 min at a flow rate of 0.6 ml/min. The appearance of morphine **11** and oripavine **6** was monitored by peak integration and normalized to the internal standard. All chemicals were obtained from a commercial source (Sigma Aldrich).

#### *D. Steady-State Kinetic Assay of Wild-Type CODM and A<sub>1</sub>A<sub>2</sub>BC<sub>1</sub>\*C<sub>2</sub> Mutant*

Assay components were used in the following final concentrations: pH 7.4 Tris-HCl (67 mM) containing 10% (v/v) glycerol and 14 mM 2-mercaptoethanol, α-ketoglutarate (6.7 mM), sodium ascorbate (6.7 mM), iron (II) sulfate (333 mM), and enzyme wild type (4.1 mg/ml) or enzyme mutant (18 mg/ml), in an assay volume of 150 ml. The enzyme concentrations were estimated by Bradford assay. Assays were initiated by addition of iron (II) sulfate and conducted at 30°C. Aliquots (25 ml) were quenched every 30 min from 1 to 3 hr with 975 ml methanol containing 500 nM ajmaline as an internal standard. Prior to analysis, methanol quenched samples were centrifuged in a microcentrifuge at 13,000 rpm

for 5 min to remove any particulates. Liquid chromatography was performed on an Acquity Ultra Performance BEH C18, 1.7  $\mu$ M, 2.1 x 100 mm column. The gradient was 10%–90% acetonitrile over 4 min with water and 0.1% formic acid in water as the second solvent. The flow rate was 0.5 ml/min. We performed electrospray ionization (ESI) with a Micromass LCT Premier TOF Mass Spectrometer in positive ionization V-mode.

Product accumulation was measured to determine the kinetic parameters. A standard curve for morphine **11**, the demethylation product of codeine **10**, was constructed to determine the kinetic parameters of codeine **10** demethylation. However, oripavine **6**, the demethylation product of thebaine **5**, was unavailable; therefore, hydromorphonone, which was available, was used as a surrogate. Plotted experimental data were fit to a Michaelis-Menten curve using SigmaPlot version 9.0. Experiments were duplicated or triplicated to ensure reproducibility.

### *E. Computational modeling*

Both wild type and mutant *PsCODM* homology models were built using *A. thaliana* anthocyanadin synthase (*AtANS*) as a template in SWISS MODEL. Primary substrates were docked into the active site at the lowest catalytically competent conformation. The location of the iron-oxo species was postulated based on the position of histidine/aspartate facial triad.

## V. References

1. Facchini, P.J., Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation and Metabolic Engineering Applications. *Annu Rev. Plant Physiol.* 2001, 52, 29-66.
2. Facchibi, P.J.; Bird, D.A.; Borgault, R.J.; Hagel, J.M.; Liscombe, D.K.; MacLeod, B.P. and Zulak, K.G. Opium Poppy: A model System to Investigate Alkaloid Biosynthesis in Plants. *Can. J. Bot.* 2005. 83, 1189-1206.
3. Hagel J.M. and Facchibi, P.J., Dioxygenases Catalyze the O-demethylation Steps of Morphine Biosynthesis in Opium Poppy. *Nat. Chem. Biol.* 2010. 6, 273-275.
4. Hagel, J.M. and Facchini, P.J., Biochemistry and Occurrence of O-demethylation in Plant Metabolism. *Front Physiol.* 2010, 1(14).
5. Hagel, J.M.; Weijei, A.M., Voge, H.J. and Facchini, P.J., Quantitative <sup>1</sup>H Nuclear magnetic Resonance Metabolite Profiling as a Functional Genomic Platform to Investigate Alkaloid Biosynthesis in Opium Poppy. *Plant Physiol.* 2008, 147(4), 1805-21.
6. Wilmouth, R.C.; Turnbull, J.J.; Welford, R.W.; Clifton, V.J.; Prescott, A.G. and Schofield, C.J., Structure and Mechanism of Anthocyanidine Synthase from *Arabidopsis thaliana*. *Structure* 2002, 10, 93-103.
7. Schwede, T.; Kopp, J.; Guex, N.; Peitsen, M.C., SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acid Res.* 2003, 31, 3381-85.
8. Pickens, L.B.; Tang, Y and Chooi, Y.-H., Metabolic Engineering for the Production of Natural Products. *Annu. Rev. Chem. Biomol. Eng.* 2011, 2, 211-36.
9. Prajapat, S.; Bajpai, S.; Singh, D.; Luthra, R.; Gupta, M.M. and Kumar, S. Alkaloid Profiles of Indian Land Races of the Opium Poppy *Papaver somniferum* L. 2002. *Genet. Resour. Crop. Evol.* 2002, 49, 183-8.
10. Hawkins, KM and Smolke, C.D., Production of Benzylisoquinoline Alkaloids in *Saccharomyces cerevisiae*. *Nat. Chem. Biol.* 2008, 4, 564-573.
11. Nakagawa, A.; Minami, H.; Kim, J.S., Koyanagi, T.; Katayama, T.; Sato, F. and Kumagai, H., A Bacterial Platform for Fermentative Production of Plant Alkaloids. *Nat. Commun.* 2011, 2, 326.
12. Minami, H.; Kim, J.-S.; Ikezawa, N.; Takemura, T.; Katayama, T.; Kumagai, H. and Sato, F. *Microbial Production of Plant Benzylisoquinoline Alkaloids.* *Proc. Natl. Acad. Sci. USA.* 2008, 105, 7393-8.



## **VI. Acknowledgements**

Dr. Weerawat Runguphan and Prof. Sarah O'Connor designed and initiated the project. Dr. Weerawat Runguphan performed all cloning and developed in-house protein expression conditions. Computational models were produced collaboratively. We thank Dr. Jill Hagel and Prof. Pete Facchini for constructs with open reading frames for DIOX1 and DIOX3. Prof. Sarah O'Connor procured funding for this project.

## Chapter 5

Conclusions and Future Work: The Alkaloids and Beyond

## I. Conclusions

Chapter 1 opens the thesis by discussing current work in the understanding and manipulation of plant natural product biosynthesis. Specifically, chapter 1 surveys current research regarding monoterpene indole alkaloids, benzyloquinoline alkaloids and the glucosinolates. These classes of compounds have been witness to heroic efforts aimed at enzyme discovery and engineering. These efforts have undoubtedly been energized by the desire to harness and improve yields of these highly bioactive compounds.

While many pathway segments of benzyloquinoline biosynthesis—most notably the morphinan pathway<sup>1</sup>—have been completely characterized, monoterpene indole alkaloid biosynthesis has not been fully elucidated in any organism. Fully elucidated pathways are more amenable to various engineering strategies. Moreover, we also have an impetus to study missing pathway steps because they may possess unique and interesting chemistry. Approximately three steps are missing in the biosynthesis of secologanin, a direct monoterpene indole alkaloid precursor in Madagascar periwinkle (Figure 5.1). Chapter 2 chronicles our efforts to discover 10-hydroxygeraniol oxidoreductase activity in periwinkle (Figure 5.1). This enzyme is positioned at the seat of iridoid biosynthesis and produces the substrate for the iridoid synthase, which assembles the characteristic iridoid molecular scaffold.<sup>2,3</sup>

Despite having an incomplete understanding of monoterpene indole alkaloid biosynthesis, our lab has demonstrated several successful engineering strategies for this pathway, most notably the incorporation of halogens.<sup>4-6</sup>

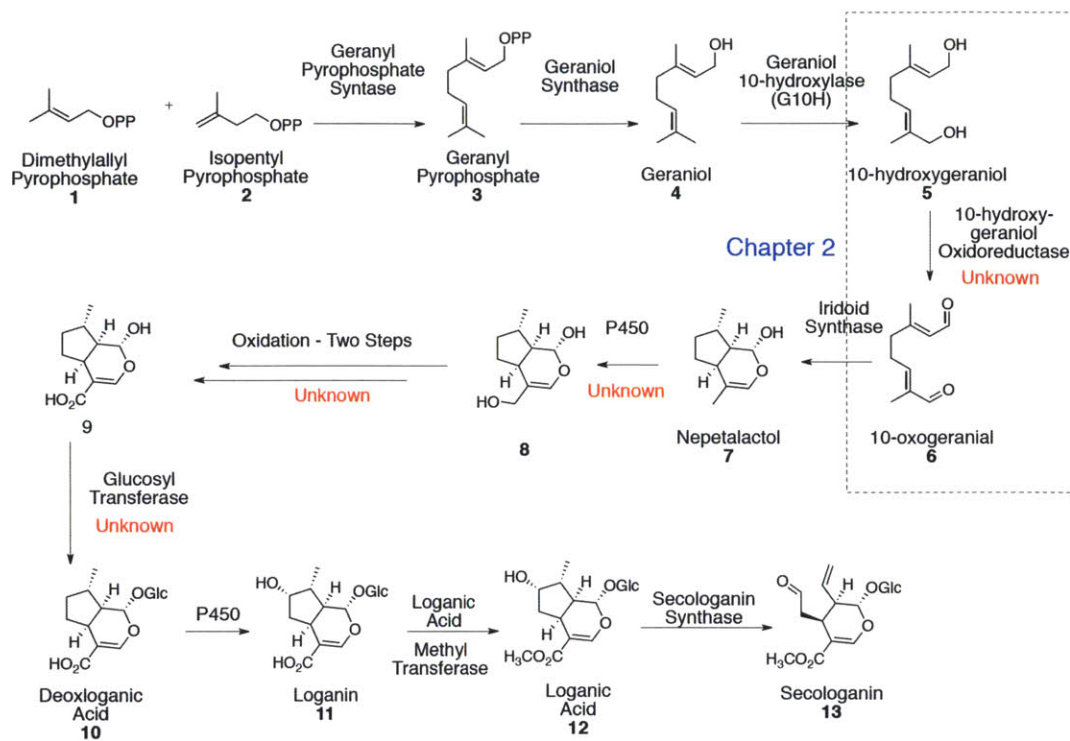


Figure 5.1: Proposed biosynthesis of secologanin **13**. Chapter 2 focuses on the discovery of 10-hydroxygeraniol oxidoreductase activity in *C. roseus*. 10-hydroxygeraniol reductase oxidizes both hydroxyl moieties of 10-hydroxygeraniol **5** to form 10-oxogeraniol **6** (dotted line box), which is the substrate for Iridoid Synthase—the enzyme that assembles the iridoid scaffold. Approximately 4 steps are missing in the biosynthesis of secologanin **13**.

Runguphan *et al.* provides a seminal study on the *de novo* production of halogenated alkaloids by introducing prokaryotic genes that code for tryptophan halogenases into periwinkle, a terrestrial plant.<sup>5</sup> This is the first example of *de novo* combinatorial biosynthesis in plants. In this study, Runguphan *et al.* produced 12-chloro-19,20-dihydroakuammicine **13** at a titer of 25 µg per gram fresh weight.<sup>5</sup> However, because of the metabolic bottleneck situated at tryptophan decarboxylase (TDC), the plant also accumulated 7-chlorotryptophan—the RebH/F product—at 50 µg per gram fresh weight (twice as high as the chlorinated alkaloid titer).<sup>5</sup> We posited that the accumulation of 7-chlorotryptophan—an analog to a canonical, proteinogenic amino acid that is structurally similar to an auxin growth hormone—caused the slow growth and browning morphology observed in Runguphan *et al.*'s initial study. Therefore, as highlighted in chapter 3, we explored two strategies to alleviate the metabolic bottleneck: (1) over-expressing endogenous TDC and (2) reengineering RebH to preferentially chlorinate tryptamine, a downstream and direct alkaloid precursor, instead of the native substrate L-tryptophan (Figure 5.2).<sup>6</sup> While over-expressing endogenous TDC failed to fracture the metabolic bottleneck, we successfully circumnavigated the bottleneck by reengineering RebH to preferentially accept tryptamine (Figure 5.2).<sup>6</sup> The tryptamine specific mutant RebH Y455W along with flavin reductase RebF was incorporated into periwinkle's metabolism, and we observed *de novo* production of 12-chloro-19,20-dihydroakuammicine **13**. In contrast to Runguphan *et al.*, no 7-chlorotryptophan was observed when this mutant was employed, demonstrating its desired substrate selectivity (Figure 5.2).<sup>6</sup>

Pathways are more amenable to sophisticated engineering strategies (or only plausible) when the targeted pathway is completely characterized. The morphinan pathway, whose elucidation was completed in 2010 with the discovery of the two dioxygenases Codeine O-Demethylase (*PsCODM*) and Thebaine 6-O-Demethylase

### Chapter 3 - Circumvent TDC Bottleneck by Installing Chlorine on Tryptamine

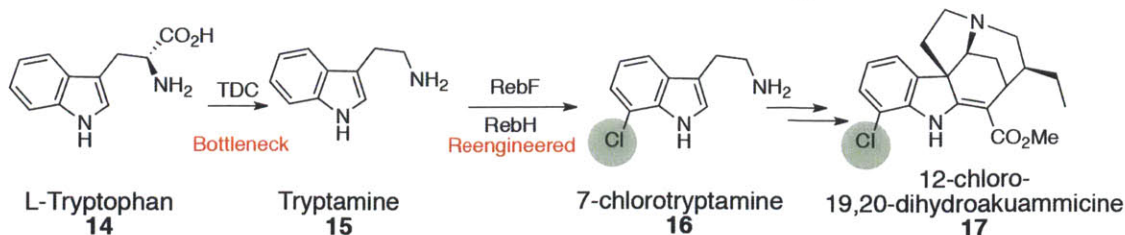


Figure 5.2: RebH was reengineered to preferentially chlorinate tryptamine **11** instead of native L-tryptophan. This strategy successfully alleviated the metabolic bottleneck by circumventing it altogether. Incorporation of this mutant into the MIA metabolism of *C. roseus* (Madagascar periwinkle) led to the *de novo* production of halogenated metabolite 12-chloro-19,20-dihydroakuammicine **13**, without the accumulation of 7-chlorotryptophan, highlighting the reengineered specificity of the enzyme.

(*PsT6ODM*) from *Papaver somniferum*, is particularly attractive as it produces codeine and morphine.<sup>1</sup> The pathway utilizes two separate routes to convert thebaine into morphine.<sup>1</sup> Chapter 4 archives our serendipitous discovery of a *PsCODM* mutant that would presumably selectively disable one pathway—adventitiously, the pathway associated with poor licit opium and seed quality—while simultaneously committing thebaine to the route with more medicinally valuable compounds.<sup>7</sup>

The *PsCODM* project began by attempting to switch the regioselectivity of *PsCODM* (which selectively demethylates the 3 position of thebaine **18** and codeine **21**) and *PsT6ODM* (which selectively demethylates the 6 position of thebaine **18** and oripavine **22**). The enzymes' primary structures are highly similar, but sequence analysis pinpointed five regions of dissimilarity. We hypothesized that mixing-and-matching the regions of dissimilarity between the two sequences would alter the enzymes' regioselectivity and, in turn, offer some insight into how regioselectivity is controlled in this class of enzyme.

*PsT6ODM* mutants were only poorly expressed; therefore, we focused our efforts on *PsCODM* mutants. While most of the 16 mutants were inactive for either thebaine or codeine, one *PsCODM* mutant (E338G I340L L341V K342E) selectively demethylates codeine. This switch in selectivity does not readily correlate to the selectivities of the parent enzymes. Nonetheless, this mutant could fulfill a pivotal role in morphinan pathway reconstitution (Figure 5.3).

**Chapter 4 - Reengineered Codeine O-demethylase (*Ps.CODM*) mutant commits thebaine exclusively to route A**

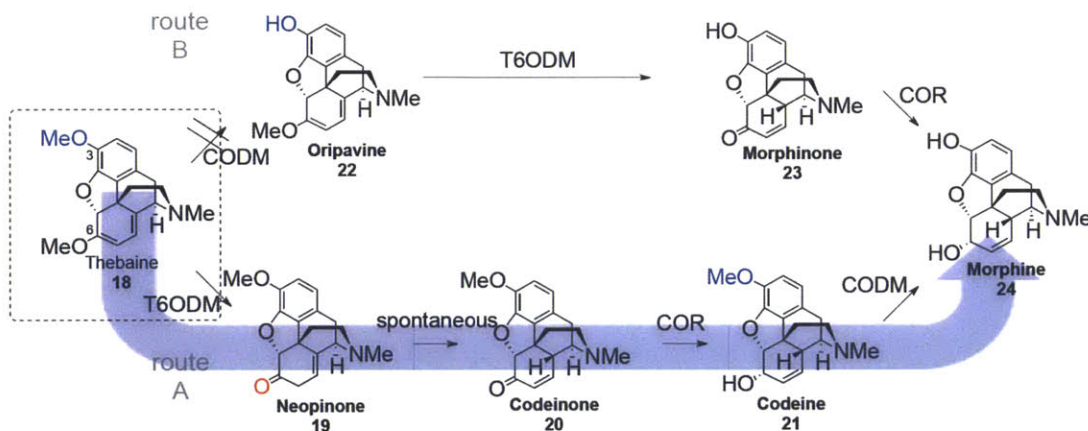


Figure 5.3: *PsCODM* mutant E338G I340L L341V K342E selectively demethylates thebaine **14**. This mutant disables route B, which has been associated with poor licit opium and seed quality and instead commits thebaine **14** to the more medically useful pathway that produces codeine **17** and leads to morphine **20**.

**II. Future Directions – The Alkaloids and Beyond**

Historically, altering metabolic pathways in plants to achieve a given end has been difficult. Metabolic engineering in plants is still in its infancy and until very recently has largely been confined to single-gene expression or silencing events in the background of endogenous plant cell metabolism. The complexity of the plant host's metabolism has been shown, in many cases, to effectively mute the engineering effort or lead to unpredictable results (Table 1.1). However, in recent years, a wealth of new approaches has expanded the capabilities of multi-gene pathway expression in both

plants and microbes and has highlighted our increasing ability to engineer the production of plant natural products in both plants and heterologous systems. The increase in available and reliable sequencing and expression data enables the (relatively) facile discovery of gene, transporter and regulatory elements, the identification of which is often a prerequisite for multi-step metabolic engineering efforts. The three case studies described in Chapter 1 (MIAs, BIAs and glucosinolates) exemplify the challenges and progress in metabolically engineering plant-derived natural products. While we have made a special effort to highlight the advantages and pitfalls of individual techniques and efforts throughout this thesis, a number of grand challenges for plant metabolic engineering remain to be tackled in the coming years.

#### *A. Effective Mining Strategies*

Effective mining strategies, such as those employed by Giddings *et al.*,<sup>8</sup> Hagel *et al.*,<sup>1</sup> Winzer *et al.*,<sup>9</sup> Liscombe *et al.*,<sup>10</sup> and Geu-Flores *et al.*,<sup>5</sup> are required to sift through the mounting data of the sequencing age. Hanson *et al.* provides a recent comprehensive review of effective mining strategies and phylogenetic analyses. Traditionally, plant enzyme discovery methods have relied heavily upon time-intensive reverse genetics based strategies. Bioinformatic techniques that engage co-expression analyses and comparative metabolite profiling to limit the gene space to be investigated are greatly accelerating the discovery process in plant systems. Moreover, a suite of new silencing tools, including VIGS,<sup>12</sup> RNAi<sup>13</sup> and the IL-60 system,<sup>14</sup> can provide rapid insight into the physiological function of plant enzymes.

#### *B. Metabolic Engineering in Native Versus Non-Native Hosts*

Many efforts aimed at improving the yield of alkaloids in native hosts have focused on feeding precursors and over-expressing transcription factors or enzymes



positioned at metabolic bottlenecks. While these efforts often result in modest improvements to yield (Table 1.1), many are often accompanied by adverse morphological effects that may significantly stunt the growth of plant and tissue cultures, highlighting the tight regulation of metabolic processes within highly organized plant cells and tissues. In native systems especially, the slow growth phenotypes may result from the depletion of cellular resources used in synthesizing a surfeit of transcripts and enzymes or from the accumulation of toxic intermediates that negatively impact growth and development. Engineering in native hosts or heterologous plant species is attractive because not having to build the starting substrates and supply the co-factors greatly simplifies engineering efforts. However, with this strategy, maintaining the balance between primary metabolism and the engineered metabolism—a feat that will likely improve growth morphologies—is complicated precisely because endogenous primary metabolite pools are expropriated for the overproduction of selected metabolites. Also, despite the advantage of minimal gene stacking, the often uncharacterized and unanticipated complex metabolism and regulatory elements of native systems can lead to engineering outcomes that are particularly difficult to control and predict. The industrial scale production of plant natural products will likely require more comprehensive engineering efforts than single-gene over-expression or silencing events in the context of native plant hosts. Engineering in faster-growing and “chemically silent” heterologous hosts may increase biomass accumulation and simplify purification.

### *C. Controlling Metabolic Flux through New Expression Constructs, Scaffolds and Tunable Regulatory Elements*

The intricate relationship between primary metabolism (i.e. glycolysis, the TCA cycle) and the native or heterologous secondary metabolism (i.e. isoprenoid and alkaloid pathways) must be considered. In plants, Park *et al.* alluded to this interplay by

demonstrating that BBE expression levels vastly affect amino acid levels.<sup>15,16</sup> As metabolic engineering strategies in plants become more sophisticated, we should also begin to consider flux analyses, taking into account that natural product pathways are evolutionarily optimized to channel intermediates toward product through a highly choreographed system of protein-protein interactions, localization and regulation.<sup>17</sup> The overall goal is to maximally channel metabolic resources to the desired products without over taxing the host system.

Co-localization through scaffolding is a proven way to channel metabolites in *E. coli*. These systems attempt to mock natural megasynthases, which efficiently shuttle metabolites between adjacent active sites. Essentially, scaffolding increases the local metabolite and enzyme concentrations and effectively lowers the  $K_m$  of the substrate. These systems are widely modular and are known to improve titers, alleviate metabolic bottlenecks and reduce metabolic loads by preventing carbon from exiting the pathway. Under conditions of low enzyme expression (decreased metabolic load), Dueber *et al.* successfully achieved a 77-fold enhancement in mevalonate production by building a scaffold based on the protein-protein interactions of GBD, SH3 and PDZ domains and their cognate ligands.<sup>18</sup> They built the scaffold on hydroxymethylglutaryl-CoA reductase, the enzymatic bottleneck of mevalonate production.<sup>18</sup> Notably, these scaffolding systems require that the enzyme at the metabolic bottleneck, the subsequent enzyme and the substrate be co-localized,<sup>18</sup> underscoring why they may be untenable for some highly compartmentalized systems. Nonetheless, the prospect of engineering metabolons into plants is exciting.

The effective metabolic engineering of plant natural products will inevitably require advanced, but easy-to-use, gene stacking techniques. Traditionally, multi-gene expression in plants has been plagued with inadvertent silencing events, the incomplete incorporation of all genes and lengthy and technically challenging procedures.<sup>19</sup> A

number of new technologies, however, are being developed to assemble and transplant large fragments of DNA.<sup>19</sup> Golden Gate cloning and USER fusion have been used to clone multiple pathway elements.<sup>19,20</sup> Additionally, synthetic plant chromosomes and the universal expression and silencing IL-60 platform both have the demonstrated capability of introducing multiple plant pathway elements into plants.<sup>19,21</sup> For example, under the transformation-free IL-60 platform, Mozes-Koch *et al.* expressed an entire bacterial operon in tomato and produced pyrollinitrin, which they observed after only two days.<sup>21</sup>

A number of RNA-based silencing systems, including RNAi, have also been engineered and applied to medicinal plants.<sup>22</sup> Notably, RNAi, which provides a permanent pheno- or chemotype, has been employed to block shunt pathways and channel metabolic resources toward a desired product, enhancing our ability to engineer in multiple dimensions.<sup>22</sup>

Lastly, promoter libraries, engineered untranslated regions (i.e. 5' untranslated regions and intergenic regions), genetic circuit designs and biosensor regulators have been tremendously helpful in microbial engineering. Applying these design principles to the metabolic engineering of plants may greatly enrich our efforts to produce valuable and chemically diverse alkaloids. Notably, a variety of constitutive and inducible plant promoters and expression systems are now widely available. Synthetic RNA elements, ribosome binding site elements and a combination of different strength promoters strategically placed in front of stacked pathway genes could theoretically enable tunable protein expression.<sup>23,24</sup> These elements could potentially limit the expression of toxic activities or the accumulation of toxic metabolites until the stationary phase (or an appropriate stage) of growth, thereby absolving the system of unsustainable metabolic burden.

#### *D. Localization and Transport – Engineering in Multiple Dimensions*

Many alkaloid biosynthetic pathways are highly compartmentalized at both the inter- and intracellular levels. For example, at least three cell types are required for the biosynthesis of many MIAs.<sup>25</sup> The impact that localization has on product yields is not currently well understood. The forward engineering of plant natural products will require sifting through the increasing amount of available sequencing and expression data and untangling the complexity of the plant cell and different tissue types. In addition to the linear design and channeling of metabolic pathways, the successful metabolic engineering of plant natural products will require engineering in the “third dimension,” namely at the level of localization and cell type.<sup>26</sup>

#### *E. Physiological Relevance of Alkaloids*

It is not entirely clear what role the alkaloids have evolved to fulfill, though it is commonly postulated that the alkaloids are defense compounds that protect the plant. Certainly that is a sensible hypothesis given that most alkaloids exhibit some degree of bioactivity. This hypothesis could potentially be tested first by studying how generalist herbivores feed on alkaloid-containing plants versus engineered plants devoid of alkaloids (e.g. the *C. roseus* tryptophan decarboxylase RNA interference line).<sup>27</sup> Increased feeding on lines devoid of alkaloids would suggest that the alkaloids are plant defense compounds. Secondly, introducing a non-canonical amino acid or radio-labeled tag into the plant may permit the direct study of what proteins are expressed in response to insect feeding. Notably, this strategy may also require that the plant's amino acyl tRNA synthetases be reengineered to accommodate these non-canonical amino acids. Therefore, comparative transcriptomic, proteomic and metabolomic analyses may be cheaper, faster and easier, but potentially less direct. Up-regulation of known alkaloid

biosynthetic enzymes and alkaloid production would strongly implicate alkaloids in plant defense strategies.

#### *F. Combinatorial Biosynthesis in Plants – Mixing and Matching Pathways and Engineering New Enzyme Specificities*

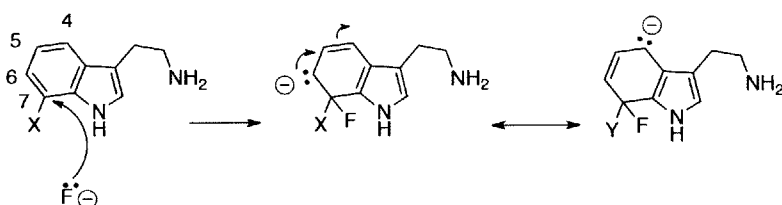
*De novo* combinatorial biosynthesis in plant systems has gone largely underexplored. Most of the few efforts to engineer unnatural natural products have utilized precursor feeding or mutasynthesis-based approaches, which can be costly and time intensive.<sup>5,27</sup> The *de novo* biosynthesis of unnatural natural products will require that constituent enzymes have reengineered or broad specificity. Notably, directed evolution has been successfully used to alter enzyme specificity. However, enzyme-engineering efforts are greatly enhanced if protein structure is known and the mechanism is well understood. Then, the enzyme can be subjected to structure-guided techniques, such as site-directed mutagenesis and domain swapping, which create smaller protein libraries enriched with functional mutants.

Halogenation—particularly fluorination—is a typical lead editing strategy.<sup>28</sup> Despite its value, chemical halogenation often suffers from low regio- and stereospecificity and is oftentimes non-catalytic.<sup>27</sup> Moreover, these reactions require harsh, anhydrous reaction conditions.<sup>27</sup> Notably, however, Halex reactions can be performed at room temperature, but still require anhydrous reaction conditions. For these reasons, enzymatic fluorination is highly attractive as it presumably would require only the addition of simple salts, and the reaction can be run under aqueous conditions.<sup>27,28</sup> Though fluoride ions are potent nucleophiles, they are tightly solvated by water, which makes them effectively inert in aqueous media.<sup>29</sup> The binding of substrate SAM in the fluorinase (5'-fluoro-5'-deoxyadenosine synthase) from *Streptomyces*

*cattleya*—the only as yet characterized fluorinase—is predicted to desolvate fluoride and enable fluoride's S<sub>N</sub>2 displacement of methionine on the SAM substrate.

No nucleophilic aromatic fluorinase has yet been reported. To the best of our knowledge, no naturally occurring aromatic fluorometabolites have been reported either, suggesting that the enzymatic repertoire for nucleophilic aromatic fluorination either does not exist in nature or is very rare. Because of our interest in medicinally useful MIAs, we include as a future challenge the development of a nucleophilic aromatic fluorinase that

#### A. Design of Substrate Activated for Nucleophilic Aromatic Fluorination



Sigma complex demonstrating why electron withdrawing groups should be placed at positions 6 (middle) and 4 (right) to balance charge

#### B. Enzymatic Nucleophilic Aromatic Fluorination

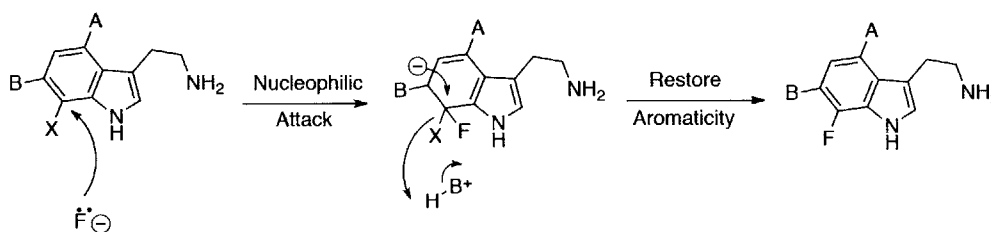


Figure 5.4: A. Design of Substrate Activated for Nucleophilic Aromatic Fluorination. The Sigma complex suggests that electron withdrawing groups should be placed at positions 6 and 4, which are ortho and para to position 7, respectively. Position 7 has been selected as the position of fluorination because RebH—the well-characterized enzyme we propose to reengineer—regioselectively chlorinates at that position, though notably through an electrophilic aromatic substitution mechanism. B. Enzymatic nucleophilic aromatic fluorination mechanism. The first step is nucleophilic attack followed by a restoration of aromaticity. X represents a good leaving group whereas A and B represent electron withdrawing groups.

accepts tryptamine 11, a direct MIA precursor. It is likely that if such an enzyme can be engineered, early engineered versions would require tryptamine analogs strongly activated for nucleophilic aromatic substitution (Figure 5.4). Therefore, we propose as one possible strategy the synthesis of a tryptamine analog with a good leaving group at position 7 (the proposed site of fluoride substitution), and strong electron withdrawing groups at positions 6 and 4, which are para and ortho to position 7, respectively, to help distribute the negative charge upon nucleophilic attack (Figure 5.4). After synthesizing a library of activated tryptamine analogs, RebH—a 7-tryptophan chlorinase that operates via electrophilic aromatic substitution—would need to be reengineered to function instead via nucleophilic aromatic substitution to accept these substrates. Alternatively, since the structure and mechanism of 5'-fluoro-5'-deoxyadenosine synthase are known and relatively well understood, it may be plausible to reengineer this fluorinase to accept non-native, aromatic substrates.

As we seek further to convert plants and microbes into the chemical factories to meet our medicinal needs, we should remember that, although many plant natural products are bioactive and serve as important lead compounds, they often require modification before making it to the clinic. Therefore, the forward engineering of “unnatural” or “new-to-nature” natural products must also be a grand challenge if plant natural products are to be shuttled from the annals of human tradition into the drug development programs and clinics of tomorrow.

### III. References

1. Hagel JM and Facchini PJ, Dioxygenases Catalyze the O-demethylation Steps of Morphine Biosynthesis in Opium Poppy. *Nat. Chem. Biol.* 2010, 6, 273-275.
2. [A] Collu, G.; Unver N.; Peltenburg-Looman, A.M.; van der Heijden, R.; Verpoorte, and Memelink, J., Geraniol 10-hydroxylase, a Cytochrome P450 Enzyme Involved in Terpenoid Indole Alkaloid Biosynthesis. *FEBS Lett.* 2001, 508 (2), 215-20. [B] Guirimand, G.; Burlat, V.; Oudin, A.; Lanoue, A; St-Pierre, B. and Courdavault, V., Optimization of the Transient Transformation of

- Catharanthus roseus* cells by Particle Bombardment and its Application to the Subcellular Localization of Hydroxymethylbutenyl-4-diphosphate Synthase and Geraniol 10-hydroxylase. *Plant Cell Rep* 2009, 28(8), 1215-34.
3. Geu-Flores, F.; Sherden, N.H.; Courdavault, V.; Burlat, V.; Glenn, W.S.; Wu, C.; Nims, E.; Cui, Yuehua and O'Connor, S. E., An Alternative Route to Cyclic Terpenes by Reductive Cyclization in Iridoid Biosynthesis. *Nature* 2012, 492 (7427), 138-42.
  4. McCoy, E. and O'Connor SE: Directed Biosynthesis of Alkaloid Analogs in the Medicinal Plant *Catharanthus roseus*. *J Am Chem Soc.* 2006, 128, 14276-7.
  5. Runguphan, W.; Qu, X. and O'Connor, S.E., Integrating Carbon-Halogen Bond Formation into Medicinal Plant Metabolism. *Nature* 2010, 18, 461-4.
  6. Glenn, W.S.; Nims, E. and O'Connor, S.E., Reengineering a Tryptophan Halogenase to Chlorinate a Direct Alkaloid Precursor. *J Am Chem Soc.* 2011, 133, 19346-9.
  7. Runguphan W, Glenn WS and O'Connor S. E: Redesign of a dioxygenase in morphine biosynthesis. *Chem Biol.* 2012, 19, 674-678.
  8. Giddings, L.-A.; Liscombe, D.K.; Hamilton, J.P, Childs, K. L.; DellaPenna, D.; Buell, C. R.; O'Connor, S. E., A Stereoselective Hydroxylation Step of Alkaloid Biosynthesis by a Unique Cytochrome P450 in *Catharanthus roseus*. *J Biol Chem.* 2011, 286, 16751-7.
  9. Winzer, T.; Gazda, V.; He, Z.; Kaminski, F.; Kern, M.; Larson, T. R.; Li, Y.; Meade, F.; Teodor, R.; Vaistij, F. E. *et al.*: A *Papaver somniferum* 10-gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. *Science* 2012, 336, 1704-8.
  10. Liscombe, D. K.; Usera, A. R. and O'Connor, S. E., Homolog of Tocopherol C Methyltransferases catalyzes N Methylation in Anticancer Alkaloid Biosynthesis. *Proc Natl Acad Sci U S A* 2010, 107: 18793-18798.
  11. Bradbury, L.M.T.; Niehaus, T.D. and Hanson, A.D., Comparative Genomics Approaches to Understanding and Manipulating Plant Metabolism. *Curr Opin Biotech.* 2013, 24(2), 278-284.
  12. Liscombe, D. K.; O'Connor, S. E.; A Virus-Induced Gene Silencing Approach to Understanding Alkaloid Metabolism in *Catharanthus roseus*. *Phytochemistry* 2011, 72: 1969-1977.
  13. Runguphan, W.; O'Connor, S. E., Metabolic reprogramming of periwinkle plant culture. *Nat Chem Biol.* 2009, 5, 151-153.
  14. Peretz, Y.; Mozes-Koch, R.; Akad, F.; Tanne, E.; Czosnek, H. and Sela, I., A Universal Expression/ Silencing Vector in Plants. *Plant Physiol.* 2007, 145: 1251-1263.
  15. Park, S.-U.; Yu, M. and Facchini, P. J., Modulation of Berberine Bridge Enzyme Levels in Transgenic Root Cultures of California Poppy Alters the Accumulation of Benzophenanthridine Alkaloids. *Plant mol biol.* 2003, 51, 153-64.
  16. Park, S.-U. and Facchini, P.J., Antisense RNA-Mediated Suppression of Benzophenanthridine Alkaloid Biosynthesis in Transgenic Cell Cultures of California Poppy. *Plant Physiol.* 2002, 128, 696-706.
  17. Weeks, A. M. and Chang, M. C. Y., Constructing de novo Biosynthetic Pathways for Chemical Synthesis in Living Cells. *Biochemistry* 2011, 50, 5404-18.



18. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J. and Keasling, J. D., Synthetic Protein Scaffolds Provide Modular Control Over Metabolic Flux. *Nat Biotechnol.* 2009, 27, 753-759.
19. Naqvi, S.; Farré, G.; Sanahuja, G.; Capell, T.; Zhu, C. and Christou P: Where More is Better: Multigene Engineering in Plants. *Trends in Plant Science* 2009, 15, 48-56.
20. Nour-Eldin, H.; Geu-Flores, F. and Halkier, B. A., USER cloning and USER fusion: *The Ideal Cloning Techniques for Small and Big Laboratories.* in *Plant Secondary Metabolism Engineering, Methods in Molecular Biology.* Edited by Fett-Netto AG. Springer Science + Business Media, LLC; 2010:185-200.
21. Mozes-Koch, R.; Gover, O.; Tanne, E.; Peretz, Y.; Maori, E.; Chernin, L., and Sela, I, Expression of an Entire Bacterial Operon in Plants. *Plant Physiol.* 2012, 158: 1883-1892.
22. Runguphan, W.; Maresh, J. J. and O'Connor, S. E., Silencing of Tryptamine Biosynthesis for Production of Nonnatural Alkaloids in Plant Cultures. *Proc Natl Acad Sci. U S A* 2009, 106: 13673-8.
23. Runguphan, W. and O'Connor, S. E., Metabolic Reprogramming of Periwinkle Plant culture. *Nat Chem Biol.* 2009, 5:151-153.
24. Chang, A. I.; Wolf, J. J. and Smolke, C. D., Synthetic RNA Switches as a Tool for Temporal and Spatial Control Over Gene Expression. *Curr Opin Biotechnol* 2012, 23; 1-10.
25. Murata, J.; Roepke, J.; Gordon, H. and De Luca, V., The leaf Epidermome of *Catharanthus roseus* Reveals its Biochemical Specialization. *Plant Cell* 2008, 20, 524-542.
26. Agapakis, C. M.; Boyle PM, Silver PA: Natural Strategies for the Spatial Optimization of Metabolism in Synthetic Biology. *Nat Chem Biol.* 2012, 8, 527-35.
27. Watson, D.A.; Su, M.; Teverovski, G.; Zhang, Y.; Garcia; Fortanet, J.; Kinzel, T.; Buchwald, S.L., Formation of ArF from LPdAr(F): Catalytic Conversion of Aryl Triflates to Aryl Fluorides. *Science* 2009, 325, 1661-4.
28. Sun, H. and DiMugno, S.G., Room-Temperature Nucleophilic Aromatic Fluorination: Experimental and Theoretical Studies. *Angew. Chem. Int. Ed.* 2006, 45, 2720-5.
29. Dong, C.; Huang, F.; Deng, H.; Schaffrath, C.; Spencer, J.B.; O'Hagan, D. and Naismith, J.H., Crystal Structure and Mechanism of a bacterial Fluorinating Enzyme. *Nature* 2004, 427, 561-5.

#### IV. Acknowledgments

Parts of this chapter are based on an opinion article published in *Curr. Opin.*

*Biotech.* Dr. Weerawat Runguphan and I collaboratively wrote this article.

## Appendix A - Chapter 2

### Discovery of 10-hydroxygeraniol Oxidoreductase Activity in *C. roseus*

I. Candidate DNA and Protein Sequences – Page 166

II. TLC Controls – Page 172

III. GC-MS Spectra – Page 174

#### Candidate 1786 open reading frame sequence from clone

```
ATGCAGATCATAACTTGCAAGGCTGTGGTGTGCTGGGCGGCCGGAGAGCCACCGGTGGT
TGAGGAGATACTGGTAGAACCTCCGAGGTCAGGCGAAGTCAGGATTAAGATTTTGTTTG
CTAGTCTTTGCCACACTGATGTCTCGCTGCAAGGGCTTCCCAACGCCCATGTTTCCT
CGAGTTCTGGGACATGAAGGTGTCGGCGTGGTGGAGTGTGTGGGTGAAGGAGTTTCAGA
ACTGAGAGAGGGAGACGTGGTGTATCCCCACATACTTGGGAGAATGCGGAGAATGTGAGA
ATTGTGAGTCAGGAAGAACGAATCTATGCCGAACCTACCCTTTGCAAGCATTACAGGC
TTAATGCCTGATGGTTCCCTCAAGAATGTCTTCCGCCAAAGGAGGGGAAATGTTGTACCA
ATTCTTAGCTGCTCCACTTGGTCTGAGTATACTGTTATTGACGCCAACTATGCCGTGA
AGATAGACTCCAGAATACCTCTGCCCATGCTAGCTTCCTTTCTTGCGGCTTCACCACT
GGGTTTGGGGCAACCTGGAAGGAAGCCAAGCTTCAAGAGGGATCCAGCACCGTTGCTGT
TCTGGGTCTTGGGGCAGTTGGACTTGGAGCTGTGGAGGGAGCTCGAGTGCAGGGAGTAA
CTCAAATAATAGGAATAGACATTAACGACAACAAACGTGAGAAAGGAGAAGCCTTCGGA
ATGACTCATTTTCATCAACCCCAAAAAAGATAATAATAAATCCATTTTCAGAATTAGTTAA
AGAGTTAACAAAAGGACAAGGTGTGGACGTCTGTTTTGAATGCACGGGAGTCCCTGACT
TGGTTAATGAAGCTCTTGAATCCACAAAGATCGGAACAGGAAATATGATAATGCTAGGA
GCAGGAACCCAGAAAAGCATGACCATAAACTTCGTTTTACTATTGGGCTGCAGAACTTT
CAAGTATTCTGTTTTTCGGCGGGGTTAAGGTCCAATCCGACCTTCCTCTCATTATTCAGA
AATGCTTAAATAAGGAAATACAGAAAATTGAGCAGCTTTTAACTCATCAAGTTCAACTG
GAAGACATAAATAGAGCCTTTGAGCTGCTTAAGGAACCTGATTGCGTGAAGGTTCTCAT
CACATTGTGA
```

#### Candidate 1786 protein sequence

```
M Q I I T C K A V V C W A A G E P P V V E E I L V E P P R S
G E V R I K I L F A S L C H T D V L A C K G F P T P M F P R
V L G H E G V G V V E C V G E G V S E L R E G D V V I P T Y
L G E C G E C E N C E S G R T N L C R T Y P L Q A F T G L M
P D G S S R M S S A K G G E M L Y Q F L S C S T W S E Y T V
I D A N Y A V K I D S R I P L P H A S F L S C G F T T G F G
A T W K E A K L Q E G S S T V A V L G L G A V G L G A V E G
A R V Q G V T Q I I G I D I N D N K R E K G E A F G M T H F
I N P K K D N N K S I S E L V K E L T K G Q G V D V C F E C
T G V P D L V N E A L E S T K I G T G N M I M L G A G T Q K
S M T I N F V S L L G C R T F K Y S V F G G V K V Q S D L P
```

L I I Q K C L N K E I Q K I E Q L L T H Q V Q L E D I N R A  
F E L L K E P D C V K V L I T L Stop

*Residues predicted to be involved in zinc binding are shown in red, bolded and underlined.*

**Candidate 26 open reading frame sequence from clone**

ATGGCCGCCATGGGTACCTCGGAAAAATATGCAGTTGTTACGGGATCAAACAAAGGTAT  
TGGATTTGAAACCTGCAAGAAATTGGCTTCTCAAGGGATCACTGTGGTCCTTACTGCTA  
GAGATGAAAAAAGAGGGCTCGATGCTCTTGAGAAGCTCAAAGAATTGGGTCTCTCTGGT  
AAGGTGCTATTTTCATCAGCTTGATGTGACCGATTTCATCCAGCGTTGCTTCCCTTGCAGA  
ATTTGTCAAGAAACAATTTGGAAGACTTGATATCTTGGTAAACAATGCAGGGGTTAATG  
GAGTGATTACTGATGTTGAAGCTGTGAAAAAGCTAAATCCTGCAGAAGATCCGGCCGAT  
GTCGACTTTAGCAAGATATAACAAGGAAACATATGAGTTGGCTGAAGAATGCATTCAAAT  
TAACTACTTTTGAACAAAAAGAACCCTGATGCACCTTCTTCCCTCTTCTCCAATTATCTG  
CATCACCAAGAATCGTAAATATTTCCCTCCATCATGGGACAGTTAAAGAACATAACCAAGT  
GAATGGGCTAAAGGAATCCTGGGAGATGCTAGCAACCTCACAGAAGATAGATTGGATGA  
GGTGATCAATAACTTCTTGAAGGACTTCAAGGAAGGATCCCTTGCAGCTAAGGGATGGC  
CTCCATCCTTTTTAGCTTATATAGTCTCAAAGTTGTGGTGAATGCCTACACAAGAATT  
CTGGCCAAGAAGTATCCCAATTTCAAGATCAATTGTGTTTGTCCAGGGTTTGCCAAGAC  
AGATTTGAATCATGGCTTAGGCTTATTAAGTGCAGAAGAAGCTGCTGAAAACCCTGTGA  
AACTCGCTTTGCTGCCTGATGATGGTCCTTCGGGTTTGTTCCTTTGATCGCAGCGAGGAG  
TCATCGTTTGAATGA

**Candidate 26 protein sequence**

M A A M G T S E K Y A V V T G S N K G I G F E T C K K L A S  
Q G I T V V L T A R D E K R G L D A L E K L K E L G L S G K  
V L F H Q L D V T D S S S V A S L A E F V K K Q F G R L D I  
L V N N A G V N G V I T D V E A V K K L N P A E D P A D V D  
F S K I Y K E T Y E L A E E C I Q I N Y F G T K R T T D A L  
L P L L Q L S A S P R I V N I S S I M G Q L K N I P S E W A  
K G I L G D A S N L T E D R L D E V I N N F L K D F K E G S  
L A A K G W P P S F S A Y I V S K V V V N A Y T R I L A K K  
Y P N F K I N C V C P G F A K T D L N H G L G L L T A E E A  
A E N P V K L A L L P D D G P S G L F F D R S E E S S F E  
Stop

**Candidate 4319 open reading frame sequence from clone**

ATGGCCAGAAAATCACCAGAAGATGAACATCCCGTGAAGGCTTACGGATGGGCCGTCAA  
AGATGGAACAACCTGGAATTCTTTCTCCCTCAAATTTTCCATAAGGGCAACAGGTGATA  
ATGATGTTCGAATCAAGATCCTCTATTGTGGAGTTTGTTCGTACCGATCTTGGCGCAACC  
AAGAACGCATTCGGGTTTCTTTCTTATCCTCTTGTGCCTGGTAGAGAGATCGTGGGAAT

AGTGAGCGAGATAGGGAAAAATGTGAAAAAAGTTAAAGTTGGAGAAAAAGTTGGAGTAG  
CCCCGCATGTGGGTAGCTGTGGCAAATGCAAGAGTTGTGTGAATGAGGTGGAGAATTTTC  
TGTCCGAAACTGATCATCCCTTATGGCACCCATAACCACGATGGTACTATTTGCTACGG  
TGGTTTCTCCAACGAGACTGTCAGAGATGAACGCTTTGTTTTTCGTTTTCTGAAAATC  
TTTCGCTGCCTGGCGGAGCTCCCTTGGTTAGTGCTGGGGTTACCACGTACGGTGCATTG  
AGAAATAATGGCCTCGACAAGCCCGGATTACACGTGGGAGTCGTCCGTCTAGGTGGACT  
AGGTCATCTGGCTGTTAAATTTGCTAAGGCTTTAGGCGTCAAAGTAACTGTTATTAGTA  
CCAATCCTAGCAAGGAGCATGATGCTATAAATGGTTTCGGTGCTGATGCCTTCATCCTC  
ACCCACCATGAGGAACAAATGAAGGCTGCCATGGGAACTTTAGATGGAATTCCTTATAC  
AGTGCCTGTTGTTTCATGCCATTGCACCATTACTTAGTCTACTGGGAAGTCAAGGGAAAT  
TTGTGTTGATTGGGGCACCATCTCAATTACTTGAGGTGCCACCTATTCAATTATTATTT  
GGTGGAAAATCTATTATTGGAAGTGCGGCTGGAAATGTGAAGCAAATCCAAGAAATGCT  
TGAATTTGCAGCAAAACATGATATAATTGCGAATGTTGAGATTATCCAAATGGATTATA  
TAAATACTGCAATGGAACGCTAGACAAAGGTGATGTTAGATATCGATTTGTAATTGAT  
ATCGAAAACCTCTCTCACTCTTCCATCAGAGGTGTGA

### Candidate 4319 protein sequence

M A R K S P E D E H P V K A Y G W A V K D G T T G I L S P F  
K F S I R A T G D N D V R I K I L Y C G V C R T D L A A T K  
N A F G F L S Y P L V P G R E I V G I V S E I G K N V K K V  
K V G E K V G V A P H V G S C G K C K S C V N E V E N F C P  
K L I I P Y G T P Y H D G T I C Y G G F S N E T V R D E R F  
V F R F P E N L S L P G G A P L V S A G V T T Y G A L R N N  
G L D K P G L H V G V V G L G G L G H L A V K F A K A L G V  
K V T V I S T N P S K E H D A I N G F G A D A F I L T H H E  
E Q M K A A M G T L D G I L Y T V P V V H A I A P L L S L L  
G S Q G K F V L I G A P S Q L L E V P P I Q L L F G G K S I  
I G S A A G N V K Q I Q E M L E F A A K H D I I A N V E I I  
Q M D Y I N T A M E R L D K G D V R Y R F V I D I E N S L T  
L P S E V Stop

### Candidate 5743 open reading frame sequence from clone

ATGACCAAGACCAATTCCCCTGCTCCTTCTGTCACTTGCAAAGCTGCTGTGGTATG  
GAAATCAGGGGAGCCACCAAAGGTGGAAGAGATAACAAGTTGATCCACCAAAGGCCTCAG  
AAGTTAGGATTAAGATGCTTTGTGCCAGTTTGTGCCACACTGATTTCCCTTGCTGCAAT  
GGCCTTCCTGTTCCATTGTTCCCTCGCATTCCCTGGACACGAAGGAGTCGGAATGATCGA  
GAGCGTTGGAGAAAATGTTACGAACCTAAAAGAAGGAGACATAGTGATGCCACTTTACT  
TGGGAGAATGTGGGGAATGCTTGAATTGCAAATCAGGAAGGACAAATTTGTGCCACAAA  
TATCCTTTAGGTTTTAGTGGATTATTGCTTGATGGAACATCAAGAATGTCAATTGGAGA  
GCAAAAAGTATATCACCATTTAGCTGTTTCGACATGGTCAGAGTATATAGTGATTGAAG  
CAGCTTATGCAGTGAAGGTAGATCCAAGGGTTTCTCTTCCACATGCTAGCTTTCTTTGC  
TGTGGATTACCACTGGTTTTGGTGCTACTTGGAGAGATGTCAATGTTGTCAAGGGCTC  
TACTGTGGCTGTTCTAGGCCTTGGTGCTGTTGGACTTGGGGCTGTGCAAGGAGCTAAAT  
CACAAGGAGCATCAAGAATTATAGGTTTGGATATCAACGACAAGAAACGTGAAAAGGGA  
GAAGCATTGGAATGACTGAATTTATAAATCCAAAAGGTTCAAACAAATCAATTTCTGA



ACTAATAAATGAAGCAACTGGTGGATTAGGACTTGACTATGTCTACGAATGCACTGGAG  
TTCCAGCTTTACTCAATGAAGCCATTGAATCCTCTAAAGTGGGACTTGGAACTGCAGTA  
TTGATTGGGGCAGGACTTGAAACAAGTGGAGAAATCAAATTCATTCCCTCTTCTGTGTGG  
TAGAACTGTTAAAGGTTCAATTTATGGTGGAGTAAGACCTAAATCAGACCTCCCCACTT  
TAATTGAAAAGTGCATAAACAAGGAAATTCATGGATGAACTAATGACTCATGAAGTT  
TCATTGTCTGAGATAAACAAGGATTTGAGTACCTTAAGCATCCTGATTGTGTCAAAGT  
TGTTATCAAGTTCTAA

### Candidate 5743 protein sequence

M T K T N S P A P S V I T C K A A V V W K S G E P P K V E E  
I Q V D P P K A S E V R I K M L C A S L **C** H T D F L A C N G  
L P V P L F P R I P G **H** E G V G M I E S V G E N V T N L K E  
G D I V M P L Y L G E **C** G E **C** L N **C** K S G R T N L C H K Y P  
L G F S G L L L D G T S R M S I G E Q K V Y H H F S C S T W  
S E Y I V I E A A Y A V K V D P R V S L P H A S F L C C G F  
T T G F G A T W R D V N V V K G S T V A V L G L G A V G L G  
A V Q G A K S Q G A S R I I G L D I N D K K R E K G E A F G  
M T E F I N P K G S N K S I S E L I N E A T G G L G L D Y V  
Y E C T G V P A L L N E A I E S S K V G L G T A V L I G A G  
L E T S G E I K F I P L L C G R T V K G S I Y G G V R P K S  
D L P T L I E K C I N K E I P M D E L M T H E V S L S E I N  
K G F E Y L K H P D C V K V V I K F Stop

*Residues predicted to be involved in zinc-binding are shown in red, bolded and underlined.*

### Candidate 7220 open reading frame from clone

ATGGAGATTAATGTTGAAGTTGCTCCAGTAAGGTATGCAGTCGTTACTGGAGCAAACAA  
GGGCATTGGTCTTGAGACTGTCCAACAGCTAGCAGCCTCAGGTGTGACTGTCTGTGTTAA  
CAGCTCGAAATGAGAAGAGAGGCATGGAGGCCACTTCTTTGCTGCATGAATCAGGTTTG  
TCAAATGTAATTTTCCATCAGCTTGATGTTCAAGACAAAGAAAGCATCAAATCATTGGC  
CGAGTTTATACAAAAAGAGTTTGGAAAGACTTGACATTTTGGTAAATAATGCCGGTGCTT  
CAGGAGTGGCAGTTGACAAAGATGGACTAAGGGCCTTAAATATAGACACTGCATCTTGG  
TTAGCCGAAAGGTTGTTAATGTGGTAGCTGATGTAATCAAACACTACATATGAAAAAGC  
CAAAGAATGTCTGGAGACCAACTATTATGGTGTAAAGGATGTAACCGAAGCTCTTCTTC  
CACTGCTGCAACTTTCAACTTCAGGAGCAAGGATTGTAAACGTCTCTTCTCTAAGGAGT  
GAATTAAGTAGGATCCCTAACGAGCAAAGAAGAAAAGTACTAGCGGATATTGAGACTCT  
AACAGAAAACAAAATCAATGAGATTCTTCAACAGTTTTTGCATGATCTGAAGCATGATG  
CTTTAGAAGCAAACGGATGGCAAAGATGTTGCCGGCCTACAGCATATCAAAGCAACA  
CTTAATGCTTATACAGAATTCTTGCAAAGAAGTATCCTCATATGTGTATAAACTGTGT  
TCATCCGGGATATGTTAACACAGACATCAATGGCATAACAGGACCATTGCCGGTGGAAG  
AGGGAGCTGCCGGACCTGTTATGCTGGCTCTTCTACCAGAAGGAGGTCTACTGGTTGC  
TACTTTGATCAGACAAAACACTATCCGAGTTTTGA

### Candidate 7220 protein sequence

M E I N V E V A P V R Y A V V T G A N K G I G L E T V Q Q L  
 A A S G V T V V L T A R N E K R G M E A T S L L H E S G L S  
 N V I F H Q L D V Q D K E S I K S L A E F I Q K E F G R L D  
 I L V N N A G A S G V A V D K D G L R A L N I D T A S W L A  
 G K V V N V V A D V I K T T Y E K A K E C L E T N Y Y G V K  
 D V T E A L L P L L Q L S T S G A R I V N V S S L R S E L S  
 R I P N E Q R R K V L A D I E T L T E N K I N E I L Q Q F L  
 H D L K H D A L E A N G W Q K M L P A Y S I S K A T L N A Y  
 T R I L A K K Y P H M C I N C V H P G Y V N T D I N W H T G  
 P L P V E E G A A G P V M L A L L P E G G P T G C Y F D Q T  
 K L S E F Stop

**Candidate 8694 open reading frame from clone**

ATGACGTCGTCATCCTCGCCGTCGCCGTTGAAGGGAAAAGCTGTGGATAAAGA  
 CGGCGATCACAAGGTGAAGAAGAAAGAGGCATTAGGATGGATGGAGTGGCTTAGAGGAT  
 GGATGTATATAGTGTACGAAATGCTGTTCCAGCGGATCATGGCCAGTCATTTATCTAAT  
 CCAATGCCTCTTCCGCCTCTGAATGAGCTTACTTTTGTAGTCACCGGCTCCACCAGCGG  
 TATTGGCCGCGAAATCGCCGTC AATTGGCAGAGTCCGGCGGGCACGTGATAATGGCTG  
 TTAGAAATACCAAGGCAGCTAATGAATTAATTCGCAAATGGCAAGAGGAATGGTCTGGT  
 CGCGGACTACCTCTTAATATTGAGGTGATGGAGCTGGATCTTCTATCATTGGATTCCGGT  
 TGTGAGATTTGCTGAGGCATTTAACGCACGTTCCGGACCTTTGAATGTGCTCATTAAACA  
 ATGCTGGCATATTTTCAATCGGAGAACCACAGAGGTTTTCAAAGGATGGTTATGAAGAA  
 CACCTGCAAGTGAATCATCTAGCTCCAGCACTGTTGTCTATATTGCTCTTACCTTCTCT  
 TATTAGAGGCTCTCCAAGCCGAATAGTTAATGTGAACTCTATAATGCATTATGTTGGAT  
 TTGTTGATACGGAAGATATGAATGTTACATCTGGGAGAAGAAAGTACAGCAGTTTAGTT  
 GGATACTCTGGCAGCAAACCTGGCAGAGGTGATGTTTCAGTAGTGTCTGCACAAACGGCT  
 GCCTGCCGAATCTGGCATAAGTGTACTATGCGTATCGCCTGGAATAGTACACACAAATG  
 TGGCTAGGGATCTTTCAAAAATTGTTCAAGCTGCTTATCATCTAATTCCTATTTTTATT  
 TTTAGTCCTGAAGAAGGCTCTAGAAGCGCACTTTTTGCAGCTACAGATCCACAAGTTCC  
 GGAGTACACTGAGATGTTAAAAGCAGATGAGTGGCCAGTTTGTGCTTTCATATCTCAAG  
 ATTGCCGTCCAACAAATCCATCTGAAGAAGCACATAATGTTGAAACTTCTTACAAAGTC  
 TGGGAGAAGACCTTGAAATGGTTGGACTTCCATCAGATGTTGTGGAGAAGCTTATAGA  
 AGGGGAAGAAGTTAAATGCAGATATGGGGCTTCTCAAGAGTAA

**Candidate 8694 protein sequence**

M T S S S S P S P S P L K G K A V D K D G D H K V K K K E A  
 L G W M E W L R G W M Y I V Y E M L F Q R I M A S H L S N P  
 M P L P P L N E L T F V V T G S T S G I G R E I A R Q L A E  
 S G G H V I M A V R N T K A A N E L I R K W Q E E W S G R G  
 L P L N I E V M E L D L L S L D S V V R F A E A F N A R S G  
 P L N V L I N N A G I F S I G E P Q R F S K D G Y E E H L Q  
 V N H L A P A L L S I L L L P S L I R G S P S R I V N V N S  
 I M H Y V G F V D T E D M N V T S G R R K Y S S L V G Y S G  
 S K L A E V M F S S V L H K R L P A E S G I S V L C V S P G  
 I V H T N V A R D L S K I V Q A A Y H L I P Y F I F S P E E  
 G S R S A L F A A T D P Q V P E Y T E M L K A D E W P V C A

F I S Q D C R P T N P S E E A H N V E T S Y K V W E K T L E  
M V G L P S D V V E K L I E G E E V K C R Y G A S Q E Stop

### Candidate 2041 open reading frame from clone

ATGGGATACTACCATTATTATATTAGACAACCACTCACCCTGACCAATTGGTTTTATC  
TCTTCCTTCTATAATGGCAGTTCATCGGCAGAAACAGCAAAGACAATCGAGGCCTATG  
GATGGGCAGCCAGAGACTCATCTGGGCTTCTCTCTCCCTTCAAGTTCAGAGACGGGCC  
ACAACGGAGCATGATGTCCAGCTCAAAATATTGTATTGTGGGATGTGCGATTGGGATCT  
ACATGTAGTCAAGAATTGGTTTGGCACCACCAACTATCCCATTGTACCTGGGCACGAGG  
CAGTGGGCGTGGTGAAGTAAATCGGCAACAAGGTACAGAAATTCAAGATTGGGGACATA  
GTAGGCGTTAGTACTTACATTCGAACATGTCCGAGCTGCGAGAGATGTAAAGAAGGTGA  
AGACAGTTACTGTCCCAGCTTAATAACAGGAGATGGAAGTTCATTTAGTGTGGAAAAG  
ATGCATTTTTCTATGATCCAAATGATGATAATAACAAAAGAGACAACAAAAACATATGGC  
TCATATTCCAATTCACAGTTGTGGATGAATATTACGTTATTTCGTTGGCCAGAAAACCTT  
TCCTTTGGCTGCTGGAGTACCTCTTCTTTGTGCTGGTACAGTTCCTTATAGTCCAATGA  
GGCACTTTGGATTTGATAAACCTGGAATTCATATTGGTGTGGTTGGATTTGGTGGGATT  
GGCAAATTAGTTGTTAAATTTGCTAAGGCTTTTGGAGTTAAAGTAACAGTGATTAGTAC  
CTCCATTGATAAGAAGCATGAAGCTATTCATGAATATGGTGCTCATGGATTCTTACTCA  
GCAAAGAACCTCAGCAGCTTCAGGCTGCTATTAATACTATGGAAGGTATAGTTGATACA  
GTTCCCTAAAGTTCACCCTATTCTTCCATTGATCAAATTTGTTGAAATTCGATGGTACCCT  
TCTTATGCTCGGAGCACCGCCGGAGCCATATGAGTTTCCAATCTCCACATTGCTTATGG  
GGAGGAAGAGGGTGGTGGGAAGTGCTGGAGCGAGCATGAAGGAAACACAAGAAATGATG  
GATTTTGCAGCGAAGCACAAACATAGTTGCAGATGTTGAATTAACCTCAGCAAGCTTGC  
GGCCGCACTCGAGCACCACCACCAACCACTGAGATCCGGCTGCTAA

### Candidate 2041 protein sequence

M G Y Y H Y Y I R Q P L T T D Q L V L S L P S I M A V P S A  
E T A K T I E A Y G W A A R D S S G L L S P F K F Q R R A T  
T E H D V Q L K I L Y C G M C D W D L H V V K N W F G T T N  
Y P I V P G H E A V G V V T E I G N K V Q K F K I G D I V G  
V S T Y I R T C R S C E R C K E G E D S Y C P S L I T G D G  
T S F S D G K D A F F Y D P N D D N T K E T T K T Y G S Y S  
N F T V V D E Y Y V I R W P E N F P L A A G V P L L C A G T  
V P Y S P M R H F G F D K P G I H I G V V G F G G I G K L V  
V K F A K A F G V K V T V I S T S I D K K H E A I H E Y G A  
H G F L L S K E P Q Q L Q A A I N T M E G I V D T V P K V H  
P I L P L I K L L K F D G T L L M L G A P P E P Y E F P I S  
T L L M G R K R V V G S A G A S M K E T Q E M M D F A A K H  
N I V A D V E L N L S K L A A A L E H H H H Q P L R S G C  
Stop

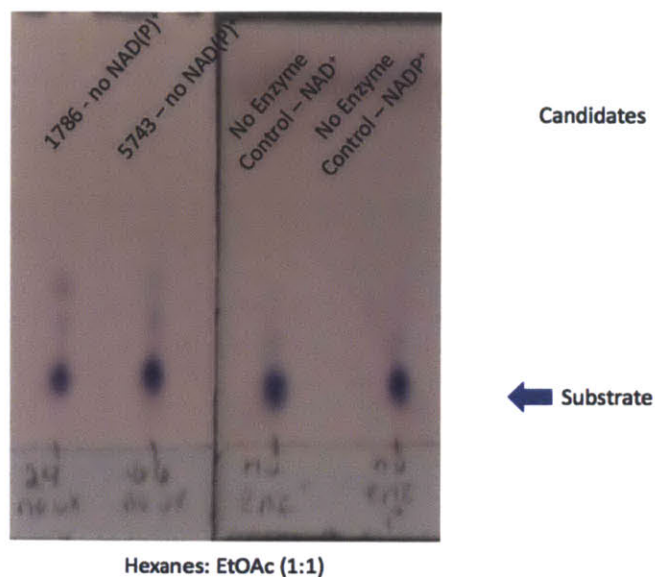


Figure A.1 – No NAD(P)<sup>+</sup> and no enzyme controls. No product formation is observed when either enzyme or NAD(P)<sup>+</sup> is missing from the assay mixture.

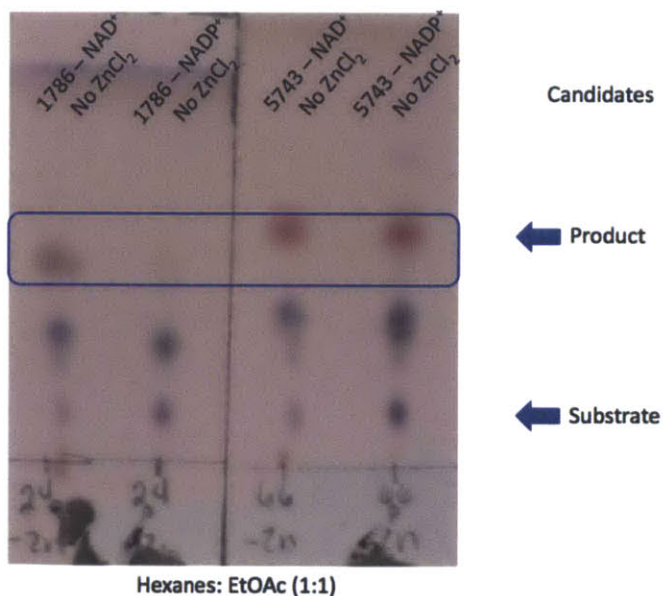


Figure A.2 – No zinc controls. Product formation is still observed when zinc is omitted from the reaction, suggesting zinc binds during protein folding.



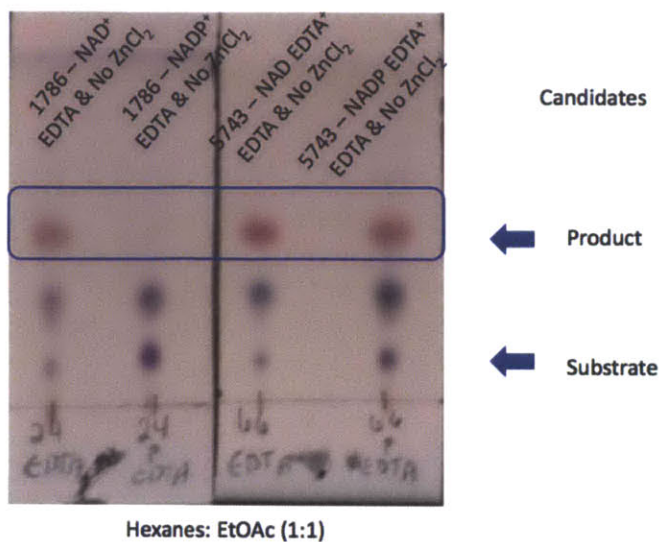


Figure A.3 – Chelator controls. 1 mM EDTA was added to the assay mixture and zinc was omitted. Product formation is only quelled with the Candidate 1786-NADP+ assay, suggesting the zinc binds less tightly under those conditions.

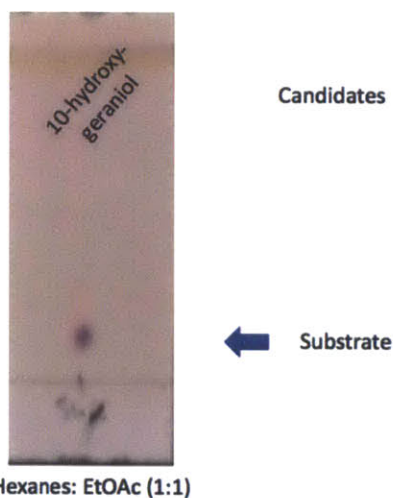


Figure A.4: Authentic standard of 10-hydroxygeraniol oxidoreductase substrate 10-hydroxygeraniol **8**.

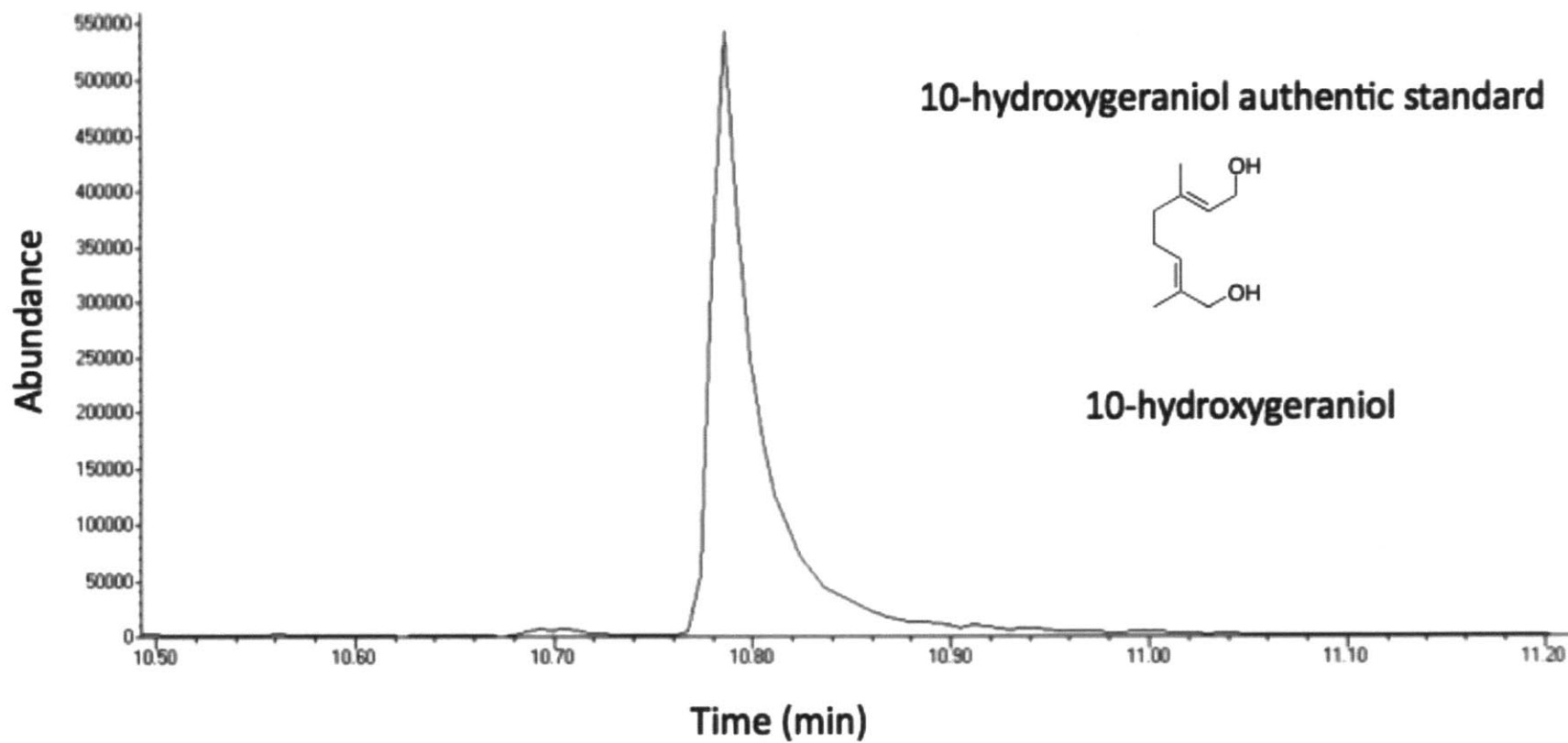


Figure A.5: GC-MS chromatogram of 10-hydroxygeraniol authentic standard. 10-hydroxygeraniol is the substrate for 10-hydroxygeraniol oxidoreductase.

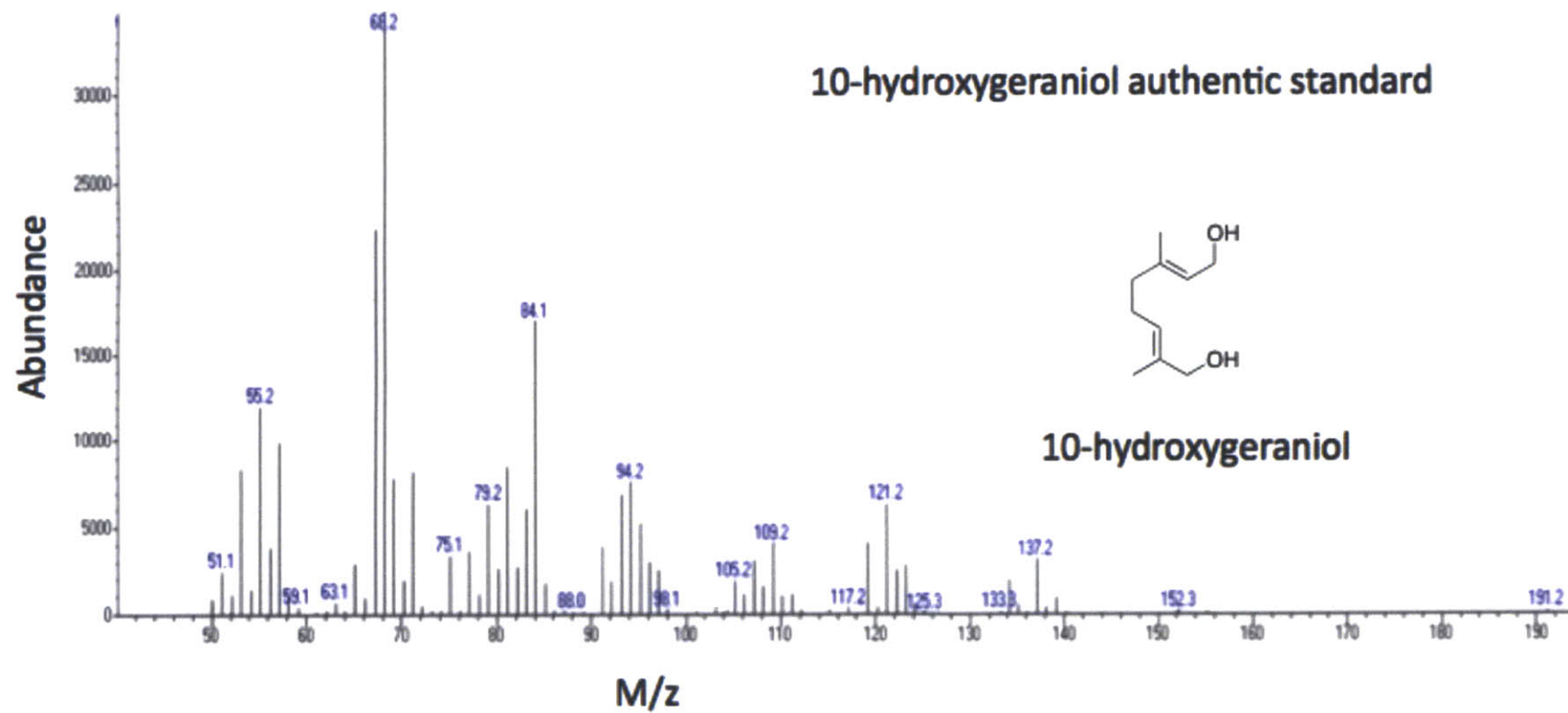


Figure A.6: Mass spectrum of 10-hydroxygeraniol authentic standard.

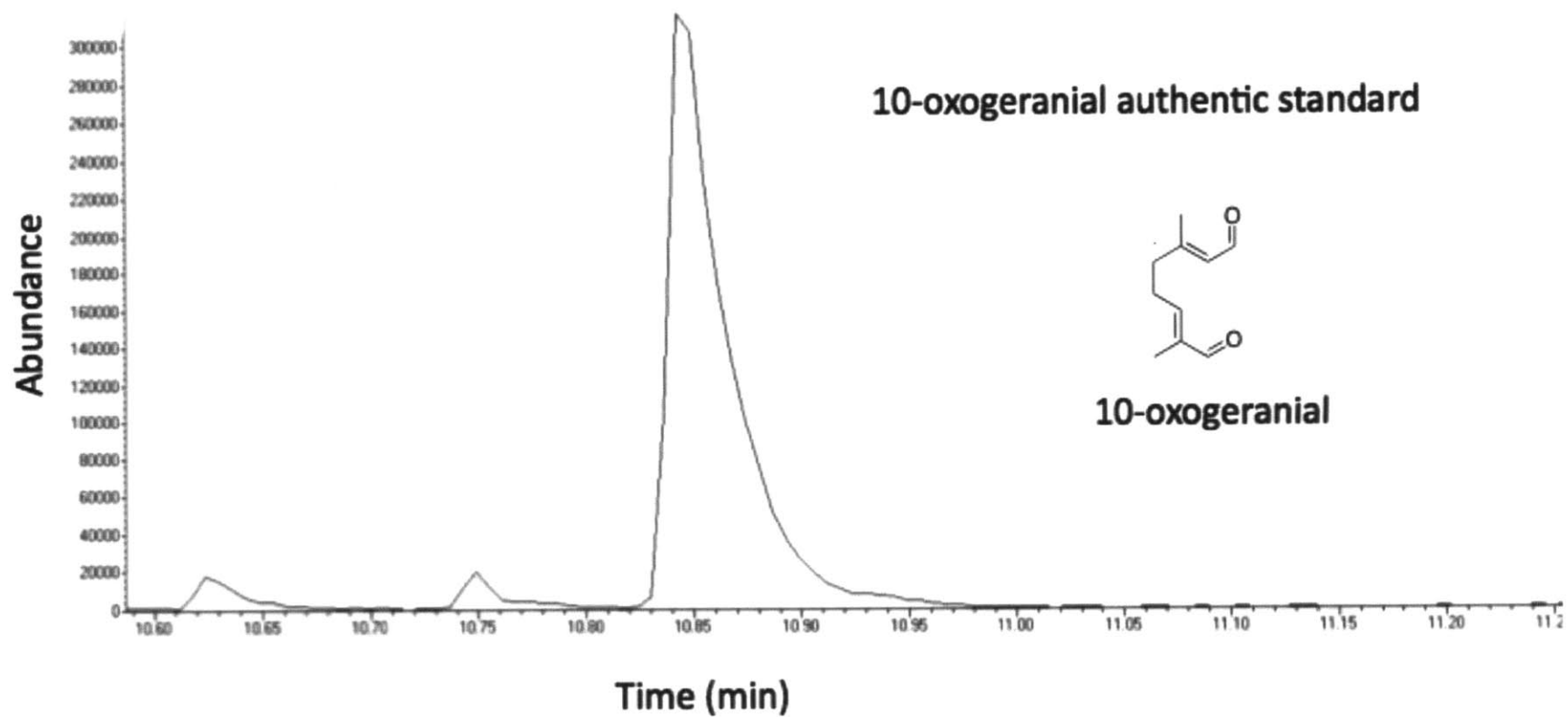


Figure A.7: GC-MS chromatogram of 10-oxogeranial authentic standard. 10-oxogeranial is the enzymatic product of 10-hydroxygeraniol oxidoreductase.

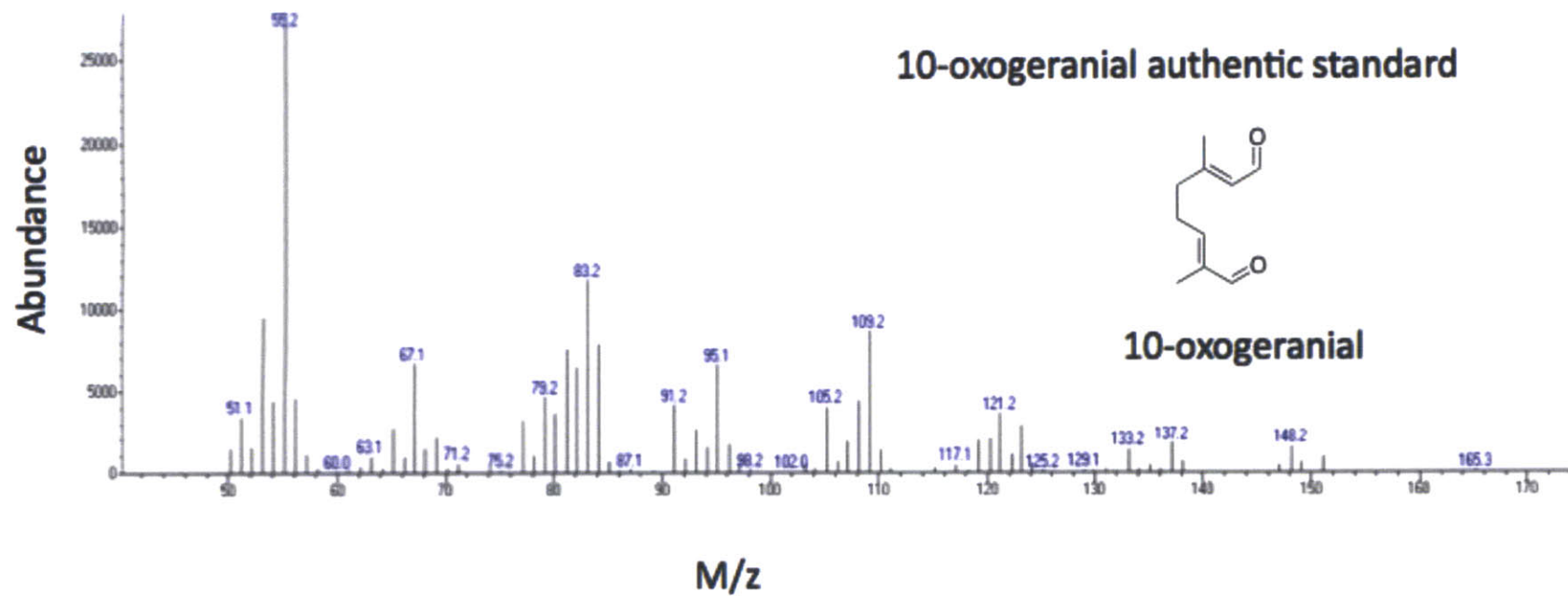


Figure A.8: Mass spectrum of 10-oxogeranial authentic standard.

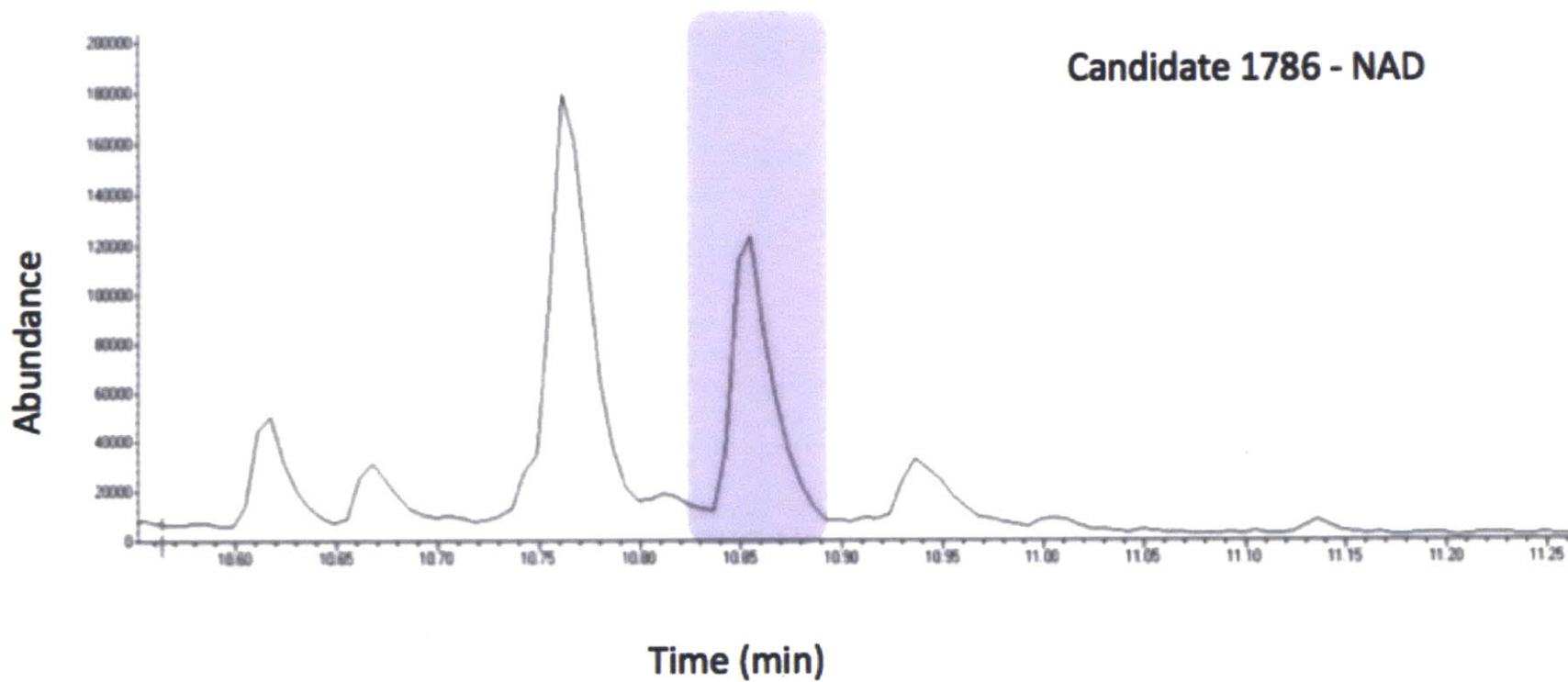


Figure A.9: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786 and cofactor NAD<sup>+</sup>. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

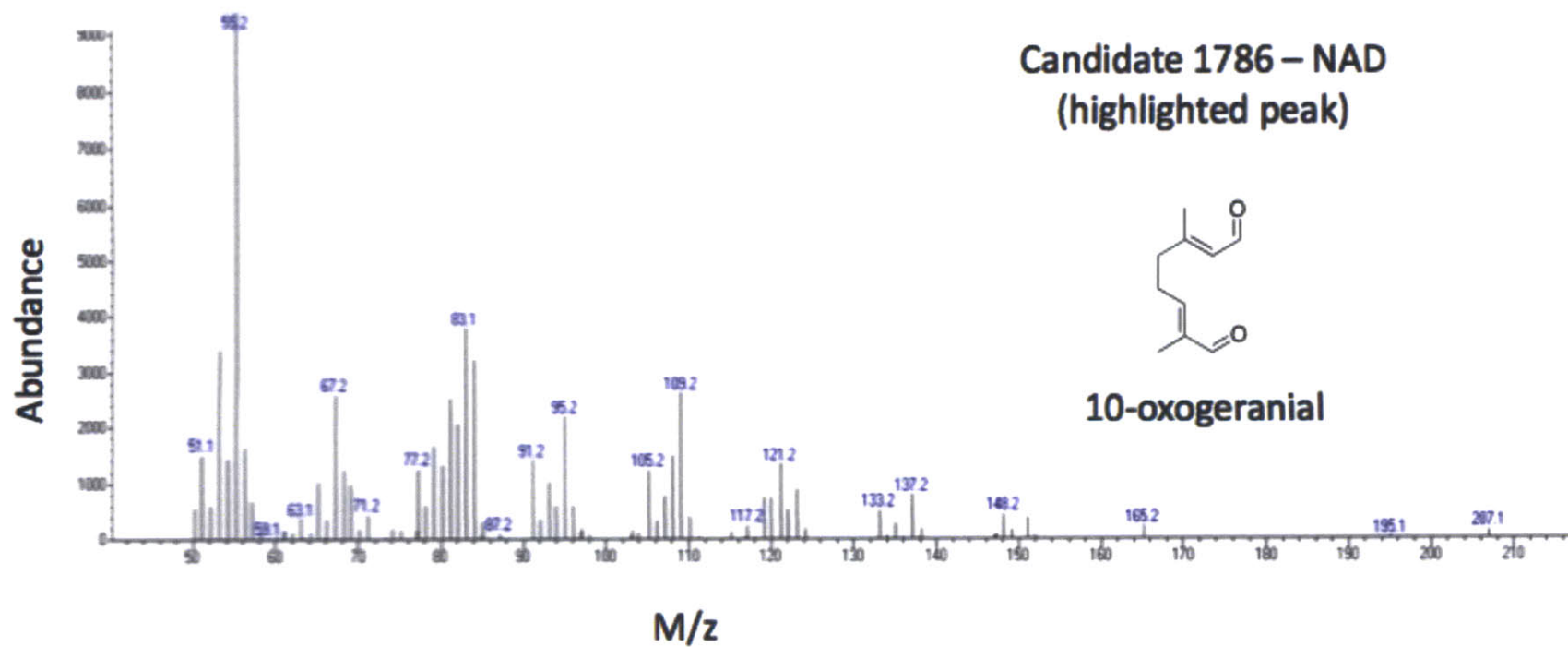


Figure A.10: Mass spectrum of highlighted peak from Candidate 1786 and cofactor NAD<sup>+</sup> assay. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

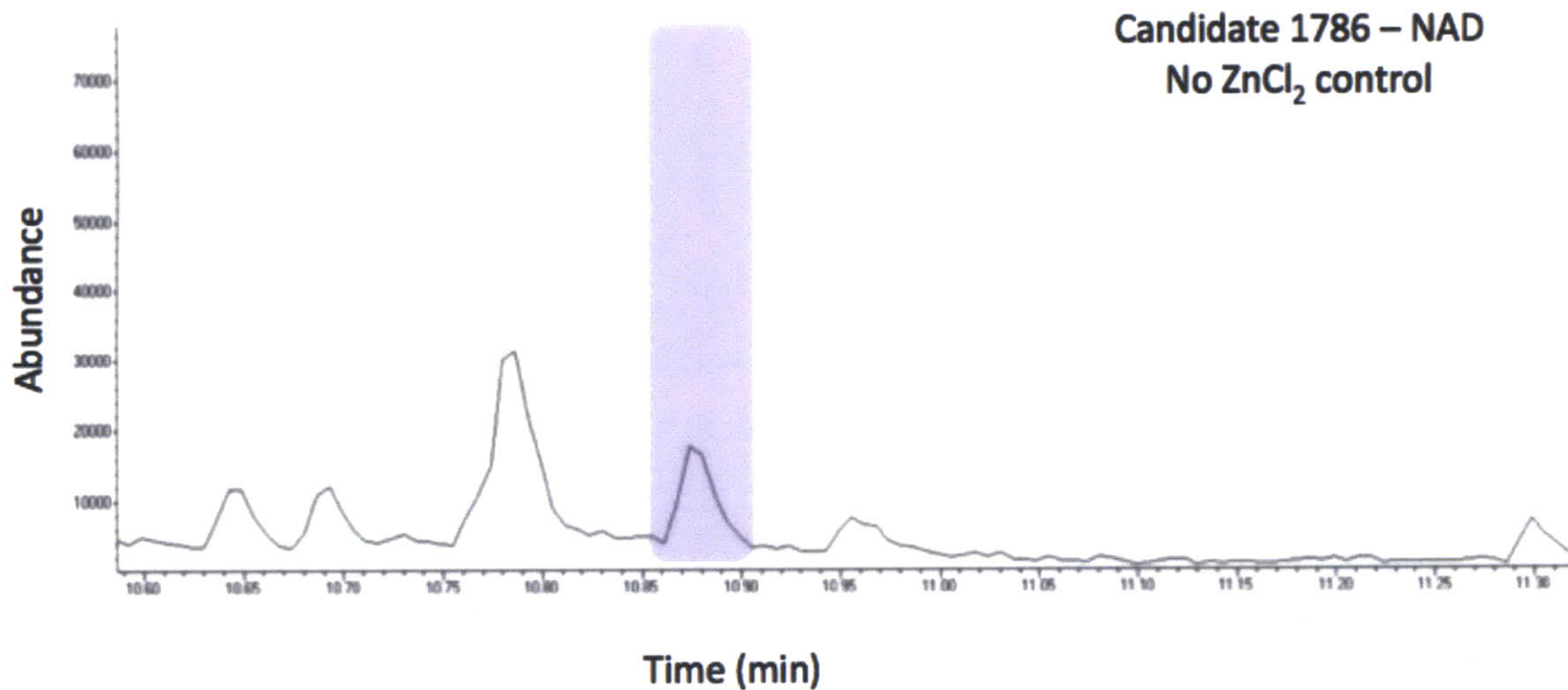


Figure A.11: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786 and cofactor NAD<sup>+</sup> without the addition of ZnCl<sub>2</sub> to the assay mixture. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.



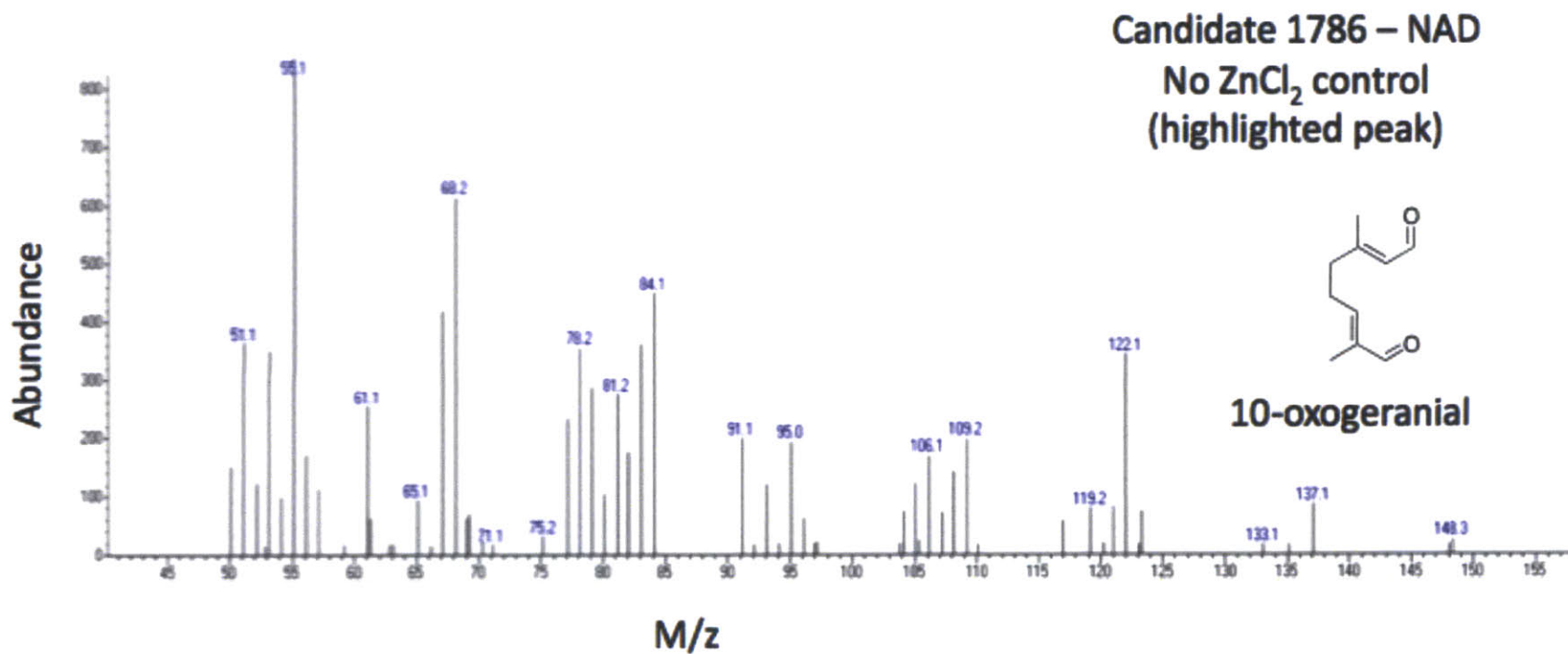


Figure A.12: Mass spectrum of highlighted peak from Candidate 1786 and cofactor NAD<sup>+</sup> assay when  $\text{ZnCl}_2$  is omitted from the assay mixture. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

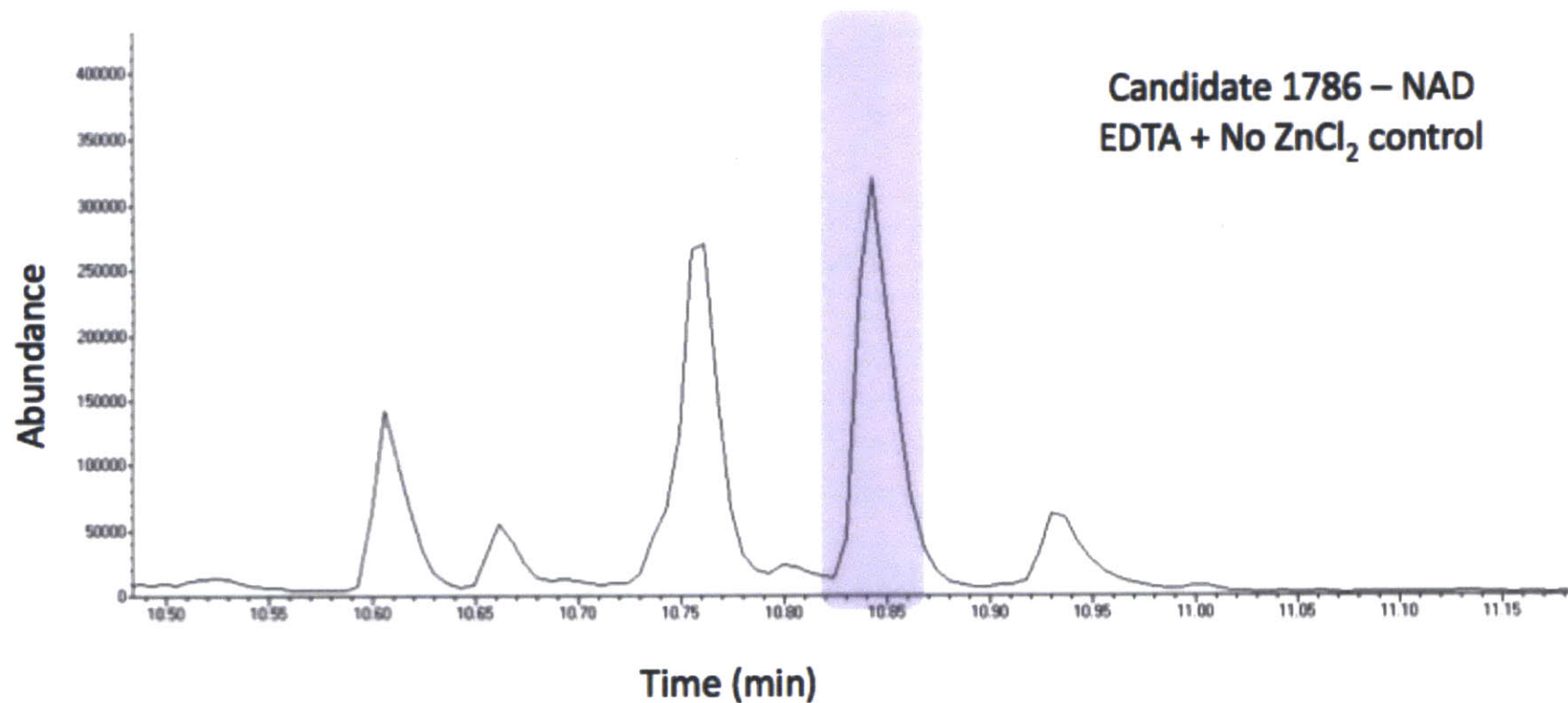


Figure A.13: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786 and cofactor NAD<sup>+</sup> when EDTA to a final concentration of 1 mM was added to the reaction mixture and ZnCl<sub>2</sub> was omitted. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

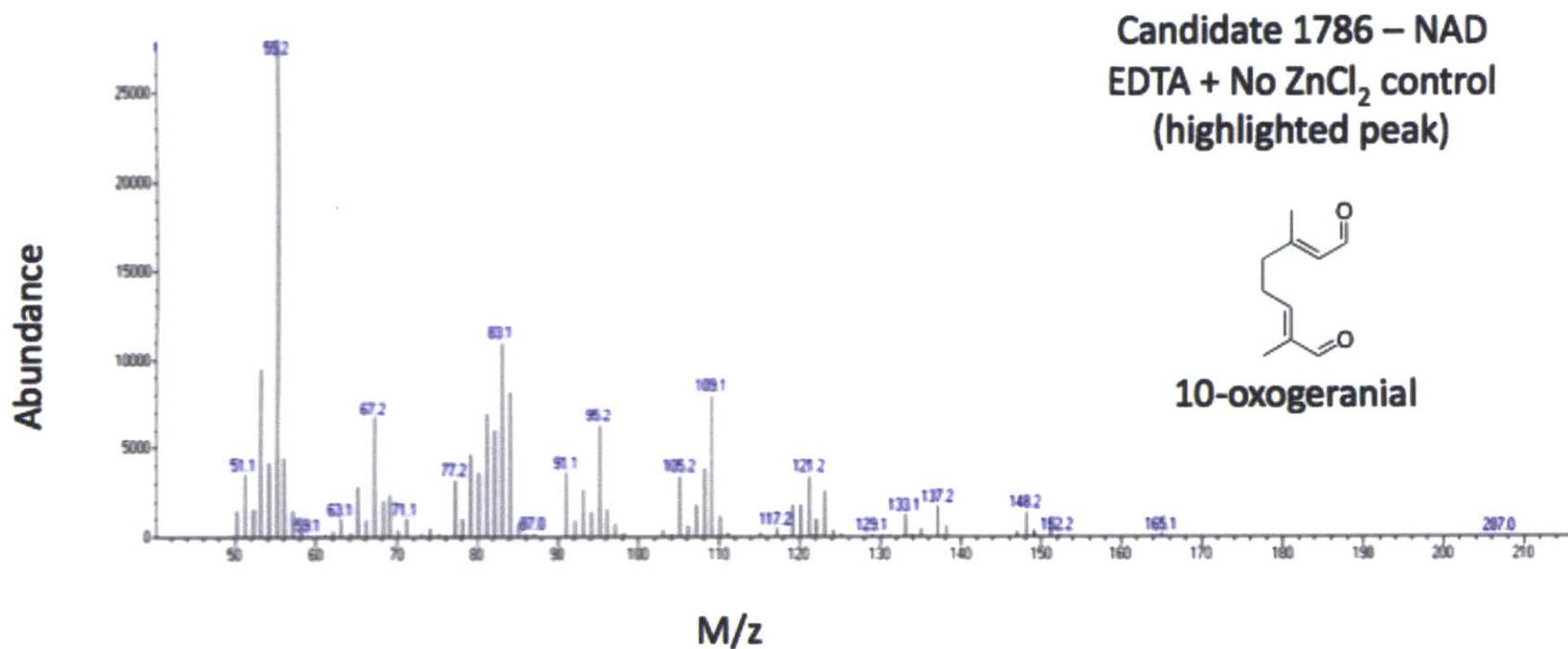


Figure A.14: Mass spectrum of highlighted peak from Candidate 1786 and cofactor NAD<sup>+</sup> assay when EDTA is added to a final concentration of 1 mM in the assay mixture and ZnCl<sub>2</sub> is omitted. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

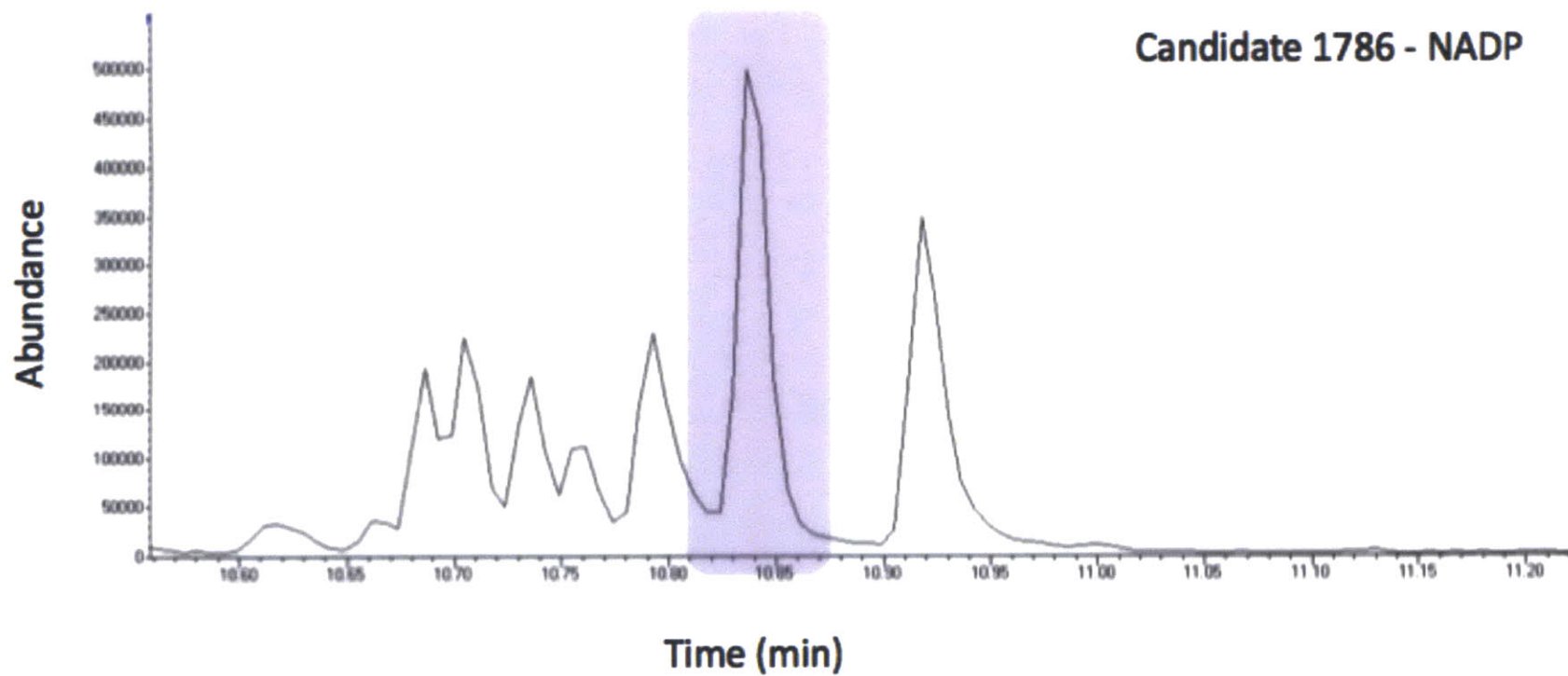


Figure A.15: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786 and cofactor NADP<sup>+</sup>. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

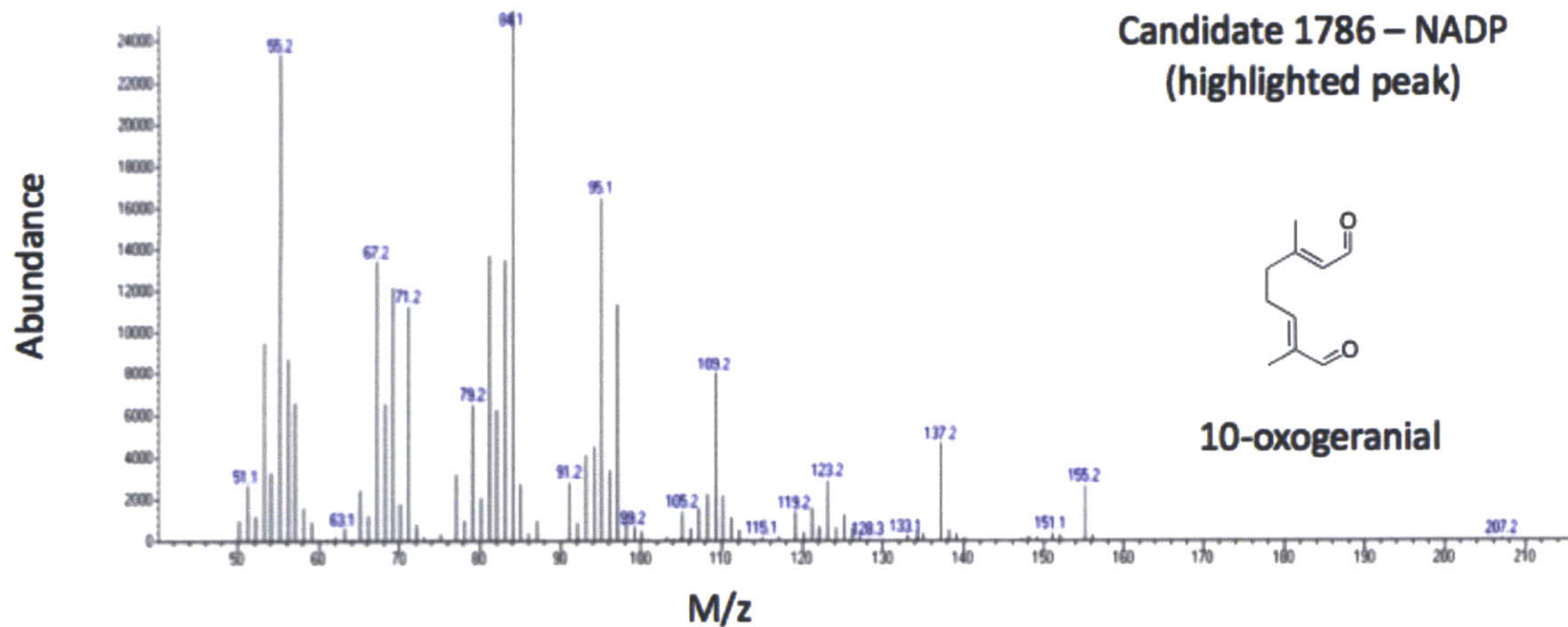


Figure A.16: Mass spectrum of highlighted peak from Candidate 1786 and cofactor NADP<sup>+</sup> assay. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

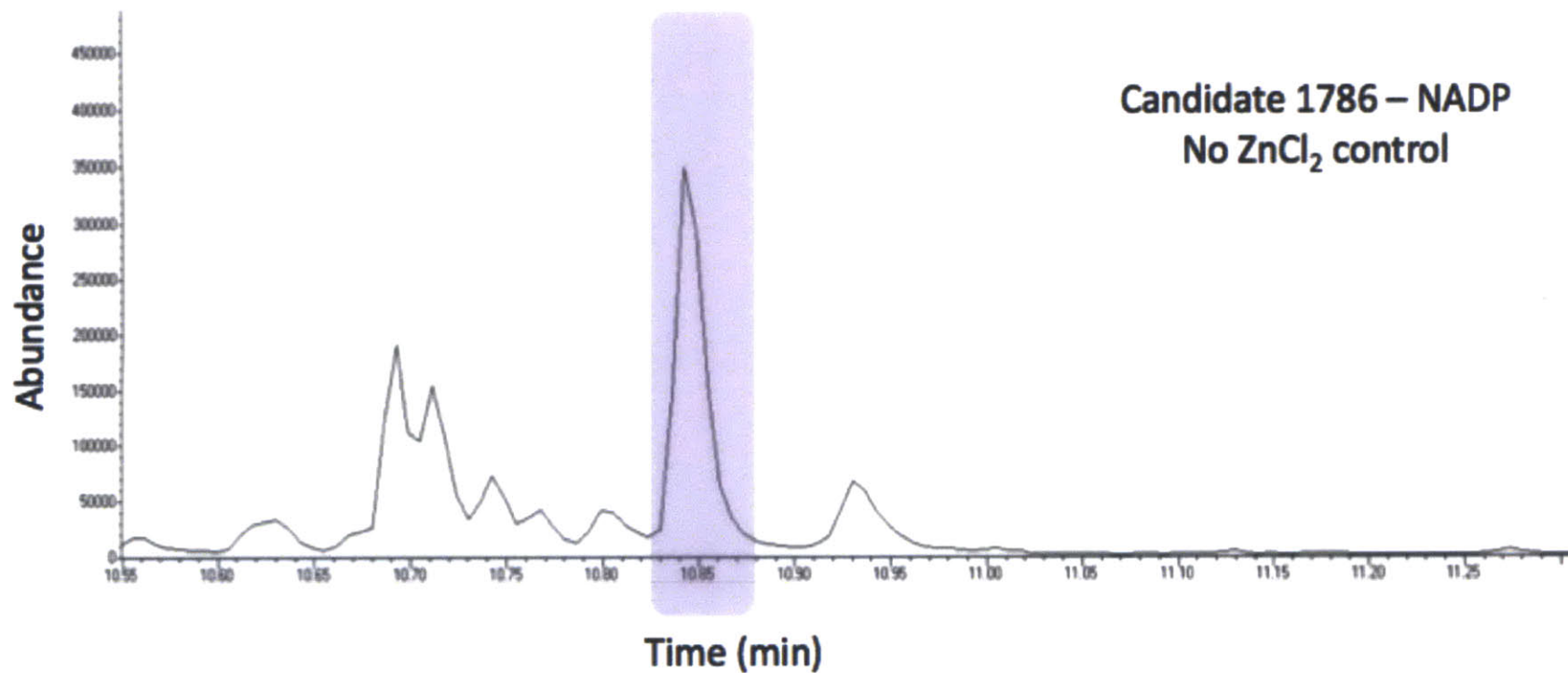


Figure A.17: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786 and cofactor NADP<sup>+</sup> without the addition of ZnCl<sub>2</sub> to the assay mixture. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

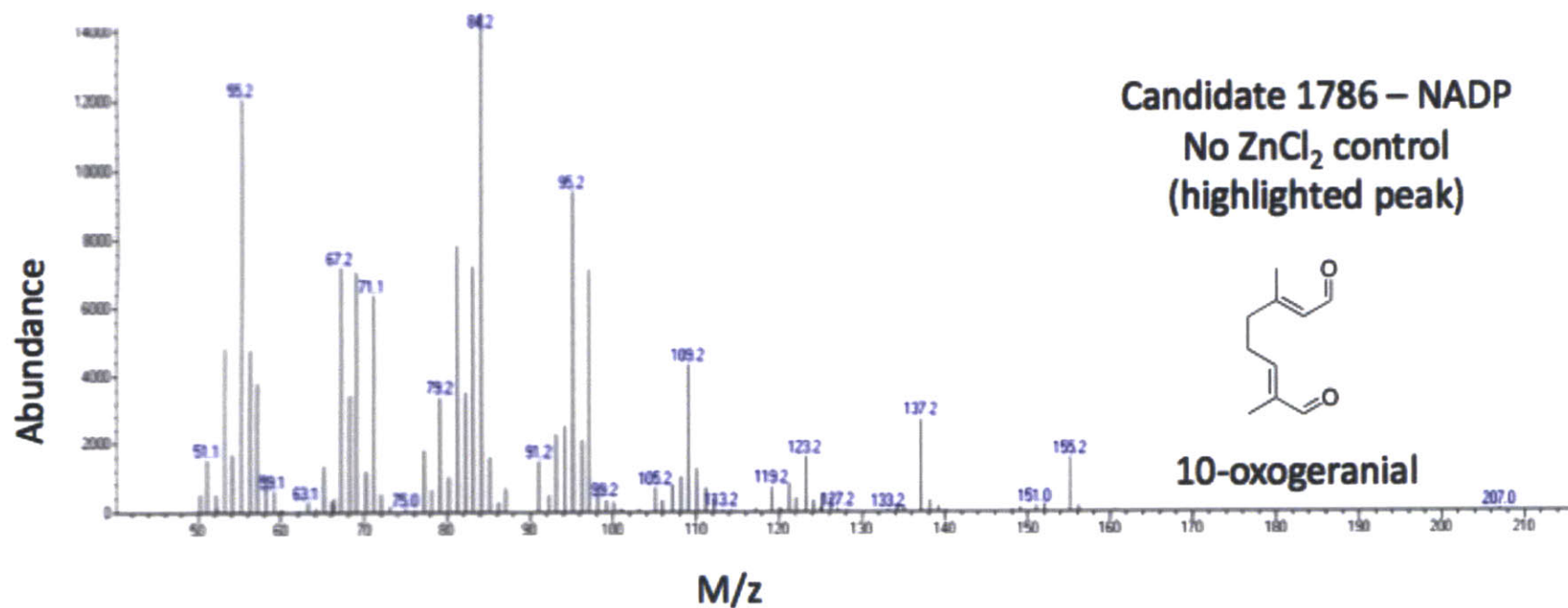


Figure A.18: Mass spectrum of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786+ and cofactor NAD when ZnCl<sub>2</sub> was omitted from the reaction. The peak that co-elutes with the authentic 10-oxogeranial standard is highlighted in purple.

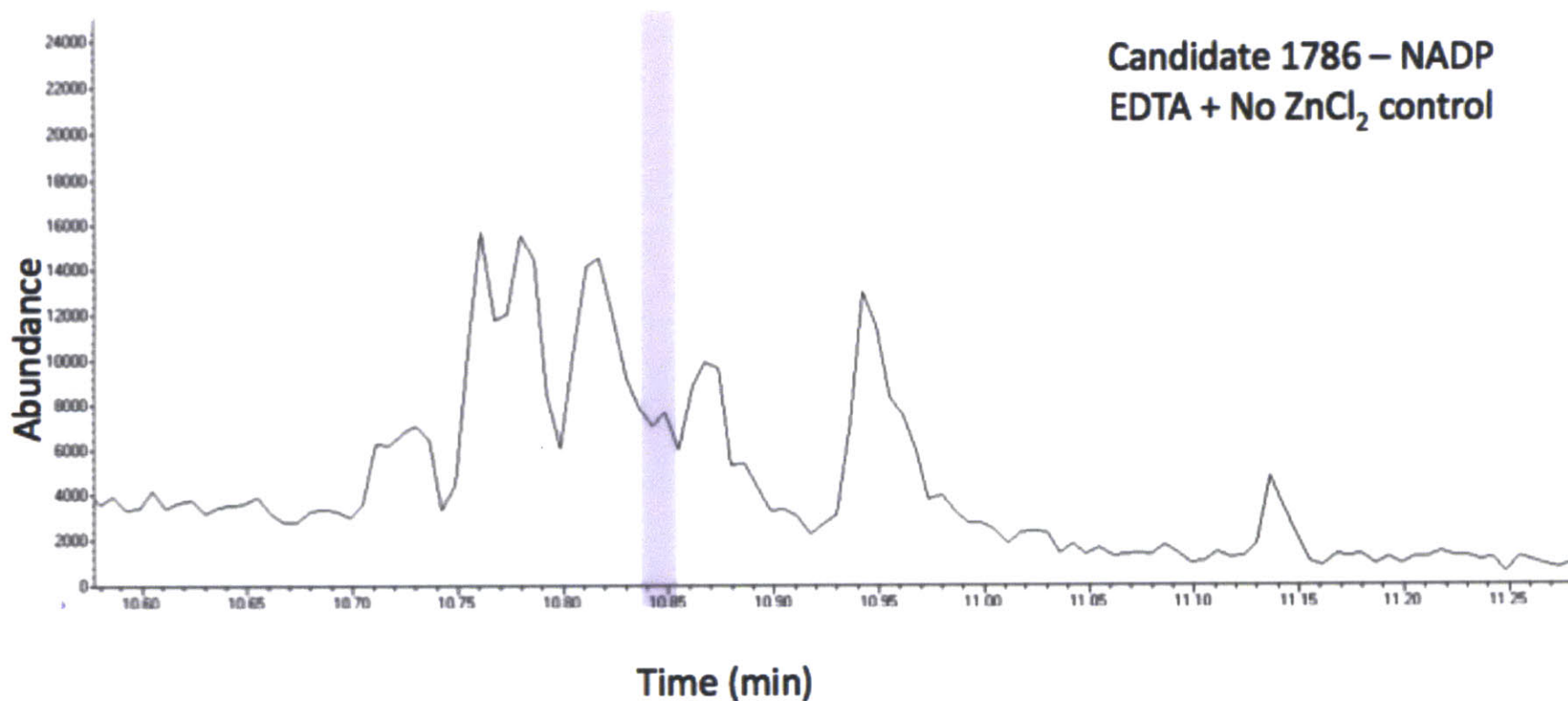


Figure A.19: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screening with Candidate 1786 and cofactor NADP<sup>+</sup> when EDTA to a final concentration of 1 mM was added to the reaction mixture and ZnCl<sub>2</sub> was omitted. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.



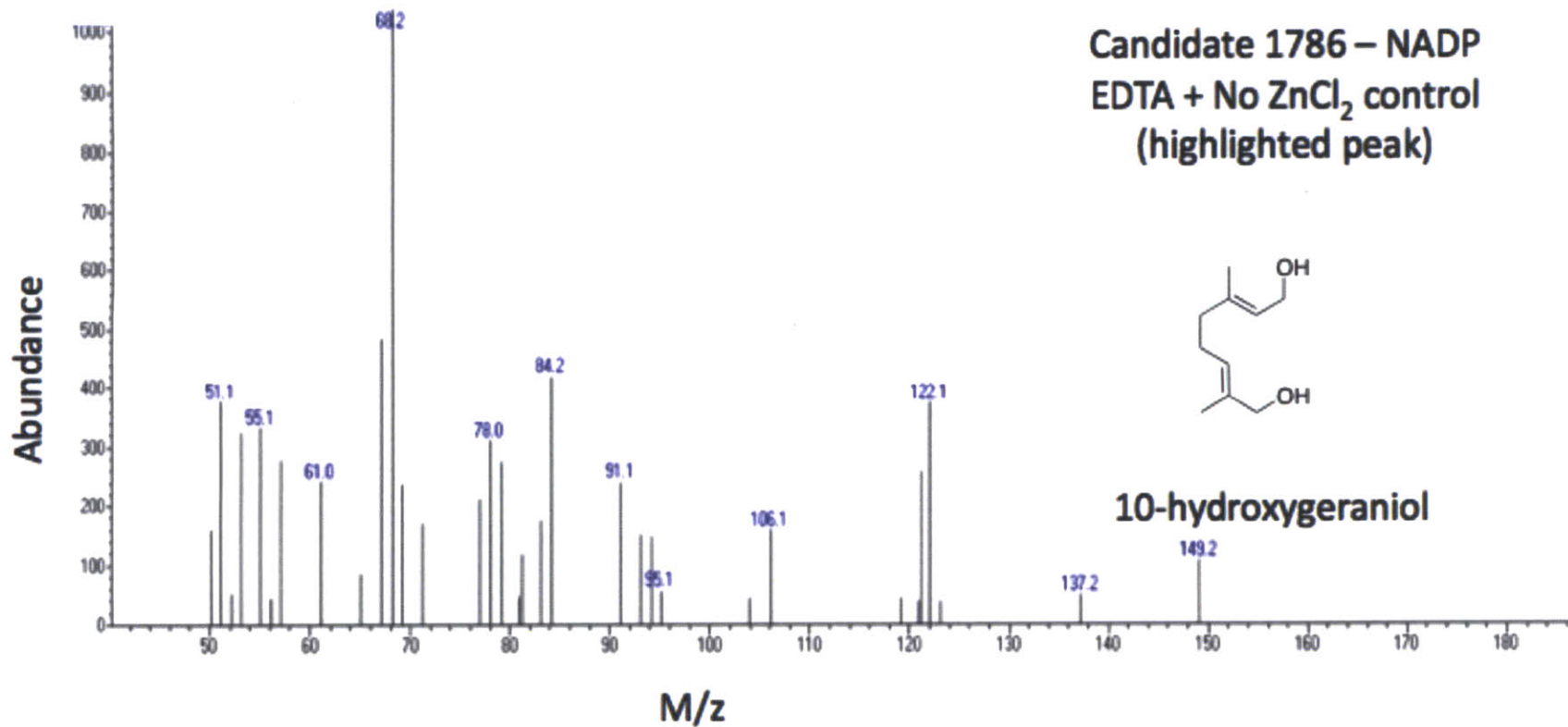


Figure A.20: Mass spectrum of highlighted peak from Candidate 1786 and cofactor NAD<sup>+</sup> assay when EDTA is added to a final concentration of 1 mM and ZnCl<sub>2</sub> is omitted. Similarity to the mass spectrum of authentic 10-hydroxygeraniol enables positive assignment as starting substrate, not product.

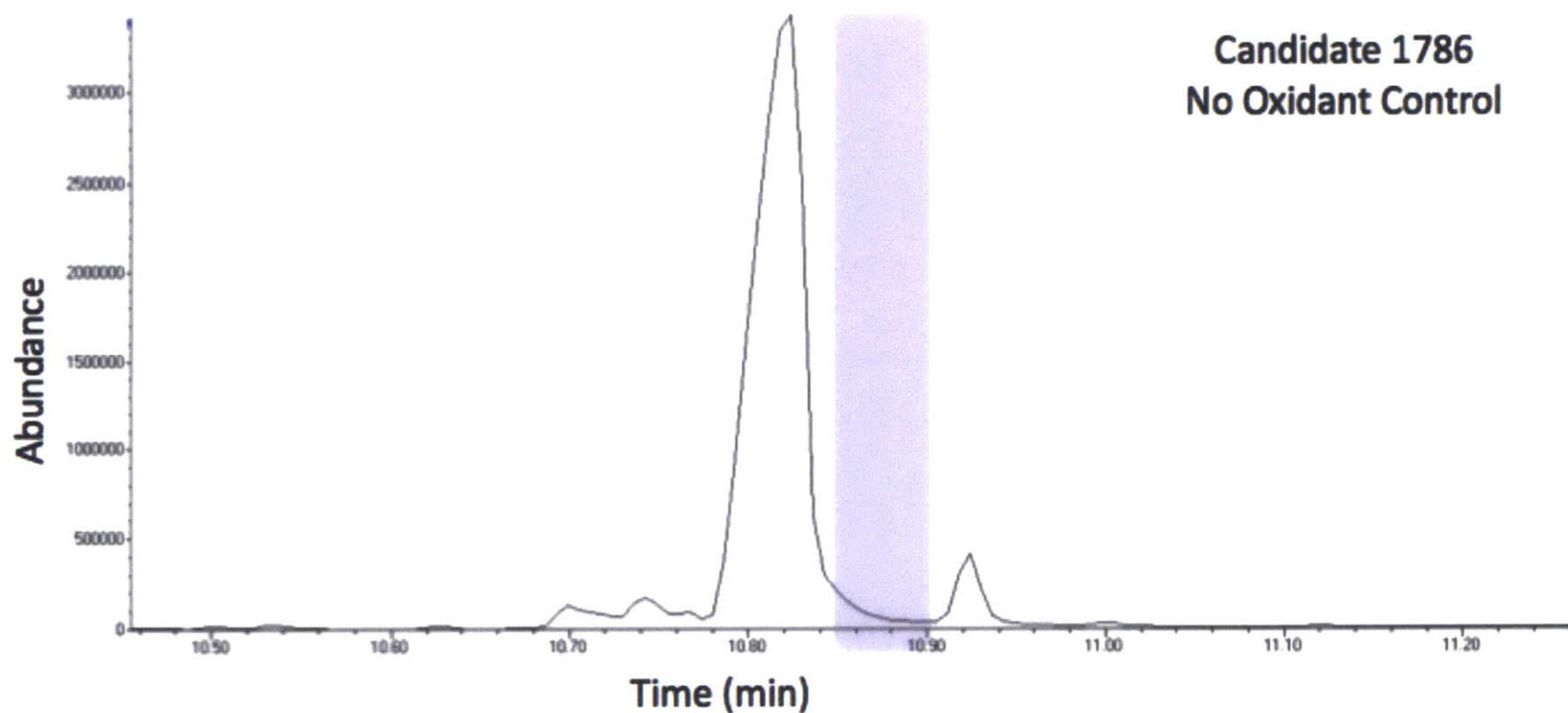


Figure A.21: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 1786 in the absence of either NAD<sup>+</sup> or NADP<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeraniol standard; the region of 10-oxogeraniol elution is highlighted in purple.

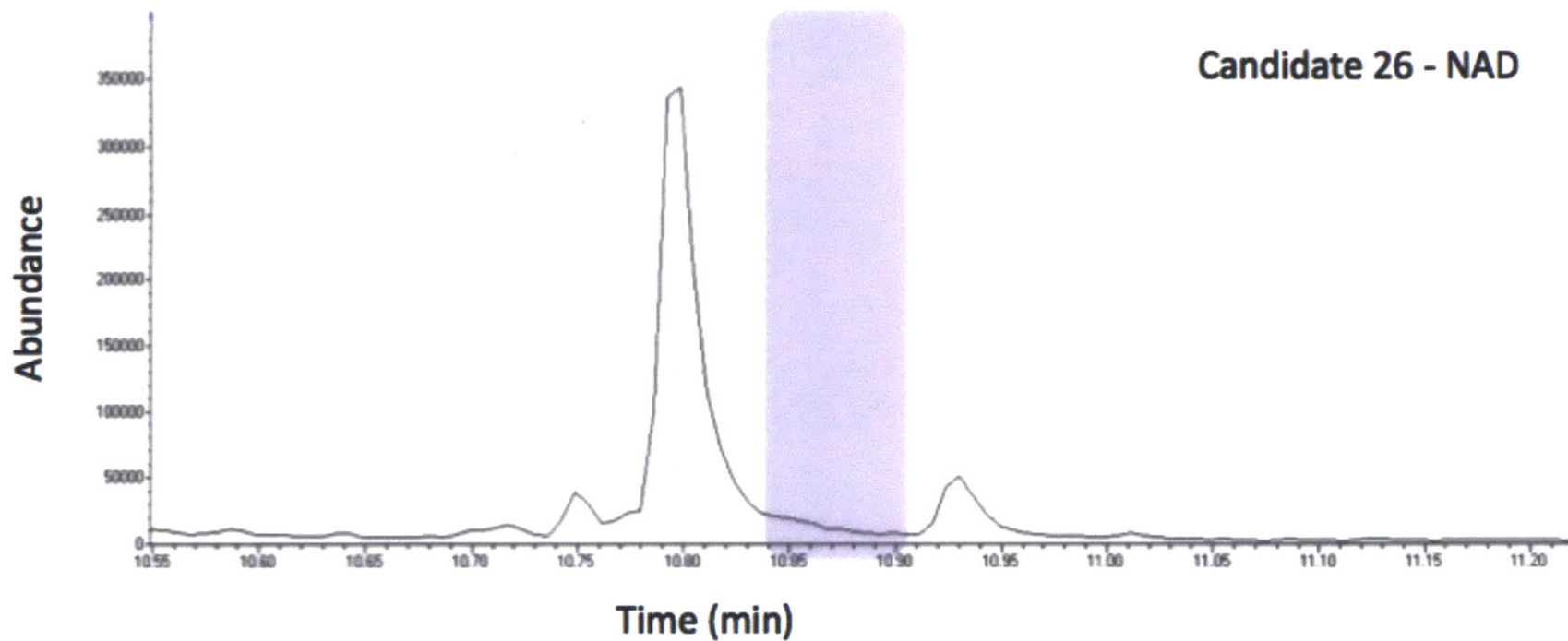


Figure A.22: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 and cofactor NAD<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeranial standard; the region of 10-oxogeranial elution is highlighted in purple.

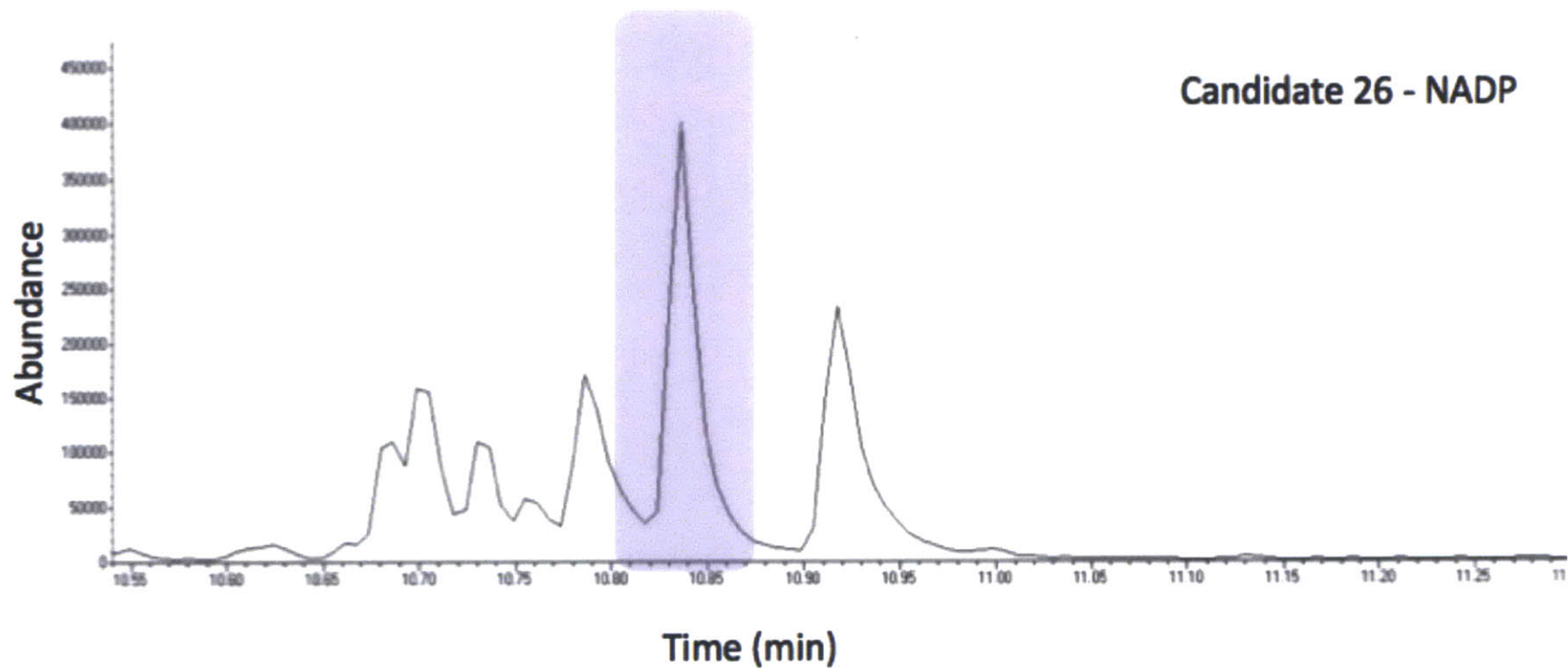


Figure A.23: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 and cofactor NADP<sup>+</sup>. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

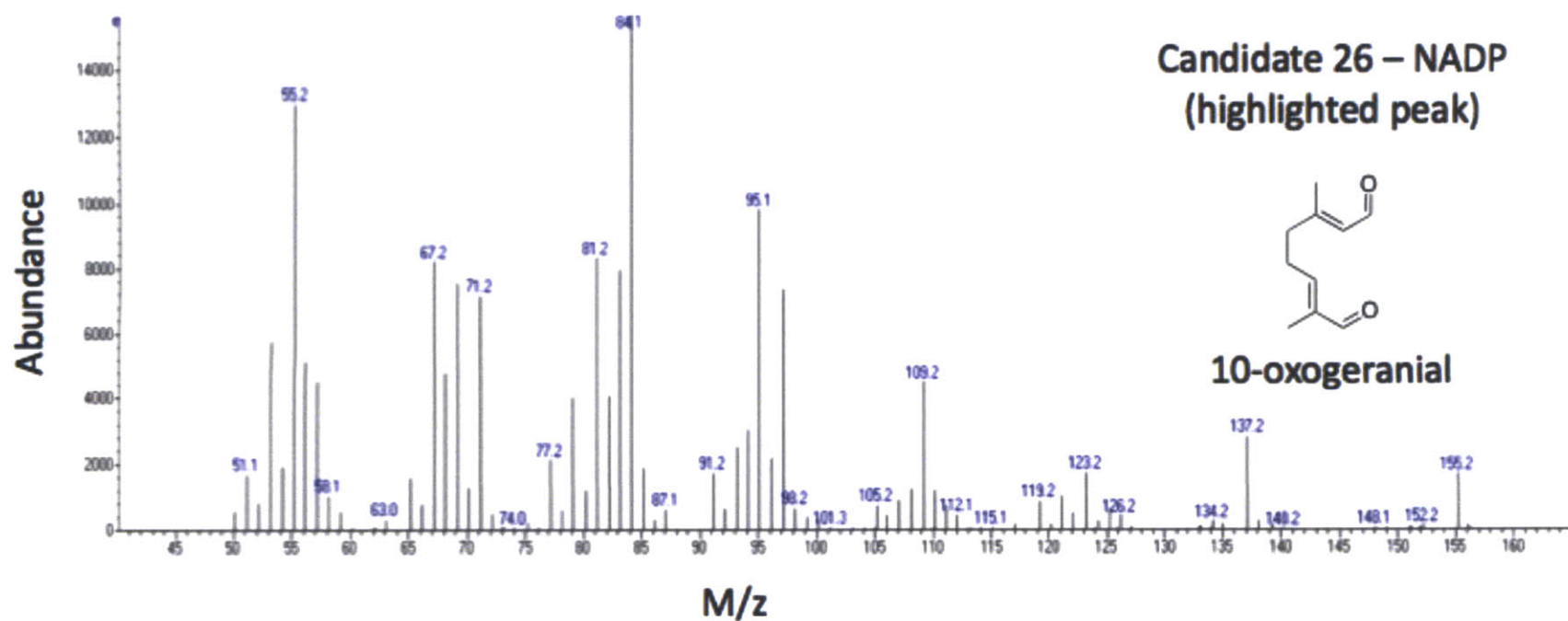


Figure A.24: Mass spectrum of highlighted peak from Candidate 1786 and cofactor NADP+ assay. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

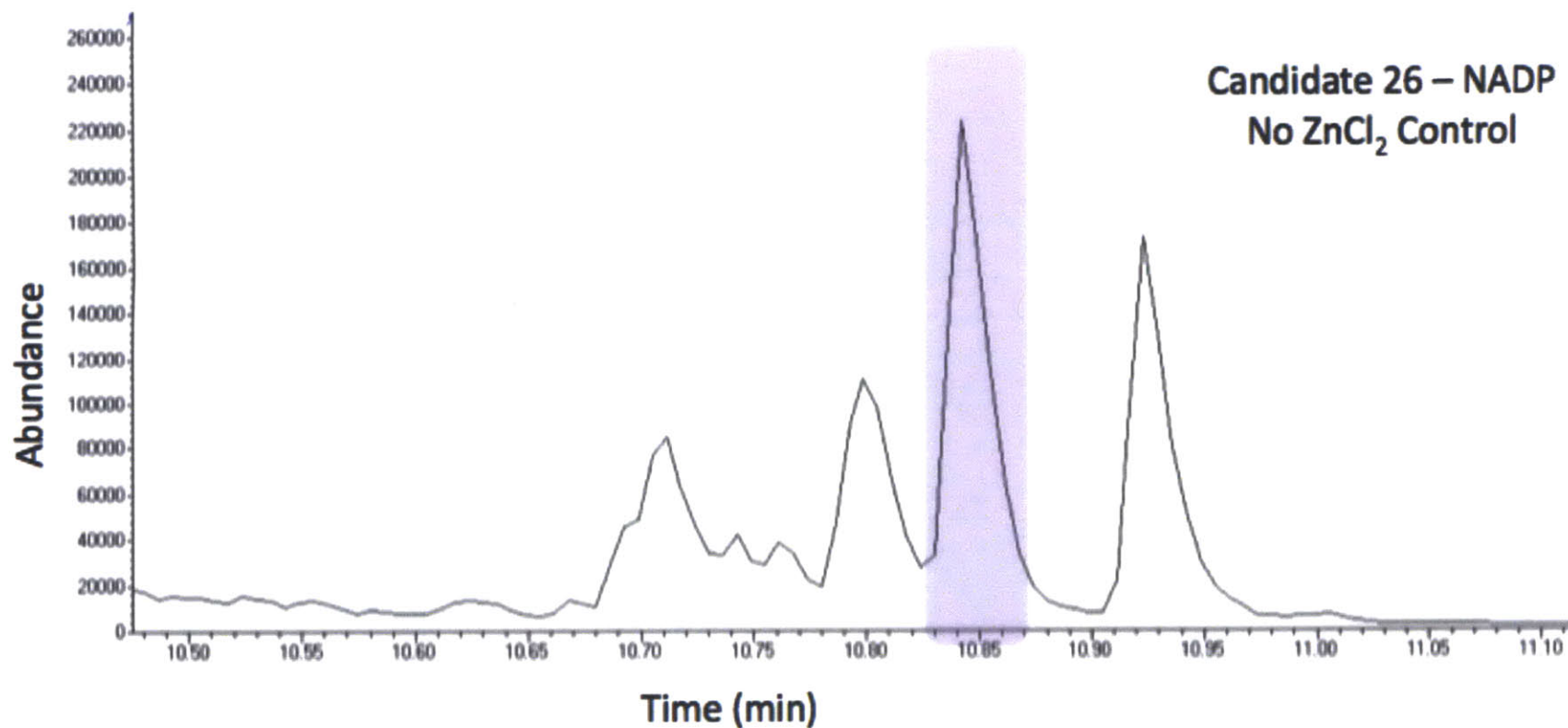


Figure A.25: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 and cofactor NADP<sup>+</sup> without the addition of ZnCl<sub>2</sub> to the assay mixture. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

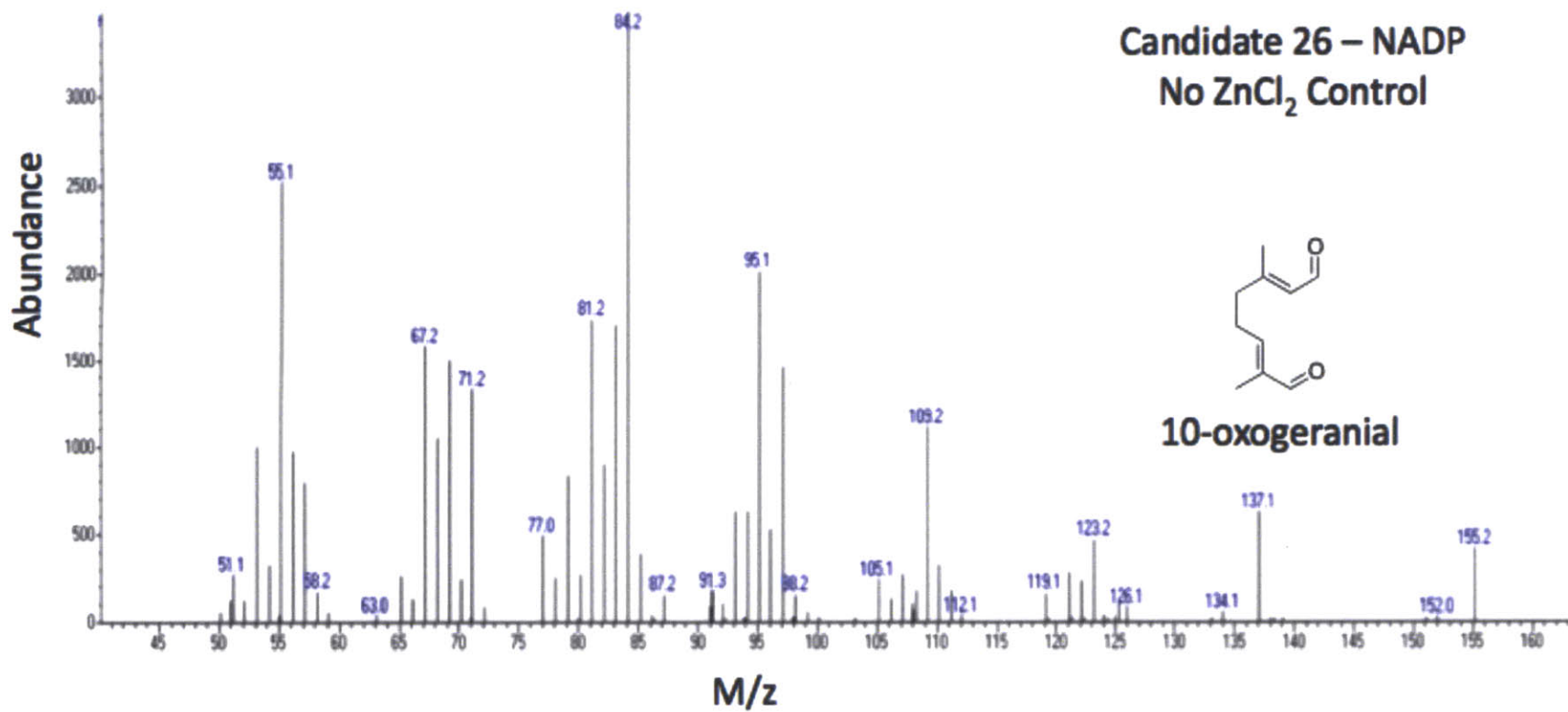


Figure A.26: Mass spectrum of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 and cofactor NADP<sup>+</sup> when ZnCl<sub>2</sub> was omitted from the assay mixture. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.



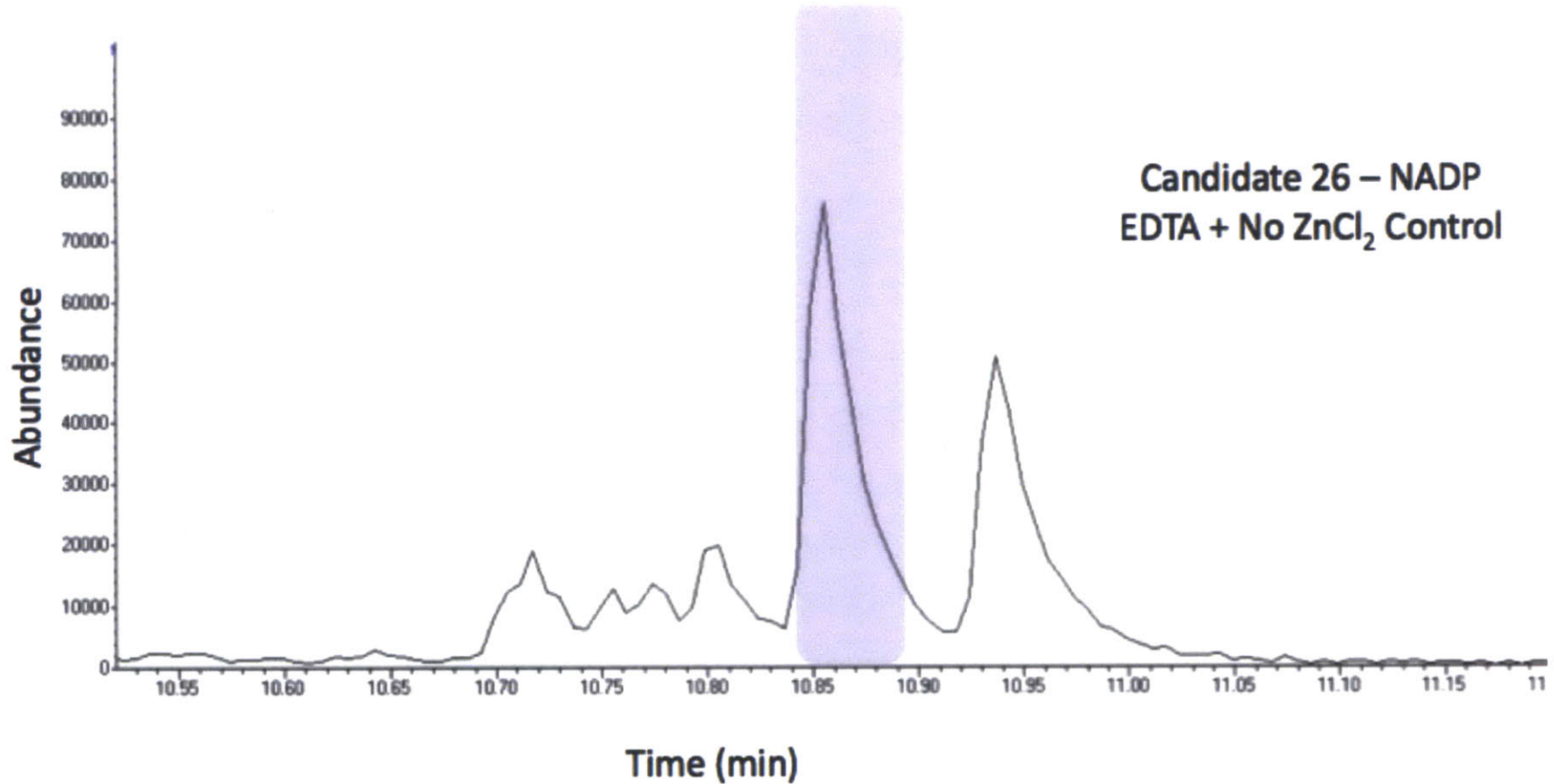


Figure A.27: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 and cofactor NADP<sup>+</sup> when EDTA to a final concentration of 1 mM was added to the reaction mixture and ZnCl<sub>2</sub> was omitted. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.



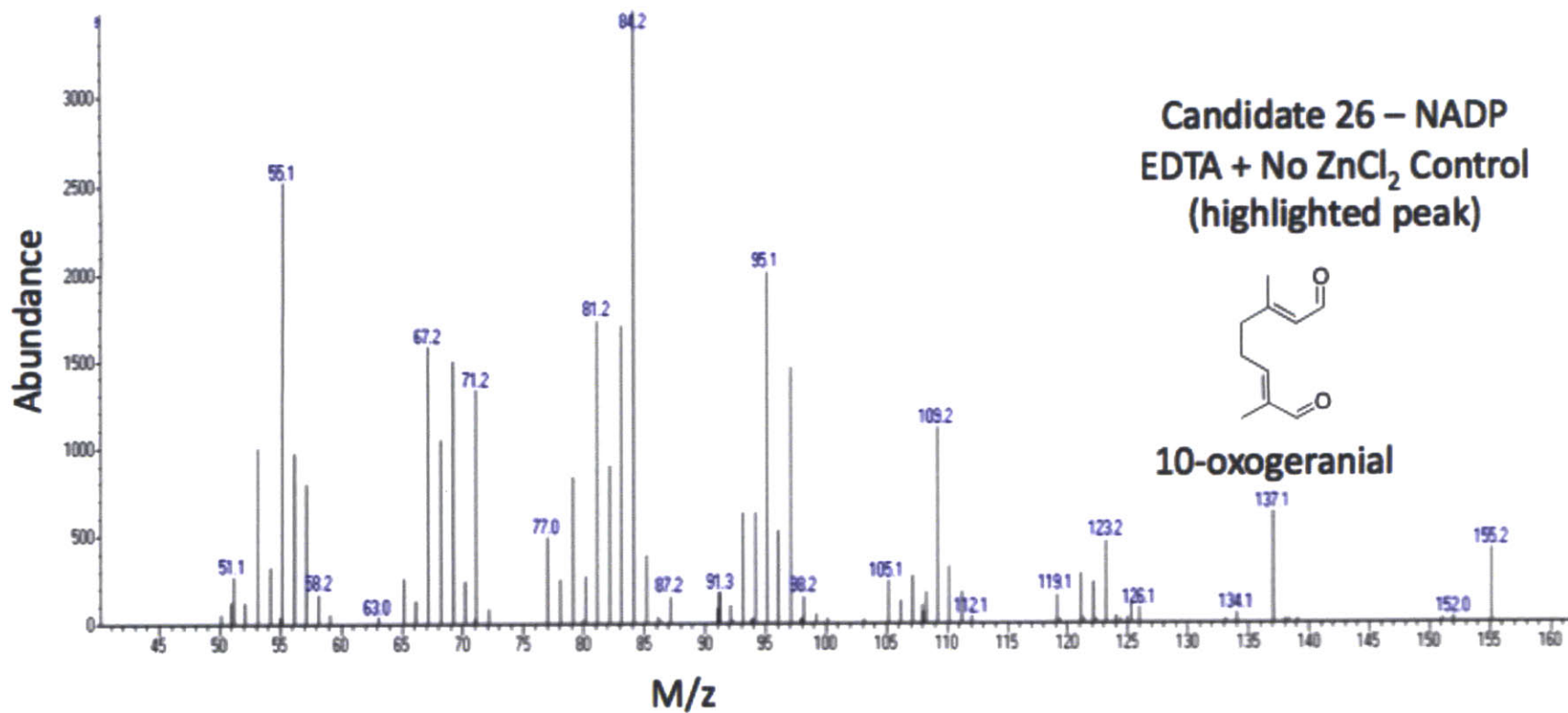


Figure A.28: Mass spectrum of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 and cofactor NADP<sup>+</sup> when ZnCl<sub>2</sub> was omitted from the assay mixture. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

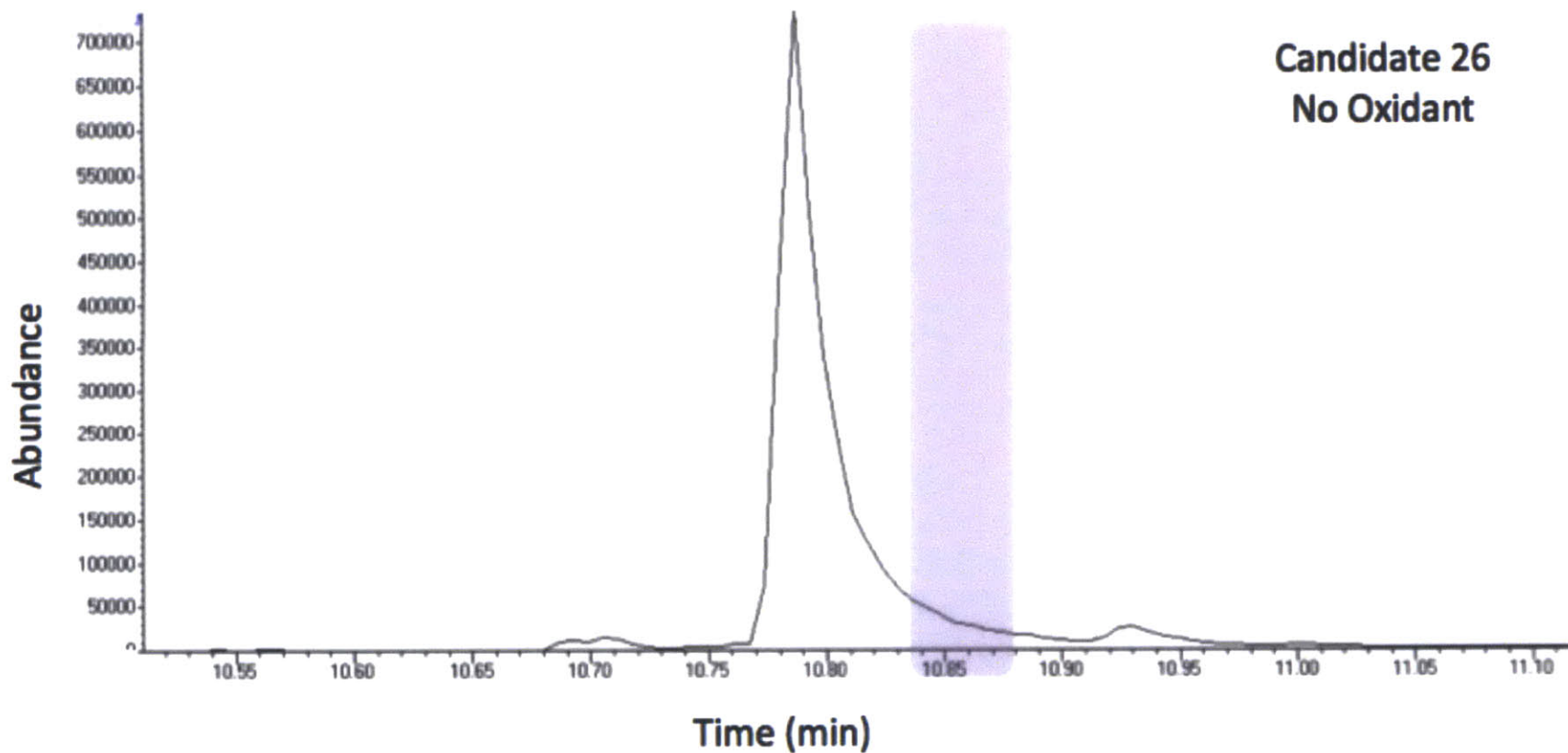


Figure A.29: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 26 in the absence of either NAD<sup>+</sup> or NADP<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeraniol standard; the region of 10-oxogeraniol elution is highlighted in purple.

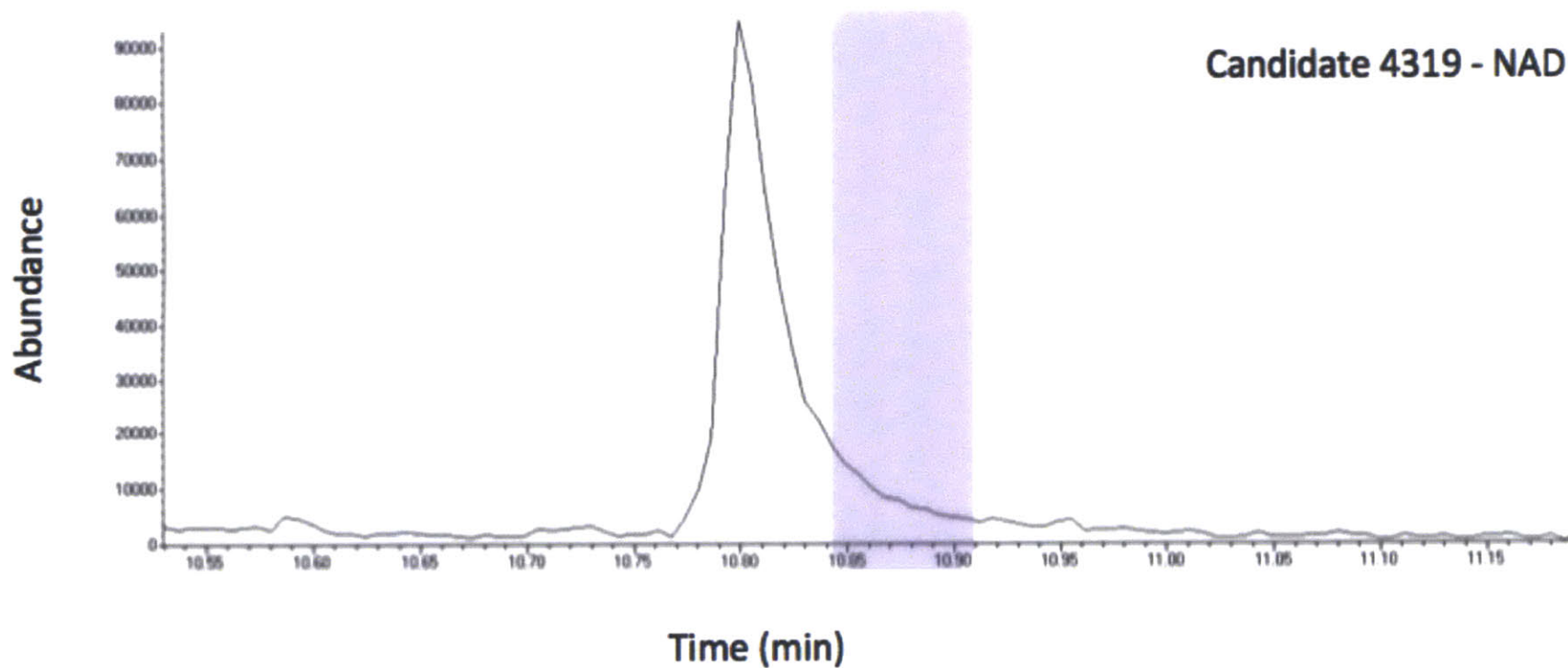


Figure A.30: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 4319 and cofactor NAD<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeraniol standard; the region of 10-oxogeraniol elution is highlighted in purple.

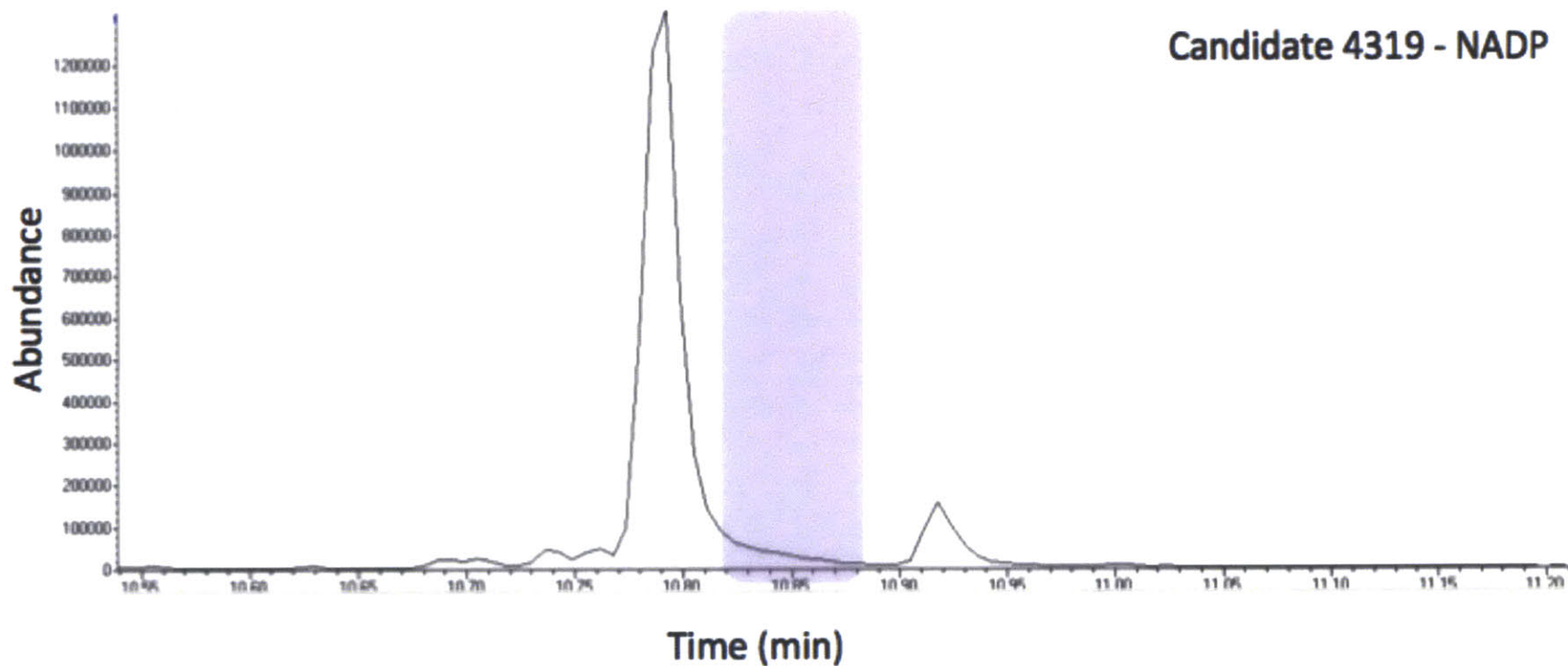


Figure A.31: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 4319 and cofactor NADP<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeraniol standard; the region of 10-oxogeraniol elution is highlighted in purple.

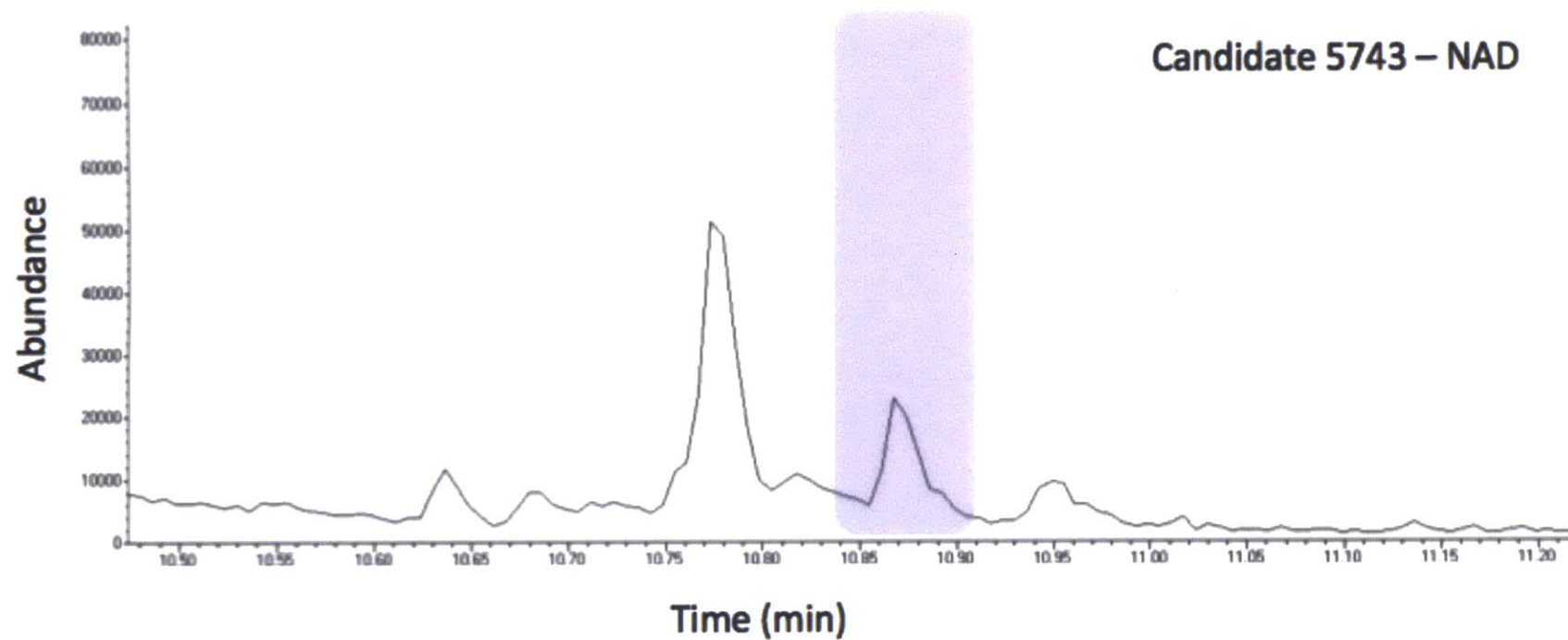


Figure A.32: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 and cofactor NAD<sup>+</sup>. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

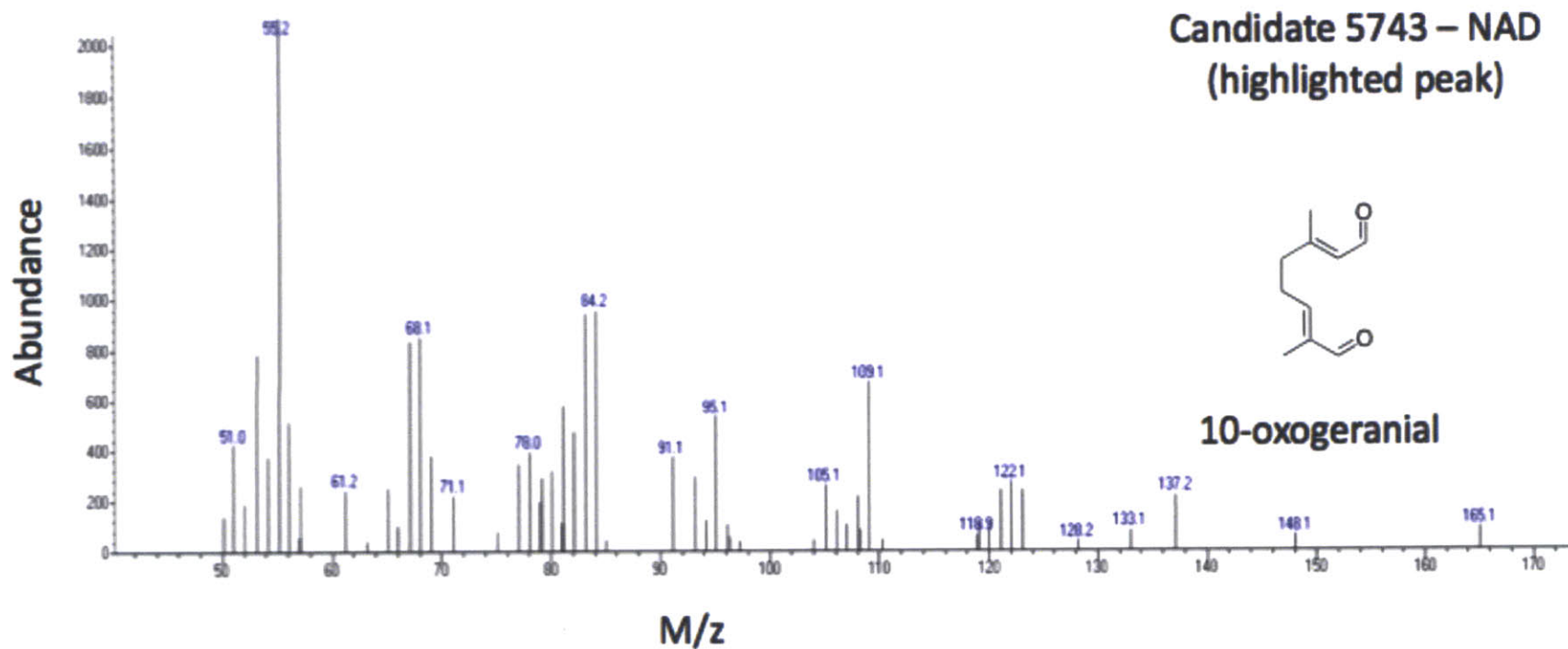


Figure A.33: Mass spectrum of highlighted peak from Candidate 5743 and cofactor NADP+ assay. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

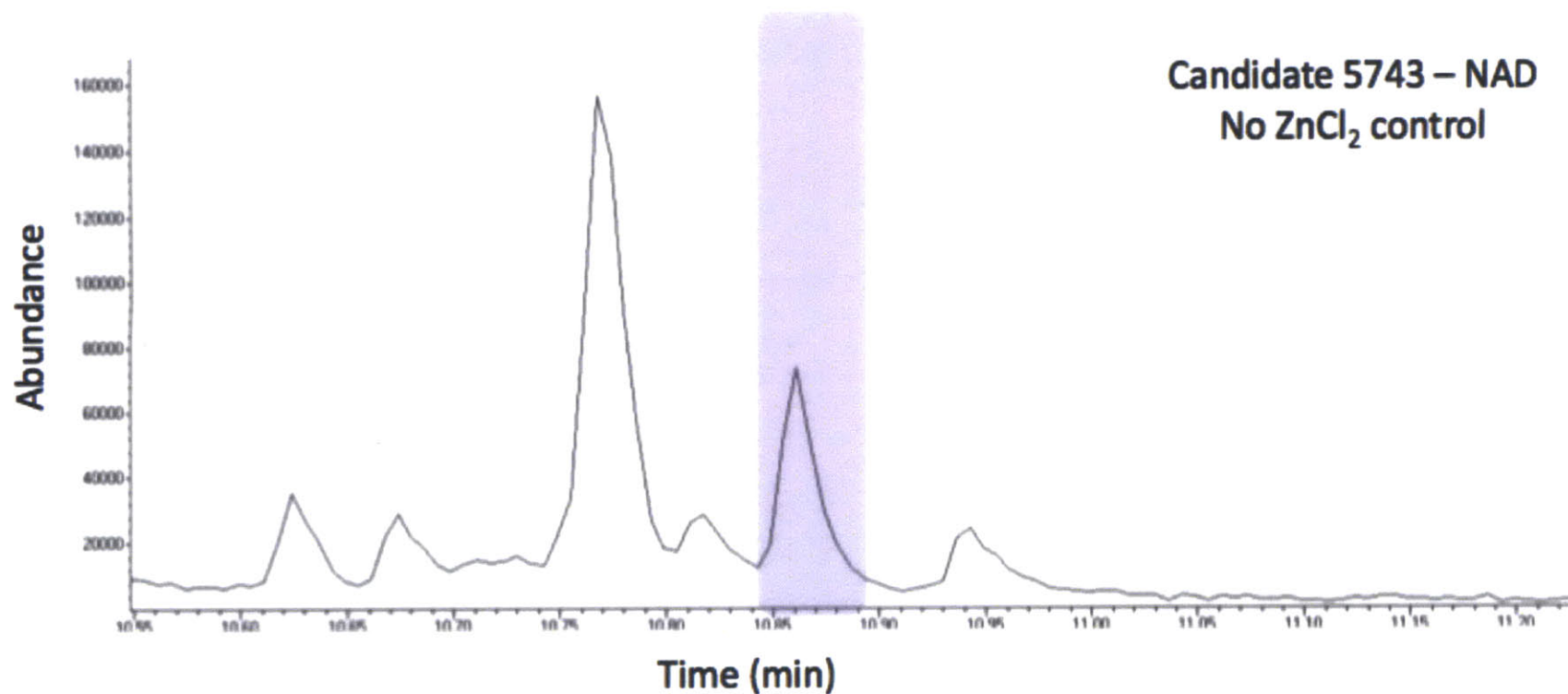


Figure A.34: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 and cofactor NAD<sup>+</sup> without the addition of ZnCl<sub>2</sub> to the assay mixture. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

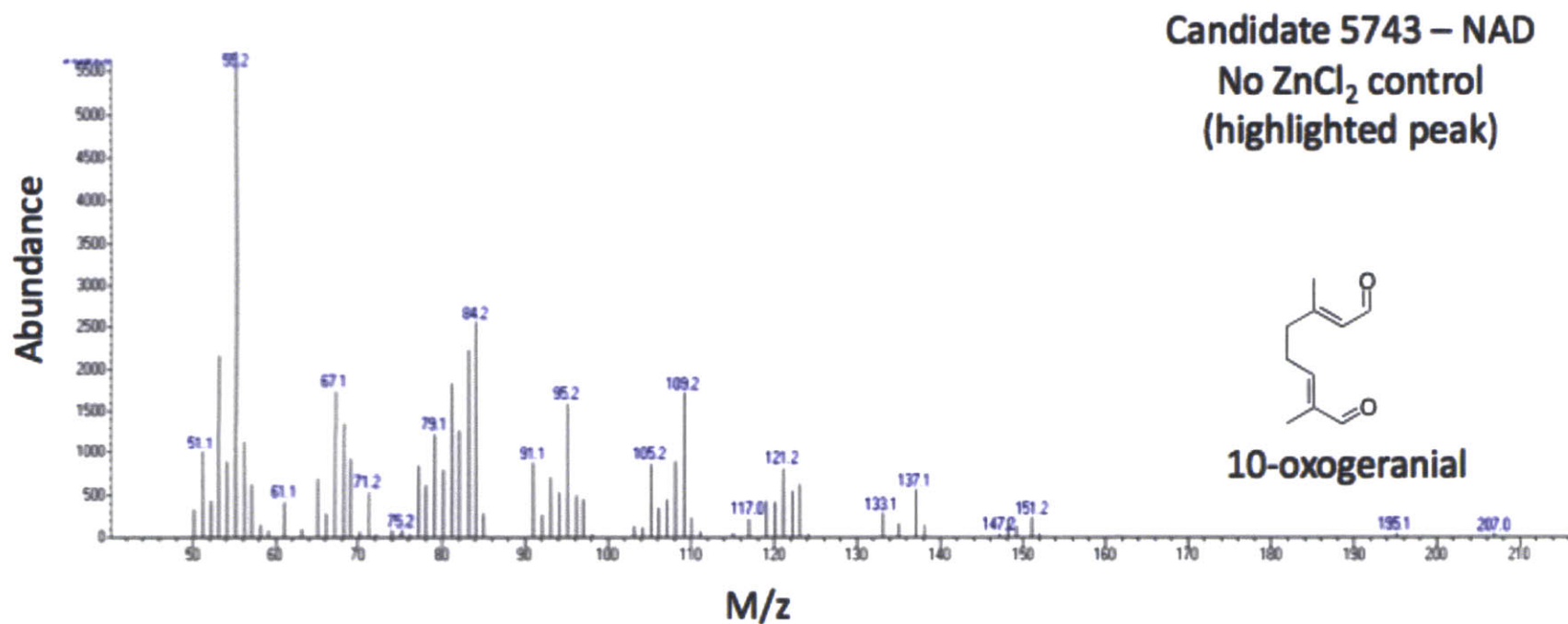


Figure A.35: Mass spectrum of highlighted peak from Candidate 5743 and cofactor NAD<sup>+</sup> assay when ZnCl<sub>2</sub> is omitted from the assay mixture. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.



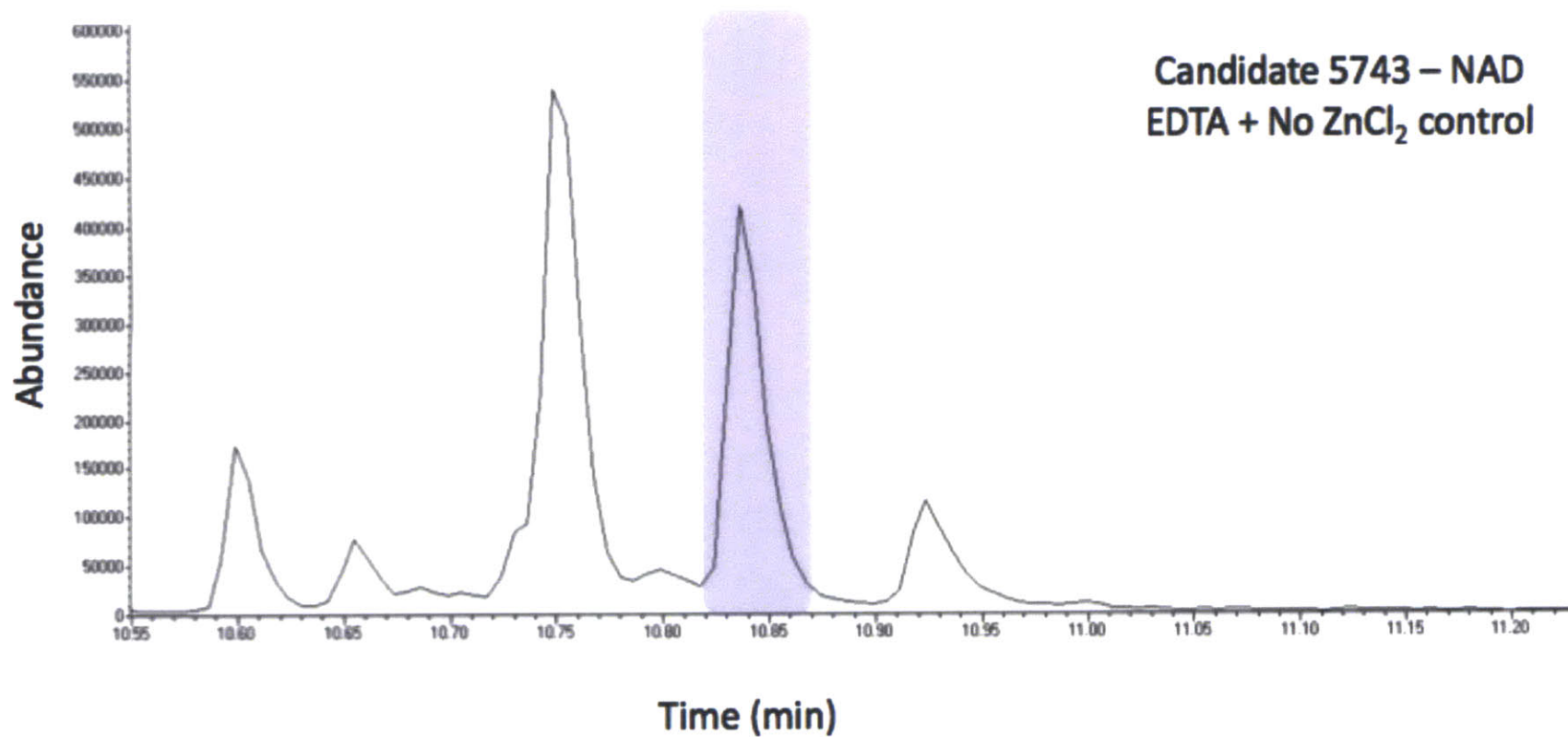


Figure A.36: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 and cofactor NAD<sup>+</sup> when EDTA is added to a final concentration of 1 mM and ZnCl<sub>2</sub> is omitted from assay mixture. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

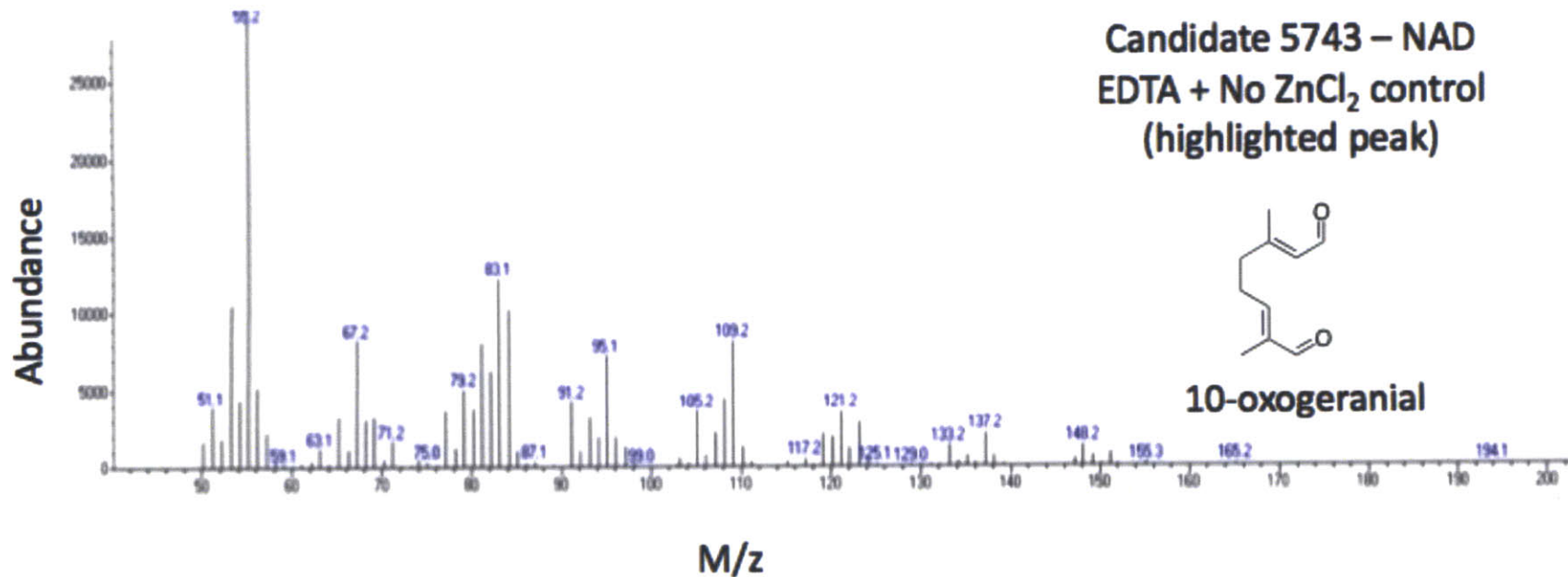


Figure A.37: Mass spectrum of highlighted peak from Candidate 5743 and cofactor NAD<sup>+</sup> assay when EDTA is added to a final concentration of 1 mM and ZnCl<sub>2</sub> is omitted from the assay mixture. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

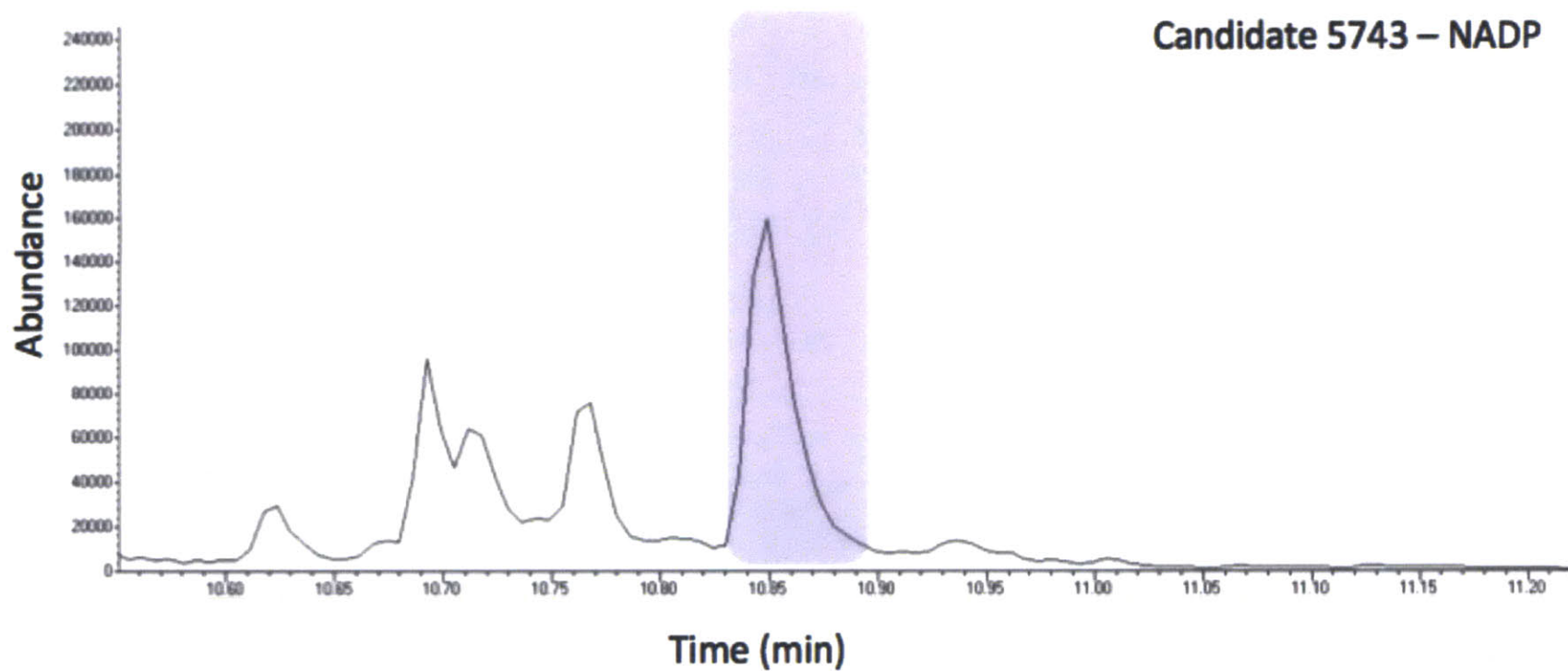


Figure A.38: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 and cofactor NADP<sup>+</sup>. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

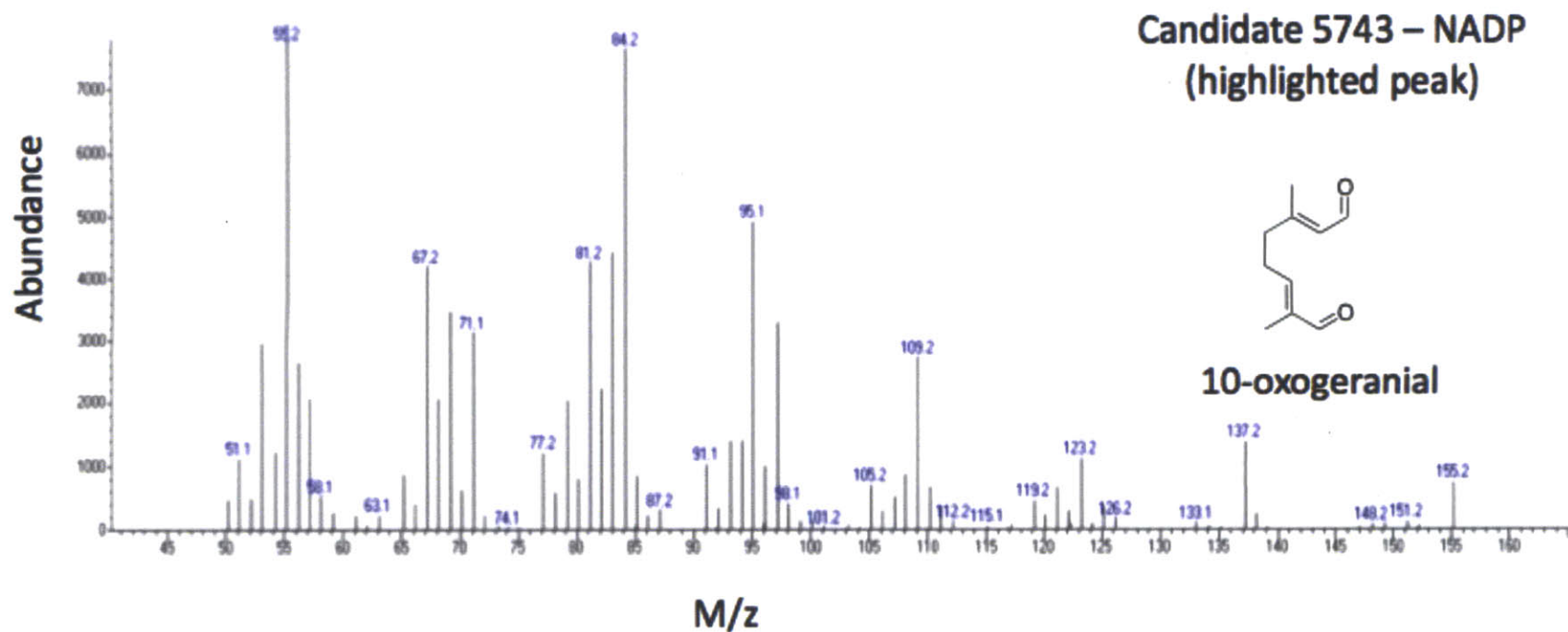


Figure A.39: Mass spectrum of highlighted peak from Candidate 5743 and cofactor NADP+ assay. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

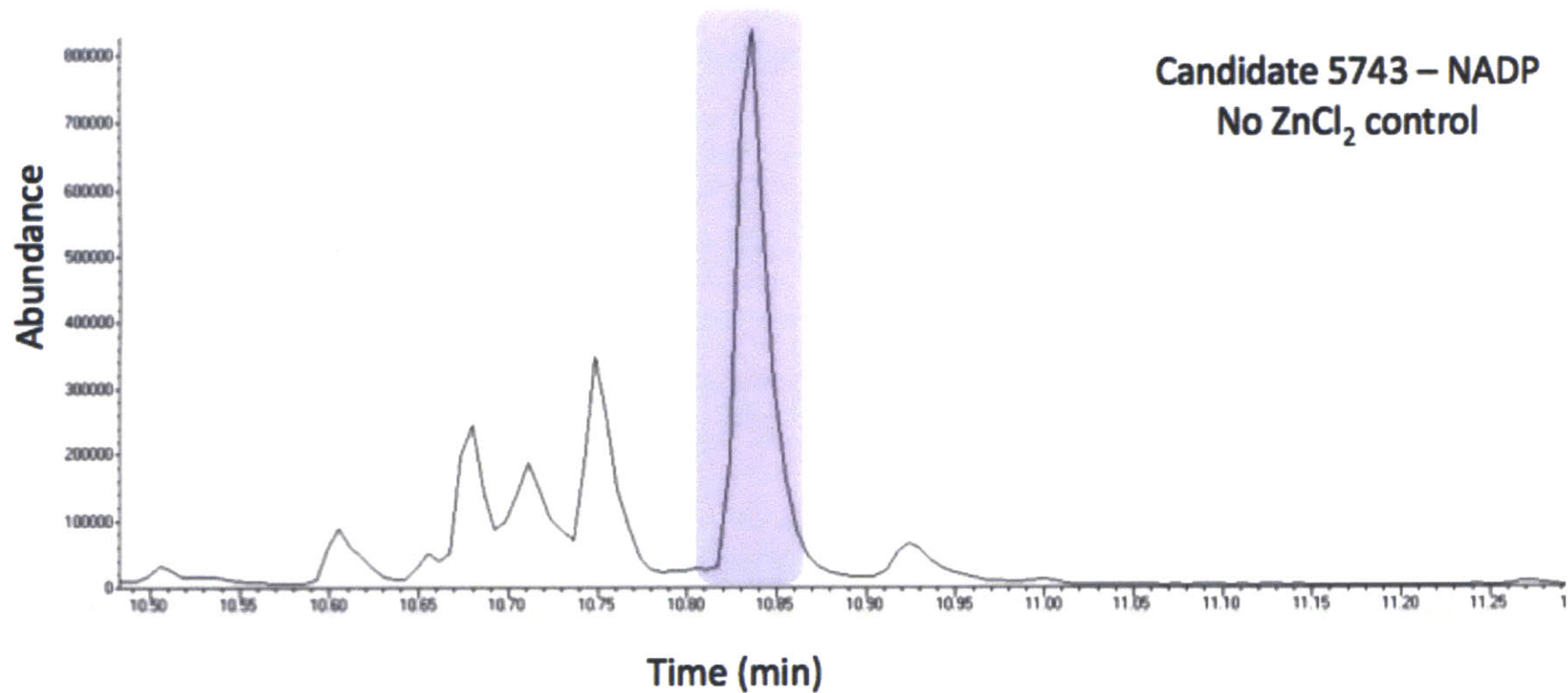


Figure A.40: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 and cofactor NADP<sup>+</sup> when ZnCl<sub>2</sub> is omitted from the assay mixture. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

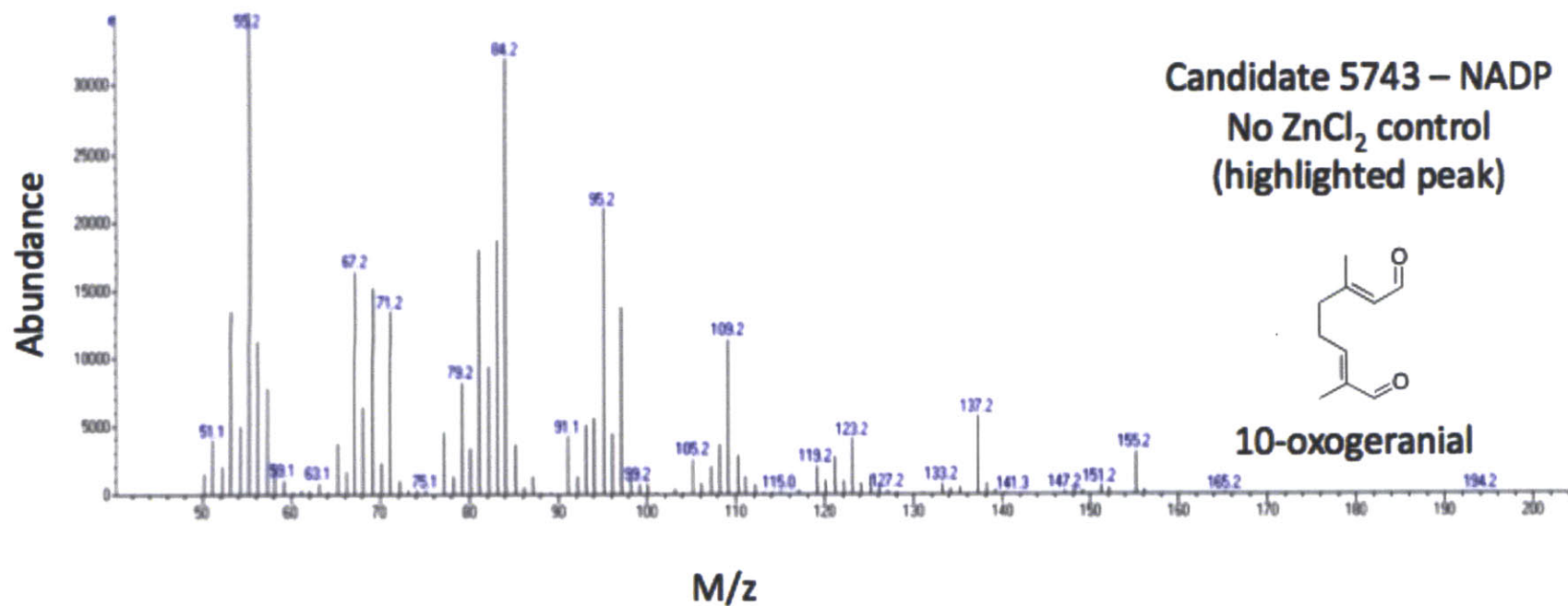


Figure A.41: Mass spectrum of highlighted peak from Candidate 5743 and cofactor NADP<sup>+</sup> assay when ZnCl<sub>2</sub> is omitted from the assay mixture. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.

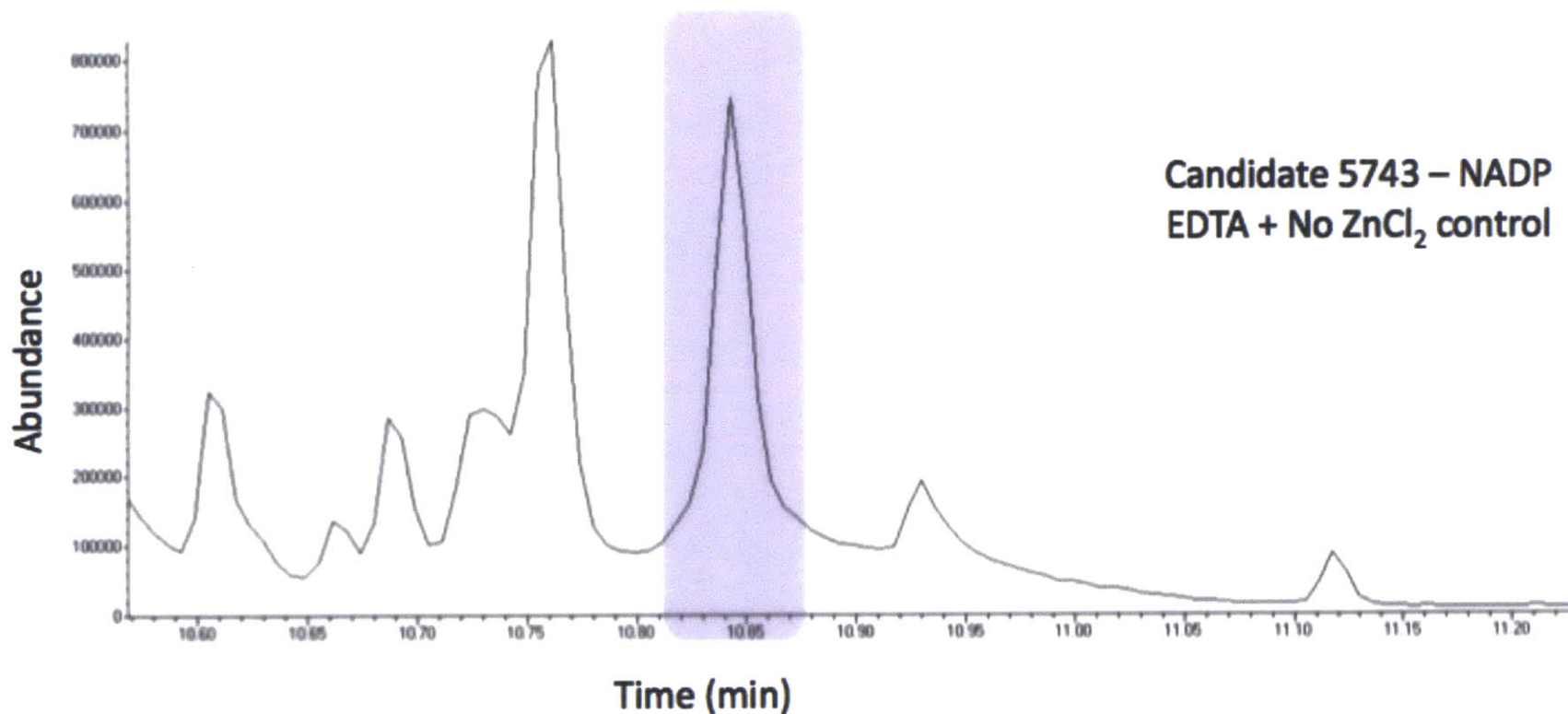


Figure A.42: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 and cofactor NADP<sup>+</sup> when EDTA is added to a final assay concentration of 1 mM and ZnCl<sub>2</sub> is omitted. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.



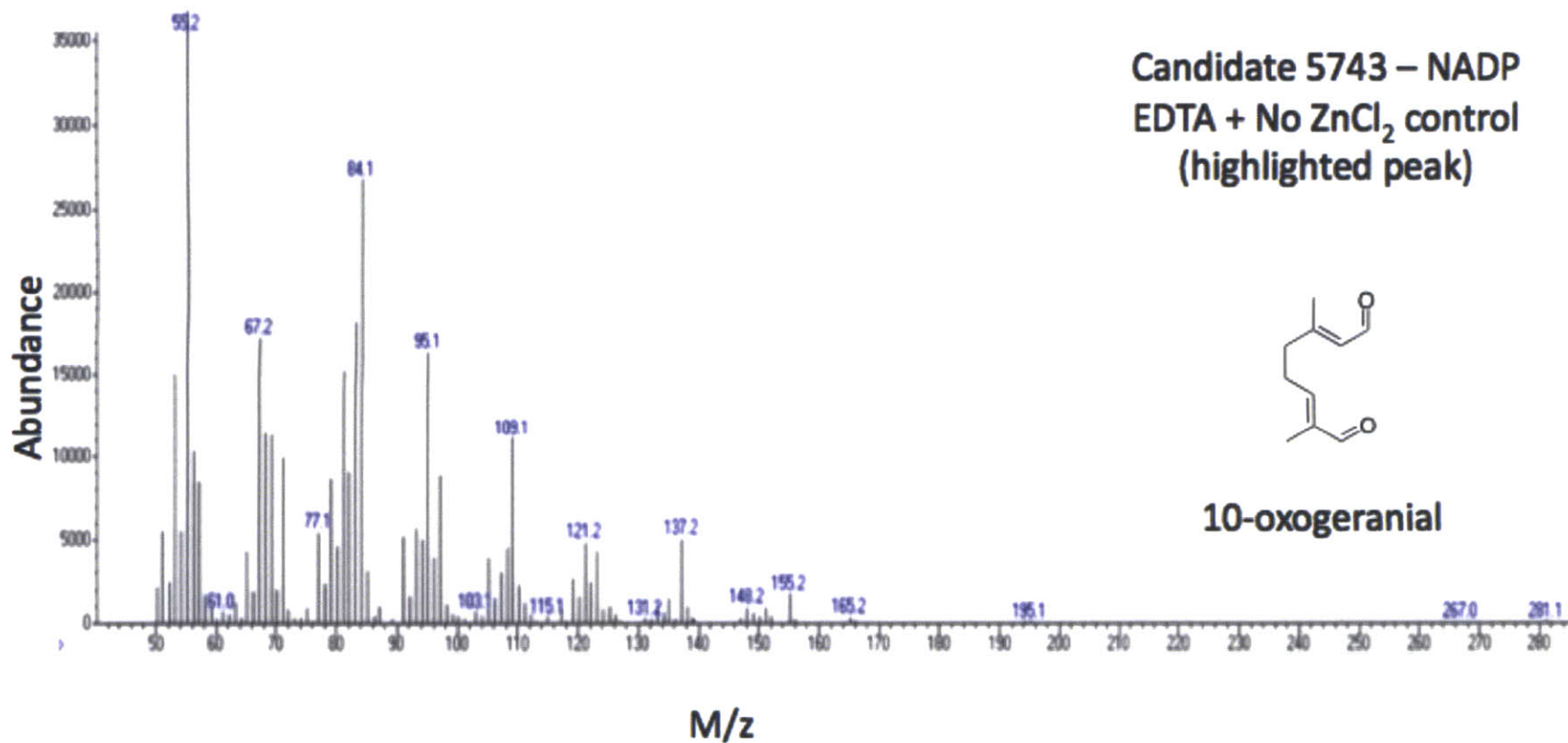


Figure A.43: Mass spectrum of highlighted peak from Candidate 5743 and cofactor NADP<sup>+</sup> assay when EDTA is added to the assay mixture to a final concentration of 1 mM and ZnCl<sub>2</sub> is omitted. Similarity to the mass spectrum of authentic 10-oxogeranial enables positive assignment.



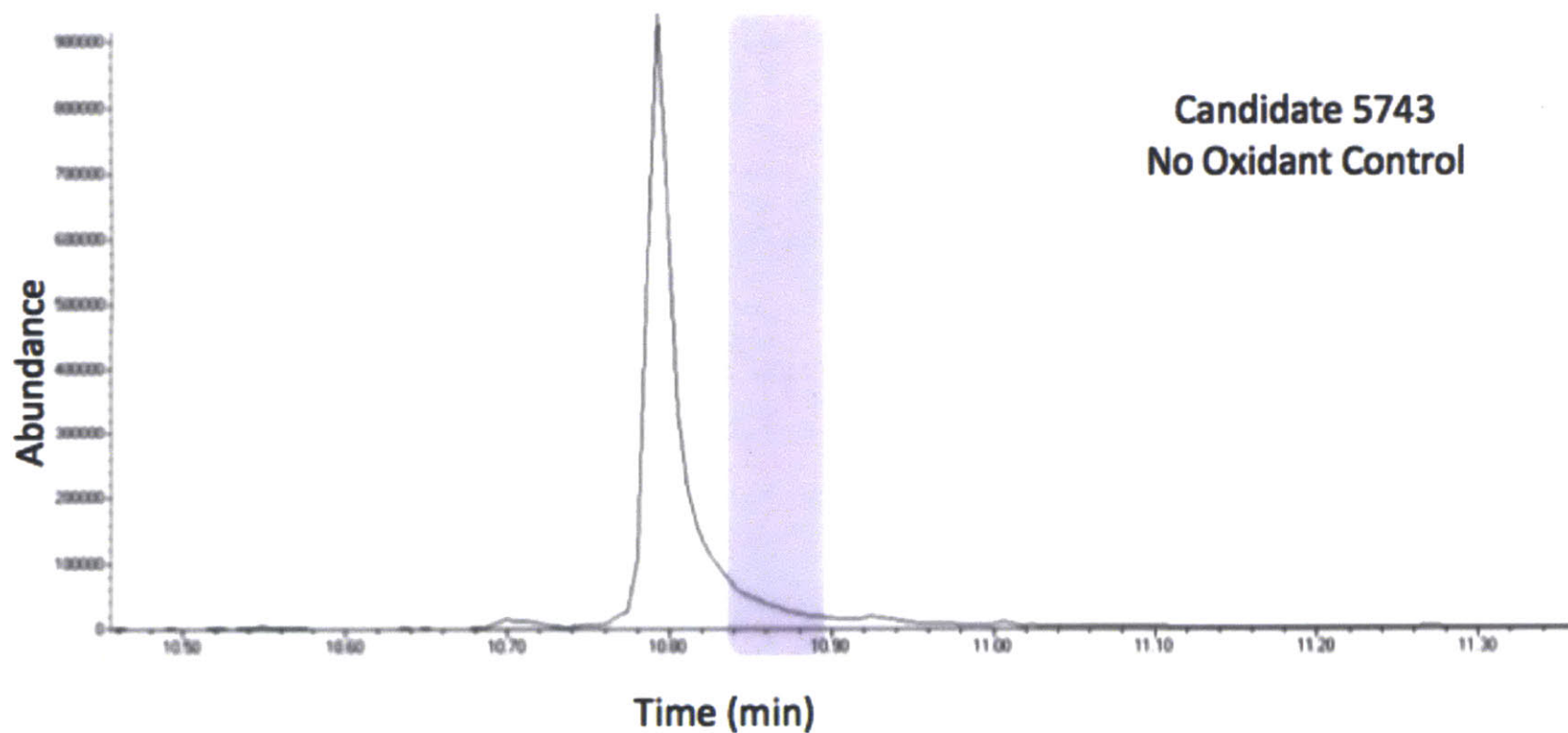


Figure A.44: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 5743 in the absence of either NAD<sup>+</sup> or NADP<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeraniol standard; the region of 10-oxogeraniol elution is highlighted in purple.

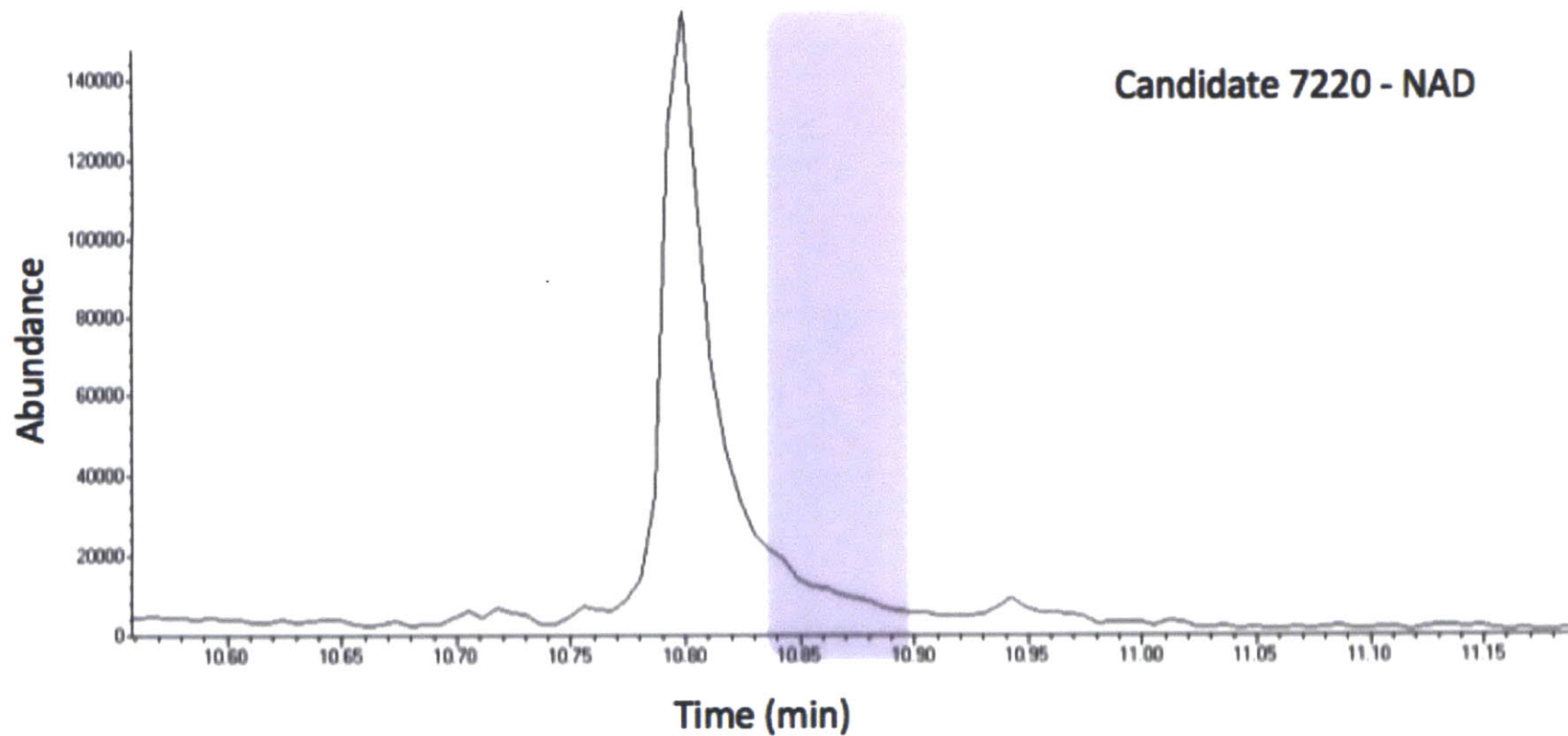


Figure A.45: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screen with Candidate 7220 and cofactor NAD<sup>+</sup>. No peak co-elutes with the authentic 10-oxogeranial standard; the region of 10-oxogeranial elution is highlighted in purple.

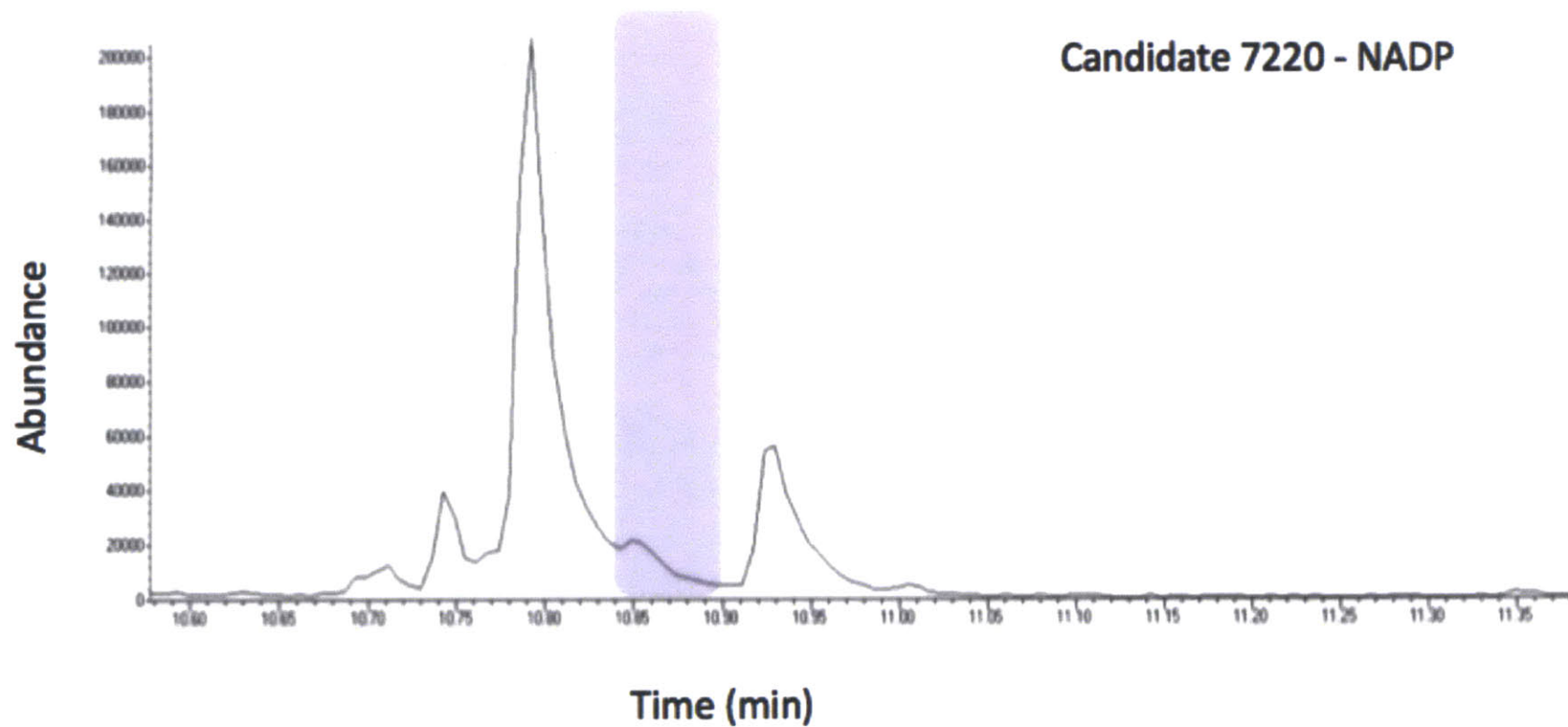


Figure A.46: GC-MS chromatogram of 10-hydroxygeraniol oxidoreductase candidate screening with Candidate 7220 and cofactor NADP<sup>+</sup>. The peak that co-elutes with the authentic 10-oxogeraniol standard is highlighted in purple.

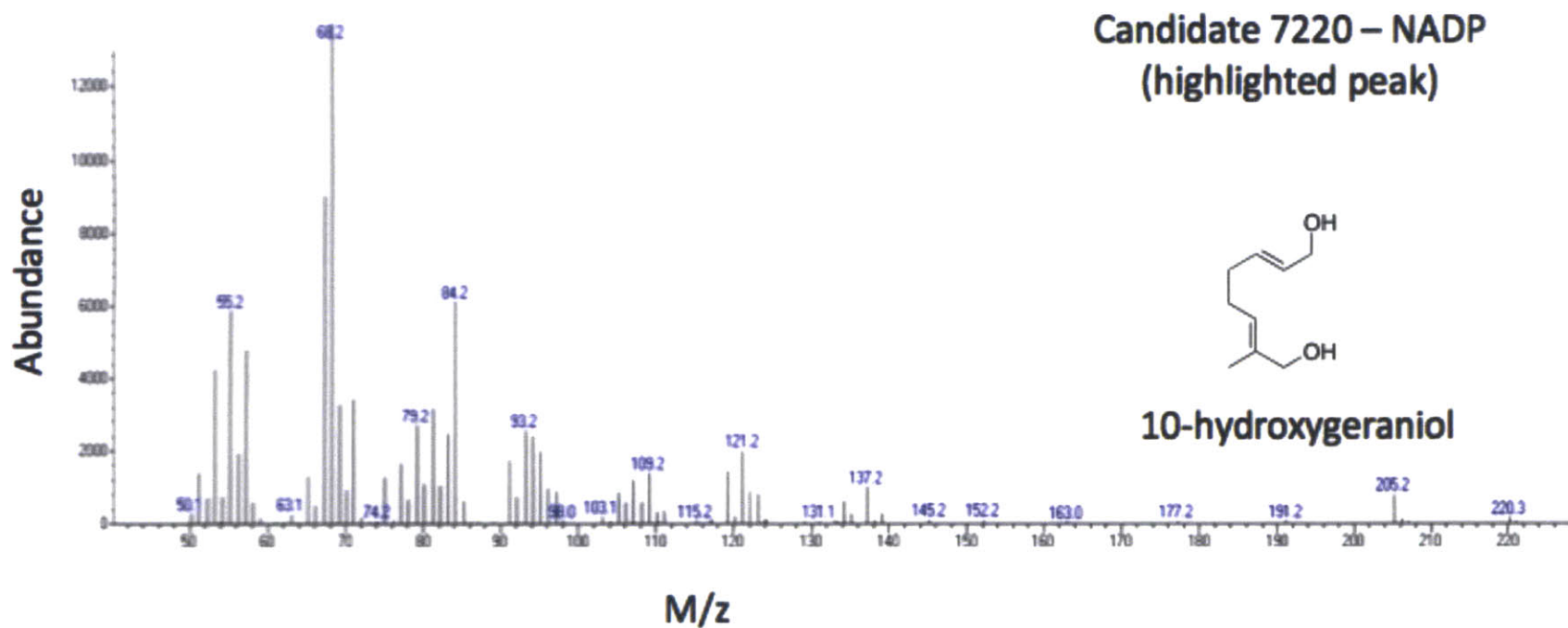


Figure A.47: Mass spectrum of highlighted peak from Candidate 7220 and cofactor NADP<sup>+</sup> assay. Similarity to the mass spectrum of authentic 10-hydroxygeraniol enables positive assignment as starting substrate, not product.

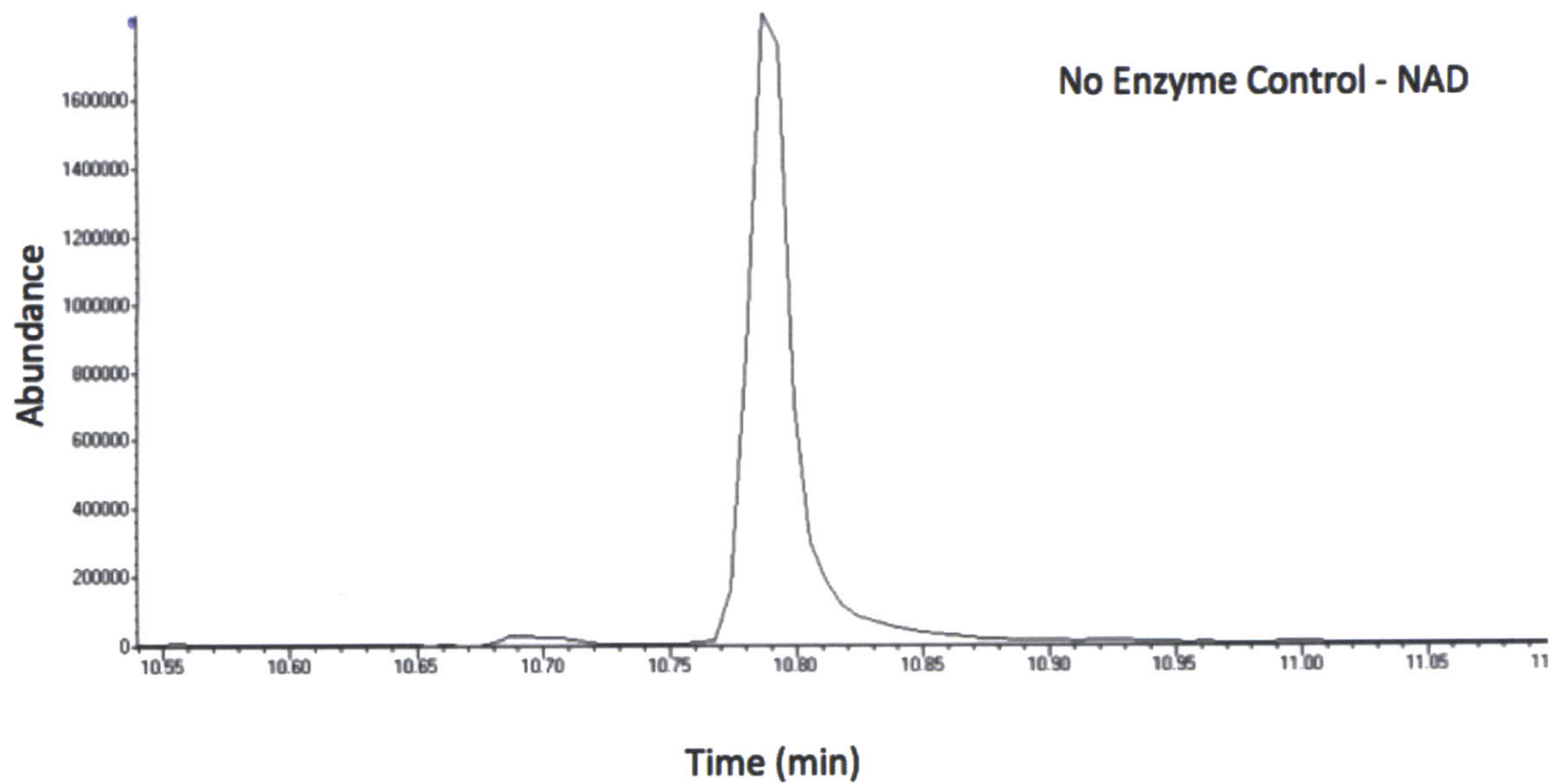


Figure A.48: No enzyme control with NAD<sup>+</sup> cofactor. No 10-oxogeraniol product is observed. The single peak shown co-elutes with substrate 10-hydroxygeraniol.

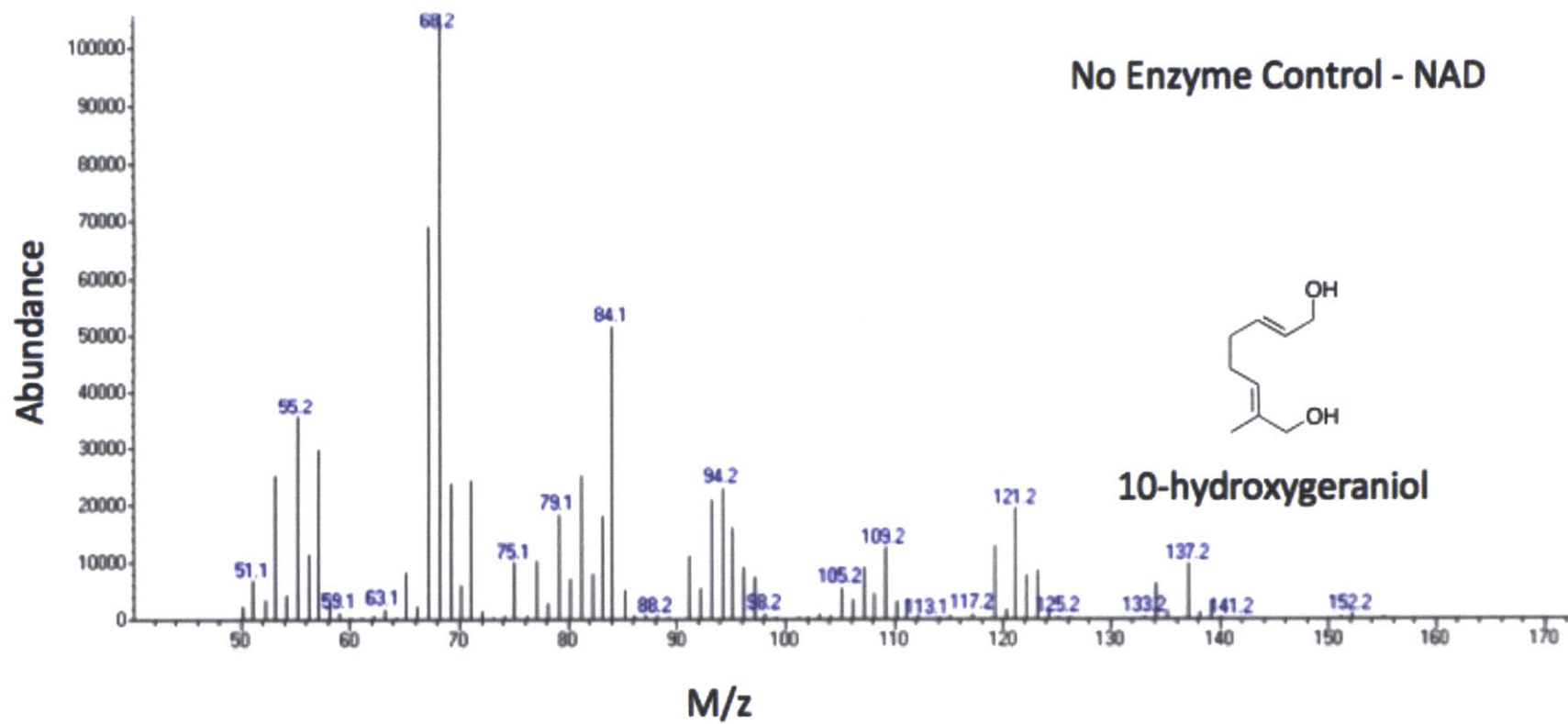


Figure A.49: Mass spectrum of no enzyme control with cofactor NAD<sup>+</sup>. Averaging of the single peak in the GC-MS chromatogram displays a mass spectrum signature consistent with 10-hydroxygeraniol, the substrate.

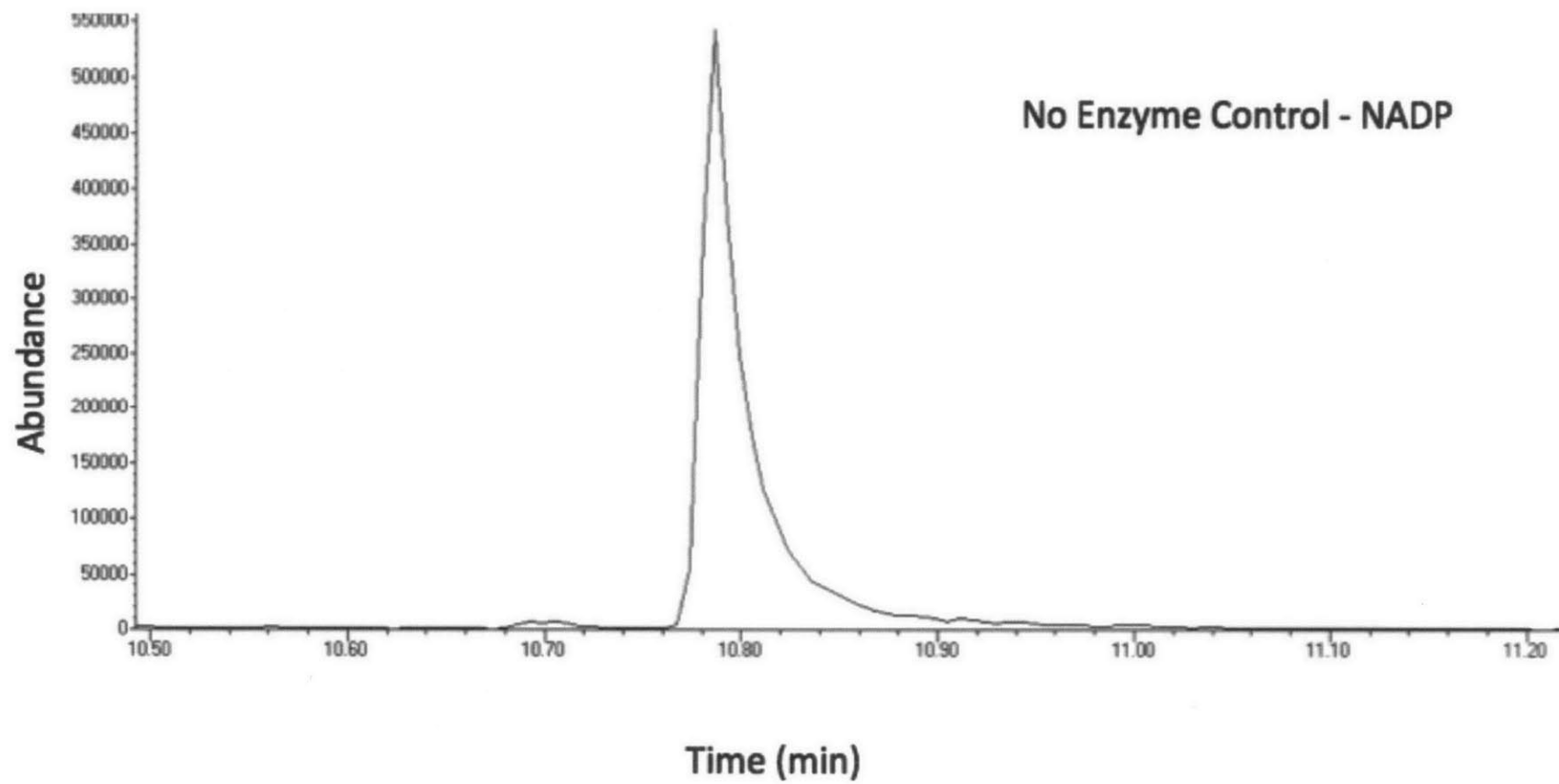


Figure A.50: No enzyme control with NADP<sup>+</sup> cofactor. No 10-oxogeraniol product is observed. The single peak shown co-elutes with substrate 10-hydroxygeraniol.

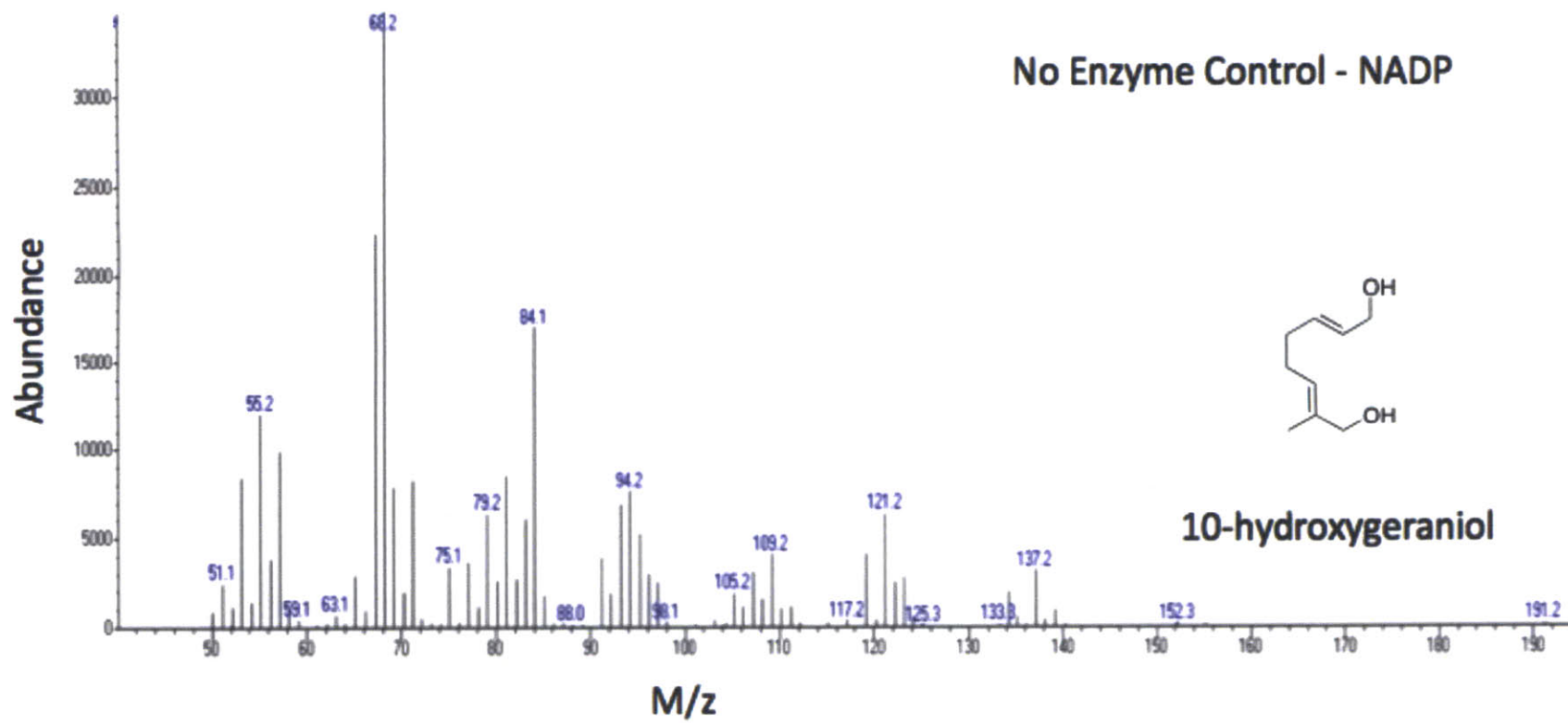


Figure A.51: Mass spectrum of no enzyme control with cofactor NADP<sup>+</sup>. Averaging of the single peak in the GC-MS chromatogram displays a mass spectrum signature consistent with 10-hydroxygeraniol, the substrate.



# Weslee S. Glenn

The John Innes Centre  
Department of Biological Chemistry  
Norwich, UK NR4 7UH

Phone: +44 07909 024 014  
E-Mail: wsglenn@mit.edu

## Education

- 2011-present**      **The John Innes Centre, Norwich, UK**  
Visiting Postgraduate Student, Department of Biological Chemistry  
Laboratory of Professor Sarah E. O'Connor
- 2008-present**      **Massachusetts Institute of Technology, Cambridge, MA**  
Ph.D. Candidate in Biological Chemistry, Department of Chemistry  
National Science Foundation Predoctoral Scholar (2010-13)
- 2004-2008**      **Hampton University, Hampton, VA**  
Bachelor of Science in Chemistry, 2008 – *Summa cum laude*  
Honors College, 2008 – *Summa cum laude*  
Cumulative GPA: 4.01      GPA in major: 4.06  
(A+ = 4.1, A = 4.0)  
National Science Foundation - Historically Black College Undergraduate  
Program Scholar (2005-2008)  
Presidential Scholar (2004-2008)

## Research Experience

- 1/2009-present**      **Laboratory of Professor Sarah E. O'Connor**  
The John Innes Institute, Department of Biological Chemistry  
Massachusetts Institute of Technology, Department of Chemistry  
*Metabolic engineering of plant alkaloid metabolism*
- 6/2008-7/2008**      **Laboratory of Professor Shanthi Paranawithana**  
Hampton University, Department of Chemistry  
*Gene expression and activity assay development of a putative  
transcriptional regulator in Mycobacterium tuberculosis*
- 6/2007-8/2007**      **Laboratory of Professor Sarah E. O'Connor**  
Massachusetts Institute of Technology, Department of Chemistry  
*Substrate specificity of Tabersonine 16-Hydroxylase from C. roseus*
- 1/2007-5/2007**      **Laboratory of Professor Edmund Ndip**  
Hampton University, Department of Chemistry  
*Computational research on two photon absorbers – optical switching*
- 6/2006-8/2006**      **Laboratory of Robert H. Grubbs**  
California Institute of Technology, Division of Chemistry and Chemical  
Engineering  
*Development of a more active and stable ruthenium catalyst for  
tetrasubstituted olefin formation*
- 9/2005-5/2007**      **Laboratory of Professor Charles Bump**  
Hampton University, Department of Chemistry

## **Publications**

- 4/2013**                    **Recent Progress in the Metabolic Engineering of Alkaloids in Plant Systems**  
Weslee S. Glenn<sup>†</sup>, Weerawat Runguphan<sup>†</sup> and Sarah E. O'Connor († - equal authorship)  
*Current Opinion in Biotechnology*, 2013, 24(2), pp. 354-65.
- 12/2012**                    **An Alternative Route to Cyclic Terpenes by Reductive Cyclization in Iridoid Biosynthesis**  
Fernando Geu-Flores, Nathaniel H. Sherden, Vincent Courdavault, Vincent Burlat, Weslee S. Glenn, Cen Wu, Ezekiel Nims, Yuehua Cui and Sarah E. O'Connor  
*Nature*, 2012, 492(7427), pp. 138-42.
- 6/2012**                    **Redesign of a Dioxygenase in Morphine Biosynthesis**  
Weerawat Runguphan<sup>†</sup>, Weslee Glenn<sup>†</sup> and Sarah E. O'Connor († - equal authorship)  
*Chemistry and Biology*, 2012, 19, pp. 674-678
- 12/2011**                    **Reengineering a Tryptophan Halogenase to Chlorinate a Direct Alkaloid Precursor**  
Weslee S. Glenn, Ezekiel Nims and Sarah E. O'Connor  
*Journal of the American Chemical Society*, 2011, 133 (48), pp. 19346-19349

## **Research and Academic Grants**

### *Massachusetts Institute of Technology*

- 9/2010-9/2013**            **National Science Foundation Predoctoral Fellowship** (bioorganic)  
**5/2010-8/2010**            **Walter L. Hughes Graduate Fellowship in Biochemistry**  
**9/2009- 5/2010**            **Henry A. Hill Fellowship**  
**9/2008-5/2009**            **Chemistry/ Biology Interface Program Training Grant**  
**9/2008-5/2009**            **Institute Fellowship**

### *Hampton University*

- 5/2005-5/2008**            **Historically Black College/ University Undergraduate Program Scholarship** (sponsored by the National Science Foundation)  
**9/2004-5/2008**            **Presidential Scholarship** (Full tuition, room and board)

## Other Awards

### *Massachusetts Institute of Technology and the John Innes Centre*

- 3/2012 **Friends of John Innes Outreach presentation voted “Most worthy of additional funding”**  
*Getting Plants to Make Medicines*
- 2/2011 **Massachusetts Institute of Technology Martin Luther King, Jr. Leadership Award**  
(awarded for work on reducing stereotype threat in the science classroom)

### *Hampton University*

- 5/2008 **Chemistry Excellence in Service Award**
- 5/2008 **Merck Index Excellence in Undergraduate Chemistry Award**
- 5/2008 **School of Science flag bearer for 2008 Commencement Exercises**
- 4/2008 **All-Virginia Collegiate Honors Council Poster Award Winner**
- 4/2008 **Honors Council Award 2008 – Chemistry**
- 4/2008 **Honors Council Award 2008 – Honors College**
- 12/2007 **ACS Award for Achievement in Physical Chemistry**
- 12/2006 **Beta Kappa Chi Scientific Honor Society Inductee**
- 12/2006 **ACS Award for Achievement in Organic Chemistry**
- 11/2006 **Golden Key International Honour Society Inductee**
- 10/2006 **Alpha Kappa Mu Honors Society Inductee**
- 3/2006 **“Future Nobel Prize Nominee – Chemistry” 2006 Honors Council Award**
- 12/2005 **ACS Award for Freshman Achievement in Chemistry**

## Professional Activities

### *Massachusetts Institute of Technology and the John Innes Centre*

- 3/2012-present **Contract science editor** (for scientists who are non-native English speakers)  
American Journal Experts
- 6/2011 **Future Faculty Workshop of MIT**
- 6/2010-8/2010 **Developed an interactive workshop to mitigate the effects of stereotype threat in teaching assistants at the Massachusetts Institute of Technology**  
Howard Hughes Medical Institute Professor Program  
Professor Catherine L. Drennan (MIT)
- 9/2009-7/2010 **Mentor for middle school-aged children interested in STEM**  
Office of Education and Outreach Programs (MIT)
- 9/2008-9/2009 **Co-chair of National Organization for the Professional Development of Black Chemists and Chemical Engineers**



## *Hampton University*

9/2007-5/2008      **Chemistry Club – Co-President**  
9/2006-5/2008      **Department of Chemistry – Head Tutor**  
9/2006-5/2007      **Chemistry Club – Special Projects Chair**  
9/2005-5/2008      **Honors College – Winner’s Circle Chair**

## **Conferences and Presentations**

3/2012      ***Getting Plants to Make Medicines***  
(Invited outreach lecture)  
Friends of John Innes

11/2011      ***Hijacking Monoterpene Indole Alkaloid Biosynthesis for the  
Production of Unnatural Natural Products***  
John Innes Institute, Department of Biological Chemistry

12/2010      ***Optimizing Engineered Halogenation in Periwinkle Plant Culture***  
(Invited lecture)  
Hampton University, Department of Chemistry

11/2010      ***Optimizing Engineered Halogenation in Periwinkle Plant Culture***  
National Organization of Black Chemists and Chemical Engineers  
Northeastern Regional Conference

10/2009      ***Engineering Halogenation into Periwinkle*** (poster)  
National Organization for the Professional Development of Black  
Chemists and Chemical Engineers Regional Conference

7/2008      ***Toward a Fluorescence Based Activity Assay of Rv 1151c in  
Mycobacterium tuberculosis***  
Hampton University Center for Research Excellence in Science and  
Technology – Internal Review

4/2008      ***Substrate Specificity of Tabersonine 16-Hydroxylase-Reductase in  
Catharanthus roseus*** (poster)  
All Virginia Honors Council Conference

4/2008      ***Substrate Specificity of Tabersonine 16-Hydroxylase-Reductase in  
Catharanthus roseus*** (Poster)  
American Chemical Society Southeast Regional Meeting

3/2008      ***Substrate Specificity of Tabersonine 16-Hydroxylase-Reductase in  
Catharanthus roseus*** (Invited lecture—selected paper in the Rohm and  
Haas National Undergraduate Competition)  
National Organization for the Professional Development of Black  
Chemists and Chemical Engineers Conference

10/2007      ***Development of a More Active and More Stable Ring-Closing  
Metathesis Catalyst***  
Historically Black College/ University Undergraduate Program  
National Conference

## Teaching and Education Experience

### *Massachusetts Institute of Technology and the John Innes Centre*

- 10/2011-5/2012**      **Postgraduate Mentor to Master's Student**  
Laboratory of Professor Sarah E. O'Connor
- 9/2010-3/2011**      **Certificate in Teaching College Science and Engineering**  
Teaching and Learning Laboratory (MIT)
- 6/2010-9/2010**      **Diversity training – Reducing Stereotype Threat**  
Professor Catherine Drennan's Howard Hughes Medical Institute  
Professor's Program
- 9/2009-5/2010**      **Middle School Counselor**  
Office of Engineering Outreach Programs (MIT)
- 8/2009-12/2009**      **General Chemistry (5.111) Teaching Assistant**

### *Hampton University*

- 6/2008-7/2008**      **Mentor to High School Science Student**  
Laboratory of Professor Shanthi Paranawithana
- 6/2008-7/2008**      **Organic Chemistry Laboratory Teaching Assistant**
- 6/2008-7/2008**      **Middle school counselor for the "Young Doctors' Program"**
- 9/2005-8/2008**      **General Chemistry and Organic Chemistry Tutor**
- 9/2005-12/2005**      **Elementary School Tutor**  
Read: For the Future

## References

**Professor Sarah E. O'Connor** (advisor)  
sarah.o'connor@jic.ac.uk (E-Mail)  
The John Innes Centre  
Department of Biological Chemistry

**Professor Catherine L. Drennan** (committee chair)  
cdrennan@mit.edu (E-Mail)  
Massachusetts Institute of Technology  
Department of Chemistry

**Professor Barbara Imperiali** (committee member)  
imper@mit.edu (E-Mail)  
Massachusetts Institute of Technology  
Department of Biology